



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

**MODELO DE CALIDAD PARA EVALUACIÓN Y MEJORA
DE VISUALIZACIONES DE INFORMACIÓN NARRATIVAS**

por

Andrea Fernanda Lezcano Airaldi

Tesis para obtener el grado académico de

DOCTOR EN INFORMÁTICA

Facultad de Ciencias Exactas y Naturales y Agrimensura

de la

UNIVERSIDAD NACIONAL DEL NORDESTE

Director de la Tesis: Emanuel Irrazábal

Codirector de la Tesis: Jorge Andrés Díaz Pace

Junio de 2025



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

Certificamos haber leído el documento de la Tesis preparada bajo nuestra dirección y recomendamos sea aceptada como parte de los requisitos para obtener el grado académico de Doctor en Informática.

Director de Tesis: Dr. Emanuel
IRRAZÁBAL

Codirector de Tesis: Dr. Jorge Andrés
DIAZ PACE

Corrientes, 3 de junio de 2025



Doctorado en Informática

DECLARACIÓN DEL AUTOR

Este documento de Tesis ha sido presentado y aprobado como parte de los requisitos para obtener el grado académico de Doctor ante la Universidad Nacional del Nordeste. Un ejemplar de este en formato electrónico (PDF) ha sido remitido al Comité Académico de dicho Doctorado para que esté disponible para su incorporación a los repositorios institucionales que corresponda.

Citas breves de este documento son permitidas sin necesidad de un permiso especial, siempre y cuando la fuente sea correctamente referida. Citas extendidas o la reproducción total o parcial de este manuscrito sólo podrán realizarse previa autorización del portador legal del derecho de propiedad intelectual del mismo.



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

A MI FAMILIA



Doctorado en Informática

AGRADECIMIENTOS

A mi familia: por el apoyo constante; por sostener lo cotidiano cuando la agenda del doctorado se volvía impredecible; y por la paciencia frente a mis cambios de humor. A Aye, por su amistad, siempre presente en este camino. A la memoria de mi abuela, cuyas meriendas fueron durante años un espacio de charla y catarsis: su afecto me acompaña en este cierre.

A Emanuel, por la dirección de este trabajo y su mentoría a lo largo de mi formación, alentándome a salir de la zona de confort e involucrarme en experiencias que enriquecieron mi desarrollo académico. A Andrés, por la codirección, el apoyo en distintas etapas y la colaboración en los estudios realizados.

A Juan, por haber compartido buena parte de este recorrido. A colegas y equipos con quienes compartí trabajo e intercambio en distintos espacios académicos, por el aprendizaje compartido y las conversaciones nerd de pasillo. Rodearse de personas tan comprometidas renueva la curiosidad y anima a ir por más.

Al CONICET, por el apoyo económico brindado a través de la beca doctoral. A la Universidad Nacional del Nordeste, por ofrecer el entorno institucional necesario para llevar adelante esta investigación. Al Servicio Alemán de Intercambio Académico (DAAD) y a la Secretaría de Educación de la República Argentina, por hacer posible la estancia de investigación en Alemania.

Este recorrido me puso frente a desafíos y me abrió oportunidades para explorar ideas, asumir responsabilidades y formar criterio propio, y, en el camino, conocer personas valiosas. En el proceso aprendí a reconocer mis miedos y a atravesarlos. Hoy cierro esta etapa con una perspectiva más clara del rumbo que quiero seguir y con gratitud por quienes la hicieron posible.



Doctorado en Informática

TABLA DE CONTENIDOS

CAPÍTULO 1. INTRODUCCIÓN	17
1.1. Motivación.....	17
1.2. Problema.....	20
1.3. Hipótesis	21
1.4. Objetivos.....	22
1.5. Contribuciones e impacto	22
1.6. Contexto de la investigación.....	23
1.7. Estructura de la tesis	24
CAPÍTULO 2. METODOLOGÍA.....	26
2.1. Investigación en Ciencia del Diseño	26
2.1.1. Explicar el problema.....	27
2.1.2. Definir objetivos y requerimientos	28
2.1.3. Diseñar y desarrollar el artefacto	29
2.1.4. Demostrar el artefacto	29
2.1.5. Evaluar el artefacto	30
2.2. Organización de las actividades	31
2.3. Conclusiones.....	32
CAPÍTULO 3. ESTADO DEL ARTE.....	33
3.1. Narrativa de datos.....	33
3.2. Buenas prácticas para la creación de visualizaciones narrativas.....	36
3.3. Criterios de evaluación	38
3.3.1. Comprensión.....	38
3.3.2. Compromiso	39



Doctorado en Informática

3.3.3.	Información	40
3.3.4.	Memorabilidad	41
3.3.5.	Usabilidad.....	42
3.4.	Metodologías de evaluación existentes	43
3.5.	Modelos de calidad en Ingeniería de Software.....	49
3.6.	Conclusiones.....	50
CAPÍTULO 4. DESARROLLO DEL ARTEFACTO		52
4.1.	Modelo propuesto: NaVi-Q	52
4.2.	Definir el alcance y el contexto de aplicación.....	53
4.2.1.	Técnicas de visualización consideradas	54
4.3.	Diseñar el metamodelo de calidad.....	55
4.4.	Identificar los requerimientos de calidad	56
4.5.	Definir las propiedades de calidad relevantes	57
4.5.1.	Criterios	57
4.5.2.	Propiedades.....	57
4.5.3.	Aplicaciones	59
4.5.4.	Relación entre criterios, propiedades y aplicaciones.....	60
4.6.	Definir las reglas de evaluación.....	62
4.6.1.	Calidad como constructo formativo	62
4.6.2.	Asignación de pesos con Proceso Analítico Jerárquico	64
4.6.3.	Definición de escala de medición.....	66
4.6.4.	Funciones de medición	66
4.7.	Instanciar el modelo para un caso de uso	68
4.8.	Conclusiones.....	69



Doctorado en Informática

CAPÍTULO 5. VALIDACIÓN DEL ARTEFACTO.....	71
5.1. Estrategia de validación.....	71
5.2. Evaluación a pequeña escala	73
5.2.1. Preguntas de investigación	73
5.2.2. Configuración del perfil de calidad	74
5.2.3. Conjunto de datos	74
5.2.4. Perfil de los evaluadores.....	76
5.2.5. Procedimiento.....	76
5.2.6. Resultados.....	77
5.3. Primera iteración de diseño del artefacto	81
5.3.1. Desarrollo de la herramienta soporte.....	82
5.4. Experimento	84
5.4.1. Participantes	85
5.4.2. Materiales	87
5.4.3. Procedimiento.....	88
5.4.4. Recolección de datos	88
5.4.5. Resultados.....	90
5.5. Segunda iteración de diseño del artefacto	102
5.5.1. Automatización de aplicaciones	103
5.5.2. Evaluación asistida por IA.....	103
5.6. Estudio de caso	106
5.6.1. Contexto	107
5.6.2. Procedimiento.....	108
5.6.3. Recolección de datos	109



Doctorado en Informática

5.6.4. Resultados.....	113
5.7. Conclusiones.....	121
CAPÍTULO 6. CONCLUSIONES	125
6.1. Objetivos de la tesis abordados	125
6.2. Contribuciones adicionales.....	128
6.3. Amenazas a la validez.....	130
6.3.1. Validez externa	130
6.3.2. Validez de constructo.....	132
6.3.3. Validez de conclusión	132
6.4. Trabajos futuros	133
6.5. Difusión de resultados	135
6.5.1. Publicaciones derivadas de la tesis.....	135
6.5.2. Otras publicaciones.....	136
6.5.3. Participación en eventos	137
6.5.4. Registro del artefacto.....	138
APÉNDICE A. ESTUDIOS PRIMARIOS DEL MAPEO SISTEMÁTICO	139
APÉNDICE B. LISTA COMPLETA DE APLICACIONES.....	147
APÉNDICE C. MATERIALES: EVALUACIÓN A PEQUEÑA ESCALA.....	156
APÉNDICE D. MATERIALES: EXPERIMENTO	160
APÉNDICE E. MATERIALES: ESTUDIO DE CASO	166
APÉNDICE F. MODELO COMPLETO NAVI-Q	172
APÉNDICE G. MANUAL DE USUARIO NAVI-Q	178
REFERENCIAS.....	187

Doctorado en Informática

ÍNDICE DE TABLAS

Tabla 1: Referencias a los estudios conducidos.....	32
Tabla 2: Buenas prácticas para la creación de visualizaciones narrativas	36
Tabla 3: Resumen comparativo de metodologías de evaluación existentes	47
Tabla 4: Propiedades de calidad	58
Tabla 5: Aplicaciones (fragmento)	59
Tabla 6: Criterios de calidad y propiedades.....	61
Tabla 7: Aplicaciones de cada propiedad	61
Tabla 8: Ejemplos de perfiles de calidad	69
Tabla 9: Peso de los criterios para el perfil de calidad Estándar	74
Tabla 10: Ejemplo de matriz de evaluación por participante	89
Tabla 11: Tabla binaria por visualización (extracto).....	89
Tabla 12: Resultados de la encuesta sobre asociaciones entre criterios y propiedades..	93
Tabla 13: Principales temas y palabras clave asociadas	100
Tabla 14: Pesos asignados a los criterios del perfil de calidad personalizado.....	108
Tabla 15: Preguntas del cuestionario de comprensión e información	111
Tabla 16: Ítems de la escala de Usabilidad (SUS) adaptada.....	112
Tabla 17: Frecuencia, utilidad, conocimiento y aplicación de las recomendaciones	113
Tabla 18: Agrupación de recomendaciones según el tipo de problema.....	114
Tabla 19: Media, mediana, rango y desviación estándar del SUS por grupo	118
Tabla 20: Requisitos para métodos de evaluación de calidad	122

Doctorado en Informática

ÍNDICE DE FIGURAS

Fig. 1: Organización de actividades y tareas desarrolladas en el marco de DSR.....	31
Fig. 2: Ejes conceptuales de la visualización narrativa	34
Fig. 3: Proceso de narración con datos según (B. Lee et al., 2015)	35
Fig. 4: Actividades abarcadas en el capítulo.....	50
Fig. 5: Foco del modelo dentro del ciclo de vida de la calidad	53
Fig. 6: Elementos y estructura del metamodelo	56
Fig. 7: Relación entre un constructo formativo y sus medidas (ISO/IEC 33003).....	63
Fig. 8: Comparación entre una jerarquía estándar en AHP (a la izquierda) y un extracto del modelo de calidad propuesto (a la derecha)	65
Fig. 9: Relación entre el modelo de calidad y el modelo de medición.....	67
Fig. 10: Actividades abarcadas en el capítulo.....	70
Fig. 11: Esquema general de validación.	72
Fig. 12: Visualizaciones seleccionadas para la evaluación a pequeña escala.....	75
Fig. 13: Fragmento de la hoja de cálculo utilizada para las evaluaciones.....	76
Fig. 14: Resumen de resultados de evaluación.....	77
Fig. 15: Cumplimiento de objetivos en la primera etapa de validación.	81
Fig. 16: Diagrama de clases de la herramienta de evaluación.....	82
Fig. 17: Interfaz de la herramienta NaVi-Q.....	83
Fig. 18: Ejemplos de visualizaciones narrativas utilizadas en el estudio	87
Fig. 19: Distribución de puntajes asignados por expertos y no expertos en cada nivel de calidad.....	90

Doctorado en Informática

Fig. 20: Puntajes asignados por expertos y no expertos a las visualizaciones (V) de cada nivel	92
Fig. 21: Puntuaciones asignadas por expertos, no expertos y evaluaciones de referencia por nivel de calidad.....	95
Fig. 22: Correlación entre las puntuaciones de los evaluadores y las evaluaciones de referencia	96
Fig. 23: Calificaciones promedio en la encuesta final.....	100
Fig. 24: Cumplimiento de objetivos en la segunda etapa de validación.....	102
Fig. 25: Ayuda contextual generada por IA durante la evaluación manual.	104
Fig. 26: Esquema del proceso de evaluación.....	105
Fig. 27: Sugerencia de mejora generada por IA	106
Fig. 28: Ejemplos de visualizaciones evaluadas en el estudio	108
Fig. 29: Utilidad, conocimiento y aplicación de las recomendaciones según tipo de problema identificado. Los valores absolutos indican la frecuencia de cada error	114
Fig. 30: Ejemplo de visualización original y rediseñada.....	116
Fig. 31: Precisión y confianza normalizada por pregunta y grupo	117
Fig. 32: Puntaje promedio por ítem y grupo.....	119
Fig. 33: Cumplimiento de objetivos en la tercera etapa de validación.....	121
Fig. 34: Actividades abarcadas en el capítulo.....	124



Doctorado en Informática

SIGLAS

Sigla	Descripción
DE	Desvío Estándar
DSR	<i>Design Science Research</i>
IA	Inteligencia Artificial
IS	Ingeniería de Software
LLM	<i>Large Language Model</i>
ML	<i>Machine learning</i>
PI	Pregunta de Investigación
SMS	<i>Systematic Mapping Study</i>
SWK	Siglas en alemán del Grupo de investigación en Métodos de Desarrollo y Construcción de Software
VI	Visualización de Información

Doctorado en Informática**RESUMEN**

Las visualizaciones narrativas combinan recursos gráficos y estructuras discursivas para facilitar la comprensión de los datos en diversos contextos. Si bien su uso se ha expandido en áreas como el periodismo, la salud pública y la educación, la evaluación de su calidad sigue siendo un desafío. Las metodologías existentes tienden a centrarse en aspectos como la precisión en la lectura de datos o el tiempo de respuesta, dejando de lado dimensiones relacionadas con la narrativa, la utilidad percibida o la experiencia general del usuario.

La presente tesis propone NaVi-Q, un modelo de calidad diseñado para evaluar y asistir en la mejora de visualizaciones narrativas en interfaces de usuario. El modelo se construyó a partir de estándares consolidados en Ingeniería del Software y principios de diseño aplicados a la visualización de información. Su estructura jerárquica organiza buenas prácticas y criterios de calidad en propiedades observables y aplicaciones evaluativas con instrumentos de medición concretos. Cada elemento cuenta con un peso relativo configurable según el objetivo de la evaluación o el contexto de uso.

El modelo fue operacionalizado en una herramienta de software que permite evaluar visualizaciones, identificar oportunidades de mejora y generar recomendaciones específicas. Su validación se realizó en tres etapas: un estudio exploratorio a pequeña escala, un experimento controlado y un estudio de caso aplicado. Los resultados muestran que NaVi-Q permite detectar errores comunes, orientar mejoras concretas y generar diagnósticos valorados como útiles por quienes diseñan o utilizan visualizaciones.

Este trabajo contribuye con un enfoque estructurado para la evaluación de visualizaciones narrativas, con una base conceptual sólida, una herramienta práctica y evidencia empírica sobre su aplicabilidad en contextos reales. El modelo se adapta a distintos objetivos y escenarios de uso mediante perfiles de calidad configurables, lo que facilita su integración en procesos de diseño, revisión o mejora continua de gráficos.

Palabras clave: *visualización de información, visualizaciones narrativas, evaluación de calidad, modelo de calidad, comunicación visual, toma de decisiones.*

Doctorado en Informática**ABSTRACT**

Narrative visualizations integrate visual design and storytelling techniques to support data interpretation in a variety of contexts. While their use has expanded in fields such as journalism, public health, and education, assessing their quality remains a challenge. Existing evaluation methods often focus on technical aspects such as data accuracy or response time, while overlooking factors related to narrative structure, perceived usefulness, or overall user experience.

This thesis presents NaVi-Q, a quality model developed to assess and improve narrative visualizations in user interfaces. The model builds on established Software Engineering standards and design principles from the field of Information Visualization. Its hierarchical structure maps best practices and quality criteria into observable properties and evaluation components, each linked to concrete measurement instruments. The model allows components to be weighted according to evaluation goals, making it adaptable to different contexts.

NaVi-Q was implemented as a software tool to assess visualizations, identify areas for improvement, and generate actionable recommendations. Its validation followed a three-stage strategy: an exploratory pilot study, a controlled experiment, and a real-world case study. Results show that NaVi-Q helps identify common design issues, support targeted improvements, and produce evaluation outputs considered useful by creators and users of visualizations.

This work contributes a structured approach to evaluating narrative visualizations, combining a solid conceptual framework, a practical software implementation, and empirical validation in applied settings. The model can be adapted to different goals and usage scenarios through configurable quality profiles, enabling its integration into design, review, or continuous improvement workflows.

Keywords: *information visualization, narrative visualization, quality assessment, quality model, visual communication, decision-making.*



Capítulo 1. Introducción

Este capítulo presenta la motivación de la tesis, el problema que pretende resolver, el objetivo principal y la estructura general del documento. Está organizado de la siguiente manera: la sección 1.1 expone la motivación; cómo las técnicas narrativas pueden contribuir a la creación de visualizaciones de información comprensibles y la importancia de poder evaluar los gráficos resultantes. La sección 1.2 describe el problema, abordando los desafíos que presenta la evaluación de visualizaciones y las limitaciones de las herramientas existentes. La sección 1.3 plantea la hipótesis central que guio el desarrollo de la investigación. La sección 1.4 define los objetivos generales y específicos de la tesis. La sección 1.5 establece la contribución principal. La sección 1.6 ofrece el contexto en el cual la tesis fue conducida y finalmente, la sección 1.7 presenta la estructura del documento.

1.1. Motivación

La visualización de información (VI) se ha consolidado como una herramienta fundamental para comprender conjuntos de datos y comunicar hallazgos. Este subcampo de la investigación en visualización (Post et al., 2003) se enfoca en la representación visual de datos abstractos (Keim et al., 2008; Munzner, 2014), con el objetivo de mejorar la comprensión y amplificar los procesos cognitivos (Card et al., 1999).

Los datos abstractos son aquellos que representan información que no tiene una forma física concreta o directa (Munzner, 2014; Ware, 2020); capturan aspectos que no se pueden observar o medir directamente en el mundo físico, pero que son fundamentales

Doctorado en Informática

para entender ciertos fenómenos o comportamientos, como, por ejemplo, variables estadísticas, indicadores de negocio, relaciones entre elementos, conceptos o ideas.

Estos datos requieren una visualización adecuada para ser comprensibles y accionables. Gráficos como mapas de calor, diagramas de dispersión y visualizaciones interactivas convierten la abstracción en representaciones tangibles e interpretables, lo que resulta útil para la toma de decisiones y la solución de problemas complejos (Dimara & Stasko, 2022).

Dentro de este marco, el *storytelling* o narrativa emerge como un método eficaz para transmitir información y conocimiento. El storytelling se define como la transmisión de una secuencia de eventos, que a menudo involucran personajes y lugares – historias – utilizando el discurso y soportes visuales (Gottschall, 2012).

Una narrativa visual es una historia transmitida a través de un medio, como ilustraciones, animaciones, video y – más recientemente – visualizaciones (Eisner, 2008; Sless, 1981). En particular, la visualización posee una inclinación natural a la comunicación debido a su forma gráfica, lo que da lugar al concepto de visualización orientada a la comunicación (Viégas & Wattenberg, 2006).

La combinación de la idea de visualización orientada a la comunicación con la narración de historias produce la noción de **visualizaciones narrativas**: técnicas de visualización que utilizan recursos narrativos para contar historias sobre datos (Segel & Heer, 2010). En contextos estructurados, estas historias apoyan la discusión, toma de decisiones y análisis de procesos (Willett et al., 2011). Este potencial comunicativo ha sido aprovechado en diversos contextos donde la dimensión afectiva o persuasiva de la información es relevante, como en periodismo, salud pública o educación. Estudios recientes destacan que las visualizaciones narrativas pueden aumentar la comprensión, generar confianza y modificar actitudes frente a ciertos temas (Fansher et al., 2025; Shao et al., 2024) lo que resalta su valor como herramienta de comunicación estratégica.

Los principios generales de diseño propuestos por Cleveland & McGill (1984), seguidos por las contribuciones de Tufte (2001) y Kosslyn (2006), ofrecen una base sólida para la

Doctorado en Informática

creación de visualizaciones de información efectivas. Sin embargo, las visualizaciones narrativas requieren consideraciones adicionales. Prácticas específicas como proporcionar contexto, según lo identificado por Figueiras (2014) y Ajani et al. (2022), estructurar un flujo narrativo efectivo como sugieren Hullman et al. (2013) y McKenna et al. (2017), y utilizar el color para transmitir tonos y emociones (Bera, 2016) (Schloss et al., 2019), pueden transformar una visualización convencional en una narrativa más comprensible y eficaz (Lezcano Airaldi et al., 2021).

Sin embargo, diversos autores señalan que estas buenas prácticas, si bien conceptualmente sólidas, no siempre resultan fáciles de aplicar en la práctica, y pueden entrar en conflicto con los objetivos comunicativos o con la diversidad del público destinatario (Errey et al., 2025). Con la creciente demanda de este tipo de visualizaciones, es cada vez más importante avanzar, estandarizar y difundir estas guías para que más usuarios, sin importar su nivel de experiencia, puedan crear visualizaciones narrativas efectivas y tomar decisiones informadas.

La aplicación de las buenas prácticas y su impacto correspondiente deben validarse mediante una evaluación precisa, un desafío central y persistente en el campo de la visualización (Carpendale, 2008; Lam et al., 2012). Aunque existen métodos cualitativos y cuantitativos para evaluar diferentes aspectos de las historias basadas en datos (Carpendale, 2008), estos suelen enfocarse únicamente en la capacidad de una visualización para comunicar hechos. Estudios recientes sugieren que las personas pueden tomar decisiones irracionales incluso cuando comprenden los datos correctamente (Dimara et al., 2017b), y un buen rendimiento en tareas analíticas no garantiza lo mismo en la toma de decisiones (Dimara et al., 2017a). Esta dificultad se agrava en el caso de las visualizaciones narrativas, donde los métodos tradicionales de evaluación muestran limitaciones para capturar aspectos como la interpretación del mensaje, la coherencia narrativa o la respuesta emocional del usuario (Errey et al., 2024; Sakarang et al., 2025). Por lo tanto, es fundamental avanzar más allá de este enfoque de evaluación para valorar la utilidad de una visualización.

Doctorado en Informática**1.2. Problema**

La evaluación de visualizaciones de información representa un desafío persistente en esta área de investigación. A lo largo de los años, se han desarrollado varias metodologías que pueden clasificarse en enfoques cuantitativos y cualitativos. Los estudios cuantitativos, también conocidos como experimentos controlados, se centran en recopilar medidas de rendimiento como el tiempo y los errores y requieren un alto grado de precisión (Carpendale, 2008). Los estudios cualitativos, por otro lado, son generalmente menos formales y consideran la interacción entre varios factores para lograr una comprensión más enriquecida (Patton, 2014).

Algunos ejemplos de evaluación cuantitativa incluyen estudios sobre percepción gráfica, como los realizados por Nowell et al. (2002), Heer & Bostock (2010) y Perin et al. (2018). En el ámbito de la interacción, se destacan las contribuciones de Yi et al. (2007) y Feng et al. (2019). Respecto a la estética en visualizaciones, algunos ejemplos incluyen los trabajos de Cawthon & Moere (2007), Quispel et al. (2016) y Lei et al. (2018). En cuanto a la evaluación cualitativa, el enfoque más común es la heurística de usabilidad, pero esta no siempre captura todas las dimensiones de calidad relevantes para las visualizaciones de información, como señalan Zhu (2007), Santos et al. (2015) y Williams et al. (2018).

Varios autores han destacado las dificultades en lograr una evaluación empírica más allá de las métricas tradicionales en el campo de la visualización. Plaisant (2004), Carpendale (2008), Lam et al. (2012) y Elmqvist & Yi (2015), señalan que la efectividad de las visualizaciones depende de factores subjetivos. Estos incluyen el conocimiento previo del usuario, como lo indican Tory & Moller (2005) y Yi (2010), así como los objetivos o el contexto (Padda et al., 2007, Hullman & Diakopoulos, 2011). Además, la interpretación de las visualizaciones puede verse influenciada por sesgos y preferencias individuales, complicando el desarrollo de criterios de evaluación objetivos y procedimientos estandarizados.

A pesar de la diversidad de enfoques existentes, se observa una carencia de herramientas que ofrezcan una evaluación sistemática y completa (Lezcano Airaldi, Irrazábal, & Diaz-

Doctorado en Informática

Pace, 2025). Las metodologías actuales tienden a enfocarse en aspectos específicos, como el soporte a la comprensión, el compromiso del usuario o el diseño, pero requieren de una operacionalización efectiva que abarque todas las dimensiones críticas.

Esto evidencia la necesidad de desarrollar un marco de evaluación más exhaustivo y práctico, que permita capturar los aspectos multidimensionales de las visualizaciones narrativas y optimizarlas para cada escenario de uso.

1.3. Hipótesis

En base a lo planteado anteriormente, la hipótesis de trabajo propone que:

La creación de un modelo de evaluación para visualizaciones narrativas que integre buenas prácticas y las implemente en un entorno de soporte semiautomatizado, contribuirá a incrementar su valor desde las fases iniciales de su desarrollo.

Esta premisa se basa en la idea de que el valor de una visualización está vinculado estrechamente con su capacidad para facilitar las tareas del usuario, como, por ejemplo, la educación, la exploración o la toma de decisiones.

Esta última ha sido identificada como uno de los objetivos principales de la visualización de información, de acuerdo con Munzner (2014) y Ware (2020). Por este motivo, el desafío radica en proporcionar un soporte efectivo para hacerlo. Para ello, es necesario identificar los principales criterios que intervienen en la evaluación de visualizaciones, como así también las técnicas existentes para su medición.

El desarrollo de un modelo de evaluación que contemple estos aspectos podría, por tanto, optimizar el proceso de creación y mejorar la eficacia de las visualizaciones narrativas en el contexto del desarrollo de software, donde estas visualizaciones se integran en interfaces, reportes, tableros o sistemas de apoyo a la decisión, y cumplen funciones clave para facilitar la interpretación y el uso de la información.

Doctorado en Informática

1.4. Objetivos

El objetivo general de la tesis es "proporcionar un modelo para evaluar la calidad y asistir en la mejora de visualizaciones narrativas utilizadas en las interfaces de usuario de sistemas software". Este se divide en objetivos específicos, que abordan los aspectos metodológicos y tecnológicos de la investigación:

OE1: Identificar las buenas prácticas para el diseño de visualizaciones narrativas efectivas. En general, una visualización se considera efectiva si ayuda a las personas a extraer información precisa (Card et al., 1999) sin mayor complejidad (Haroz & Whitney, 2012).

OE2: Construir un modelo de evaluación con diferentes niveles de abstracción y métricas, basado en los patrones y buenas prácticas encontrados en el OE1.

OE3: Desarrollar una herramienta de software como soporte tecnológico del modelo para asistir en la evaluación de visualizaciones y proporcionar recomendaciones de mejora.

OE4: Validar el modelo mediante su implementación y uso en distintos escenarios, incluyendo evaluaciones controladas y aplicaciones en contextos reales.

Estos objetivos buscan mejorar la forma en que se presentan y se entienden las visualizaciones de información en las interfaces de usuario a través de la implementación de técnicas narrativas, integrando conocimientos prácticos en una herramienta que sea útil tanto para investigadores como para profesionales en el campo.

1.5. Contribuciones e impacto

La contribución principal de la tesis es un modelo de evaluación de calidad para visualizaciones de información narrativas que capture la complejidad inherente a este tipo de visualizaciones y permita una evaluación sistemática e integral.

Este modelo está destinado a asistir a investigadores y profesionales de diversas áreas en el proceso de creación de visualizaciones narrativas. Además de ofrecer recomendaciones

Doctorado en Informática

de mejora, permitirá a los equipos de trabajo analizar las buenas prácticas implementadas, identificar oportunidades de optimización, establecer objetivos claros y tomar decisiones de diseño fundamentadas en la literatura. El modelo será operacionalizado mediante una herramienta software, cuyo objetivo es ampliar su alcance y facilitar su aplicación.

Se prevé que la aplicación del modelo de calidad como herramienta de evaluación permita diseños estandarizados y facilite la comunicación y colaboración entre equipos técnicos y no técnicos. La evidencia de una demanda creciente, tanto en la comunidad de investigación (Matzen et al., 2018; Wall et al., 2019) como en la industria (Errey et al., 2024), de enfoques de evaluación más estructurados y rigurosos señala la relevancia del modelo propuesto. Al proveer un mecanismo que facilite la realización de evaluaciones rápidas e iterativas este modelo se sitúa como un aporte significativo y necesario para incorporar prácticas de evaluación sistemática en entornos de diseño y desarrollo de visualizaciones de datos.

1.6. Contexto de la investigación

Esta investigación doctoral se condujo en el marco de los siguientes proyectos y grupos de investigación:

- Proyectos de investigación PI-17F018 "Metodologías y herramientas emergentes para contribuir con la calidad del software", y PI-21F001 "Desarrollo de estudios empíricos en Ingeniería del Software" para los períodos 2018-2021 y 2022-2025 respectivamente, ambos acreditados por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE).
- Beca interna doctoral del Consejo Nacional de Investigaciones Científicas y Técnicas de la Argentina (CONICET) otorgada por la RESOL-2021-154- APN-DIR#CONICET para el período 2021-2025 con lugar de trabajo el Grupo de Investigación en Calidad de Software (GICS) de la UNNE y con estancias de investigación en el Instituto Superior de Ingeniería de Software e Inteligencia Artificial (ISISTAN).

Doctorado en Informática

- Beca ALEARG para investigaciones doctorales de corta duración en Alemania otorgada por la Secretaría de Educación de la República Argentina y el Servicio de Intercambio Alemán a través de IF-2024-31870727-APN-DNCI#ME trabajando en el Grupo de Investigación en Métodos de Desarrollo y Construcción de Software (SWK) de la Universidad de Hamburgo.

1.7. Estructura de la tesis

Capítulo 1. Introducción: este capítulo presenta la motivación de la tesis, el planteamiento del problema, la hipótesis y los objetivos generales y específicos. Además, detalla la contribución principal y la estructura general del documento.

Capítulo 2. Metodología: este capítulo detalla la metodología de investigación seguida, en el marco de la Ciencia del Diseño. Se describen los métodos y técnicas empleados para cada una de las fases.

Capítulo 3. Estado del arte: este capítulo presenta los fundamentos teóricos y antecedentes necesarios para comprender el desarrollo del modelo propuesto.

Capítulo 4. Desarrollo del artefacto: este capítulo describe cómo se construyó la versión final del artefacto propuesto y contribución principal de la tesis: un modelo de calidad para la evaluación de visualizaciones de información narrativas.

Capítulo 5. Validación del artefacto: Este capítulo expone las evaluaciones practicadas sobre la solución propuesta. Las mismas están centradas en validar la viabilidad y utilidad del modelo de calidad.

Capítulo 6. Conclusiones: este capítulo concluye la tesis. Se discute cómo se alcanzaron los objetivos propuestos y se presentan los principales hallazgos. También se abordan las limitaciones y líneas de investigación futuras.

Apéndice A. Estudios primarios del mapeo sistemático: Apéndice con las referencias a los estudios primarios seleccionados en el mapeo sistemático descrito en el Capítulo 1.



Doctorado en Informática

Apéndice B. Aplicaciones del modelo: presenta la lista completa de las 55 aplicaciones iniciales definidas en el modelo NaVi-Q, junto con su descripción, instrumentos asociados y escalas de valoración.

Apéndice C. Materiales: evaluación a pequeña escala: presenta los instrumentos aplicados, hojas de cálculo, scripts y formularios empleados en dicha etapa.

Apéndice D. Materiales: experimento: reúne el conjunto de visualizaciones empleadas en el estudio experimental, junto con sus metadatos y condiciones de evaluación.

Apéndice E. Materiales: estudio de caso: contiene las visualizaciones analizadas, el perfil de calidad utilizado y los instrumentos de recolección de datos aplicados.

Apéndice F. Modelo completo NaVi-Q: Incluye el modelo definitivo propuesto en la tesis, con el detalle de sus criterios, propiedades, aplicaciones evaluativas y perfiles de calidad.

Apéndice G. Manual de usuario NaVi-Q: incluye un manual para la herramienta desarrollada, con instrucciones de instalación, uso y ejemplos de aplicación.

Referencias: este apartado lista las referencias bibliográficas utilizadas a lo largo del documento.

Capítulo 2. Metodología

Este capítulo describe el enfoque metodológico adoptado en el desarrollo de la tesis. La sección 2.1 expone las actividades principales y las tareas realizadas en cada una de ellas, mientras que la sección 2.2 muestra cómo se organizaron dichas tareas a lo largo del proceso.

2.1. Investigación en Ciencia del Diseño

El desarrollo de la tesis se llevó a cabo siguiendo el marco de Investigación en Ciencia del Diseño (DSR, por sus siglas en inglés), un enfoque para la creación de artefactos —modelos, métodos y sistemas— para resolver problemas en un contexto determinado, tomando como referencia a Johannesson & Perjons (2014) y Wieringa (2014).

A diferencia de la investigación empírica, que busca describir, explicar y predecir fenómenos, DSR además aspira a transformar la realidad mediante la creación de soluciones. En este marco, los resultados de la investigación son dobles: el artefacto desarrollado y el conocimiento generado sobre él. Este conocimiento va más allá del artefacto en sí, e incluye una comprensión de su impacto en el entorno. Esta es la diferencia entre el diseño tradicional y la ciencia del diseño: en el primero, basta con crear una solución novedosa para un contexto puntual, mientras que en DSR los resultados deben ser generalizables a otros contextos similares.

Según Johannesson y Perjons (2014), “los investigadores adoptan una postura intencional en el sentido de que ven un artefacto como algo que debería apoyar a las personas en una práctica”. Este enfoque resulta adecuado para el desarrollo de la tesis, ya que el objetivo principal es construir un artefacto —un modelo de evaluación para visualizaciones de

Doctorado en Informática

información narrativas— que contribuya a mejorar una práctica específica: el desarrollo de productos software. Asimismo, se busca generar conocimiento sobre el diseño y uso de este tipo de visualizaciones.

Cabe señalar que muchos proyectos de DSR no abordan todas las actividades por igual, sino que se enfocan en una o varias actividades específicas. En este caso, el enfoque fue en las actividades de diseño, desarrollo y evaluación del artefacto. En línea con las categorías propuestas por Johannesson y Perjons (2014), esta investigación puede caracterizarse como un “proyecto de DSR con enfoque en desarrollo y evaluación”.

Las siguientes secciones describen brevemente cómo se llevó a cabo cada actividad y sus tareas para el desarrollo esta tesis.

2.1.1. Explicar el problema

El objetivo de esta actividad es definir el problema con precisión, justificar su importancia e investigar sus causas. Consiste en el conocimiento descriptivo sobre las características y el entorno del problema.

Como primera tarea, se realizó un estudio de caso (Lezcano Airaldi et al., 2021) siguiendo las pautas de (Yin, 2017) y (Runeson & Höst, 2009) para obtener un conocimiento más profundo del tema. El mismo se definió en términos de las limitaciones actuales en la evaluación de visualizaciones, y la necesidad de las organizaciones y los equipos de desarrollo de adoptar buenas prácticas para el diseño de gráficos efectivos que asistan en el proceso de toma de decisiones. El público potencial del modelo se definió como profesionales, consultores, profesores e investigadores.

La segunda tarea fue establecer el estado del arte, con el fin de definir la base de conocimiento para la investigación. Para ello se llevó a cabo un Estudio de Mapeo Sistemático (SMS, por sus siglas en inglés) siguiendo las pautas propuestas en (Kitchenham & Charters, 2007; Petersen et al., 2015).

Doctorado en Informática**2.1.2. Definir objetivos y requerimientos**

Esta actividad tiene como propósito establecer de manera explícita el objetivo y los requerimientos del artefacto a desarrollar, los cuales pueden abarcar tanto características propias del artefacto como aspectos de su contexto de aplicación.

En primer lugar, se selecciona y describe el tipo del artefacto a diseñar, que puede ser un constructo, un modelo, un método o una instanciación. Los modelos se diferencian de otros artefactos por su capacidad de ser instanciados en sistemas funcionales. Mientras que los constructos y métodos aportan marcos teóricos y procedimientos, los modelos proporcionan una base concreta que puede ser aplicada y evaluada en contextos prácticos.

De acuerdo con esta clasificación, el artefacto construido en la tesis es un modelo prescriptivo. Este modelo describe o explica un fenómeno —en este caso, la calidad de una visualización narrativa— y propone una solución a un problema práctico: su evaluación. Al ser prescriptivo, establece una estructura para abordar dicho problema y puede servir de base para el diseño de nuevos artefactos.

La primera tarea en esta actividad es definir el objetivo del modelo. El mismo se estableció como la evaluación y mejora de la calidad para visualizaciones de información narrativas utilizadas en aplicaciones de software. El modelo busca facilitar la toma de decisiones informadas por parte de los usuarios a partir de este tipo de visualizaciones.

La segunda tarea es la identificación de requerimientos. El primer requerimiento identificado fue que el modelo resultara útil para su público objetivo, enfocándose en aportar valor práctico en su aplicación, en particular, para usuarios no expertos. En segundo lugar, se consideró la necesidad de flexibilidad, de modo que el modelo pudiera adaptarse a distintos contextos de uso. Estos aspectos se tratan en el Capítulo 4, donde se describe el alcance del modelo, incluyendo la audiencia objetivo y los escenarios de uso, y se discute su adaptabilidad a través de perfiles de calidad. El tercer requerimiento fue que el modelo fuera confiable, es decir, que proporcionara resultados consistentes y reproducibles entre distintos usuarios. Los estudios de validación correspondientes se detallan en el Capítulo 5.

Doctorado en Informática

Tanto el objetivo como los requerimientos fueron derivados del análisis de necesidades realizado en la actividad anterior, a partir de los hallazgos obtenidos en el estudio de caso y el mapeo sistemático. Esta etapa fue fundamental para asegurar que el diseño del modelo respondiera a los desafíos identificados y estuviera alineado con las demandas reales del contexto.

2.1.3. Diseñar y desarrollar el artefacto

En esta etapa del marco de DSR se lleva a cabo el diseño y desarrollo del artefacto, con el objetivo de satisfacer los requerimientos establecidos para abordar el problema identificado.

La primera tarea consistió en construir una versión inicial del modelo. Para ello, se realizaron actividades grupales e individuales orientadas a estimular la generación de ideas, lo que permitió consolidar una base conceptual sobre la cual construir el modelo. En este proceso se identificaron y definieron los elementos de calidad relevantes, junto con sus respectivas métricas.

A partir del esquema inicial, el modelo fue refinado progresivamente mediante adaptaciones sucesivas hasta llegar a su versión definitiva. Esta versión integra los elementos de calidad y establece un proceso de evaluación que facilita su aplicación práctica (ver Capítulo 4).

El modelo también fue ajustado en función de comentarios internos del grupo de investigación y de la retroalimentación externa recibida durante la participación en eventos académicos a lo largo del desarrollo de la tesis.

2.1.4. Demostrar el artefacto

Esta fase tiene como objetivo demostrar la viabilidad del artefacto desarrollado mediante su aplicación en un caso real, a modo de prueba de concepto. Este tipo de validación permite verificar que el artefacto puede efectivamente resolver una instancia concreta del problema identificado.

Doctorado en Informática

La primera tarea en esta actividad es elegir o diseñar el caso al cual aplicar el artefacto. Se optó por realizar una Evaluación a Pequeña Escala, siguiendo las pautas de Robson (2017) y Wohlin & Rainer (2022). Este enfoque es apropiado en fases iniciales, ya que permite realizar evaluaciones preliminares con un número reducido de fuentes de datos.

La segunda tarea es aplicar el artefacto. Se seleccionaron usuarios con distintos niveles de experiencia y conocimiento en visualizaciones narrativas, quienes utilizaron el modelo para evaluar un conjunto de visualizaciones. El diseño del estudio, así como el desarrollo del proceso y los resultados obtenidos, se detallan en el Capítulo 5.

2.1.5. Evaluar el artefacto

Esta actividad tiene como propósito determinar en qué medida el artefacto es capaz de resolver el problema identificado y cumplir con los objetivos establecidos. Su finalidad es proporcionar retroalimentación al proceso de construcción y desarrollo, demostrando la utilidad, calidad y eficacia del artefacto propuesto (Hevner et al., 2004). Como resultado, la evaluación contribuye a la generación de conocimiento disciplinar, al ofrecer evidencia sobre el desempeño del artefacto en contextos específicos.

Para evaluar la utilidad y fiabilidad del modelo de calidad propuesto, se llevó a cabo un cuasi experimento. Este tipo de estudio, según Wohlin et al. (2012) y Privitera & Lynn (2018), es apropiado cuando se busca una alta validez externa, es decir, cuando el contexto experimental refleja situaciones reales.

De manera complementaria, se realizó un estudio de caso con el objetivo de analizar la aplicabilidad y relevancia del modelo en un entorno práctico. De acuerdo con Yin (2017), los estudios de caso son adecuados para investigar fenómenos complejos en contextos reales, especialmente cuando se pretende comprender cómo y por qué funciona un artefacto en determinadas condiciones. En este caso, el modelo fue aplicado en el grupo de investigación *Software Engineering and Construction Methods* (SWK), de la Universidad de Hamburgo, en el marco de una estancia académica de investigación. Los resultados y análisis de ambos estudios se presentan en el Capítulo 5.

Doctorado en Informática

2.2. Organización de las actividades

La Fig. 1 presenta la organización general de las tareas realizadas a lo largo de la tesis. Las actividades principales se indican en la parte superior, mientras que las tareas se disponen debajo, alineadas horizontalmente según la actividad a la que pertenecen y verticalmente de acuerdo con su secuencia de ejecución. Al final de cada capítulo se muestra el mismo esquema para sintetizar las tareas tratadas en el capítulo actual, las completadas previamente y las tareas aún pendientes.

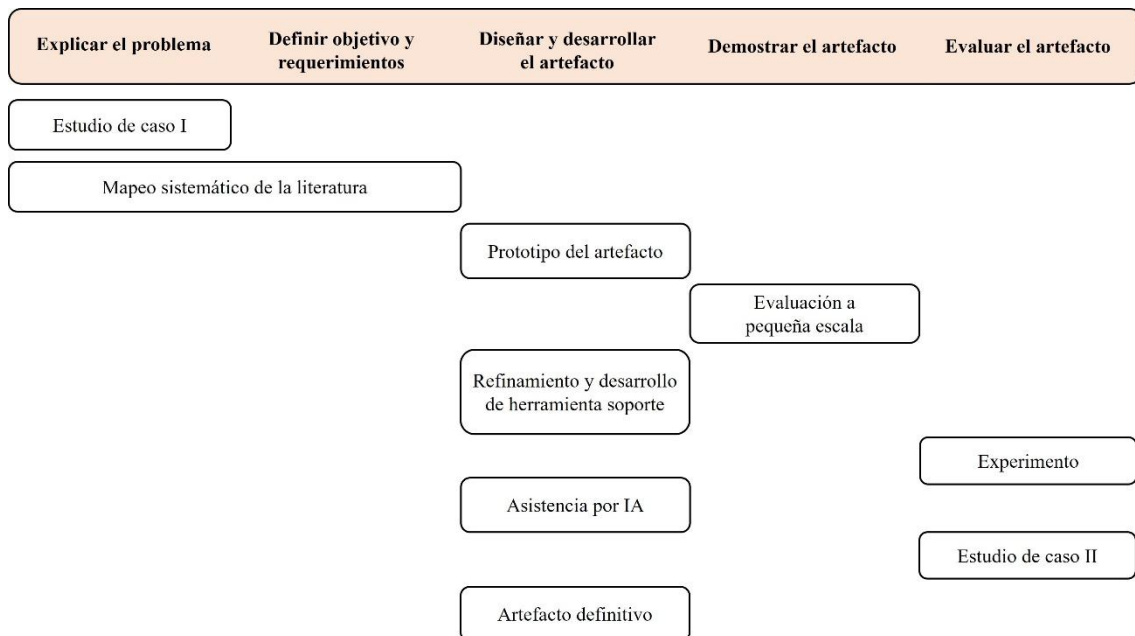


Fig. 1: Organización de actividades y tareas desarrolladas en el marco de DSR

La Tabla 1 resume las secciones del documento en las que se abordan los distintos estudios realizados durante el desarrollo de esta tesis. Su objetivo es facilitar la navegación por el documento y contextualizar las tareas representadas anteriormente.

Doctorado en Informática

Tabla 1: Referencias a los estudios conducidos

Estudio	Referencia
Estudio de Caso I	Sección 1.1 - 1.2, 3.2
Mapeo Sistemático de la Literatura	Sección 3.2 - 3.4
Evaluación a Pequeña Escala	Sección 5.2
Experimento de Validación	Sección 5.4
Estudio de Caso II	Sección 5.6

2.3. Conclusiones

Este capítulo proporcionó el marco metodológico que guía el desarrollo de la tesis, detallando las actividades llevadas a cabo y su organización a lo largo del proceso. En el capítulo siguiente se presentan los fundamentos teóricos y antecedentes necesarios para contextualizar el modelo propuesto.

Capítulo 3. Estado del Arte

Este capítulo introduce los fundamentos teóricos necesarios para el desarrollo de la tesis. La sección 3.1 resume las definiciones sobre narrativa de datos, con el objetivo de unificar su uso y clarificar su aplicación en el contexto de este trabajo. La sección 3.2 describe las principales pautas y recomendaciones para la creación de visualizaciones efectivas. La sección 3.3 examina los criterios existentes para evaluar visualizaciones, mientras que la Sección 3.4 sintetiza las metodologías actuales de evaluación tanto para visualizaciones de información tradicionales como narrativas. Finalmente, la sección 3.5 introduce el concepto de modelos de calidad en Ingeniería de Software y explora su adaptación al ámbito de la visualización de información.

3.1. Narrativa de datos

El concepto de narrativa de datos o *data storytelling*, se presenta bajo una variedad de términos, como “historia con datos”, “visualización narrativa” o “historia dirigida por datos”, entre otros. Cada uno aporta una perspectiva y un enfoque particular. Un elemento común en todas estas definiciones es el uso de estructuras narrativas en la presentación de datos para enriquecer la experiencia del usuario y facilitar la comprensión de información compleja.

Diversos autores coinciden en que la visualización narrativa surge de la convergencia de tres componentes: los datos, la narrativa y los elementos visuales, como se muestra en la Fig. 2. Esta tríada ha sido ampliamente reconocida como base conceptual de la narrativa con datos (Dykes, 2019; Edmond & Bednarz, 2021), y más recientemente retomada y

Doctorado en Informática

ampliada por Z. Zhao & Elmqvist (2023), quienes destacan que los medios empleados en la narración también influyen en la forma en que se estructura y percibe el mensaje.

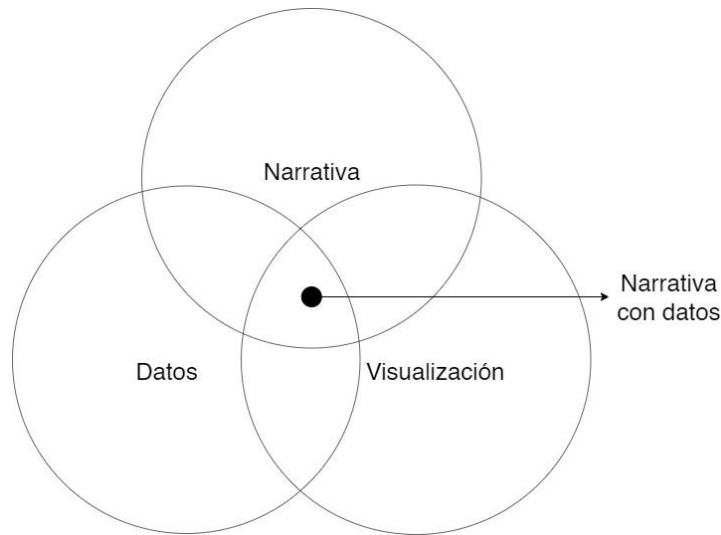


Fig. 2: Ejes conceptuales de la visualización narrativa

Según Kosara y McKinlay (2013), la narrativa de datos consiste en una secuencia ordenada de visualizaciones, complementadas con texto e imágenes, basadas en datos. Por su parte, Riche et al. (2018) conceptualizan las “historias de datos” como narrativas que contienen datos, los cuales están representados gráficamente para clarificar, informar y contextualizar.

Otra definición, presentada por B. Lee et al. (2015) describe la narrativa de datos como una serie de "fragmentos de una historia" respaldados por datos, que se visualizan para apoyar uno o varios mensajes clave. Se destaca el uso de anotaciones, etiquetas y texto para resaltar puntos importantes y evitar ambigüedades, presentando estas piezas en un orden que soporte los objetivos del autor. Además, los autores analizan el proceso de la narrativa de datos e identifican los pasos necesarios en cada fase (ver Fig. 3).

Doctorado en Informática

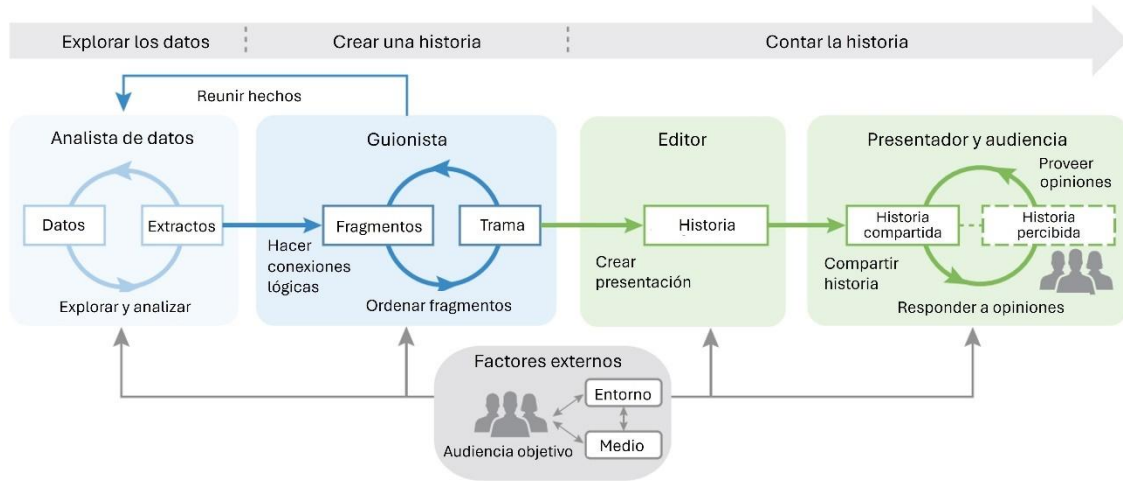


Fig. 3: Proceso de narración con datos según (B. Lee et al., 2015)

La primera fase consiste en explorar y comprender los datos, es decir, identificar los fragmentos que pueden dar forma a una historia, lo cual puede realizarse mediante herramientas de análisis visual. Estos hallazgos se utilizan luego como piezas en la fase de creación que involucra decisiones sobre la selección e interacción de los componentes, dependiendo del género y el medio.

Esta visión ha sido ampliada por Sakarang et al. (2025), quienes describen la visualización narrativa como un proceso iterativo que incluye la definición de requisitos, el diseño, la implementación y la evaluación. Esta perspectiva refuerza la idea de que la visualización narrativa no debe entenderse como un producto estático, sino como un artefacto comunicativo construido de manera progresiva a lo largo de distintas etapas metodológicas.

A partir de estas perspectivas, en esta tesis se adopta el término “visualización narrativa” para referirse a una representación gráfica de datos que combina técnicas de visualización con estructuras narrativas, con el objetivo de comunicar datos de forma comprensible y relevante para el público destinatario (Sakarang et al., 2025).

3.2. Buenas prácticas para la creación de visualizaciones narrativas

A través del SMS (Lezcano Airaldi, Irrazábal, & Diaz-Pace, 2025) se identificaron 21 buenas prácticas para la creación de visualizaciones de información, tradicionales y narrativas. Las mismas se detallan en la Tabla 2 y abarcan una serie de consideraciones, desde la simplificación de conceptos complejos hasta la adaptación del diseño a las necesidades específicas de la audiencia. La columna “ID” asigna una identificación a cada práctica para facilitar su discusión. La columna “Referencias” indica los estudios de los que se derivaron. La lista completa de estudios primarios incluidos en el mapeo puede consultarse en la Tabla A. 1 del apéndice A.

Tabla 2: Buenas prácticas para la creación de visualizaciones narrativas

ID	Buena práctica	Referencias
BP1	Simplificar ideas complejas.	S03, S08, S16, S91, S100, S102, S108, S113
BP2	Proporcionar información contextual.	S23, S41, S42, S52, S85, S97
BP3	Asignar señales visuales a significados implícitos.	S03, S23, S96, S112
BP4	Seleccionar técnicas de visualización apropiadas para los datos y las tareas.	S01, S12, S15, S20, S21, S24, S35, S38, S40, S42, S45, S46, S51, S65, S67, S69, S74, S75, S77, S84, S87, S91, S93, S94, S96, S98, S105, S115
BP5	Usar una línea base común para facilitar comparaciones.	S37, S50, S115
BP6	Mapear la información y las dimensiones de los datos a las características más salientes.	S03, S12, S22, S23, S30, S32, S33, S40, S42, S51, S54, S64, S74, S79, S84, S90, S99, S110
BP7	Proporcionar redundancia para mejorar la comprensión y la memorización.	S16, S57, S69
BP8	Enfocar los puntos de datos más importantes.	S02, S03, S30, S47, S76, S78, S81, S90, S95, S101, S102, S113, S114
BP9	Eliminar el desorden.	S10, S13, S19, S39, S37, S42, S55, S81, S96
BP10	Maximizar la proporción datos-tinta.	S13, S17, S34, S82, S83, S39, S67, S72
BP11	Usar texto, etiquetas y anotaciones.	S02, S19, S23, S25, S49, S50, S57, S60, S92, S23, S100, S109, S113, S115
BP12	Evitar ocultar información.	S03, S29, S34, S43, S50, S56, S83, S88

Doctorado en Informática

ID	Buena práctica	Referencias
BP13	Proporcionar créditos por la procedencia de los datos y transparencia en el diseño.	S03, S08, S23, S80, S91, S96
BP14	Evitar omitir información importante.	S03, S24
BP15	Incorporar sensaciones tangibles o situadas para evocar experiencias.	S23, S34, S43, S56, S57, S81, S102, S107, S116
BP16	Comunicar una narrativa de forma clara.	S02, S03, S08, S23, S62, S100, S102, S107, S116
BP17	Disponer los elementos gráficos y de la historia de forma lógica.	S25, S26, S62, S75, S109
BP18	Mantener la coherencia a lo largo de la historia.	S02, S47, S99
BP19	Incluir técnicas de interacción que permitan la exploración.	S02, S03, S04, S08, S29, S62, S68, S98, S100, S103, S112
BP20	Hacer la información accesible para todos los usuarios.	S23, S25, S78, S95, S111
BP21	Usar el color estratégicamente.	S12, S22, S23, S30, S64, S74, S79, S90, S98, S99, S102, S109

A partir del análisis realizado se pudo observar que las recomendaciones están dispersas en la literatura, siendo algunas poco difundidas. También se evidencia que algunas buenas prácticas presentan recomendaciones contrastantes, creando una dicotomía que resalta la importancia de adoptar un enfoque sensible al contexto al aplicarlas.

Por ejemplo, la BP9 sugiere eliminar elementos innecesarios para reducir el desorden visual, mientras que la BP7 alienta a incluir redundancia para mejorar la comprensión y la memorización. Esto podría implicar repetir ciertos elementos o información, aumentando así la complejidad o la sobrecarga de la visualización, lo cual indica que es necesario encontrar un equilibrio entre la claridad y exhaustividad en la transmisión de información.

Por otro lado, la BP10 que promueve maximizar la proporción datos/tinta —es decir, utilizar la menor cantidad posible de elementos gráficos que no aporten información (Tufte, 2001)—, enfatiza un enfoque minimalista en la representación visual. Esta práctica puede entrar en conflicto con la BP15, que sugiere incorporar elementos visuales que evoquen experiencias tangibles para fortalecer el impacto narrativo.

Los ejemplos mencionados ilustran las complejidades inherentes y las dependencias contextuales al establecer recomendaciones, indicando que la efectividad de una práctica

Doctorado en Informática

puede variar dependiendo de los objetivos específicos, la audiencia y la naturaleza de los datos en cada proyecto de visualización.

Esta necesidad de un enfoque contextualizado fue identificada en las etapas iniciales de esta investigación, mediante un estudio de caso desarrollado durante la pandemia de COVID-19 (Lezcano Airaldi et al., 2021), en el que se analizó el impacto de incorporar técnicas narrativas las visualizaciones de datos. Los resultados mostraron mejoras significativas en la forma en que los usuarios interpretaban y recordaban la información presentada, en comparación con versiones que no aplicaban dichas prácticas. El estudio también demostró que muchas de estas recomendaciones pueden implementarse sin generar demoras significativas, especialmente cuando se consideran desde las etapas iniciales del diseño.

Este caso respalda la importancia de adoptar un enfoque sistemático y ajustado al contexto para la aplicación de buenas prácticas, particularmente en escenarios donde la comprensión rápida y precisa de la información resulta crítica para la toma de decisiones.

3.3. Criterios de evaluación

La evaluación de visualizaciones de información ha sido abordada desde múltiples enfoques, debido a la ausencia de criterios estandarizados. Diversos autores han propuesto criterios centrados en determinados aspectos de calidad. A partir del análisis de la literatura, se identificaron cinco criterios comunes que aparecen de manera recurrente: comprensión, compromiso, información, memorabilidad y usabilidad, los cuales se describen a continuación.

3.3.1. Comprensión

Este criterio, señalado por diversos estudios como el objetivo central de las visualizaciones de datos, busca facilitar la extracción de conocimiento. Para ello, las representaciones visuales deberían aprovechar las capacidades perceptuales humanas. Varios autores han estudiado la comprensión gráfica y la definen como “la habilidad de entender e interpretar un gráfico” (Friel et al., 2001). En este sentido, se ha propuesto un

Doctorado en Informática

marco basado en tres niveles de lectura (Bertin, 2010; Carswell, 1992; Wainer, 1992): el nivel elemental, donde el usuario puede identificar valores específicos; nivel intermedio, donde se reconocen relaciones o tendencias; y nivel avanzado, donde se infieren conclusiones que van más allá de lo que está explícitamente representado.

La comprensión no depende exclusivamente del diseño del gráfico, sino también de cómo este interactúa con el conocimiento previo y los objetivos del usuario (Shah & Hoeffner, 2002). Esta relación está vinculada con el concepto de alfabetización visual, entendido como “la capacidad de extraer, interpretar y dar significado a la información presentada en una visualización” (S. Lee et al., 2017). A medida que el valor y la disponibilidad de los datos continúan creciendo, también aumenta la necesidad de garantizar que los usuarios comprendan lo que están observando.

En el caso de las visualizaciones narrativas, evaluar la comprensión plantea desafíos adicionales. Estas visualizaciones no se limitan a representar datos aislados, sino que articulan una historia que puede ser interpretada de diferentes maneras según el público. La comprensión en este contexto implica no solo identificar valores o patrones, sino también captar el mensaje global, seguir la secuencia narrativa y detectar inferencias implícitas (Errey et al., 2024; Shao et al., 2024).

Ante esta complejidad, algunos estudios han propuesto complementar las métricas tradicionales —como la precisión o el tiempo de respuesta— con evaluaciones cualitativas que indaguen si los usuarios comprenden el argumento narrativo o extraen la conclusión deseada (Fansher et al., 2025). En este tipo de visualizaciones, la coherencia del relato y el entendimiento del mensaje principal se consideran indicadores relevantes de una buena comprensión, especialmente cuando el objetivo es provocar reflexión o modificar percepciones.

3.3.2. Compromiso

Este criterio hace referencia al grado en que un usuario se involucra con la historia presentada mediante la visualización (Bach et al., 2018). Puede manifestarse a través de la atención sostenida, la interacción activa con los elementos visuales o la respuesta

Doctorado en Informática

emocional que provoca. Según Cherubini y Nielsen (2016), si el compromiso se define como el tiempo que el usuario dedica al contenido, la interacción se convierte en un componente relevante para su medición. En contextos donde la información es abundante, captar y mantener la atención representa un desafío (Norman, 2005). Por este motivo, el compromiso resulta un aspecto relevante en la evaluación de visualizaciones narrativas.

Estudios recientes han señalado que los métodos convencionales de evaluación, centrados en métricas como el tiempo de permanencia o la cantidad de clics, no capturan adecuadamente el compromiso emocional o narrativo del usuario (Errey et al., 2024, 2025). Estas métricas cuantitativas pueden reflejar interés superficial, pero no necesariamente indican si el usuario comprendió la historia o experimentó un cambio en su percepción (Fansher et al., 2025). Este tipo de hallazgos sugiere la necesidad de desarrollar indicadores específicos para narrativas, que contemplen factores como la coherencia del relato, la evocación emocional o el impacto actitudinal del contenido.

En este sentido, el compromiso se ha caracterizado como un constructo compuesto, asociado a factores como la estética, la exploración y el control por parte del usuario. En escenarios donde la audiencia no posee conocimientos técnicos específicos, puede facilitar la comprensión de los contenidos.

3.3.3. Información

Este criterio se refiere a la procedencia de los datos, las decisiones metodológicas adoptadas durante su procesamiento y los posibles intereses de quienes los presentan (Diakopoulos, 2018). Incluye aspectos como la fuente original, el grado de precisión y completitud, y la intencionalidad detrás de su recopilación y comunicación.

La transparencia en el uso de datos puede promoverse mediante la citación o el enlace a las fuentes correspondientes (Hullman & Diakopoulos, 2011). Sin embargo, también es necesario considerar si es apropiado utilizar determinados conjuntos de datos, lo cual implica considerar cómo fueron obtenidos y si su uso respeta principios éticos básicos.

Por otro lado, las transformaciones gráficas aplicadas durante el diseño pueden afectar la interpretación de la información. Por ejemplo, en los gráficos de barras suele ser necesario

Doctorado en Informática

comenzar el eje en cero para que la longitud de las barras represente con precisión los valores. En cambio, en gráficos de líneas o de dispersión, un eje que no inicia en cero puede resaltar mejor determinadas variaciones en los datos. En este contexto, Tufte (2001) introduce el concepto de integridad gráfica, que se refiere a la correspondencia entre la representación visual y la magnitud real de los datos. Identificar y evitar técnicas de distorsión se vuelve relevante para garantizar la confiabilidad del mensaje.

En visualizaciones narrativas, donde los datos se combinan con componentes discursivos, estos aspectos adquieren una dimensión adicional (Hullman & Diakopoulos, 2011). Las decisiones de selección, omisión o énfasis pueden influir no solo en la comprensión, sino también en la percepción general del contenido.

3.3.4. Memorabilidad

La memorabilidad se refiere a la capacidad de retener y recordar información (J. Brown et al., 1977). Este criterio está relacionado con funciones cognitivas superiores como la comprensión (Borkin et al., 2013), y ha sido señalado como un aspecto relevante en el diseño de visualizaciones narrativas. En este contexto, no se trata únicamente de que una visualización sea recordada, sino de que los elementos retenidos correspondan a la información clave y no a detalles irrelevantes o erróneos.

Estudios previos han identificado factores que pueden influir en la memorabilidad, como el uso de títulos, colores, disposición gráfica y recursos simbólicos (Borkin et al., 2016). En particular, elementos visuales distintivos o emocionalmente cargados pueden aumentar la probabilidad de que una visualización sea recordada (Lahrache et al., 2018; Zdanovic et al., 2022). Sin embargo, también se ha observado que una mayor memorabilidad no implica necesariamente una mejor comprensión o una evaluación positiva de la visualización.

En el caso de visualizaciones narrativas, la memorabilidad puede estar asociada no solo a elementos gráficos aislados, sino también a la estructura narrativa general y al mensaje que se pretende transmitir. Garretón et al. (2023) demostraron que las visualizaciones embebidas en historias periodísticas pueden generar un mayor impacto emocional y

Doctorado en Informática

favorecer el cambio de actitud del lector, especialmente cuando se integran con una narrativa clara. Asimismo, se ha sugerido que ciertas representaciones visuales, como ilustraciones o metáforas visuales, pueden facilitar la retención de contenido, aunque también pueden distorsionar la memoria si no están alineadas con el mensaje central (Lan et al., 2024).

En este sentido, evaluar la memorabilidad implica considerar tanto la retención de información factual como la persistencia del mensaje narrativo, particularmente cuando las visualizaciones buscan influir en la percepción o la toma de decisiones.

3.3.5. Usabilidad

Este criterio se refiere al grado en que una visualización permite a los usuarios alcanzar sus objetivos con efectividad, eficiencia y satisfacción percibida en un contexto de uso específico (ISO 9241-11, 2018). La usabilidad está influida por múltiples factores, entre ellos la claridad del diseño, la adecuación al público objetivo, la interacción, la accesibilidad y la estética. Si la visualización facilita la tarea del usuario, es más probable que su experiencia sea positiva; de lo contrario, pueden surgir dificultades que afecten tanto la comprensión como el abandono prematuro de la tarea (Norman, 2005; Rezaie et al., 2024).

La literatura ha señalado la importancia de componentes perceptivos como los atributos visuales preatentivos¹ —por ejemplo, el color, la forma o el tamaño— que permiten guiar la atención de manera eficiente sin requerir esfuerzo consciente. Su uso adecuado puede reducir la carga cognitiva en etapas tempranas del procesamiento visual (Barrera-Leon et al., 2023; Healey et al., 1993) y favorecer la localización de patrones relevantes, especialmente en usuarios no expertos.

La interacción y el control del usuario también juegan un papel relevante. Cuando los usuarios enfrentan dificultades —como leyendas confusas, codificaciones ambiguas o relaciones poco claras entre componentes— tienden a desarrollar estrategias exploratorias

¹ Se refieren a las propiedades visuales que el sistema perceptual humano puede procesar de forma rápida y automática, sin necesidad de atención consciente focalizada (Healey et al., 1993).

Doctorado en Informática

(por ejemplo, leer etiquetas, probar combinaciones o utilizar pistas contextuales). El éxito de estas estrategias depende en gran medida del diseño y de su capacidad para ofrecer retroalimentación intuitiva (Rezaie et al., 2024).

La accesibilidad es otro aspecto central de la usabilidad. Estudios recientes han mostrado que usuarios con baja visión experimentan barreras específicas que no siempre son consideradas en los estándares visuales tradicionales (Wang et al., 2024). Estas incluyen problemas de contraste, tipografía o distancia entre elementos, así como la necesidad de diseños flexibles que se integren con tecnologías de asistencia. En este sentido, la usabilidad no solo se relaciona con la facilidad de uso general, sino también con la capacidad de adaptación a distintas condiciones perceptuales y cognitivas.

En conjunto, estos hallazgos refuerzan la idea de que la usabilidad es un criterio multidimensional. Evaluarla implica considerar tanto el diseño gráfico como la interacción, la percepción y el contexto de uso. En visualizaciones narrativas, donde la comprensión del mensaje depende de múltiples factores combinados, la usabilidad permite que el contenido sea interpretado y utilizado de manera efectiva por audiencias diversas.

En conjunto, estos cinco criterios permiten abordar la calidad de las visualizaciones narrativas desde una perspectiva amplia que contempla tanto aspectos cognitivos como perceptuales, éticos y funcionales. La sección siguiente examina los métodos utilizados para evaluar estos aspectos en estudios empíricos y aplicaciones prácticas.

3.4. Metodologías de evaluación existentes

Diversos autores han abordado la falta de herramientas sistemáticas para evaluar visualizaciones, proponiendo técnicas y modelos con distintos enfoques. Entre los primeros intentos se encuentra el modelo Q-VIS (Haase, 1998), que define la calidad en seis categorías ponderadas por su importancia y evaluadas según su grado de cumplimiento. Aunque se trata de una propuesta teórica sin validación empírica, introduce la posibilidad de ajustar el modelo según las prioridades del contexto.

Doctorado en Informática

En (Lezcano Airaldi, Irrazábal, & Diaz-Pace, 2025), se identificaron métodos que van desde cuestionarios de autoevaluación hasta métricas de rendimiento. Entre ellos, el enfoque más frecuente es el uso de heurísticas, una técnica de inspección de usabilidad ampliamente reconocida (Nielsen, 1994). Algunos conjuntos de heurísticas fueron diseñados para abordar aspectos específicos del diseño visual y la toma de decisiones (Barcellos et al., 2018; Dowding & Merrill, 2018; Forsell & Johansson, 2010). Wall et al. (2019), por ejemplo, propusieron un conjunto basado en el marco de valor de Stasko (2014), orientado a estimar el valor potencial de una visualización. Lan et al. (2021) clasificaron sus heurísticas en función de dos dimensiones: usabilidad y expresividad, con el objetivo de analizar cómo diferentes aspectos del diseño influyen en la experiencia del usuario.

Aunque estas herramientas permiten una evaluación rápida, dependen en gran medida del juicio experto, y no siempre abordan todos los aspectos relevantes, como señalan Zhu (2007), Santos et al. (2015) y Williams et al. (2018). En la mayoría de los casos, la adaptabilidad al contexto no está contemplada de manera explícita, aunque podría inferirse a partir de la aplicación selectiva de las heurísticas relevantes.

Además de los enfoques generales basados en heurísticas, se identificaron métodos orientados a criterios específicos. Zhu et al. (2007) propusieron evaluar la complejidad de una visualización, considerando factores asociados a la carga cognitiva. Esta se mide mediante subcriterios como la dificultad de interpretación y la eficiencia en la búsqueda visual, con el objetivo de ayudar a los desarrolladores a seleccionar opciones de diseño que faciliten la comprensión.

De forma complementaria, Padda et al. (2008) plantearon un marco para evaluar la comprensión, basado en tres dimensiones: percepción, cognición y presentación. Sin embargo, este enfoque no establece métricas para la cuantificación, por lo que depende del juicio experto en su aplicación.

En otro estudio, Bai et al. (2009) propusieron la métrica de densidad de inteligencia visual para estimar la cantidad de información útil que ofrece una visualización en relación con la toma de decisiones. Posteriormente, presentaron una taxonomía de visualización “útil”

Doctorado en Informática

(Bai et al., 2010), estructurada en siete áreas: representación visual, presentación de la información, psicología del observador, calidad de los datos, impacto visual, estilo de diseño y rendimiento.

Para abordar el compromiso del usuario, Hung & Parsons (2017) desarrollaron *VisEngage*, un cuestionario de autoevaluación que incluye veintidós ítems distribuidos en once subcriterios. El instrumento utiliza una escala Likert de siete puntos y está diseñado para ser aplicado tras la interacción del usuario con la visualización.

Otros métodos se centran en las habilidades del usuario. Por ejemplo, las herramientas propuestas por Boy et al. (2014) y S. Lee et al. (2017) evalúan el nivel de alfabetización visual mediante cuestionarios que contemplan distintos tipos de visualización y tareas. Estos instrumentos pueden ser utilizados en una amplia variedad de usuarios, independientemente de su experiencia previa con visualizaciones.

En contraste con los métodos anteriores, Wang et al. (2018) propusieron un enfoque comparativo. Su algoritmo asigna un puntaje de consistencia para seleccionar entre gráficos de líneas o de dispersión, según la alineación entre los datos y su representación gráfica. Sin embargo, el método no especifica cómo integra el contexto o las necesidades del usuario en la elección.

Además de las metodologías identificadas en el SMS, se han documentado otras propuestas relevantes. Evergreen (2012) desarrolló un instrumento de evaluación centrado en cinco áreas del diseño de visualizaciones de información. Este se presenta en formato de lista de verificación y emplea una escala cualitativa de tres niveles: completamente cumplido, parcialmente cumplido y no cumplido. El puntaje final se calcula como el porcentaje de ítems evaluados con la categoría más alta. Aunque el método permite obtener una medida cuantitativa general, su aplicación puede verse limitada por el número reducido de ítems y el formato de la escala de medición, lo que afecta la precisión y profundidad de la evaluación.

En el ámbito profesional, Errey et al. (2024) examinaron cómo los equipos de la industria evalúan visualizaciones narrativas en la práctica. El estudio evidenció una ausencia de

Doctorado en Informática

procedimientos sistemáticos y la prevalencia de enfoques informales, basados en la experiencia individual o el consenso entre pares. En respuesta, los autores propusieron un marco heurístico basado en la práctica, estructurado en torno a tres categorías que abordan el diseño visual y la credibilidad de los datos.

Por otra parte, investigaciones recientes han comenzado a explorar el uso de técnicas de inteligencia artificial (IA) para automatizar la evaluación de visualizaciones. La revisión de Wu et al. (2018) distingue entre enfoques basados en reglas, aprendizaje automático y métodos híbridos, e identifica la evaluación como una de las tareas centrales de la inteligencia artificial aplicada a la visualización. En este sentido, en relación con la memorabilidad, Fu et al. (2019) propusieron un enfoque automatizado basado en aprendizaje automático (ML, por las siglas en inglés de *machine learning*). Su modelo combina codificadores variacionales y técnicas de ordenamiento para predecir la memorabilidad de visualizaciones a partir de sus características visuales. Los resultados sugieren que ciertos atributos gráficos pueden asociarse sistemáticamente con mayores niveles de recuerdo, lo que permite identificar elementos visuales que influyen en la retención de información sin recurrir a evaluaciones manuales.

Dibia (2023) propone un sistema que combina generación, evaluación y corrección automática de visualizaciones mediante modelos de lenguaje (LLMs, por las siglas en inglés *large language models*). En este enfoque, se utiliza GPT-4 para puntuar visualizaciones en función de criterios como estética, alineación con el objetivo y calidad de mapeo, lo que permite escalar procesos de evaluación manteniendo una estructura analítica.

De manera complementaria, Lo & Qu (2024) examinaron la capacidad de modelos multimodales para detectar visualizaciones engañosas. Sus experimentos muestran que estos pueden identificar errores comunes y explicar posibles sesgos, aunque su desempeño depende del diseño de las instrucciones utilizadas. Por su parte, Shin et al. (2024) proponen un enfoque que combina mecanismos perceptuales —como contraste de color, distribución espacial y legibilidad— con directrices de diseño estructuradas. Su sistema genera sugerencias accionables que permiten a los diseñadores revisar y mejorar

Doctorado en Informática

sus visualizaciones de forma iterativa. Estas recomendaciones refuerzan varias de las buenas prácticas identificadas en la sección 3.2, como el uso estratégico del color (BP21), el empleo de texto y anotaciones (BP11) y la disposición lógica de los elementos (BP17). Además, los autores introducen una práctica emergente orientada al diseño reflexivo: la integración de mecanismos de retroalimentación durante el proceso de creación. A diferencia de los enfoques tradicionales, que evalúan la visualización una vez finalizada, los LLMs pueden generar sugerencias en tiempo real o en etapas intermedias del diseño, promoviendo de esta manera una revisión continua. Estos avances sugieren que la IA generativa puede ser utilizada como herramienta auxiliar para detectar problemas o generar retroalimentación, especialmente en contextos donde no se cuenta con evaluadores expertos.

La Tabla 3 resume las características de los métodos presentados. La columna Nombre identifica el instrumento y Referencia indica la fuente correspondiente. Año señala el momento de publicación. La columna Criterio describe el aspecto principal de calidad que busca evaluar cada método. Jerarquía indica si el método estructura los atributos de calidad en niveles o subcomponentes. Modelo de medición señala si el enfoque incluye métricas explícitas para cuantificar los resultados. Finalmente, Herramienta clasifica el tipo de instrumento propuesto, como modelos teóricos, cuestionarios, algoritmos o sistemas basados en inteligencia artificial.

Tabla 3: Resumen comparativo de metodologías de evaluación existentes

Referencia Nombre	Año	Criterio	Jerarquía	Modelo de medición	Herramienta
Haase Q-Vis	1998	Resolución de datos, Semántica, Mapeo, Imagen, Presentación, Usuario	Si	Si	Modelo teórico
Zhu et al.	2007	Complejidad	Si	Si	Modelo teórico
Padda et al..	2008	Comprensión	Si	No	Modelo teórico
Bai et al.	2009	Soporte a la toma	Si	Si	Modelo teórico
Bai et al.	2010	Utilidad	Si	No	N/A
Forsell & Johansson	2010	Usabilidad	No	No	Heurísticas

Doctorado en Informática

Referencia Nombre	Año	Criterio	Jerarquía	Modelo de medición	Herramienta
Evergreen	2012	Diseño	Si	Si	Lista de verificación y herramienta soporte
Boy et al.	2014	Comprensión	No	Si	Cuestionario
Hung & Parsons VisEngage	2017	Compromiso	Si	Si	Cuestionario de autoevaluación
S. Lee et al. VLAT	2017	Comprensión	No	Si	Cuestionario
Wang et al.	2018	Consistencia entre datos y gráfico	No	Si	Algoritmo
Dowding & Merrill	2018	Usabilidad	No	Si	Heurísticas
Barcellos et al.	2018	Usabilidad	No	Si	Heurísticas
Wall et al.	2019	Valor	Si	Si	Heurísticas
Fu et al.	2019	Memorabilidad	No	Si	Algoritmo ML
Dibia LIDA	2023	Estética, Mapeo, Alineación objetivo	Sí	Sí	LLM
Lo & Qu	2024	Información, Veracidad	No	No	LLM
Shin et al.	2024	Diseño	No	No	LLM

En conjunto, los métodos revisados permiten identificar buenas prácticas y criterios útiles, pero muestran limitaciones en cuanto a su aplicabilidad práctica y adaptabilidad. En particular, los enfoques automatizados basados en inteligencia artificial ofrecen herramientas para apoyar ciertos aspectos del proceso evaluativo, como la detección de errores comunes o la generación de retroalimentación inicial. Sin embargo, su uso actual debe entenderse como un complemento a la evaluación humana, no como un reemplazo, especialmente en contextos donde el juicio experto y la sensibilidad al contexto son necesarios.

Esto refuerza la necesidad de un marco integral y flexible para evaluar visualizaciones narrativas, orientado a optimizar su diseño en función del público objetivo. En la sección siguiente se exploran modelos de calidad desarrollados en Ingeniería de Software (IS), con el fin de analizar en qué medida pueden informar la construcción de un modelo específico para visualizaciones narrativas.

3.5. Modelos de calidad en Ingeniería de Software

En el ámbito de la IS, la evaluación de la calidad ha sido abordada desde distintas perspectivas. Entre ellas, el estándar ISO/IEC 25010 (2023) define un modelo de calidad como un conjunto de características interrelacionadas que proporciona la base para especificar requisitos y evaluar su cumplimiento. La idea de descomponer la calidad en componentes evaluables se remonta al modelo de Boehm (1978) y ha sido retomada por diversas propuestas a lo largo del tiempo.

Este modelo organiza la calidad en una jerarquía de propiedades, acompañada de pautas para su cuantificación y evaluación. Sin embargo, este tipo de modelos no cubre todos los aspectos relevantes para diferentes contextos. Un estudio de Wagner et al. (2009) muestra que solo una parte de las organizaciones adopta modelos estandarizados, mientras que muchas optan por adaptar o desarrollar modelos propios, lo que sugiere una necesidad de personalización.

Con el objetivo de reducir la brecha entre aspectos abstractos y mediciones concretas, Wagner et al. (2015) propusieron el modelo Quamoco. Más recientemente, Siebert et al. (2022) desarrollaron un modelo de calidad para sistemas de aprendizaje automático, que también adopta una estructura sistemática y jerárquica.

En el caso de las visualizaciones narrativas, estos modelos no pueden aplicarse de manera directa. Las particularidades del medio visual y de los objetivos narrativos exigen una adaptación tanto en la definición del objeto de evaluación como en la selección de atributos pertinentes. Es necesario identificar qué elementos caracterizan una visualización narrativa de calidad y cómo pueden medirse.

En Visualización de Información, distintos trabajos han destacado la necesidad de equilibrar claridad visual con precisión informativa. Card et al. (1999) señalan la importancia de una presentación intuitiva que facilite la interpretación, mientras que Zhu (2007) y Bai et al. (2010) destacan criterios como utilidad, eficiencia, relevancia y capacidad de acción.

Doctorado en Informática

En este contexto, adaptar marcos provenientes de la IS puede ofrecer una base estructurada para definir modelos de evaluación aplicables a visualizaciones narrativas. Esta adaptación permite incorporar criterios generales de calidad junto con consideraciones específicas del medio visual y su función comunicativa.

3.6. Conclusiones

Este capítulo presentó los fundamentos conceptuales que sustentan el desarrollo del modelo propuesto. Se describieron distintas perspectivas sobre la narrativa de datos y las características específicas de las visualizaciones narrativas, así como buenas prácticas y criterios y metodologías existentes para su evaluación.

A partir del análisis de la literatura, se evidenció la necesidad de contar con herramientas flexibles y adaptables al contexto, que permitan evaluar múltiples dimensiones de calidad de forma sistemática. Si bien existen enfoques parciales y modelos consolidados en otros dominios, su aplicación en el caso de visualizaciones narrativas requiere una reformulación que considere tanto sus aspectos visuales como narrativos.

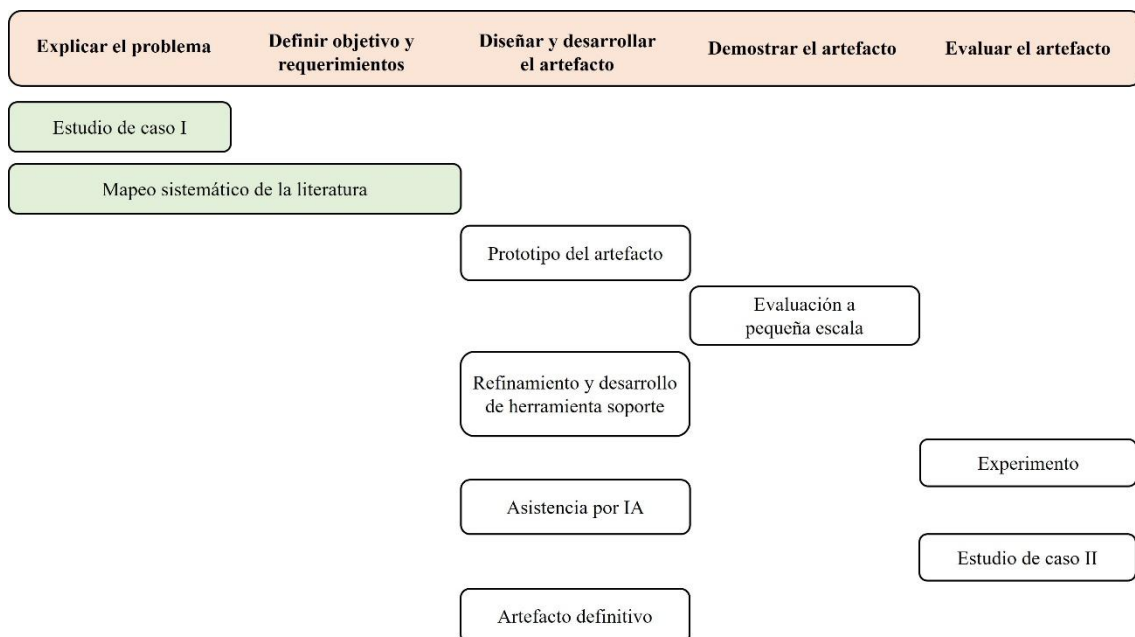


Fig. 4: Actividades abarcadas en el capítulo



Doctorado en Informática

El siguiente capítulo describe el proceso de construcción del modelo de calidad propuesto. La Fig. 4 sintetiza las tareas abordadas hasta el momento en el marco de DSR, destacadas en verde, y aquellas que serán desarrolladas en los capítulos siguientes, indicadas en gris.

Capítulo 4. Desarrollo del Artefacto

Este capítulo presenta la construcción del artefacto propuesto y la contribución principal de la tesis: un modelo de calidad para la evaluación y mejora de visualizaciones de información narrativas. La sección 4.1 introduce el modelo, describe sus fundamentos conceptuales y resume el proceso seguido para su desarrollo. Las secciones siguientes abordan en detalle cada uno de los pasos definidos.

4.1. Modelo propuesto: NaVi-Q

Para construir un modelo de calidad es necesario definir los escenarios de uso: quiénes lo utilizarán, con qué propósitos y en qué contextos. También es necesario identificar qué propiedades de calidad se evaluarán y cómo se medirán. El modelo propuesto, denominado NaVi-Q (*Narrative Visualization Quality*, por sus siglas en inglés) está basado en estándares y modelos consolidados en IS, principalmente ISO/IEC 25010 (2023), ISO/IEC 33003 (2015), y trabajos como Wagner et al. (2015) y Siebert et al. (2022). Estos enfoques ofrecen marcos jerárquicos y sistemáticos para representar atributos de calidad y vincularlos con métodos de evaluación.

El proceso de construcción del modelo está compuesto por seis pasos, cuyas actividades se desarrollan de manera iterativa. Estos pasos fueron adaptados de Siebert et al. (2022) y reformulados según las necesidades específicas del dominio de la visualización narrativa.

- 1) **Definir el alcance y el contexto de aplicación:** Se definieron los escenarios de aplicación del modelo, incluyendo los tipos de visualización a evaluar y los perfiles de usuario previstos.

Doctorado en Informática

- 2) **Diseñar el metamodelo de calidad:** Se estableció la estructura general del modelo, incluyendo los niveles jerárquicos y los componentes necesarios para representar propiedades y medidas de calidad.
- 3) **Identificar los requerimientos de calidad:** A partir del SMS y otras fuentes analizadas en el capítulo anterior, se incorporaron los requerimientos que debe cumplir una visualización narrativa efectiva.
- 4) **Definir las propiedades de calidad relevantes:** Se definieron los criterios específicos, propiedades e interrelaciones, estructurándolos en función de las buenas prácticas identificadas en la literatura, y alineándolos con los requerimientos previamente establecidos.
- 5) **Establecer las reglas de evaluación:** Se definió el método de evaluación, especificando cómo se interpretan y combinan los resultados de la medición.
- 6) **Instanciar el modelo para un caso de uso:** Finalmente, el modelo fue instanciado en un escenario concreto, ajustando la ponderación de los elementos para reflejar las prioridades específicas del contexto.

4.2. Definir el alcance y el contexto de aplicación

La norma ISO/IEC 25010 (2023) distingue entre cuatro dimensiones de calidad: calidad de proceso, propiedades internas, propiedades externas y calidad en uso. Cada una se asocia con diferentes formas de medición y evaluación (ver. Fig. 5).

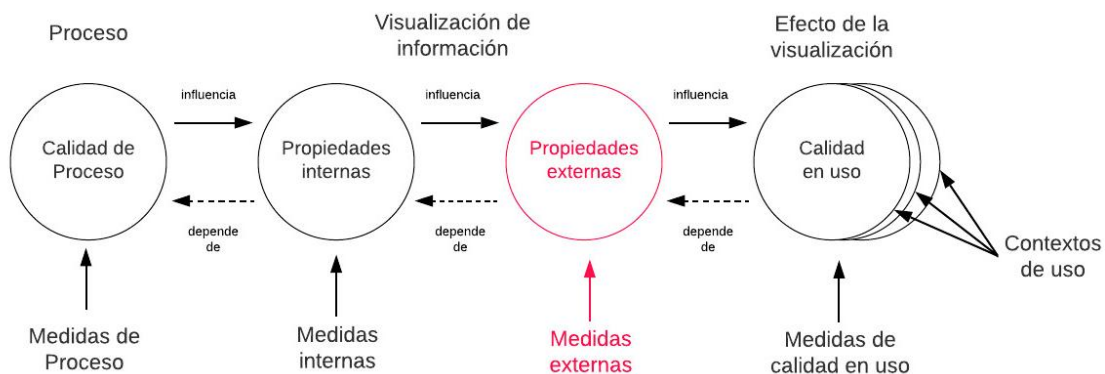


Fig. 5: Foco del modelo dentro del ciclo de vida de la calidad

Doctorado en Informática

En el contexto de visualización narrativa, la calidad de proceso se refiere a las actividades llevadas a cabo durante el diseño y desarrollo de la visualización. Puede incluir prácticas como revisiones iterativas o pruebas con usuarios. Las propiedades internas comprenden características técnicas no observables directamente, como la organización del código, la modularidad del sistema o la estructura de los datos. La calidad en uso se relaciona con los resultados obtenidos al utilizar la visualización, por ejemplo, su efecto sobre la comprensión, la toma de decisiones o el aprendizaje en un público determinado.

El modelo propuesto se enfoca en las propiedades de calidad externas, es decir, aquellas características observables en la visualización como artefacto final. Estas propiedades comprenden aspectos como la estructura narrativa, la representación de los datos y la interacción visual, sin abordar su implementación técnica ni su impacto en el entorno.

4.2.1. Técnicas de visualización consideradas

En el campo de la VI, existen diversas técnicas con distintos niveles de complejidad, que van desde visualizaciones simples, como gráficos de líneas, hasta otras más complejas, como coordenadas paralelas (Johansson & Forsell, 2016) o gráficos de barras radiales (Waldner et al., 2019).

Börner et al. (2016) señalaron que muchas personas presentan dificultades para interpretar visualizaciones con las que no están familiarizadas a través de medios, libros o sitios web. Estudios como S. Lee et al. (2017) y Battle et al. (2017) identificaron que los gráficos de líneas, barras, dispersión y pastel se encuentran entre los más frecuentes en medios digitales. Teniendo en cuenta estos hallazgos, el modelo fue diseñado para respaldar la evaluación de técnicas comunes de visualización, con el objetivo de equilibrar exhaustividad y aplicabilidad, reconociendo que el componente narrativo es transversal al tipo de gráfico. De esta manera, se espera que el modelo pueda ser aplicado por distintos perfiles de usuarios, incluyendo aquellos sin experiencia previa con representaciones visuales complejas.

En cuanto al formato, el modelo se aplica a visualizaciones tipo documento, ya sean estáticas o interactivas, como gráficos individuales o paneles de control. Quedan fuera

Doctorado en Informática

del alcance otros estilos como los videos animados o cómics, ya que presentan características particulares que requieren criterios de evaluación distintos (J. Zhao et al., 2021). Por ejemplo, los videos animados incluyen transiciones y dinámicas temporales que no son relevantes en visualizaciones estáticas, mientras que los data comics integran elementos narrativos y gráficos secuenciales que exigen consideraciones adicionales para su evaluación.

4.3. Diseñar el metamodelo de calidad

El elemento central del modelo es una **propiedad**; es decir, un atributo medible que caracteriza una visualización e indica su calidad. Como señala Wagner et al. (2015), el concepto de propiedad es general y se puede aplicar en diferentes niveles de abstracción.

Se diferencian dos tipos de propiedades: **criterios de calidad** y **propiedades de visualización**. Los criterios de calidad expresan objetivos de calidad abstractos, como la memorabilidad o la comprensión. Las propiedades de visualización, a su vez, se refieren a atributos observables de componentes específicos de visualización. Esto permite describir la calidad desde un nivel abstracto hasta mediciones concretas.

Cada propiedad de visualización se vincula con una o más **aplicaciones**, entendidas como descripciones concretas sobre cómo cuantificar la propiedad en un contexto determinado. Por ejemplo, para la propiedad “integridad”, las aplicaciones pueden incluir la atribución de fuentes o el uso de nombres de variables no ambiguos.

Cada aplicación cuenta con un **instrumento**, que describe cómo se implementa la aplicación. Este instrumento incluye opciones con valores asociados según su adecuación respecto a la propiedad evaluada. Se asigna un valor de 1 a la opción que representa el mayor grado de cumplimiento, y un valor de 0 a las opciones consideradas inadecuadas. En algunos casos, puede seleccionarse la opción *N/A* si la aplicación no es pertinente para una visualización específica.

Doctorado en Informática

Por ejemplo, el instrumento para la aplicación de “atribución de fuentes” consiste en verificar si las fuentes de datos están citadas o enlazadas. Si están presentes, se asigna un valor de 1; si no están presentes, se asigna 0.

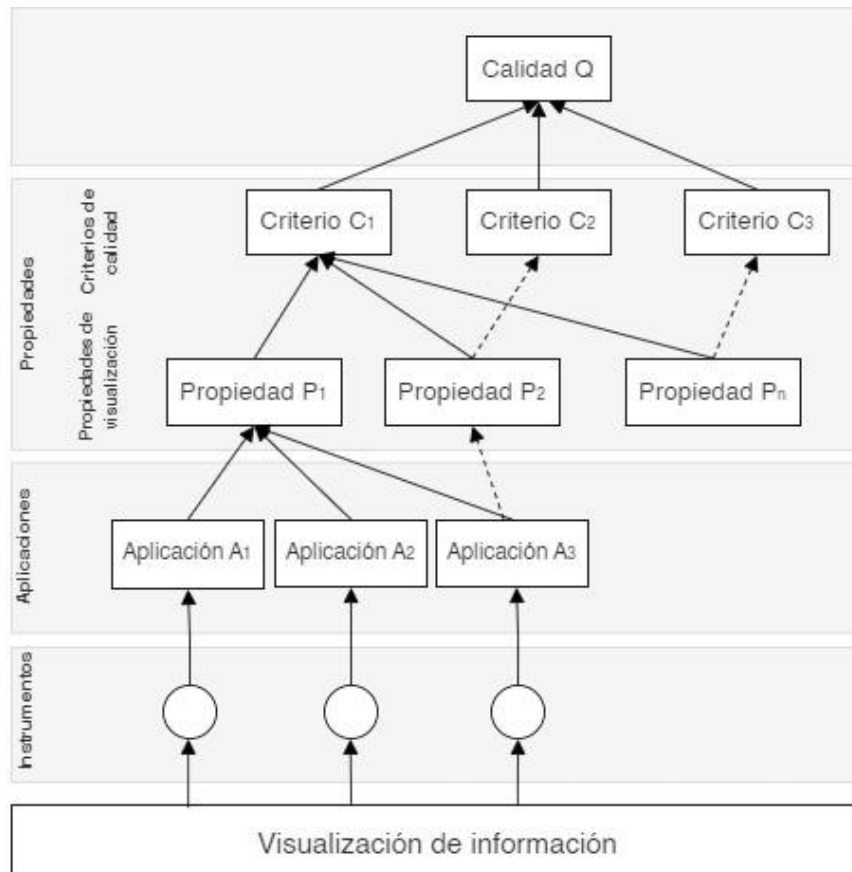


Fig. 6: Elementos y estructura del metamodelo

Cada elemento del modelo (criterio, propiedad, aplicación) tiene asignado un **peso** que indica su importancia relativa en el proceso de evaluación. Estos pesos permiten ajustar la contribución de cada componente en función del caso de uso. La Fig. 6 presenta la estructura jerárquica del metamodelo de calidad.

4.4. Identificar los requerimientos de calidad

Los requerimientos de calidad que orientan el modelo fueron identificados mediante el SMS presentado en el Capítulo 1. Este análisis abarcó 116 estudios primarios en visualización de información y narrativa de datos, y permitió reunir un conjunto de buenas

Doctorado en Informática

prácticas (ver Tabla 2) y criterios utilizados en la evaluación de visualizaciones (descritos en la sección 3.3).

Estos elementos sirvieron como base para establecer los objetivos del modelo y derivar un conjunto estructurado de propiedades, descritas en la siguiente sección.

4.5. Definir las propiedades de calidad relevantes

Esta sección describe cómo los criterios de calidad definidos anteriormente se operacionalizan mediante propiedades observables y aplicaciones concretas. Mientras que los criterios representan objetivos generales de evaluación, las propiedades permiten descomponer dichos objetivos en dimensiones específicas, y las aplicaciones proporcionan mecanismos tangibles para su medición.

4.5.1. Criterios

Los cinco criterios presentados en la sección 3.3 fueron seleccionados como objetivos centrales del modelo: comprensión, compromiso, información, memoria y usabilidad. Estos criterios permiten organizar las propiedades de calidad en torno a diferentes metas evaluativas.

4.5.2. Propiedades

Cada criterio se descompone en propiedades específicas que describen características evaluables de la visualización narrativa. Estas propiedades surgen de vincular las buenas prácticas recopiladas en el SMS con los objetivos de cada criterio, transformando lineamientos orientados a la acción en atributos estructurados que pueden ser observados y medidos.

Por ejemplo, la práctica de “usar una línea base común” se asocia a la propiedad “precisión”, mientras que “proporcionar información contextual” se vincula con la propiedad “contexto”. De manera similar, la recomendación de “eliminar elementos innecesarios” se refleja en la propiedad “claridad”. La Tabla 4 presenta el listado completo de propiedades y su descripción.

Doctorado en Informática

Tabla 4: Propiedades de calidad

ID	Propiedad	Descripción
P01	Precisión	Uso de escalas adecuadas para representar los datos, evitando distorsiones o alteraciones que puedan inducir a error.
P02	Saliencia	Asociación de la información relevante con elementos visuales destacados.
P03	Color	Uso estratégico del color, considerando su cantidad y función.
P04	Disposición	Organización de los elementos visuales, como proporción, contraste y espacios en blanco.
P05	Orden visual	Ausencia de elementos visuales superfluos que no contribuyen al mensaje y pueden dificultar la interpretación.
P06	Enfoque	Uso de señales visuales (por ejemplo, color o contraste) para resaltar patrones o datos importantes.
P07	Accesibilidad	Facilidad para diferentes tipos de usuarios, con distintos niveles de experiencia o capacidades.
P08	Fiabilidad	Grado de transparencia y documentación de la fuente y calidad de los datos.
P09	Redundancia	Representación de la misma información mediante múltiples recursos, facilitando la comprensión sin generar ruido.
P10	Integridad	Representación completa y veraz de la información, sin omisiones ni distorsiones.
P11	Claridad	Capacidad de transmitir ideas complejas de forma concisa y sin elementos superfluos.
P12	Contexto	Inclusión de información de fondo relevante para facilitar la comprensión.
P13	Narrativa	Uso de técnicas de narración para construir un mensaje estructurado y coherente.
P14	Guía para el usuario	Claridad en la navegación e interpretación, especialmente en visualizaciones complejas o con múltiples paneles.
P15	Metonimia	Uso de elementos visuales que representen datos de forma intuitiva, en línea con construcciones mentales comunes.
P16	Consistencia	Uniformidad en el uso de colores, tipografías e iconografía.
P17	Interactividad	Capacidad de interacción del usuario con los datos, como zoom, filtros o navegación.

Doctorado en Informática

4.5.3. Aplicaciones

Por último, se identificaron 55 aplicaciones que forman la base de la evaluación. La Tabla 5 presenta un fragmento de las mismas. Como se detalla en la sección 4.3, cada aplicación cuenta con un instrumento que ofrece una serie de opciones para registrar datos de evaluación. La columna ID indica su identificador único; Aplicación, el nombre asignado; Descripción, una explicación breve de su finalidad; Opciones, las alternativas disponibles para su valoración; y Valor, el puntaje asociado a cada opción. La opción "N/A" permite marcar aquellas aplicaciones que no son relevantes en determinados casos. La lista completa se puede consultar en el Apéndice A.

Tabla 5: Aplicaciones (fragmento)

ID	Aplicación	Descripción	Opciones	Valor
AP03	Títulos de ejes	Los ejes deben tener una nomenclatura apropiada.	Títulos de ejes descriptivos	1
			Títulos de ejes genéricos	0.5
			Sin títulos de ejes	0
AP10	Etiquetas de datos	Los datos deben estar etiquetados directamente	Los datos están etiquetados directamente.	1
			Los datos están etiquetados por leyendas.	0.5
			Los datos están etiquetados de ambas maneras.	0.25
			Los datos no están etiquetados.	0
AP19	Colores limitados	La cantidad de colores debe estar dentro de los límites de la memoria.	Hasta 2 colores	1
			Entre 3 y 5 colores	0.5
			Más de 5 colores	0
AP23	Mensaje principal	Los mensajes principales deben hacerse explícitos en los títulos y subtítulos.	El título contiene el mensaje principal	1
			El título no contiene el mensaje principal	0
			N / A	
AP32	Líneas de grilla	Las líneas de grilla deben ser útiles y discretas.	No se utilizan líneas de grilla	1
			Se utilizan líneas de grilla suaves	0.5

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
			Se utilizan líneas de grilla moderadas	0.2
			Se utilizan líneas de grilla oscuras	0
			N/A	
AP46	Mensajes y resúmenes	Se deben utilizar elementos de texto para resumir y explicar la información.	Se cumple	1
			Se cumple parcialmente	0.5
			No se cumple	0
			N/A	
AP47	Ilustraciones	Las ilustraciones, si están presentes, deben ser relevantes para el tema.	Las ilustraciones están relacionadas con los conceptos.	1
			Las ilustraciones son decorativas.	0.5
			Las ilustraciones son irrelevantes o confusas.	0
			N/A	

Las opciones fueron definidas a partir de la literatura y estructuradas para facilitar una evaluación precisa y tangible. En situaciones donde la evaluación pueda resultar compleja o poco práctica, se incorpora una escala unidimensional con tres niveles: “cumplida”, “parcialmente cumplida” y “no cumplida”. Esta escala busca facilitar la cuantificación de la aplicación.

4.5.4. Relación entre criterios, propiedades y aplicaciones

A continuación, se presentan la relación entre los distintos niveles del modelo de calidad. La Tabla 6 agrupa las propiedades bajo cada uno de los cinco criterios definidos, mientras que la Tabla 7 muestra las aplicaciones asociadas a cada propiedad. Esta estructura permite visualizar cómo se descompone cada objetivo general en elementos evaluables de menor nivel.

Doctorado en Informática

Tabla 6: Criterios de calidad y propiedades

ID	Descripción	Propiedades asociadas
C01	Comprensión	P01; P02; P03; P05; P06; P07; P11; P12; P13
C02	Compromiso	P03; P05; P11; P12; P13; P14; P16; P17
C03	Información	P01; P04; P06; P08; P09; P10; P12
C04	Memorabilidad	P03; P06; P09; P11; P12; P13; P14; P15
C05	Usabilidad	P02; P04; P07; P08; P11; P14; P16; P17

Tabla 7: Aplicaciones de cada propiedad

ID	Descripción	Aplicaciones asociadas
P01	Precisión	AP01; AP03; AP04; AP05; AP08
P02	Prominencia	AP14; AP16; AP20
P03	Color	AP18; AP19; AP20; AP21; AP22
P04	Disposición	AP08; AP28; AP29; AP35; AP36; AP37; AP38; AP39
P05	Desorden visual	AP06; AP19; AP28; AP29; AP32; AP33; AP34; AP35; AP36; AP37; AP38; AP39
P06	Enfoque	AP16; AP20; AP34
P07	Accesibilidad	AP10; AP15; AP21; AP22; AP23; AP26; AP27; AP29; AP31; AP35; AP36; AP37; AP38; AP39; AP55
P08	Fiabilidad	AP07; AP09; AP10; AP11; AP12; AP13; AP40; AP41; AP42; AP43; AP44
P09	Redundancia	AP03; AP10; AP15; AP16; AP21; AP23; AP30; AP46; AP47
P10	Integridad	AP02; AP07; AP33; AP40; AP42; AP43; AP44
P11	Claridad	AP06; AP07; AP17; AP23; AP24; AP35; AP36; AP37; AP38; AP39; AP46; AP47
P12	Contexto	AP03; AP23; AP25; AP30; AP44; AP46; AP47; AP50
P13	Narrativa	AP16; AP21; AP23; AP25; AP30; AP45; AP46; AP47; AP50; AP51; AP52; AP54

Doctorado en Informática

ID	Descripción	Aplicaciones asociadas
P14	Guía al usuario	AP11; AP30; AP34; AP45; AP46; AP48; AP49; AP50; AP51; AP54
P15	Metonimia	AP15; AP16; AP21; AP47
P16	Consistencia	AP14; AP16; AP17; AP18; AP19
P17	Interactividad	AP48; AP49; AP52; AP53; AP54

Como primera aproximación, los pesos de las aplicaciones dentro de una propiedad se asignan de forma equitativa con el fin de obtener un perfil de calidad inicial, genérico y ajustable. Por ejemplo, si una propiedad incluye tres aplicaciones, cada una recibe un peso de 1/3. En los casos en que una o más aplicaciones no resulten aplicables a una visualización específica, sus pesos se redistribuyen entre las aplicaciones restantes. Esta estrategia permite mantener una evaluación balanceada y coherente, al tiempo que evita sesgos en la valoración de las propiedades. Tal como se abordará más adelante, este esquema de ponderación puede ser adaptado para reflejar mejor las prioridades o características particulares de contextos específicos.

4.6. Definir las reglas de evaluación

Cada elemento del modelo debe asociarse con un enfoque para sintetizar e interpretar los datos de medición recopilados de la visualización. Mientras que el metamodelo define qué aspectos o características de la visualización se deben evaluar y cómo se relacionan entre sí, el método de evaluación determina cómo se lleva a cabo esa evaluación en la práctica. En los siguientes apartados, se detalla la metodología de evaluación adoptada para este modelo.

4.6.1. Calidad como constructo formativo

Las reglas de evaluación del modelo propuesto se basan en los lineamientos de la norma ISO/IEC 25010 (2023), que establece la relación entre un modelo de calidad y su método de evaluación, y en la familia de normas ISO/IEC 33000, que describe prácticas para el diseño de modelos de medición.

Doctorado en Informática

La calidad, entendida como un constructo multidimensional, puede representarse mediante diferentes tipos de modelos (Bagozzi, 2011; Diamantopoulos et al., 2008; Diamantopoulos & Winklhofer, 2001; Edwards & Bagozzi, 2000). La norma ISO/IEC 33003 (2015) distingue entre modelos de medición reflectivos y formativos. En los modelos reflectivos, el constructo es la causa de las medidas observadas, mientras que en los formativos, el constructo se define como una combinación de dichas medidas.

En este trabajo se adopta un modelo formativo dado que la calidad de una visualización narrativa se construye en base a los criterios de calidad y sus propiedades correspondientes, en lugar de causarlos. Esto permite una evaluación más precisa de cómo cada criterio impacta en la calidad general de una visualización, conduciendo a resultados informativo y accionables. Este tipo de modelo se utiliza cuando la investigación implica el desarrollo de un constructo que no está bien definido o comprendido (Diamantopoulos et al., 2008), como es el caso de la calidad de una visualización.

La Fig. 7 muestra la estructura del modelo de medición formativa, donde los constructos se representan como óvalos, las medidas observadas como rectángulos, las relaciones causales como flechas simples, y las correlaciones como flechas dobles.

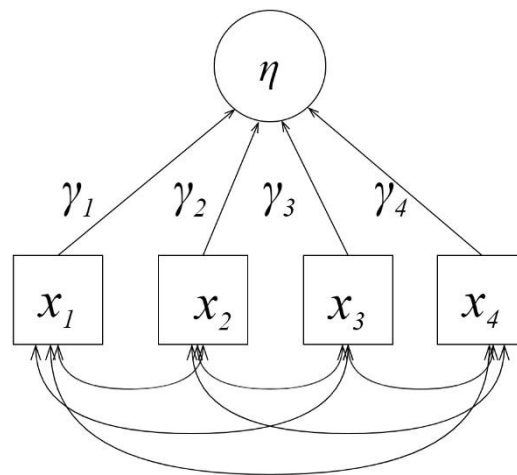


Fig. 7: Relación entre un constructo formativo y sus medidas (ISO/IEC 33003)

Este enfoque se modela como una combinación ponderada de indicadores formativos:

$$C = \gamma_1 x_1 + \dots + \gamma_q x_q$$

Doctorado en Informática

donde C es la calidad estimada, x_i las medidas formativas, y γ_i los coeficientes que indican la contribución relativa de cada medida.

Este modelo se alinea con las técnicas de decisión multiatributo (MADM, por sus siglas en inglés), en las cuales una medida compuesta se construye a partir de atributos ponderados (Bollen & Ting, 2000; Yoon & Hwang, 2011; Zeleny, 1982). En particular, se adopta un enfoque compensatorio, que permite que las fortalezas en un criterio compensen debilidades en otro (Hwang & Yoon, 1981). Esto resulta adecuado para la definición de perfiles de calidad adaptados a diferentes contextos, ya que permite a los evaluadores establecer compensaciones entre criterios de calidad.

Entre los métodos compensatorios posibles, los más adecuados para modelos formativos son el método aditivo simple (SAW, por sus siglas en inglés) y el producto ponderado (WP) (Jung, 2013; Yoon & Hwang, 2011). Ambos permiten calcular puntuaciones compuestas a partir de los pesos asignados a los elementos del modelo.

4.6.2. Asignación de pesos con Proceso Analítico Jerárquico

El Proceso Analítico Jerárquico (AHP, por sus siglas en inglés) (Saaty, 1980) es un marco de toma de decisiones que puede utilizarse para asignar pesos relativos en métodos de agregación como SAW o WP. Este enfoque consiste en descomponer un problema en una jerarquía de criterios y alternativas, seguido de una serie de comparaciones por pares para determinar su prioridad relativa.

Uno de los principales desafíos del AHP tradicional es mantener la tasa de consistencia dentro de los umbrales aceptables (Saaty, 1980). Esta dificultad no siempre depende del conocimiento del evaluador, como señalan Tavana et al. (2021), sino que puede surgir del número y complejidad de los elementos a comparar. A medida que se incrementa el número de criterios, el proceso se vuelve más lento debido a la cantidad de comparaciones requeridas, del orden $n(n - 1)/2$ para n elementos.

Para abordar estas limitaciones, se han propuesto diversas variantes del método, como AHP difuso (Aliyev et al., 2020; Helmy et al., 2021), que incorpora grados de

Doctorado en Informática

incertidumbre, o AHP por Votos (Liu & Hai, 2005), que combina múltiples juicios individuales mediante técnicas de agregación de preferencias.

Entre estas alternativas, AHP-Express (Leal, 2020) ofrece una simplificación del método clásico. Esta versión reduce el número de comparaciones necesarias a $(n - 1)$, manteniendo la lógica jerárquica y la estimación relativa de pesos, pero con una carga cognitiva menor. Esta característica lo hace especialmente útil para contextos donde se requiere eficiencia en la toma de decisiones sin comprometer la trazabilidad del razonamiento.

La principal ventaja de AHP Express es que requiere una sola comparación por cada criterio. El proceso consta de tres pasos: (a) realizar una única comparación de alternativas para cada criterio, (b) calcular el valor recíproco y su suma, y (c) aplicar una normalización. Luego, los valores se combinan en una matriz y se sigue el mismo procedimiento de toma de decisiones que en el AHP clásico.

La Fig. 8 muestra una comparación entre una jerarquía estándar en AHP (izquierda) y una vista parcial del modelo de calidad propuesto (derecha).

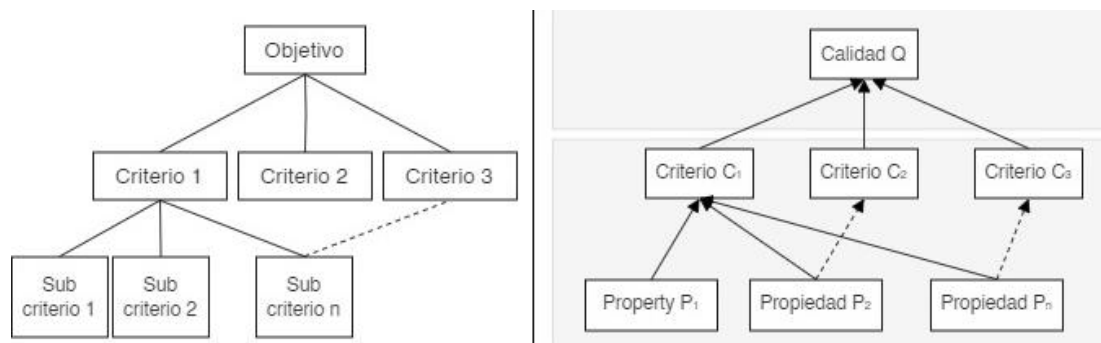


Fig. 8: Comparación entre una jerarquía estándar en AHP (a la izquierda) y un extracto del modelo de calidad propuesto (a la derecha)

Como puede observarse, ambas estructuras comparten una base conceptual similar. En el modelo de calidad, los criterios representan aspectos abstractos de la calidad, mientras que las propiedades corresponden a atributos concretos y medibles. De manera análoga, en AHP los criterios de nivel superior se descomponen en subcriterios más específicos.

Doctorado en Informática

Esta similitud refuerza la elección de AHP como herramienta para derivar pesos dentro del modelo de calidad.

4.6.3. Definición de escala de medición

Para medir el grado de cumplimiento de las propiedades de calidad, es necesario seleccionar una escala de medición adecuada. Según Stevens (1951), existen cuatro tipos de escalas, ordenadas de menor a mayor nivel de medición: nominal (los atributos se nombran sin orden), ordinal (los atributos presentan un orden significativo), de intervalo (las diferencias entre valores son comparables), y de razón (las diferencias son comparables y existe un valor cero absoluto que representa la ausencia total del atributo).

Las transformaciones entre escalas solo son posibles en dirección descendente: una escala de razón puede transformarse en una de intervalo, ordinal o nominal; una de intervalo en una ordinal o nominal; y una ordinal en una nominal. No es posible realizar transformaciones en sentido contrario.

En este modelo, las propiedades de calidad se definen mediante una escala de razón con valores entre 0 y 1, ya que se emplean para calcular el porcentaje de cumplimiento de cada propiedad. Esta escala permite una interpretación consistente: el valor 0 representa ausencia o incumplimiento total de la propiedad, mientras que el valor 1 representa su cumplimiento completo. Este enfoque es coherente con lo propuesto en otros estudios (Fayers & Hand, 2002). En el caso de las aplicaciones, el valor asignado (entre 0 y 1) se obtiene a través del instrumento correspondiente definido para cada una.

4.6.4. Funciones de medición

Los elementos del modelo de calidad pueden cuantificarse mediante funciones de medición. Estas funciones consisten en secuencias de operaciones que combinan medidas de calidad de menor nivel, desde los valores obtenidos directamente de la visualización hasta los valores compuestos que representan la calidad general. La Fig. 9 ilustra la relación entre el modelo de calidad y el modelo de medición.

Doctorado en Informática

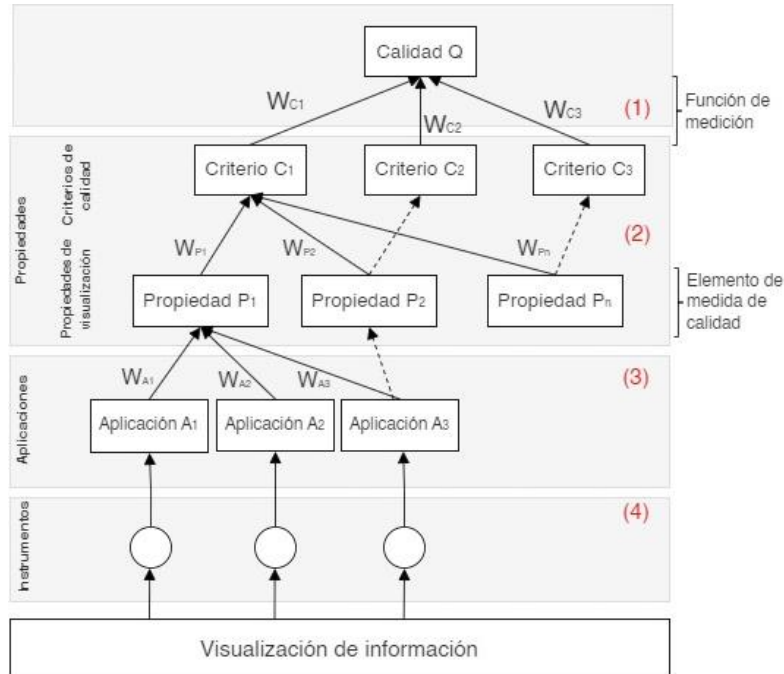


Fig. 9: Relación entre el modelo de calidad y el modelo de medición

Se distinguen cuatro niveles de funciones, que siguen la jerarquía del modelo:

- 1) La función de medición de calidad, que agrega los elementos de medidas de calidad dados por los valores de los criterios ($V_{C1}, V_{C2}, \dots, V_{Cn}$) y tiene la forma:

$$fQ(C_1, C_2, \dots, C_n) = W_{C1} \cdot V_{C1} + W_{C2} \cdot V_{C2} + \dots + W_{Cn} \cdot V_{Cn}$$

- 2) La función de medición de criterios, que agrega los elementos de medidas de calidad dados por los valores de las propiedades ($V_{P1}, V_{P2}, \dots, V_{Pn}$) y tiene la forma:

$$fC(P_1, P_2, \dots, P_n) = W_{P1} \cdot V_{P1} + W_{P2} \cdot V_{P2} + \dots + W_{Pn} \cdot V_{Pn}$$

- 3) La función de medición de propiedades, que agrega los elementos de medidas de calidad dados por los valores de las aplicaciones ($V_{A1}, V_{A2}, \dots, V_{An}$) y tiene la forma:

$$fP(A_1, A_2, \dots, A_n) = W_{A1} \cdot V_{A1} + W_{A2} \cdot V_{A2} + \dots + W_{An} \cdot V_{An}$$

- 4) La función de medición de aplicaciones, que toma como elementos de medidas de calidad los datos obtenidos directamente de la visualización a través de los instrumentos.

Doctorado en Informática

Estas funciones se basan en la técnica de SAW, y los pesos utilizados (denotados como W) se calculan con el método AHP Express. La estructura refleja relaciones causales entre niveles:

- La calidad de una visualización depende de la calidad de sus criterios.
- La calidad de un criterio depende de la calidad de sus propiedades asociadas.
- La calidad de una propiedad depende de la calidad de sus aplicaciones.
- La calidad de una aplicación se determina por los datos observables mediante instrumentos.

4.7. Instanciar el modelo para un caso de uso

Si bien el metamodelo proporciona una base estructural para la evaluación, no todos los elementos del modelo tienen la misma relevancia en todos los contextos. La importancia relativa de los criterios y propiedades de calidad varía según las necesidades y prioridades de las partes involucradas. Por lo tanto, antes de aplicar el modelo es necesario personalizarlo para reflejar las particularidades del caso de uso y focalizar la evaluación en los aspectos más pertinentes.

Con este fin, se propone el concepto de perfiles de calidad. Un perfil de calidad representa una configuración de pesos adaptada a un contexto o caso de uso específico (Morisio et al., 2002). Por ejemplo, un perfil estándar podría asignar igual importancia a todos los criterios, mientras que un perfil orientado a la toma de decisiones podría dar mayor peso a criterios como comprensión e información.

La definición de perfiles de calidad permite realizar evaluaciones más enfocadas y relevantes, así como identificar oportunidades de mejora en función del objetivo de la visualización. También brinda una guía para desarrolladores y diseñadores al momento de priorizar propiedades en función de los requisitos de su público objetivo.

Para construir estos perfiles, se sugiere un enfoque orientado a objetivos, en el cual cada perfil se define en función del propósito de la visualización. De acuerdo con (Amini, 2020), la importancia relativa de los criterios de evaluación puede cambiar dependiendo

Doctorado en Informática

de la intención del autor. Por ejemplo, la comprensión puede ser clave si el mensaje es enseñar, pero el compromiso puede ser más importante para fomentar la reflexión sobre un tema específico. La Tabla 8 presenta ejemplos ilustrativos de perfiles de calidad, indicando los criterios prioritarios para distintos contextos.

Tabla 8: Ejemplos de perfiles de calidad

Perfil de calidad	Descripción	Criterios prioritarios
Estándar	Cobertura general para diferentes tipos de visualizaciones.	Todos los criterios
Toma de decisiones	Énfasis en facilitar la interpretación para decisiones informadas.	Comprensión, Información
Educación	Enfocado en favorecer la comprensión y el recuerdo.	Comprensión, Compromiso, Memorabilidad
Comunicación	Mejora de la claridad y efectividad del mensaje.	Comprensión, Compromiso, Memorabilidad
Exploración	Apoyo a la exploración de patrones y descubrimiento de datos.	Usabilidad, Información

Estos perfiles se presentan como orientación inicial y pueden ajustarse según el caso. Los pesos de cada elemento se asignan utilizando el método AHP-Express, permitiendo reflejar su importancia relativa dentro del contexto evaluado.

4.8. Conclusiones

Este capítulo presentó el modelo de calidad propuesto para la evaluación de visualizaciones narrativas. Se describieron sus fundamentos, estructura jerárquica, criterios, propiedades, aplicaciones e instrumentos de medición, así como las reglas para su implementación y personalización. El modelo se construyó a partir de normas y enfoques consolidados en IS, adaptados al dominio de la visualización.

La Fig. 10 actualiza la vista general del proceso metodológico seguido en la tesis, señalando en verde los pasos desarrollados hasta el momento y en gris aquellos que serán abordados en los capítulos siguientes. El próximo capítulo describe la validación empírica del modelo.

Doctorado en Informática

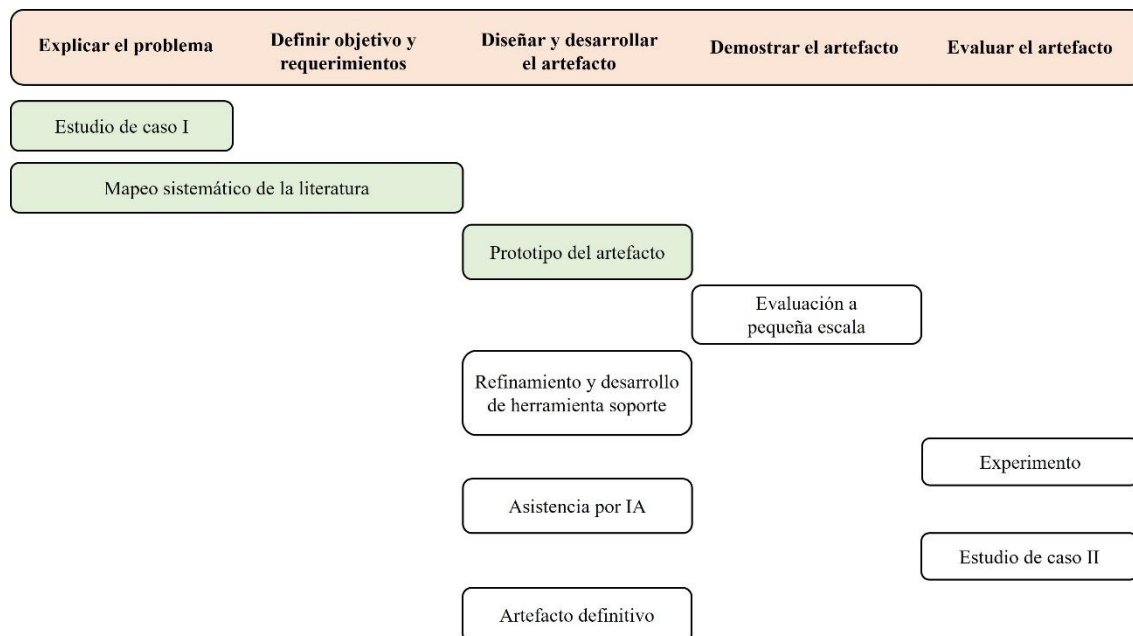


Fig. 10: Actividades abarcadas en el capítulo

Capítulo 5. Validación del artefacto

Este capítulo presenta la estrategia de validación del modelo de calidad propuesto. La misma se estructuró en tres etapas sucesivas, enfocadas en distintos aspectos del artefacto. La sección 5.1 resume el esquema general del proceso de validación. La sección 5.2 describe una evaluación a pequeña escala, centrada en la viabilidad del modelo. La sección 5.3 presenta la primera iteración del artefacto y su implementación en una herramienta operativa. La sección 5.4 continúa con un experimento controlado, diseñado para examinar la capacidad del modelo para generar evaluaciones consistentes y diferenciar niveles de calidad en un entorno estructurado. La sección 5.5 introduce una segunda iteración del artefacto, que incorpora funcionalidades de asistencia automática y evaluación basada en IA. Finalmente, la sección 5.6 presenta un estudio de caso aplicado, que permitió observar el funcionamiento del modelo en un entorno real.

5.1. Estrategia de validación

La Fig. 11 muestra la estrategia seguida para validar el modelo de calidad. El proceso se organizó en tres etapas consecutivas, con el objetivo de analizar distintos aspectos del artefacto en forma progresiva.

Cada fase se diseñó como un estudio independiente, con objetivos, preguntas e instrumentos específicos. En la primera se formularon preguntas de investigación (PI) exploratorias; en la segunda, hipótesis de trabajo; y en la tercera, se definieron PI orientadas al uso aplicado del modelo. No existe una correspondencia directa entre ellas, ya que cada estudio aborda un nivel distinto de análisis en la validación.

Doctorado en Informática

La primera etapa consistió en una evaluación a pequeña escala, orientada a explorar la viabilidad inicial del modelo. Se aplicó utilizando una planilla de cálculo y un conjunto reducido de visualizaciones estáticas e interactivas. Participaron evaluadores con distintos niveles de experiencia. Esta etapa se centró en observar la aplicabilidad del modelo, su estructura y la percepción de uso por parte de los participantes.

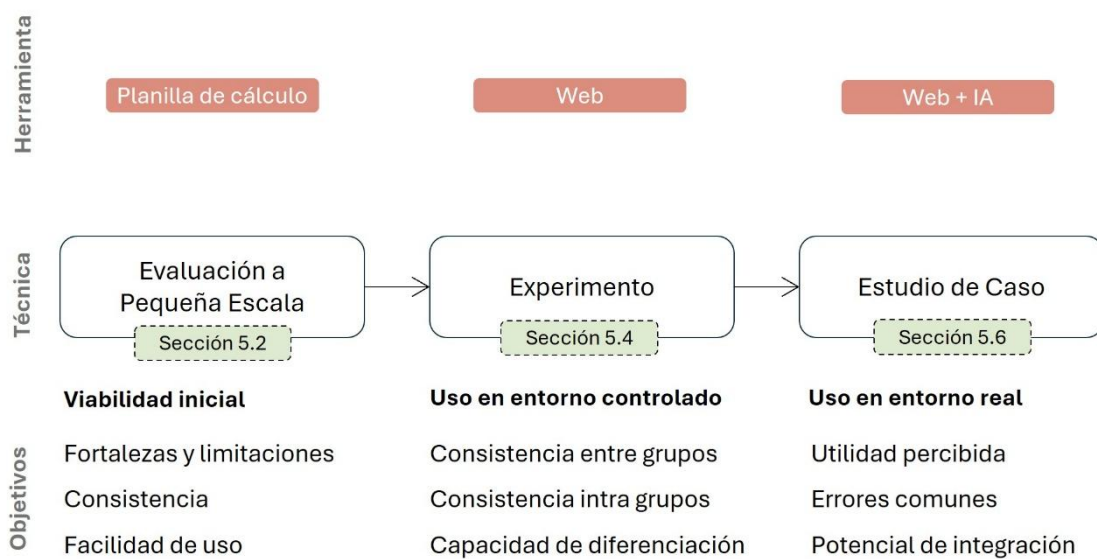


Fig. 11: Esquema general de validación.

La segunda etapa consistió en un experimento, diseñado para analizar el comportamiento del modelo en condiciones controladas. Se utilizó una herramienta web para aplicar el modelo a un conjunto ampliado de visualizaciones, clasificadas previamente según niveles de calidad. Participaron personas con distintos niveles de experiencia, agrupadas para examinar diferencias entre perfiles. Esto permitió explorar la consistencia de las evaluaciones y la capacidad del modelo para diferenciar distintos niveles de calidad.

La tercera etapa consistió en un estudio de caso, orientado a observar el uso del modelo en un entorno de trabajo real. Se implementó en el marco de un proyecto de visualización desarrollado por un grupo de investigación, utilizando un perfil de calidad personalizado. Esta instancia permitió explorar la integración del modelo en el flujo de trabajo y las percepciones del equipo sobre su utilidad en decisiones concretas de rediseño.

Doctorado en Informática

Entre cada una de las etapas se realizaron iteraciones sobre el modelo, incorporando ajustes conceptuales y operativos en función de los resultados y de la retroalimentación recibida. Estas modificaciones incluyeron la reorganización de componentes, la reducción del número de aplicaciones y el desarrollo progresivo de una herramienta operativa.

5.2. Evaluación a pequeña escala

Con el objetivo de comprobar la viabilidad inicial del modelo de calidad, se llevó a cabo una evaluación a pequeña escala siguiendo las recomendaciones metodológicas de Wohlin & Rainer (2022) y Robson (2017). Este tipo de estudios resulta apropiado en etapas preliminares, ya que permite realizar valoraciones tempranas del artefacto a partir de un número reducido de fuentes de datos.

5.2.1. Preguntas de investigación

La evaluación se diseñó para responder las siguientes PI:

PI-1: ¿Cuáles son las fortalezas y limitaciones del modelo para evaluar la calidad de visualizaciones narrativas?

El propósito de esta pregunta fue identificar componentes del modelo que resultaran útiles o problemáticos durante la aplicación. Las respuestas se analizaron a partir de comentarios registrados por los participantes.

PI-2: ¿Qué nivel de consistencia se observa entre evaluadores?

Esta pregunta se enfocó en medir el grado de acuerdo entre evaluaciones individuales sobre un mismo conjunto de visualizaciones, con el fin de estimar la estabilidad del modelo.

PI-3: ¿Cuál es la dificultad percibida al aplicar el modelo?

Finalmente, esta pregunta buscó registrar la carga de trabajo asociada a la evaluación, en función del perfil de cada participante. Se combinaron observaciones cuantitativas y cualitativas para capturar tanto el grado de dificultad como sus causas percibidas.

Doctorado en Informática**5.2.2. Configuración del perfil de calidad**

Se aplicó el método AHP-Express descrito en el capítulo anterior para establecer los pesos de los criterios en el perfil de calidad Estándar, con la participación de 51 personas (18 hombres y 33 mujeres), cuyas edades estaban comprendidas entre los 20 y los 35 años, en el marco de un taller de visualización.

El procedimiento constó de dos etapas: en primer lugar, cada participante seleccionó el criterio que consideraba más importante; en segundo lugar, asignó valores de comparación por pares (entre 1 y 9) para los demás criterios, indicando su importancia relativa respecto al criterio principal. Por ejemplo, si un participante consideraba que la comprensión era considerablemente más importante que el compromiso, podía asignar un valor de 5 (moderadamente más importante) o 7 (mucho más importante).

Una vez recopiladas todas las comparaciones, se calcularon las prioridades normalizadas para cada criterio. La Tabla 9 presenta los valores finales obtenidos.

Tabla 9: Peso de los criterios para el perfil de calidad Estándar

ID Criterio	Descripción	Peso
C01	Comprensión	0,23
C02	Compromiso	0,15
C03	Información	0,18
C04	Memorabilidad	0,26
C05	Usabilidad	0,18

Estos pesos normalizados conformaron el perfil de calidad Estándar. Por ejemplo, la comprensión obtuvo un peso de 0.23, mientras que el compromiso alcanzó un valor de 0.15. Posteriormente, estos valores se aplicaron en las fórmulas del modelo de medición (ver sección 4.6.4) para calcular la puntuación de calidad final de cada visualización.

5.2.3. Conjunto de datos

La evaluación se realizó sobre un conjunto de 18 visualizaciones, compuesto por 12 visualizaciones estáticas y 6 interactivas. Para las visualizaciones estáticas, se utilizó el

Doctorado en Informática

conjunto MASSVIS (Borkin et al., 2016), una referencia común en estudios sobre percepción y efectividad visual. Este conjunto de visualizaciones se divide en tres categorías: ciencia, gobierno y noticias. Para esta evaluación se seleccionaron únicamente visualizaciones de las categorías de noticias y gobierno, dado que las pertenecientes a ciencia suelen requerir conocimientos especializados y carecen de un componente narrativo claro, que es el foco del modelo. La selección se automatizó mediante un script en Python que aplicó muestreo aleatorio simple sobre cada categoría.

Las visualizaciones interactivas fueron extraídas de galerías de ejemplos de distintas bibliotecas de visualización, herramientas en línea y portales reconocidos por su componente narrativo. La inclusión de este tipo de visualizaciones responde a la necesidad de analizar aspectos vinculados a la interacción, considerados clave en experiencias narrativas interactivas (Saket et al., 2018) como animaciones, transiciones o exploración progresiva.

La Fig. 12 muestra el conjunto final de visualizaciones utilizadas en el estudio, junto con sus identificadores correspondientes. La fila superior representa visualizaciones del grupo “noticias”, la fila intermedia del grupo “gobierno”, y la fila inferior corresponde a visualizaciones interactivas.

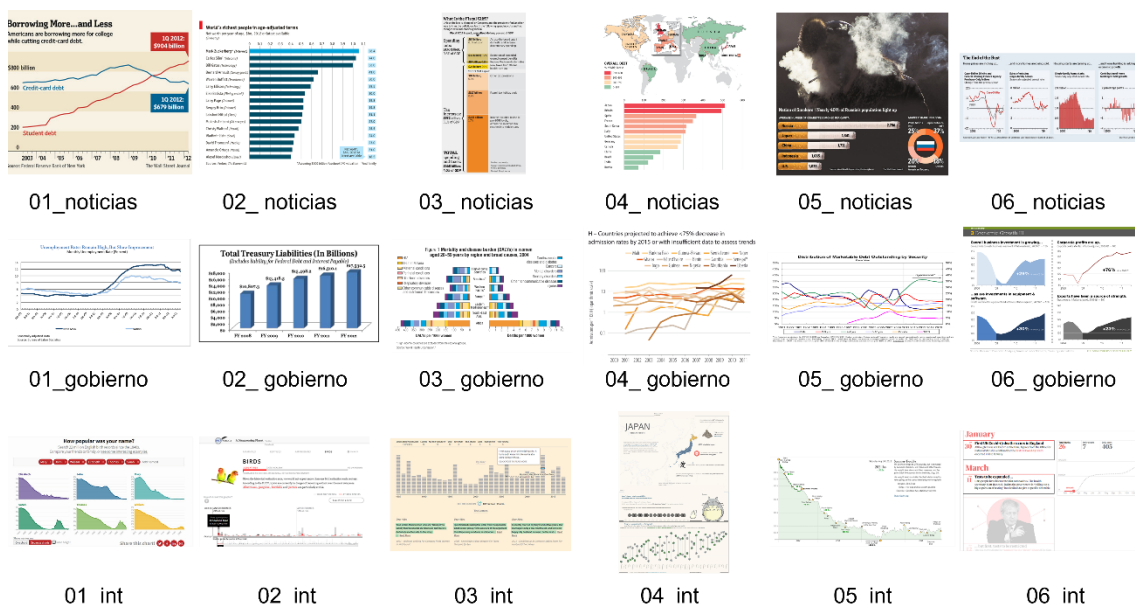


Fig. 12: Visualizaciones seleccionadas para la evaluación a pequeña escala

Doctorado en Informática

5.2.4. Perfil de los evaluadores

Participaron 12 evaluadores con distintos niveles de experiencia en visualización. Todos eran estudiantes o egresados recientes de la Universidad Nacional del Nordeste, con edades entre 22 y 36 años ($M = 30$, $DE = 4.7$). La experiencia profesional variaba desde estudiantes sin experiencia hasta profesionales con 14 años de trayectoria ($M = 6$, $DE = 3.9$). Esta diversidad permitió capturar múltiples perspectivas y reducir posibles sesgos en la evaluación.

5.2.5. Procedimiento

Se preparó una planilla de cálculo con los criterios, propiedades y aplicaciones del modelo, incluyendo las fórmulas para calcular automáticamente sus puntajes. Se utilizó el perfil Estándar, aplicando los pesos determinados anteriormente.

ID	Application	Description	Answer
AP01	Zero baseline	Bar charts axis should start at zero.	Non-zero baseline
AP02	Dual Y axis	The use of dual Y-axis should be justified.	Zero baseline Non-zero baseline Unclear
AP03	Axes titles	Axes should be named appropriately.	N/A
AP04	Scale changes	Changes in data scale should be visually clear	Changes in scale are explicit
AP05	Scale breaks	Values across scale breaks should not be connected	Values are connected
AP06	Limited series	The number of series should be limited to working memory capacity.	Between 3 and 5 series
AP07	Unambiguous variables	Variables should be defined in an accurate, unambiguous manner.	Variables are accurate
AP08	Aspect ratio	Aspect ratio should comply to the banking at 45° rule	N/A
AP09	Data ordering	Data should follow a specific order.	Data is ordered intentionally
AP10	Data labels	Data should be labeled directly.	Data is labeled by legends

Fig. 13: Fragmento de la hoja de cálculo utilizada para las evaluaciones.

Para cada visualización, los evaluadores respondieron preguntas (una por cada aplicación) seleccionando opciones predefinidas desde listas desplegables, como se muestra en la Fig. 13. Los puntajes se calcularon automáticamente y se registraron en una

Doctorado en Informática

hoja de resultados. Una vez completadas todas las evaluaciones, se agregaron los resultados individuales y se procesaron para el análisis posterior.

Finalizadas las evaluaciones, los participantes completaron una encuesta basada en una escala de Likert de cinco puntos, con el fin de recoger su percepción sobre la experiencia de uso y dificultad del modelo.

Los materiales utilizados en esta etapa, incluido el script de Python, las hojas de cálculo de evaluación y el formulario de encuesta, se encuentran en el Apéndice A.

5.2.6. Resultados

La Fig. 14 muestra los puntajes de calidad asignados por los evaluadores a cada visualización, agrupados por tipo: noticias, gobierno e interactivas. Cada panel representa una categoría, y dentro de cada uno se presentan diagramas de caja (*box plots*) para las seis visualizaciones evaluadas en ese grupo. Estos diagramas permiten observar la distribución de los puntajes, así como sus valores promedio y la variabilidad entre evaluadores.

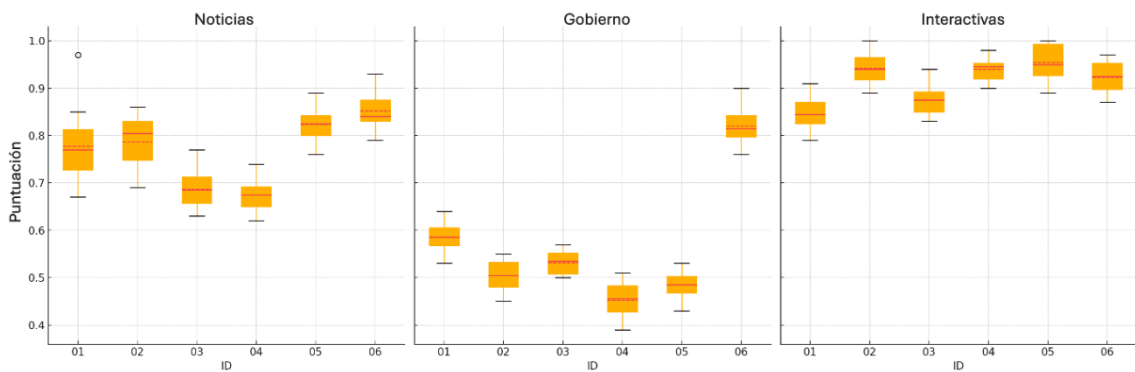


Fig. 14: Resumen de resultados de evaluación

En el grupo de noticias se registró mayor variabilidad, con algunas visualizaciones que obtuvieron puntajes altos y otras más bajos. Esto sugiere diferencias en la medida en que cada visualización se ajustaba a los criterios del modelo. Las visualizaciones del grupo gobierno presentaron puntajes sistemáticamente más bajos y con menor dispersión. En varios casos, se trataba de gráficos que mostraban datos de forma sintética, sin recursos que facilitaran su interpretación. En contraste, las visualizaciones interactivas obtuvieron

Doctorado en Informática

puntajes altos y consistentes. Este resultado podría estar asociado a la naturaleza misma de este tipo de visualizaciones, que suelen incorporar elementos como vistas múltiples, retroalimentación o navegación progresiva, que permiten evaluar un mayor número de aspectos contemplados por el modelo.

5.2.6.1. Fortalezas y limitaciones del modelo (PI-1)

Para abordar esta pregunta, se analizaron de manera cualitativa los comentarios registrados en la encuesta posterior a la evaluación. Las respuestas se organizaron según el tipo de visualización evaluada y los componentes del modelo utilizados.

En el caso de las visualizaciones estáticas, los participantes indicaron que el modelo permitió identificar aspectos relacionados con el etiquetado, la escala, la cantidad de elementos visuales y el uso del color. Aplicaciones como AP01 (Línea base en cero), AP03 (Títulos de ejes), AP06 (Series limitadas) y AP19 (Colores limitados) se utilizaron con frecuencia en estas visualizaciones, y permitieron evaluar elementos considerados relevantes para transmitir el mensaje de forma clara.

Para las visualizaciones interactivas, las observaciones se centraron en aspectos vinculados a la navegación, la retroalimentación del sistema y el control del usuario sobre la exploración. Se aplicaron aplicaciones como AP48 (Interactividad sugerida), AP49 (Retroalimentación en la interacción), AP50 (Visión general primero; detalles bajo demanda), AP53 (Navegación) y AP54 (Progreso), orientadas a registrar mecanismos comunes en experiencias interactivas con énfasis narrativo.

El componente narrativo también fue reportado en ambas categorías. La propiedad Narrativa (P13), que incluye aplicaciones como AP47 (Ilustraciones relevantes) y AP51 (Secuencia de eventos), se aplicó en la mayoría de los casos y permitió capturar elementos estructurales del mensaje visual.

Entre los comentarios registrados, varios evaluadores indicaron que el modelo permitió identificar con precisión elementos a mejorar en cada visualización. También señalaron que la estructura detallada del modelo facilitó un análisis desagregado por componentes. No obstante, algunos mencionaron que la cantidad de aplicaciones a evaluar requiere una

Doctorado en Informática

dedicación considerable. El tiempo promedio de evaluación por visualización fue de aproximadamente 10 minutos, lo cual podría limitar su uso en escenarios donde se requiere evaluar múltiples visualizaciones en poco tiempo.

5.2.6.2. Consistencia entre evaluadores (PI-2)

Se calculó el coeficiente de correlación intraclase (ICC, por sus siglas en inglés) para estimar el nivel de acuerdo entre evaluadores. Esta medida permite cuantificar la consistencia en las puntuaciones asignadas por distintos participantes a un mismo conjunto de objetos de evaluación, en este caso, visualizaciones narrativas.

Se utilizó el modelo ICC(2,1), correspondiente a un diseño de efectos aleatorios bidireccional con acuerdo absoluto y medida individual, de acuerdo con la clasificación de Koo & Li (2016). Este modelo fue seleccionado por tres motivos. En primer lugar, el diseño de efectos aleatorios considera tanto a los evaluadores como a las visualizaciones como muestras extraídas de una población mayor, lo cual permite extender los resultados más allá del conjunto específico analizado. En segundo lugar, el uso del acuerdo absoluto evalúa coincidencias exactas en los puntajes asignados, sin corregir por diferencias entre evaluadores. En tercer lugar, la opción de medida individual se ajusta al enfoque adoptado en este estudio, donde las evaluaciones fueron realizadas de forma independiente y no se combinaron entre sí.

El coeficiente obtenido fue $ICC = 0.984$, calculado sobre la totalidad de evaluadores y visualizaciones. Este valor sugiere un alto nivel de coincidencia entre los puntajes registrados, indicando que las instrucciones, el instrumento de evaluación y la formulación de las aplicaciones fueron interpretadas de manera similar por los participantes. No obstante, dado que el número de evaluadores fue reducido y la selección de visualizaciones estuvo limitada a un conjunto específico, este resultado debe considerarse como un primer indicio. Para confirmar su estabilidad, son necesarios estudios adicionales con mayor variabilidad en los perfiles de evaluadores y en los tipos de visualización analizados.

5.2.6.3. Dificultad percibida (PI-3)

Doctorado en Informática

La dificultad percibida al aplicar el modelo varió entre los participantes. Esta información se obtuvo a través de la encuesta posterior a la evaluación, que incluyó una escala de Likert de cinco puntos y preguntas abiertas. La distribución de respuestas en la escala mostró que la mayoría valoró la experiencia como “algo fácil” o “ni fácil ni difícil”. Solo un número reducido seleccionó opciones asociadas a mayor dificultad.

En términos de tiempo, los evaluadores con mayor experiencia (formación específica en visualización o más de cinco años de práctica) completaron cada evaluación en menos de 10 minutos. Los participantes con experiencia intermedia reportaron un promedio cercano a los 12 minutos, y los sin experiencia específica en visualización, alrededor de 15 minutos por visualización.

En los comentarios abiertos, los evaluadores más experimentados indicaron que la estructura del instrumento y la formulación de las preguntas eran consistentes con prácticas habituales, lo cual facilitó su aplicación. Quienes tenían experiencia intermedia señalaron que algunas aplicaciones contenían términos poco comunes o requerían interpretación adicional, y sugirieron incluir ejemplos o referencias visuales. Los participantes sin experiencia mencionaron dificultades para responder a ciertas aplicaciones sin un marco de referencia claro, y propusieron agregar una opción de respuesta que permitiera registrar la ausencia de correspondencia.

A pesar de estas diferencias, los resultados mostraron un alto nivel de consistencia entre evaluadores (ver sección anterior). Esto indica que las instrucciones y la formulación de las aplicaciones fueron comprendidas de forma similar, incluso por quienes tenían menor experiencia. No obstante, las respuestas cualitativas sugieren que la carga cognitiva percibida podría reducirse mediante ajustes en el instrumento, sin modificar el contenido evaluado.

De esta manera concluye la primera etapa en el plan de validación. La Fig. 15 sintetiza los objetivos considerados y su estado al finalizar la evaluación. Los tres objetivos de viabilidad inicial fueron cumplidos y se construyó una planilla de cálculo como primera aproximación para operacionalizar el modelo.

Doctorado en Informática

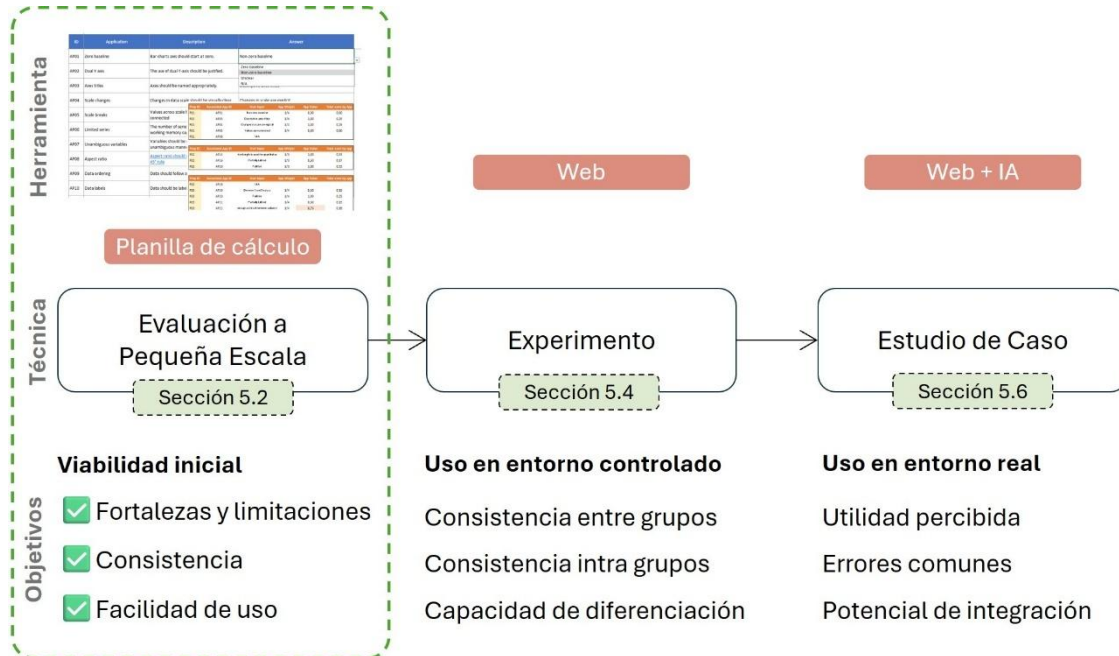


Fig. 15: Cumplimiento de objetivos en la primera etapa de validación.

5.3. Primera iteración de diseño del artefacto

A partir de los resultados obtenidos en la evaluación a pequeña escala y de los comentarios registrados en la encuesta posterior, se realizaron ajustes al modelo orientados a facilitar su uso en etapas posteriores.

En primer lugar, se reorganizó el conjunto de aplicaciones, reduciendo su número de 55 a 36. Esta modificación se llevó a cabo mediante la eliminación de redundancias, la agrupación de aplicaciones con enfoques similares y la reformulación de algunas opciones de respuesta.

En segundo lugar, se desarrolló una herramienta software para operacionalizar el modelo y gestionar el proceso de evaluación. La misma incorporó varias de las sugerencias indicadas por los evaluadores, incluyendo ejemplos visuales y mecanismos de ayuda contextual. Su funcionamiento se describe a continuación.

Doctorado en Informática

5.3.1. Desarrollo de la herramienta soporte

La herramienta implementa los componentes principales del modelo y organiza el proceso de evaluación en etapas secuenciales. Fue desarrollada con el *framework* Django, utilizando Python como lenguaje principal y se encuentra disponible en línea². Dado que las etapas de validación siguientes incluyeron participantes de distintos países, la interfaz fue construida en inglés para facilitar su aplicación en contextos diversos. El sistema permite aplicar el modelo a través de una interfaz web que cubre todo el flujo de trabajo, desde la carga de la visualización hasta la obtención del reporte final.

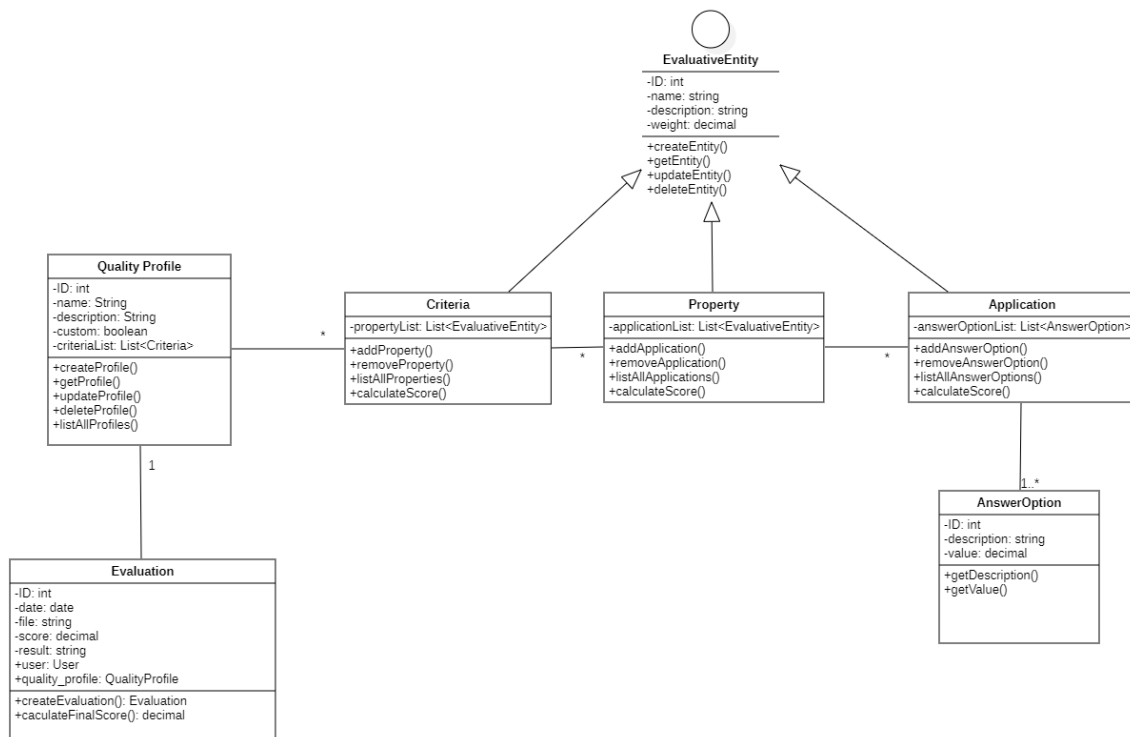


Fig. 16: Diagrama de clases de la herramienta de evaluación

La implementación técnica sigue una estructura jerárquica coherente con el modelo conceptual. La Fig. 16 presenta un diagrama simplificado de clases que incluye las entidades principales del sistema: Evaluación, Perfil de calidad, Criterios, Propiedades y Aplicaciones.

² <https://naviqmodel.pythonanywhere.com/>

Doctorado en Informática

recomendaciones de mejora asociadas a las aplicaciones con menor puntuación. Las recomendaciones se determinan calculando, para cada aplicación, la diferencia entre su puntaje obtenido y el valor esperado según el perfil de calidad. Este valor, denominado delta, permite identificar qué aspectos tuvieron mayor impacto negativo en la evaluación y priorizar oportunidades de mejora.

La Fig. 17 muestra una vista general del sistema en funcionamiento. Se observa la visualización y los datos de evaluación cargados (panel izquierdo), y la aplicación activa con las opciones de respuesta disponibles y el acceso a la ayuda contextual (panel derecho). Esta herramienta fue utilizada en las siguientes etapas de validación. Finalizada su implementación, se llevó a cabo un experimento con una muestra ampliada de visualizaciones y participantes. La siguiente sección describe su diseño, ejecución y resultados.

5.4. Experimento

Este estudio tuvo como objetivo analizar el comportamiento del modelo de calidad en condiciones experimentales controladas, evaluando su capacidad para generar puntajes consistentes y diferenciar distintos niveles de calidad en visualizaciones narrativas. A diferencia de la evaluación exploratoria descrita en la sección 5.1, esta instancia incorporó una muestra más amplia de visualizaciones y participantes, y aplicó un diseño factorial para examinar distintos escenarios de uso.

La variable dependiente fue el puntaje de calidad asignado a cada visualización. Este puntaje se calculó mediante el modelo NaVi-Q, utilizando el perfil Estándar definido en la sección 5.2.2. Para establecer un punto de comparación, primero se generaron evaluaciones de referencia. Cada visualización fue evaluada por un evaluador externo con experiencia en visualización, aplicando las 36 aplicaciones del modelo. A partir de estos puntajes, se clasificaron las visualizaciones en tres niveles de calidad: alta, media y baja. Esta categorización se utilizó posteriormente para definir el conjunto de visualizaciones del experimento, compuesto por 30 visualizaciones en total, 10 por nivel.

Doctorado en Informática

El diseño experimental combinó un factor intersujetos (nivel de experiencia del evaluador) y un factor intrasujetos (nivel de calidad de las visualizaciones). Los participantes fueron agrupados según su experiencia previa en el uso de visualizaciones, y cada uno evaluó visualizaciones de los tres niveles de calidad.

Este enfoque se fundamenta en trabajos previos sobre validación de modelos de calidad (Bansiya & Davis, 2002; Lichtenthäler & Wirtz, 2022), que proponen analizar tanto la capacidad discriminativa del modelo como su consistencia en contextos con diferentes perfiles de usuario. Se formularon las siguientes hipótesis de trabajo:

- **H1:** El modelo permite distinguir entre visualizaciones de alta, media y baja calidad.
- **H2:** Las evaluaciones generadas por el modelo son consistentes dentro de cada grupo (expertos y no expertos), en relación con las evaluaciones de referencia.
- **H3:** Los puntajes generados por los participantes presentan una correlación positiva con los puntajes de referencia.
- **H4:** Existen diferencias significativas entre las evaluaciones realizadas por participantes expertos y no expertos.

5.4.1. Participantes

Previo a la ejecución del experimento, se realizó una prueba piloto con 12 participantes, utilizando el procedimiento descrito más adelante. El objetivo fue comprobar la viabilidad del diseño y determinar el tamaño muestral adecuado mediante un análisis de potencia estadística. Este análisis se realizó utilizando el software G*Power (Heradio et al., 2022) considerando un tamaño del efecto (f) de 0,25, un nivel de significancia de 0,05 y un poder estadístico deseado de 0,80. Bajo estos parámetros, el análisis indicó que se requerían al menos 40 participantes para alcanzar una potencia adecuada y detectar posibles interacciones entre los factores del experimento (nivel de calidad de la visualización y nivel de experiencia del evaluador).

Doctorado en Informática

Los participantes fueron estudiantes y miembros de la comunidad universitaria de la Universidad Nacional del Nordeste. Esta población fue seleccionada por su diversidad de perfiles, que permitió incluir personas con distintos grados de exposición a visualizaciones en contextos académicos, profesionales o personales. La muestra se dividió en dos grupos, según el nivel de experiencia en el uso de visualizaciones:

Grupo de expertos: compuesto por personas con experiencia profesional en el uso de visualizaciones, aunque no necesariamente con formación en visualización de datos. En general, estos participantes estaban familiarizados con principios y guías generales sobre buenas prácticas, adquiridos a través de su actividad laboral.

Grupo de no expertos: integrado por personas con exposición académica o profesional a visualizaciones, pero sin una participación continua o especializada en su análisis o construcción. Muchos de ellos habían trabajado con visualizaciones en entornos de formación o en actividades puntuales, pero no de manera frecuente.

Para anticipar posibles abandonos, se reclutaron inicialmente 45 personas. Durante la planificación, dos participantes del grupo de expertos informaron conflictos de agenda que les impidieron asistir. Se convocó entonces a dos participantes suplentes del mismo grupo, quienes completaron la actividad, alcanzando así el total previsto de 40 evaluaciones completas.

Además de la clasificación por nivel de experiencia, se recolectaron datos demográficos y de formación. La edad de los participantes osciló entre 22 y 47 años, con una media de 35,9 años ($DE = 7,4$). En el grupo de expertos, la experiencia profesional vinculada al uso de visualizaciones fue de 12 años en promedio ($DE = 6,7$). En cuanto a la formación académica, la mayoría de los participantes contaba con estudios universitarios completos o en curso. También se incluyeron personas con formación de posgrado, tanto a nivel de especialización como de maestría o doctorado. Respecto al dominio de herramientas tecnológicas, más de la mitad de los participantes se identificó con un nivel alto, seguido por niveles medio y bajo.

Doctorado en Informática

5.4.2. Materiales

Visualizaciones: El conjunto utilizado en esta etapa estuvo compuesto por 30 visualizaciones seleccionadas a partir de fuentes heterogéneas. Se incluyeron ejemplos del conjunto MASSVIS (Borkin et al., 2016), de la galería pública de Tableau³, así como entradas de blogs y repositorios compartidos por profesionales del área⁴. La selección consideró una variedad de técnicas, incluyendo gráficos de barras, líneas, pastel, dispersión y mapas coropléticos. La Fig. 18 presenta ejemplos de cada grupo.

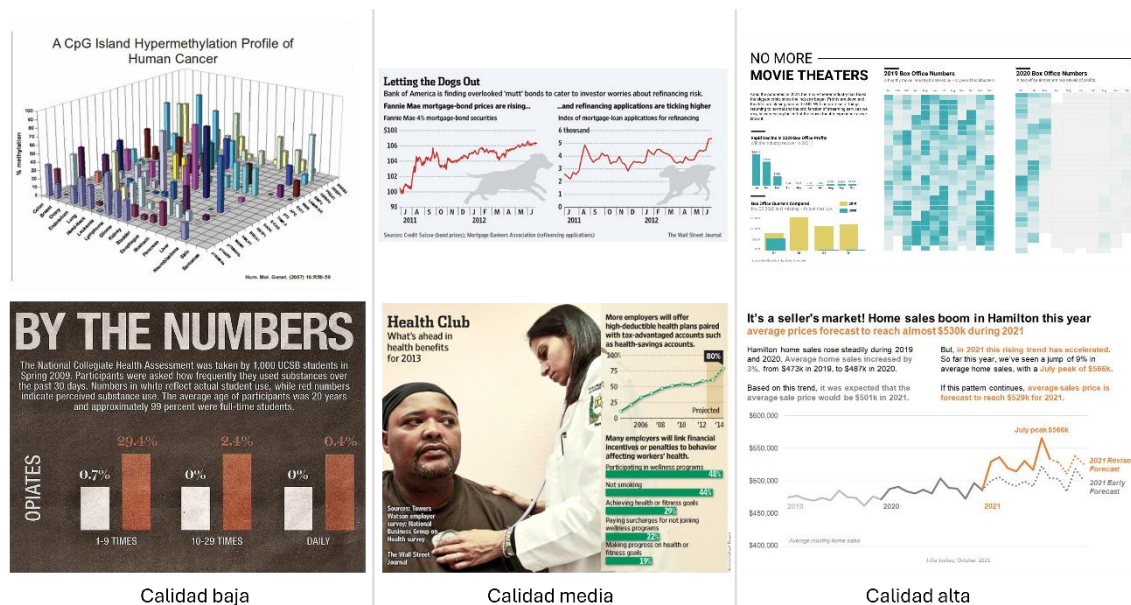


Fig. 18: Ejemplos de visualizaciones narrativas utilizadas en el estudio

Modelo de evaluación: Se utilizó la herramienta web desarrollada para operacionalizar el modelo (ver sección 5.3.1). Para cada visualización, los participantes respondieron preguntas asociadas a las 36 aplicaciones del modelo.

Instrumentos complementarios: Al finalizar las evaluaciones, los participantes completaron dos encuestas. La primera encuesta solicitó que califiquen la asociación de las distintas propiedades del modelo (por ejemplo, Color, Narrativa, Enfoque) con los cinco criterios de calidad definidos en el Capítulo 1. Las asociaciones se valoraron en una

³ www.tableau.com/viz-gallery

⁴ www.storytellingwithdata.com

Doctorado en Informática

escala de Likert de cinco puntos, donde 1 indicaba ausencia de relación y 5 una relación muy fuerte. La segunda encuesta reunió información sobre la experiencia de uso de la herramienta, incluyendo preguntas cerradas sobre claridad, utilidad y percepción general del proceso, y una pregunta abierta para registrar sugerencias.

5.4.3. Procedimiento

Cada participante evaluó 6 de las 30 visualizaciones disponibles, lo que resultó en un total de 240 evaluaciones (6 visualizaciones $\times$ 40 participantes). Para asegurar una cobertura balanceada de niveles de calidad, se aplicó un diseño por bloques. Cada persona evaluó dos visualizaciones de cada nivel (alta, media, baja). El orden de presentación se aleatorizó para reducir posibles efectos de secuencia.

Fase preevaluación: Se realizaron sesiones virtuales breves para presentar el modelo de calidad. Cada sesión tuvo una duración aproximada de 20 minutos y se realizó por grupo. Durante la presentación se explicó el funcionamiento de la herramienta y se mostró un ejemplo de evaluación. Los participantes tuvieron la oportunidad de realizar consultas antes de comenzar.

Fase de evaluación: Las evaluaciones se llevaron a cabo en días posteriores a la etapa introductoria. Los participantes utilizaron la herramienta web para analizar las visualizaciones asignadas.

Fase posevaluación: Al finalizar la actividad, los participantes completaron las dos encuestas mediante formularios en línea. Las mismas se detallan en el 0.

5.4.4. Recolección de datos

Durante el experimento, se registraron dos tipos de información: las respuestas individuales obtenidas mediante la herramienta de evaluación y los resultados de las encuestas posteriores.

Para cada participante se generó una matriz que incluía las visualizaciones asignadas, los puntajes calculados con el modelo y los puntajes de referencia correspondientes. La Tabla 10 muestra un ejemplo simplificado de esta matriz. En ella se incluyen los valores

Doctorado en Informática

generados por el modelo para cada visualización evaluada por el participante, junto con el puntaje de referencia establecido previamente por el evaluador externo.

Tabla 10: Ejemplo de matriz de evaluación por participante

	V4	V8	V13	V11	V23	V25
E1	0,2	0,31	0,63	0,55	0,89	0,86
Base	0,15	0,34	0,55	0,47	0,78	0,8

Dado que una misma visualización podía ser evaluada por múltiples participantes, se construyeron tablas resumen por visualización. Estas tablas registraron las respuestas seleccionadas por cada evaluador en las aplicaciones correspondientes. La primera columna indica la aplicación evaluada y sus opciones de respuesta; las columnas siguientes reflejan las selecciones individuales (E = evaluador). La Tabla 11 muestra un extracto representativo.

Tabla 11: Tabla binaria por visualización (extracto)

	V3 (base)	E2	E7	E12	E17	E21	E26	E31	E36
Línea base cero									
Línea base cero								.	.
Línea base distinta de cero	.								
Series limitadas									
Hasta 2 series									
Entre 3 y 5 series									
Más de 5 series	.	.	.	.	.	.	.	.	.
Orden de los datos									
Los datos están ordenados intencionalmente									
Los datos carecen de un orden discernible	.	.	.	.	.	.	.	.	.
Títulos cortos									
Los títulos son cortos y concisos									
Los títulos son largos o poco claros				.					
No hay título	.	.	.	.	.	.	.	.	.
Anotaciones									
Apropiadas					.				
Excesivas		.	.	.			.	.	
...									
Puntuación	0,15	0,05	0,07	0,11	0,08	0,06	0,07	0,08	0,06

Doctorado en Informática

5.4.5. Resultados

Esta sección presenta los resultados obtenidos en el experimento, organizados en torno a las hipótesis formuladas. Como se explicó en la sección 5.4.1, la prueba piloto permitió establecer el tamaño muestral necesario, y sus datos no fueron considerados en los análisis que se presentan a continuación. Todos los resultados reportados corresponden a las 240 evaluaciones realizadas por los 40 participantes en el experimento principal.

5.4.5.1. Distribución de puntajes

Antes de abordar las hipótesis de trabajo, se analizaron las distribuciones generales de puntajes obtenidos por los participantes. La Fig. 19 muestra los histogramas correspondientes a cada combinación de grupo (expertos y no expertos) y nivel de calidad de las visualizaciones (alta, media, baja).

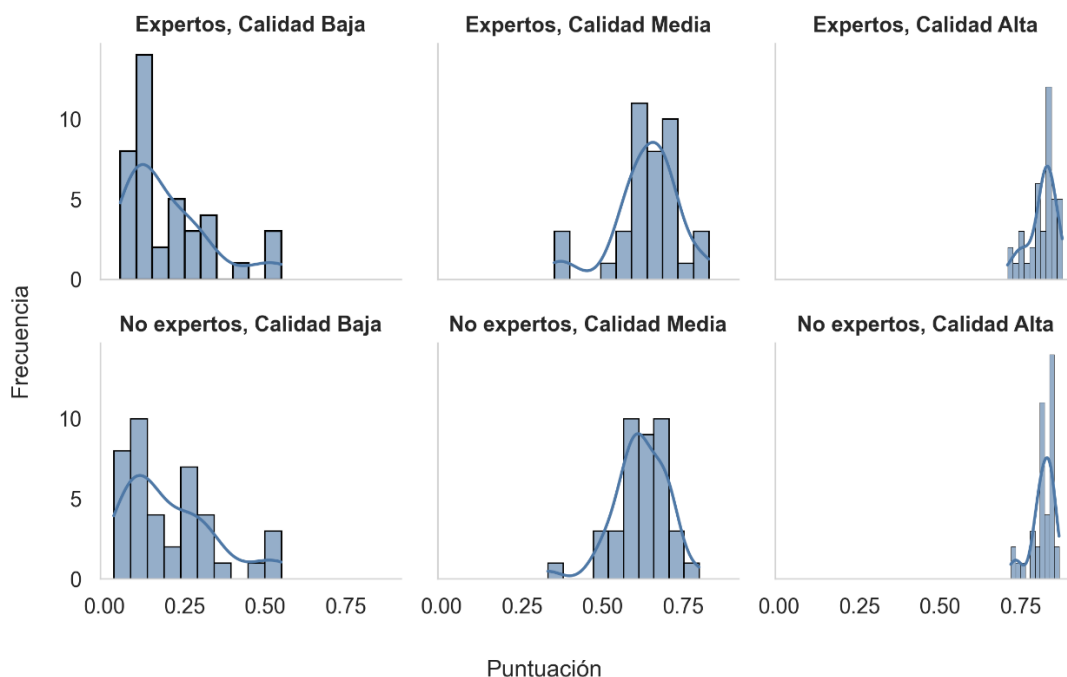


Fig. 19: Distribución de puntajes asignados por expertos y no expertos en cada nivel de calidad

Las distribuciones se presentan organizadas en dos filas: la superior corresponde al grupo de no expertos y la inferior al grupo de expertos. Las columnas representan los tres niveles de calidad.

Doctorado en Informática

Para evaluar la normalidad de las distribuciones, se aplicó la prueba de Shapiro-Wilk (Hanusz et al., 2016) en cada subconjunto. En todos los casos se rechazó la hipótesis de normalidad, lo que motivó el uso de pruebas no paramétricas en los análisis siguientes. El nivel de significancia considerado fue $\alpha = 0,05$.

De manera adicional, se aplicó la prueba de Levene para comprobar la homogeneidad de varianzas (M. B. Brown & Forsythe, 1974). No se encontraron diferencias significativas entre los grupos ($p = 0,785$), ni entre los niveles de calidad. Esto sugiere que el conjunto de datos presenta una varianza homogénea, adecuada para el análisis de hipótesis.

En términos descriptivos, se observó que las visualizaciones de baja calidad recibieron puntajes bajos de forma consistente por ambos grupos, con distribuciones sesgadas hacia la izquierda. En cambio, las visualizaciones de calidad media y alta fueron evaluadas con mayor amplitud de puntajes, especialmente por parte del grupo de no expertos. Los expertos tendieron a asignar puntajes más concentrados, particularmente en las visualizaciones de alta calidad.

5.4.5.2. Discriminación entre niveles de calidad

Esta sección analiza la hipótesis H1, que postula que el modelo permite distinguir entre visualizaciones de alta, media y baja calidad. Es decir, se espera que las visualizaciones de mayor calidad reciban puntajes más altos que las de calidad media o baja, que reflejen de manera consistente sus diferencias. La formulación estadística fue la siguiente:

- **H1₀:** No existen diferencias significativas entre los puntajes asignados a visualizaciones clasificadas como alta, media o baja calidad.
- **H1₁:** Existen diferencias significativas entre los puntajes asignados a los tres niveles de calidad.

El puntaje promedio obtenido por las visualizaciones clasificadas como de alta calidad fue de 0,818 (DE = 0,039), las de calidad media alcanzaron un valor de 0,628 (DE = 0,093), y las de baja calidad un promedio de 0,2 (DE = 0,134). Las medianas fueron 0,830, 0,625 y 0,150 respectivamente. Cada grupo incluyó 80 evaluaciones. Tal como se observa

Doctorado en Informática

en la Fig. 20, los puntajes tienden a aumentar de forma progresiva con el nivel de calidad asignado, tanto en el grupo de expertos como en el de no expertos.

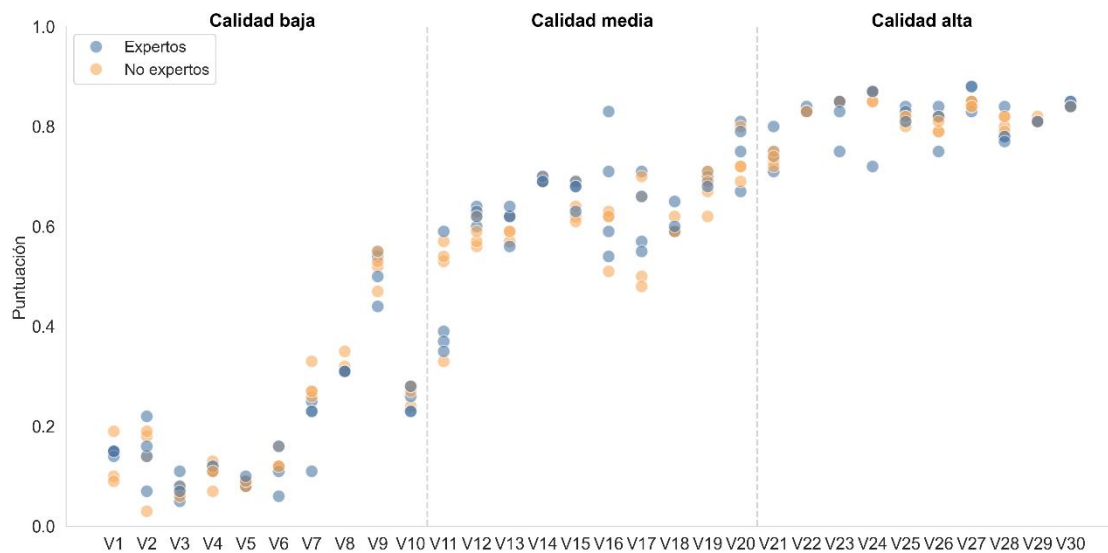


Fig. 20: Puntajes asignados por expertos y no expertos a las visualizaciones (V) de cada nivel

Para verificar si las diferencias eran estadísticamente significativas, se aplicó la prueba no paramétrica de Kruskal-Wallis (McKight & Najab, 2010). El resultado fue $H = 206,95$ con un valor $p < 0,001$ ($p = 1,15e-45$). Dado que el valor p se encuentra por debajo del umbral de significancia ($\alpha = 0,05$), se rechaza la hipótesis nula. Esto respalda la hipótesis $H1_1$ y sugiere que el modelo permite distinguir entre los niveles de calidad definidos.

De forma complementaria, se analizó también la validez de las asociaciones entre criterios y propiedades del modelo. Para ello, se utilizó la encuesta completada por los participantes al finalizar la evaluación. El coeficiente de concordancia de Kendall (Field, 2005) se utilizó para estimar el nivel de acuerdo entre participantes. Los valores obtenidos fueron $W = 0,94$ en el grupo de expertos y $W = 0,89$ en el grupo de no expertos, lo que indica una concordancia elevada en ambos casos. En 35 de las 40 combinaciones evaluadas (87,5 %), el rango intercuartil (RI) fue igual o menor a 1, lo que indica una alineación consistente en las valoraciones, independientemente del nivel de experiencia.

Para considerar una propiedad como válida se definieron dos criterios: (a) que la mediana de respuestas fuera igual o superior a 4, y (b) que al menos el 65 % de las respuestas se

Doctorado en Informática

ubicaran en los valores superiores de la escala. Las propiedades que cumplieron ambos criterios se consideraron validadas; aquellas que cumplieron solo el segundo, parcialmente validadas; y las que no cumplieron ninguno, no validadas. La Tabla 12 resume los resultados, incluyendo la mediana y el RI para cada combinación. Las celdas en gris claro indican propiedades parcialmente validadas; las celdas en gris con borde, propiedades no validadas; y las celdas sin sombreado indican las propiedades completamente validadas.

Tabla 12: Resultados de la encuesta sobre asociaciones entre criterios y propiedades

	Comprensión	Compromiso	Información	Memorabilidad	Usabilidad
Color	3 (3, 3)	3 (3, 3)		3 (3, 3)	
Orden visual	4 (4, 4)	3 (2, 3)			
Enfoque	5 (5, 5)		3 (2, 3)	3 (2, 4)	
Accesibilidad	4 (4, 4)				4 (3, 5)
Claridad	5 (5, 5)	3 (3, 4)		4 (3, 4)	4 (3, 5)
Contexto	3 (3, 4)	3 (3, 3)	3 (3, 3)	3 (3, 3)	
Narrativa	4 (3, 4)	5 (5, 5)		4 (3, 4)	
Precisión	5 (5, 5)		5 (5, 5)		
Saliencia	3 (3, 4)				3 (3, 3)
Guía al Usuario		3 (3, 4)		3 (2, 3)	3 (3, 4)
Metonimia				2 (2, 3)	
Interactividad		4 (4, 5)			3 (3, 4)
Disposición			2 (2, 3)		3 (3, 4)
Fiabilidad			5 (5, 5)		3 (3, 3)
Redundancia			2 (2, 2)	3 (2, 3)	
Integridad			5 (5, 5)		
Consistencia		3 (2, 3)			3 (3, 4)

Las propiedades mejor valoradas en cuanto a su asociación con los criterios fueron Precisión, Enfoque y Narrativa, que recibieron puntuaciones altas de forma consistente. En contraste, Redundancia (asociada al criterio Información) y Metonimia (asociada a Memorabilidad) mostraron valoraciones más dispersas, en algunos casos por debajo del umbral de validación. En particular, Redundancia presentó una mediana inferior al umbral, lo que sugiere una percepción negativa respecto a su vínculo con Información, aunque su asociación con Memorabilidad fue más neutral.

Doctorado en Informática

Del total de 17 propiedades, 16 fueron validadas para al menos un criterio. Enfoque, Claridad, Narrativa, Precisión, Confiabilidad e Integridad cumplieron ambos criterios en más de un caso. Solo una propiedad, Metonimia, no fue validada en ninguna de las dimensiones evaluadas. La lista final de criterios, propiedades y aplicaciones puede consultarse en el Apéndice A.

5.4.5.3. Consistencia entre evaluaciones dentro de los grupos

La hipótesis H2 postula que las evaluaciones realizadas por participantes expertos y no expertos son consistentes con los puntajes de referencia generados previamente por el modelo. La intención es examinar si el uso del modelo produce resultados comparables, independientemente del nivel de experiencia del evaluador. Esto permite evaluar la robustez del modelo frente a distintos perfiles de usuario, y su aplicabilidad en contextos donde no siempre se cuenta con expertos. La formulación estadística fue la siguiente:

- **H2₀:** No existen diferencias significativas entre los puntajes asignados por los participantes y los puntajes de referencia, tanto para expertos como para no expertos.
- **H2₁:** Existen diferencias significativas entre los puntajes asignados por los participantes y los puntajes de referencia, ya sea en el grupo de expertos o en el de no expertos.

Doctorado en Informática

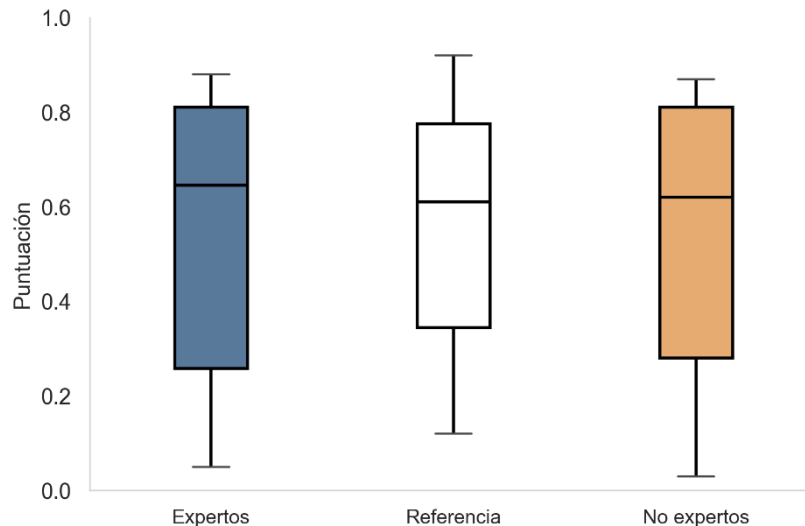


Fig. 21: Puntuaciones asignadas por expertos, no expertos y evaluaciones de referencia por nivel de calidad

Para cada participante, se calcularon los puntajes de calidad asignados a las visualizaciones, y se los comparó con los puntajes de referencia previamente establecidos. El promedio de los puntajes asignados por el grupo de expertos fue de 0,550 (DE = 0,280), mientras que en el grupo de no expertos fue de 0,548 (DE = 0,274). Las medianas respectivas fueron 0,555 y 0,550. La Fig. 21 muestra la distribución de los puntajes de cada grupo en relación con los puntajes de referencia, diferenciados por nivel de calidad.

Se aplicó la prueba de Wilcoxon (Woolson, 2005) para muestras relacionadas para contrastar las evaluaciones de cada grupo con los puntajes de referencia. Esta prueba resulta adecuada cuando se comparan dos mediciones realizadas sobre los mismos elementos. En el grupo de expertos, el estadístico de la prueba fue 3401,50 con un valor p de 0,655. En el grupo de no expertos, el estadístico fue 2986,50 con un valor p de 0,410. En ambos casos, los valores p superan el umbral de significancia ($\alpha = 0,05$), por lo que no se identificaron diferencias estadísticamente significativas entre los puntajes asignados por los participantes y los puntajes de referencia.

Estos resultados respaldan la hipótesis nula (H_0), y sugieren que el modelo produce evaluaciones consistentes cuando es utilizado por distintos tipos de usuarios. La

Doctorado en Informática

coincidencia entre los puntajes indica que tanto personas con experiencia como sin experiencia previa en visualización pueden aplicar el modelo con resultados similares.

5.4.5.4. Correlación con las evaluaciones de referencia

La hipótesis H3 propone que los puntajes asignados por los participantes (expertos y no expertos) guardan una relación directa con los puntajes de referencia generados por el modelo. El objetivo es observar si, a medida que aumenta el puntaje de referencia de una visualización, los participantes tienden a asignar también puntajes más altos, lo que indicaría un patrón de evaluación coherente. La formulación estadística fue la siguiente:

- **H3₀**: No existe correlación significativa entre los puntajes asignados por los participantes y los puntajes de referencia.
- **H3₁**: Existe una correlación significativa entre los puntajes asignados por los participantes y los puntajes de referencia.

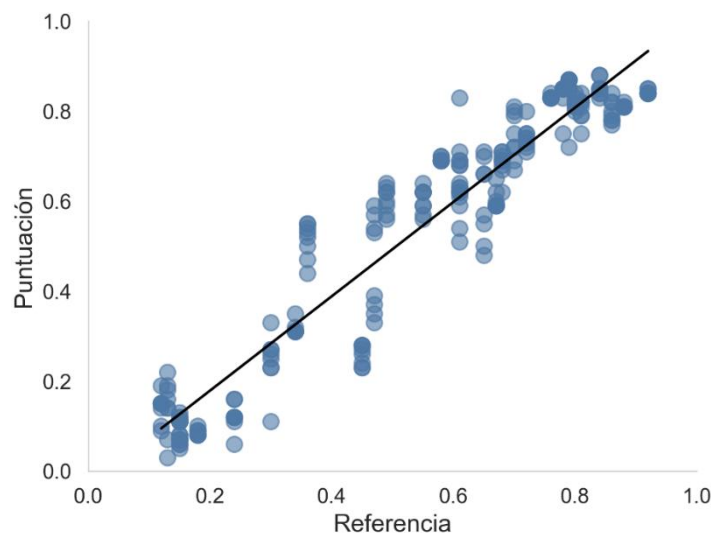


Fig. 22: Correlación entre las puntuaciones de los evaluadores y las evaluaciones de referencia

Se utilizó el coeficiente de correlación por rangos de Spearman (Sedgwick, 2014) para estimar el grado de asociación entre ambos conjuntos de puntajes. Esta medida permite evaluar la fuerza y dirección de la relación entre dos variables ordinales o cuantitativas. El valor obtenido fue $\rho = 0,931$ con un valor $p = 3,21e-106$. Este resultado indica una

Doctorado en Informática

correlación positiva muy fuerte entre los puntajes asignados por los participantes y los puntajes de referencia. La Fig. 22 ilustra esta relación, mostrando cómo a mayor puntaje de referencia, tiende a observarse un mayor puntaje asignado por los evaluadores.

Dado el valor elevado del coeficiente y la significancia estadística obtenida ($p < 0,05$), se rechaza la hipótesis nula (H_{30}). Estos resultados sugieren que existe una relación entre las evaluaciones realizadas por los participantes y los puntajes definidos por el modelo, lo que refuerza la validez de las estimaciones generadas mediante su aplicación.

5.4.5.5. Diferencias entre expertos y no expertos

Finalmente, la hipótesis H_4 plantea que pueden existir diferencias significativas entre los puntajes asignados por participantes expertos y no expertos. El objetivo es determinar si el nivel de experiencia influye en la forma en que se utiliza el modelo, más allá de su consistencia con los puntajes de referencia.

La formulación estadística fue la siguiente:

- **H_{40} :** No existen diferencias significativas entre los puntajes asignados por expertos y no expertos.
- **H_{41} :** Existen diferencias significativas entre los puntajes asignados por expertos y no expertos.

Para contrastar ambos grupos, se aplicó la prueba de Mann-Whitney U (McKnight & Najab, 2010) para muestras independientes. El resultado fue un estadístico $U = 7344,5$ con un valor $p = 0,789$. Este valor se encuentra por encima del umbral de significancia ($\alpha = 0,05$), por lo que no se identificaron diferencias estadísticamente significativas entre los puntajes asignados por expertos y no expertos.

Este resultado respalda la hipótesis nula (H_{40}), e indica que el modelo puede ser utilizado de manera similar por personas con distintos niveles de experiencia, generando evaluaciones comparables en cuanto a los puntajes asignados.

Doctorado en Informática**5.4.5.6. Influencia del nivel de experiencia en los puntajes**

Para complementar los análisis anteriores, se utilizó un modelo estadístico de efectos mixtos. Este tipo de modelo permite estudiar varios factores al mismo tiempo, considerando también las posibles diferencias entre quienes respondieron y las visualizaciones que fueron evaluadas. De esta manera, es posible estimar efectos generales sin que las características individuales de los participantes o de las visualizaciones distorsionen los resultados.

En este caso, se analizaron tres aspectos principales: el grupo al que pertenecía cada participante (experto o no experto), el nivel de calidad asignado previamente a cada visualización (baja, media o alta), y la combinación entre ambos (es decir, si la relación entre el nivel de calidad y el puntaje era distinta según la experiencia del evaluador).

Además, el modelo incorporó dos fuentes de variación adicionales: por un lado, el hecho de que cada persona tiene una forma particular de puntuar (aunque use el mismo modelo), y por otro, que algunas visualizaciones pueden resultar más fáciles o difíciles de evaluar. Estas diferencias se incluyeron como efectos aleatorios, lo cual permite ajustar los resultados sin que estas variaciones individuales interfieran con las conclusiones generales.

El modelo estimó un puntaje promedio de 0,817 para las visualizaciones de alta calidad evaluadas por participantes expertos ($p < 0,001$). La diferencia entre expertos y no expertos fue de 0,002 puntos ($p = 0,928$), sin evidencia estadística de una diferencia entre grupos.

En cambio, el nivel de calidad de las visualizaciones sí tuvo un efecto sobre los puntajes. Las visualizaciones clasificadas como de calidad baja recibieron, en promedio, 0,621 puntos menos que las de calidad alta ($p < 0,001$) y las de calidad media 0,181 puntos menos ($p < 0,001$). Estas diferencias indican que el modelo asignó puntajes diferenciados según el nivel de calidad predefinido, lo que respalda su capacidad para discriminar entre visualizaciones de distinta calidad.

Doctorado en Informática

La interacción entre grupo y nivel de calidad no resultó significativa. Los valores p para las combinaciones grupo–calidad fueron 0,831 para visualizaciones de calidad baja y 0,507 para las de calidad media, lo que indica que ambos grupos asignaron puntajes similares frente a visualizaciones del mismo nivel de calidad.

Por último, se observó que las diferencias entre visualizaciones explicaban una parte pequeña pero existente de la variación total (varianza = 0,007), mientras que las diferencias entre personas fueron mínimas (varianza = 0,000249).

En conjunto, los resultados indican que los puntajes estuvieron asociados principalmente al nivel de calidad de las visualizaciones, sin diferencias significativas entre expertos y no expertos. Tampoco se observaron interacciones entre estos factores, lo que muestra un patrón de evaluación similar en ambos grupos.

5.4.5.7. Resultados de la encuesta final

Esta sección presenta los resultados obtenidos en la segunda encuesta completada por los participantes, centrada en la experiencia de uso de la herramienta de evaluación. La encuesta incluyó preguntas cerradas con escalas de Likert y una pregunta abierta destinada a recoger comentarios adicionales. La Fig. 23 muestra el promedio de respuestas para cada ítem cuantitativo, junto con la desviación estándar correspondiente.

La pregunta sobre la comprensión del modelo, que hacía referencia a la claridad de las instrucciones y los enunciados, obtuvo un promedio de 4,35 (DE = 0,74). La utilidad percibida de las aplicaciones fue evaluada con un promedio de 3,95 (DE = 0,60), mientras que la claridad en la relación entre criterios y aplicaciones recibió un promedio de 4,03 (DE = 0,66). El ítem sobre el apoyo recibido durante la evaluación obtuvo un promedio de 3,7 (DE = 0,72), y la satisfacción general fue calificada con un promedio de 3,8 (DE = 0,46).

Doctorado en Informática

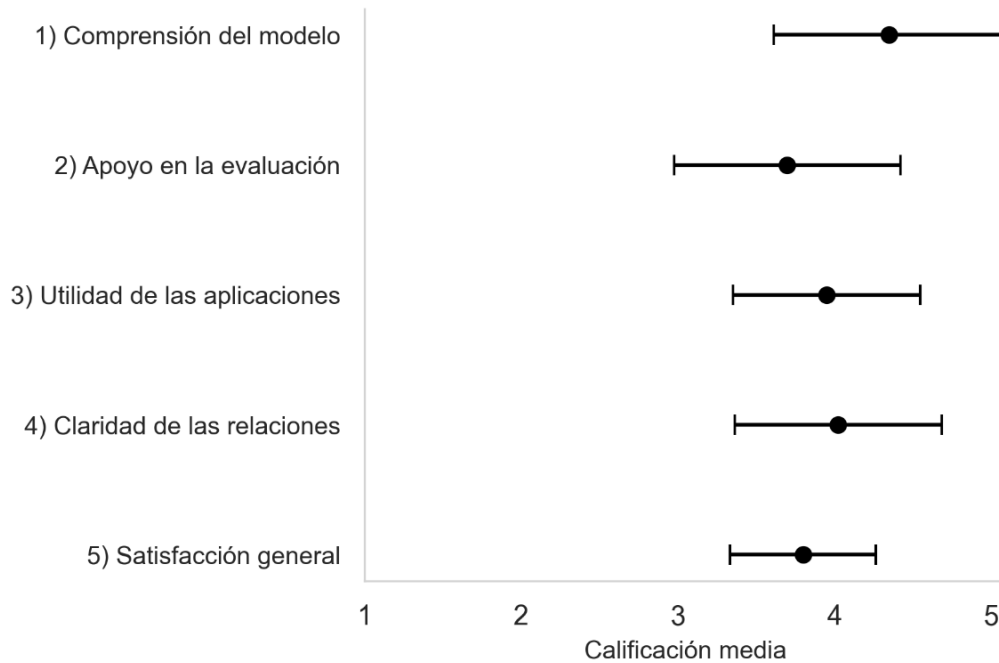


Fig. 23: Calificaciones promedio en la encuesta final

Para analizar las respuestas de la pregunta abierta, se aplicó un modelo de tópicos basado en asignación latente de Dirichlet (Blei & Lafferty, 2009). Esta técnica permite identificar temas recurrentes a partir de patrones léxicos en los textos. El modelo se configuró para extraer cinco tópicos, cada uno representado por un conjunto de palabras clave. La Tabla 13 resume los resultados.

Tabla 13: Principales temas y palabras clave asociadas

Tópico	Palabras clave
Claridad de las explicaciones	Pensar, explicaciones, evaluar, eficazmente, detallado, referencia, directamente, acción, evaluaciones, útil
Mejoras de la interfaz	Simple, interfaz, mejorar, hacer, diferente, navegación, complejo, secciones, función, tiempo real
Proceso de evaluación	Evaluación, mejor, ayuda, usuarios, interactivo, podría, entender, añadir, sería, criterios
Terminología	Diferente, sería, estándares, expectativas, incluir, útil, bueno, visualización, términos, glosario
Contextos de aplicación	Aplicar, completado, eficazmente, contextos, útil, como, escenarios, más claro, visual, evaluar

Doctorado en Informática

El primer tópico se vinculó con la claridad de las explicaciones. Algunos participantes sugirieron incluir ejemplos interactivos dentro de las secciones de ayuda o mostrar casos comunes de errores en visualizaciones. Estos comentarios coinciden con la mayor dispersión observada en el ítem de apoyo durante la evaluación.

El segundo tópico agrupó comentarios sobre la interfaz. Entre las sugerencias se mencionaron mejoras en la navegación entre secciones, así como la incorporación de retroalimentación en tiempo real o indicadores visuales durante el proceso de evaluación.

El tercer tópico se relacionó con el proceso de evaluación. Varias respuestas propusieron incluir ejemplos comparativos o un breve tutorial inicial. Algunas personas indicaron que no tuvieron dificultades, mientras que otras expresaron interés en contar con elementos adicionales de apoyo.

Los últimos dos tópicos abordaron cuestiones relacionadas con la terminología utilizada y con los contextos en los que podrían aplicarse los criterios. En varios casos se solicitó un glosario o ejemplos más concretos sobre cómo interpretar los términos utilizados en las preguntas.

Estos resultados permitieron avanzar hacia una nueva iteración del artefacto, centrada en facilitar su aplicación y reducir la carga cognitiva durante la evaluación. La Fig. 24 resume el estado de avance al cierre de esta segunda etapa. Los tres objetivos de uso en entorno controlado fueron cumplidos y se construyó una aplicación web para operacionalizar el modelo.

Doctorado en Informática

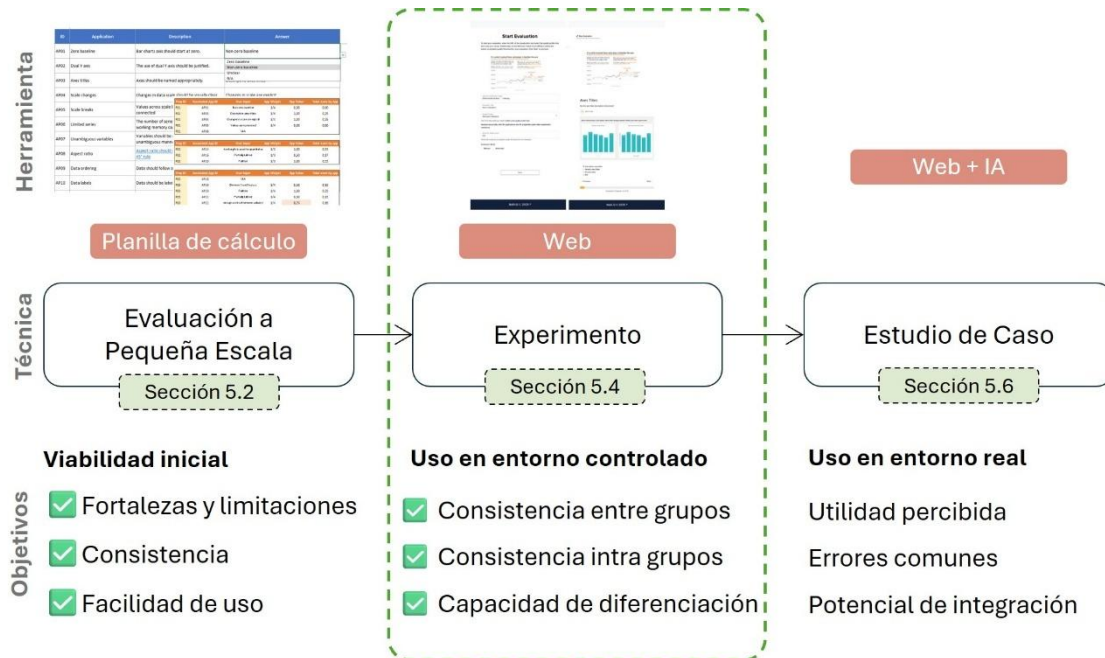


Fig. 24: Cumplimiento de objetivos en la segunda etapa de validación.

5.5. Segunda iteración de diseño del artefacto

Los resultados del experimento ofrecieron evidencia empírica sobre la capacidad del modelo para diferenciar niveles de calidad y generar evaluaciones consistentes entre distintos perfiles de usuario. Al mismo tiempo, revelaron oportunidades de mejora vinculadas con la carga cognitiva del proceso, la claridad de enunciados y la necesidad de asistencia durante la evaluación. A partir de estos hallazgos, se llevó a cabo una segunda iteración del artefacto, orientada a facilitar su aplicación mediante recursos de apoyo adicionales.

En esta etapa se incorporaron mecanismos complementarios de asistencia: por un lado, verificaciones automáticas basadas en algoritmos, y por otro, sugerencias generadas mediante modelos de lenguaje. Ambos recursos están diseñados para asistir al usuario durante la evaluación, sin reemplazar su juicio.

Doctorado en Informática**5.5.1. Automatización de aplicaciones**

Se identificaron aplicaciones del modelo cuya evaluación podía automatizarse mediante técnicas programáticas. En particular, se implementaron verificaciones automáticas para las aplicaciones de contraste entre elementos y accesibilidad de colores para personas con discapacidades visuales.

Estas funcionalidades procesan la imagen cargada por el usuario utilizando librerías de análisis de imágenes, y estiman la opción más adecuada dentro de las alternativas definidas en el modelo. Los resultados se presentan como información de apoyo dentro de la interfaz de evaluación manual, permitiendo al usuario responder de manera confiable preguntas donde la interpretación puede ser más compleja.

5.5.2. Evaluación asistida por IA

Estudios recientes han demostrado que los modelos de lenguaje multimodales (LLMs con capacidad de procesar texto e imagen) pueden asistir de manera efectiva en la evaluación de visualizaciones, ya sea brindando retroalimentación sobre aspectos de diseño (Shin et al., 2024), detectando elementos engañosos (Lo & Qu, 2024) o interpretando gráficamente información visual (He et al., 2024). Estos trabajos destacan el potencial de estos modelos para ofrecer sugerencias comprensibles y accionables, especialmente cuando se combinan con filtros perceptuales u otras técnicas automáticas.

Esta evidencia, sumada a los hallazgos del experimento (donde varios participantes manifestaron la necesidad de apoyo adicional para interpretar algunas aplicaciones) motivó la incorporación de mecanismos de asistencia basados en IA utilizando un modelo de lenguaje multimodal (GPT-4).

En primer lugar, se integró un sistema de ayuda contextual en la modalidad de evaluación manual para asistir al usuario cuando no sabe cómo responder una pregunta específica. Para ello, el sistema considera el enunciado de la aplicación evaluada, su texto de ayuda y la imagen de la visualización cargada. A partir de esa información, el modelo sugiere la opción más adecuada entre las disponibles. La instrucción enviada al modelo tiene la siguiente estructura:

Doctorado en Informática

System prompt:

Eres un/a experto/a en la evaluación de visualizaciones narrativas. Basándote en la imagen de la visualización y su descripción, selecciona la opción de respuesta más adecuada para la siguiente pregunta y justifica brevemente tu elección siguiendo el formato indicado.

User prompt:

Pregunta: <Nombre de la aplicación>

Texto de ayuda: <Texto de ayuda de la aplicación>

Opciones de respuesta:

<Primera opción>

<Segunda opción>

<Tercera opción>

...

Selecciona solo una opción exactamente como está escrita en la lista y justifica tu elección. La salida debe seguir este formato exacto:

<Respuesta seleccionada>

Explicación: <Justificación>

La Fig. 25 ilustra cómo se presenta esta funcionalidad en la interfaz, con una sugerencia basada en el contenido analizado por el modelo.

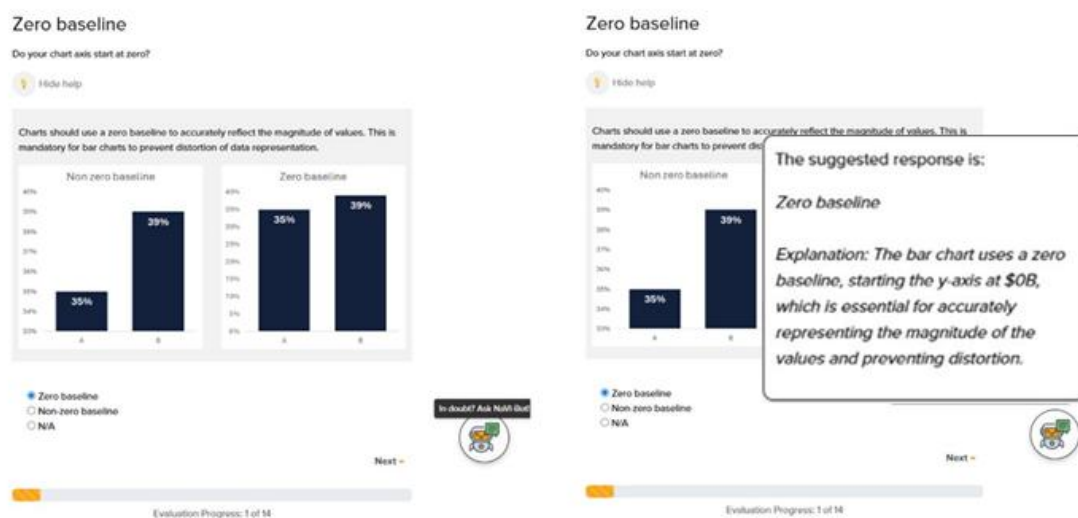


Fig. 25: Ayuda contextual generada por IA durante la evaluación manual.

Doctorado en Informática

Por otro lado, se desarrolló una modalidad de evaluación automática. En este caso, además de la imagen, se incluye una descripción contextual proporcionada por el usuario que indica el propósito de la visualización, sus características principales y otros aspectos relevantes. A partir de esta información, el modelo analiza la visualización y sugiere respuestas para cada una de las aplicaciones del modelo de calidad. La estructura de las instrucciones sigue un formato similar, iterando por cada aplicación.

Las respuestas sugeridas se almacenan y se utilizan para calcular automáticamente el puntaje final, aunque el usuario puede revisarlas y modificarlas si decide editar la evaluación. Esto permite obtener un reporte completo sin intervención humana directa, manteniendo la posibilidad de ajuste manual posterior.

A partir de estas mejoras, el sistema contempla dos modalidades de uso. En la evaluación manual, el usuario responde cada pregunta con apoyo visual y sugerencias automatizadas puntuales cuando lo requiere. En la modalidad asistida, el sistema genera una propuesta completa de respuestas a partir de la información provista. En ambos casos, el proceso finaliza con la generación de un reporte que resume los puntajes obtenidos y brinda recomendaciones de mejora. Este flujo se resume en la Fig. 26.

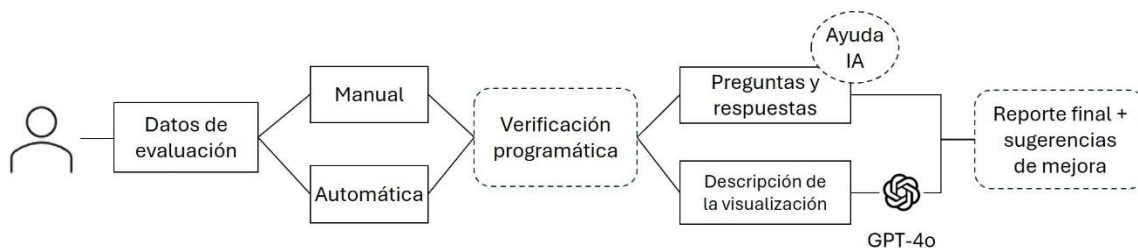


Fig. 26: Esquema del proceso de evaluación

Una vez obtenido el reporte, si se identificaron aspectos de mejora, el sistema permite solicitar una sugerencia específica para cada recomendación. Estas sugerencias son generadas mediante IA a partir de la visualización evaluada y la sugerencia obtenida, con el objetivo de guiar al usuario en la implementación concreta de los ajustes. La Fig. 27 muestra un ejemplo de esta funcionalidad, donde el modelo genera una propuesta concreta de mejora visual en respuesta a una recomendación del sistema.

Doctorado en Informática

Road to your **Minimum Viable Score**

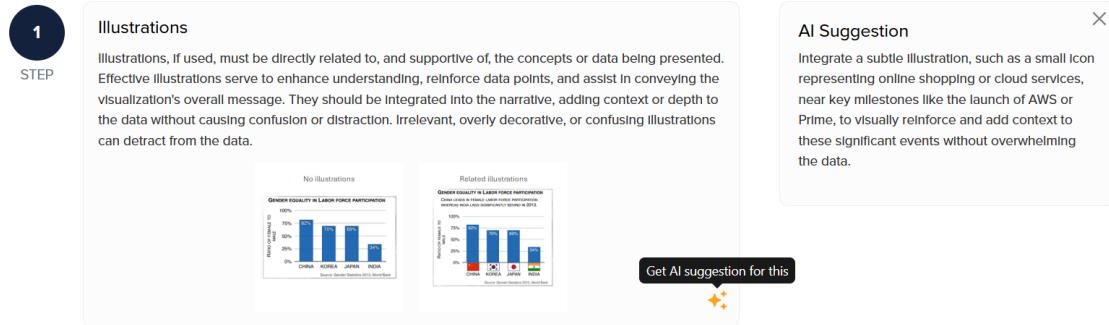


Fig. 27: Sugerencia de mejora generada por IA

Las instrucciones enviadas al modelo, tanto para la generación de respuestas como para las sugerencias de mejora, fueron ajustadas iterativamente a lo largo del desarrollo para mejorar la precisión y coherencia de las respuestas generadas. Ejemplos completos de cada modalidad pueden consultarse en el repositorio del proyecto⁵.

Con estas mejoras implementadas, se avanzó hacia la siguiente etapa de validación del modelo. La sección que sigue presenta el estudio de caso diseñado para observar su funcionamiento en un entorno aplicado, utilizando visualizaciones reales y sin intervención experimental directa.

5.6. Estudio de caso

Este estudio tuvo como objetivo observar el funcionamiento de NaVi-Q en un entorno aplicado, mediante su integración en un proyecto real de visualización. Para ello, se siguieron las pautas metodológicas propuestas por Yin (2017) para el análisis de tecnologías en contextos naturales y se formularon las siguientes PI:

PI-1: ¿Qué grado de utilidad tienen las recomendaciones de mejora proporcionadas por NaVi-Q para el equipo de trabajo? Esta pregunta busca explorar cómo percibe el

⁵ https://github.com/andreelezcano/naviq_ai

Doctorado en Informática

equipo la utilidad práctica de las recomendaciones generadas, considerando su aplicabilidad en el contexto del proyecto.

PI-1.a: ¿Qué errores o aspectos a mejorar son identificados con mayor frecuencia?

Se pretende caracterizar los problemas más comunes detectados en el trabajo del equipo, a partir de las evaluaciones realizadas mediante el modelo.

PI-2: ¿Cuál es el impacto de aplicar las mejoras sugeridas por el modelo sobre la comprensión, la información y la usabilidad de las visualizaciones? El objetivo es validar si las recomendaciones generadas por NaVi-Q tienen un efecto tangible sobre criterios considerados prioritarios en el dominio de aplicación.

PI-3: ¿Cuál es la percepción general sobre el uso del modelo? Esta pregunta apunta a explorar de qué manera los equipos de trabajo incorporarían NaVi-Q en su flujo habitual, y qué valor perciben en su utilización.

5.6.1. Contexto

El estudio se llevó a cabo en colaboración con el grupo de investigación SWK de la Universidad de Hamburgo, en el marco de un proyecto orientado a mejorar la gestión de deuda técnica en un entorno de desarrollo real. Este proyecto fue implementado en una empresa de tecnología y combinó talleres, entrevistas y análisis de datos para establecer un proceso de gestión de deuda técnica. Una de sus fases se centró en el uso de visualizaciones como herramienta de apoyo a la toma de decisiones.

En esta instancia, el equipo diseñó y discutió múltiples visualizaciones interactivas para dos tipos de público: el equipo de desarrollo de la empresa y la gestión técnica. Las mismas abordaban aspectos como la priorización de elementos técnicos en la lista de pendientes (también conocido como *backlog*), la identificación de oportunidades de mejora de bajo costo y el seguimiento de decisiones de mantenimiento. La Fig. 28 muestra una selección representativa de las visualizaciones utilizadas.

Doctorado en Informática

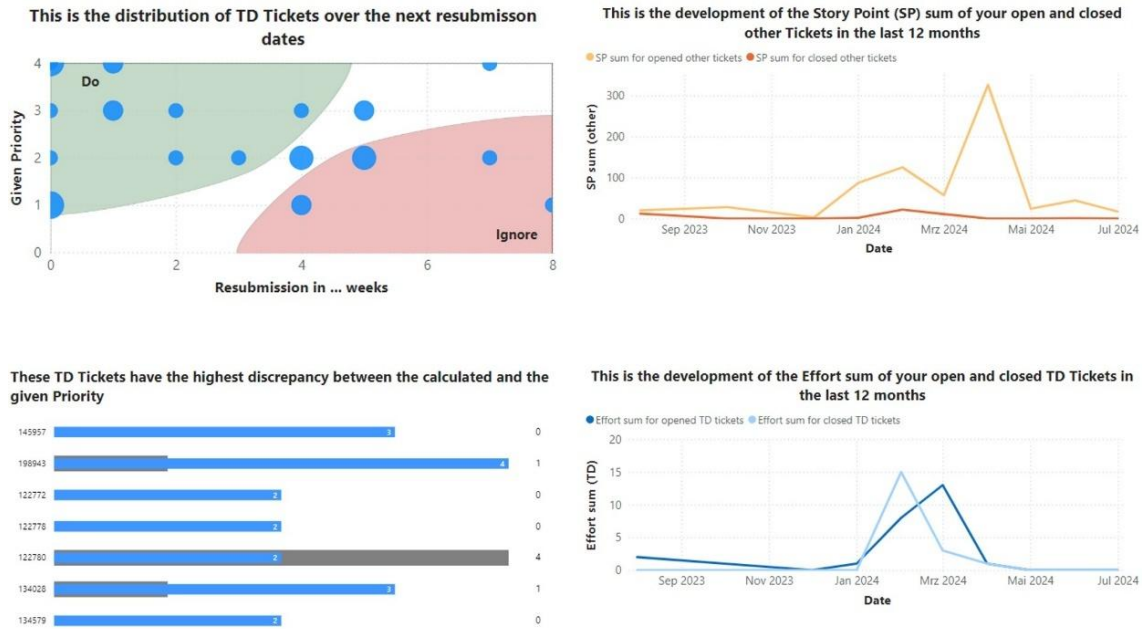


Fig. 28: Ejemplos de visualizaciones evaluadas en el estudio

5.6.2. Procedimiento

El estudio se estructuró en tres fases consecutivas: la evaluación inicial de las visualizaciones utilizando el modelo, el análisis colaborativo de las sugerencias generadas, y la implementación de mejoras sobre las visualizaciones. Previo a la evaluación, se definió un perfil de calidad personalizado, adaptado al dominio específico de aplicación. Para ello, los miembros del equipo completaron formularios donde indicaron qué criterios y propiedades consideraban más relevantes y establecieron su importancia relativa. Se aplicó el método AHP-Express para obtener los pesos finales de cada criterio y propiedad.

La Tabla 14 muestra los pesos asignados a cada criterio de calidad. El detalle completo del perfil se encuentra disponible en el Apéndice E.

Tabla 14: Pesos asignados a los criterios del perfil de calidad personalizado

ID Criterio	Descripción	Peso
C01	Comprensión	0,30
C02	Compromiso	0,10
C03	Información	0,29

Doctorado en Informática

ID Criterio	Descripción	Peso
C04	Memorabilidad	0,11
C05	Usabilidad	0,20

La evaluación se llevó a cabo utilizando la modalidad manual. Un integrante del grupo de investigación SWK aplicó la herramienta para analizar cada una de las visualizaciones desarrolladas en el proyecto. Durante el proceso, respondió las preguntas correspondientes a cada aplicación del modelo y, en los casos en que surgieron dudas, consultó la ayuda contextual asistida por IA disponible en la interfaz. Si bien la evaluación fue realizada por una sola persona, las respuestas fueron consensuadas con los otros integrantes del equipo, por lo que se considera representativa del grupo. Cabe señalar que este estudio no tuvo como objetivo analizar la consistencia entre evaluadores, aspecto abordado en el experimento presentado en la sección 5.4.

A partir de las respuestas registradas, el sistema identificó las aplicaciones con menor puntuación y generó sugerencias de mejora correspondientes. Las mismas fueron discutidas de manera conjunta por los miembros del equipo, quienes decidieron cuáles implementar en función del tiempo y el esfuerzo requeridos.

Con las visualizaciones mejoradas, se aplicaron distintos instrumentos para evaluar el impacto y la utilidad de las mejoras, identificar patrones de errores comunes y relevar la percepción general sobre el uso del modelo.

5.6.3. Recolección de datos

Para responder a la PI-1 y a la PI-1.a, una vez finalizadas las evaluaciones y generadas todas las sugerencias de mejora, se consolidaron las recomendaciones en una lista única. Cada sugerencia fue presentada a los miembros del equipo junto con su respectiva propuesta concreta de implementación, generada por el sistema mediante asistencia automatizada. De esta forma, se distinguió entre la sugerencia general —correspondiente a una aplicación del modelo con baja puntuación— y la acción específica sugerida para mejorar la visualización en base a ella. El equipo calificó la sugerencia como “útil”, “parcialmente útil” o “no útil”, indicó si ya conocía la buena práctica asociada al problema

Doctorado en Informática

señalado y si la mejora sería implementada en el rediseño de las visualizaciones. A partir de esta misma lista, se realizó una clasificación cualitativa de los aspectos de mejora detectados, lo que permitió identificar patrones comunes y agrupar las recomendaciones según el tipo de problema señalado por el modelo.

Para la PI-2, se llevó a cabo una evaluación comparativa orientada a analizar el impacto de las mejoras sugeridas por el modelo sobre la comprensión, la información y la usabilidad de las visualizaciones. Estos criterios concentraron el 79 % del peso total en el perfil de calidad definido para el estudio.

La evaluación se realizó con dos grupos de participantes, conformados por 24 estudiantes avanzados de grado en Informática de la Universidad de Hamburgo, asignados aleatoriamente a cada grupo. Uno de los grupos interactuó con las visualizaciones originales desarrolladas en el proyecto (grupo control), mientras que el otro trabajó con las versiones mejoradas tras la implementación de las recomendaciones generadas por el modelo (grupo experimental). Ambos grupos completaron las tareas de forma individual, sin asistencia externa, y en condiciones similares. Se aplicaron dos instrumentos para recoger evidencia sobre los efectos de las mejoras:

Cuestionario de comprensión e información, compuesto por preguntas de opción múltiple con una única respuesta correcta, centradas en tareas concretas de lectura, identificación y comparación, ajustadas al contenido específico de las visualizaciones utilizadas en el estudio. El formato y tipo de tarea se basaron en el cuestionario de evaluación de alfabetización visual desarrollado por S. Lee et al. (2017). La Tabla 15 muestra las preguntas incluidas.

Doctorado en Informática

Tabla 15: Preguntas del cuestionario de comprensión e información

N.º	Enunciado de la pregunta	Tipo de tarea
1	¿En qué mes se registró el mayor esfuerzo para cerrar tiques de DT?	Encontrar un extremo
2	Según el gráfico de “Oportunidades fáciles de aprovechar”, ¿qué combinación de prioridad y esfuerzo caracteriza a estos ítems?	Caracterizar la distribución
3	¿Cuál es la tendencia general del esfuerzo acumulado en los tiques de DT abiertos en los últimos 12 meses?	Encontrar correlaciones o tendencias
4	¿Cuál es el nivel de discrepancia más alto entre prioridad dada y calculada en los tiques de DT?	Encontrar un extremo
5	¿Qué visualización permite detectar qué tiques tienen un retorno de inversión más temprano?	Recuperar un valor

* DT: deuda técnica

Para cada pregunta, se solicitó a los participantes que calificaran su nivel de confianza en la respuesta seleccionada, mediante una escala de cinco puntos (1 = nada confiado, 5 = completamente confiado). Esta información adicional se utilizó para analizar cambios en la certeza de los usuarios en función de la versión de visualización utilizada.

Escala de Usabilidad (SUS) adaptada, diseñada para estimar la usabilidad percibida del conjunto de visualizaciones. La escala se administró luego de que cada participante completara su interacción con el tablero, y mantuvo la estructura original de Brooke (1996), compuesta por diez ítems formulados en términos afirmativos y negativos. Los enunciados fueron adaptados al contexto del estudio. Cada ítem se respondió mediante una escala tipo Likert de cinco puntos, y el puntaje final se calculó siguiendo el procedimiento estándar propuesto por el autor. La Tabla 16 presenta el conjunto de ítems utilizados. La columna Formulación indica si el ítem se expresa de manera afirmativa o si se trata de una afirmación invertida (es decir, redactado negativamente), lo que permite calcular correctamente el puntaje mediante la inversión de las puntuaciones.

Doctorado en Informática

Tabla 16: Ítems de la escala de Usabilidad (SUS) adaptada

N.º	Enunciado del ítem	Formulación
1	Me parece que usar el tablero fue sencillo.	Afirmativa
2	Considero que necesitaría ayuda técnica para poder usar este tablero.	Invertida
3	Creo que las diferentes partes del tablero están bien integradas.	Afirmativa
4	Me pareció que había demasiada complejidad en el uso del tablero.	Invertida
5	Me sentí seguro/a usando el tablero.	Afirmativa
6	Considero que las visualizaciones eran consistentes entre sí.	Afirmativa
7	Me gustaría usar un sistema como este con frecuencia.	Afirmativa
8	Encontré que las visualizaciones eran confusas.	Invertida
9	Pude aprender a usar el sistema rápidamente.	Afirmativa
10	Necesitaría aprender muchas cosas nuevas antes de poder usar el tablero.	Invertida

Finalmente, para la PI-3, una vez completadas todas las evaluaciones, los miembros del grupo de investigación respondieron una breve encuesta orientada a relevar su percepción sobre el uso de NaVi-Q en contextos de trabajo reales. La encuesta fue administrada de manera individual y consistió en tres preguntas abiertas, diseñadas para captar distintos aspectos de su experiencia:

1. **¿En qué etapa del proceso de creación aplicaría el modelo?**
Motivación: identificar el momento del flujo de trabajo donde se percibe mayor valor en la incorporación de una evaluación estructurada de calidad.
2. **¿En qué tarea o decisión cree que el modelo es más útil?**
Motivación: explorar qué tipo de necesidades o desafíos del proceso de diseño son mejor asistidos mediante el uso del modelo.
3. **¿Qué obstáculos cree que podrían dificultar el uso del modelo?**
Motivación: detectar posibles barreras para la adopción del modelo en escenarios de uso real.

Doctorado en Informática

Las respuestas obtenidas en los distintos instrumentos fueron analizadas de manera cuantitativa y cualitativa, según correspondiera, y sus resultados se presentan en la siguiente sección.

Los materiales utilizados en estudio, incluyendo las visualizaciones evaluadas, el perfil de calidad y los distintos cuestionarios se pueden consultar en el Apéndice E.

5.6.4. Resultados

Esta sección presenta los resultados obtenidos a partir de la aplicación del modelo. El análisis se estructura en tres partes de acuerdo con las preguntas de investigación: la evaluación de la utilidad percibida y los problemas detectados (PI), el impacto de las mejoras aplicadas (PI-2) y la percepción del equipo sobre el uso del modelo (PI-3).

5.6.4.1. Utilidad percibida y patrones de problemas detectados

En total se identificaron 41 recomendaciones asociadas a distintas aplicaciones del modelo. Tal como se describió en la sección 5.6.3, cada recomendación fue evaluada por los miembros del equipo en términos de utilidad percibida, conocimiento previo y aplicabilidad. La Tabla 17 muestra la cantidad de veces que cada recomendación fue generada por el modelo (frecuencia), así como el número de ocasiones en que fue considerada útil, conocida previamente o aplicada, según la evaluación realizada por los miembros del equipo.

Tabla 17: Frecuencia, utilidad, conocimiento y aplicación de las recomendaciones

Recomendación	Frecuencia	Útil	Conocida	Aplicada
Atribución de fuentes	7	1	7	0
Elementos asociados	7	7	2	5
Línea base en cero	6	0	6	0
Ilustraciones	7	2	0	0
Etiquetas de serie	4	4	4	4
Cambios en los datos	4	4	0	4
Mensajes y resúmenes	2	2	0	1
Títulos de ejes	1	1	0	1
Marcas y etiquetas de escala	1	1	0	0
Estructura y flujo	1	1	0	0
Resaltado visual	1	1	0	1

Doctorado en Informática

Como resultado del proceso de clasificación cualitativa mencionado anteriormente, las recomendaciones fueron organizadas en cuatro categorías, según el tipo de problema que buscaban abordar. La Tabla 18 presenta estas categorías junto con los aspectos de mejora incluidos en cada una y una breve justificación de su agrupación.

Tabla 18: Agrupación de recomendaciones según el tipo de problema

Problema detectado	Recomendaciones incluidas	Justificación
Narrativa visual débil	Ilustraciones, Cambios en los datos, Estructura y flujo	No se representan bien los cambios, o falta apoyo visual que guíe la lectura narrativa del gráfico.
Codificación ambigua	Elementos asociados, Etiquetas de serie, Resaltado visual	Elementos visuales mal diferenciados o poco legibles; dificultan la interpretación correcta.
Falta de contexto	Atribución de fuentes, Mensajes y resúmenes	Faltan elementos para entender de qué trata la visualización, cómo leerla o qué quiere decir.
Escala y ejes deficientes	Base en cero, Títulos de ejes, Marcas y etiquetas	Problemas en los ejes, escalas o etiquetas que afectan la precisión y comparabilidad.

Esta clasificación permitió observar diferencias en la utilidad, el conocimiento previo y la aplicación de las sugerencias según el tipo de problema. La Fig. 29 resume esta información.

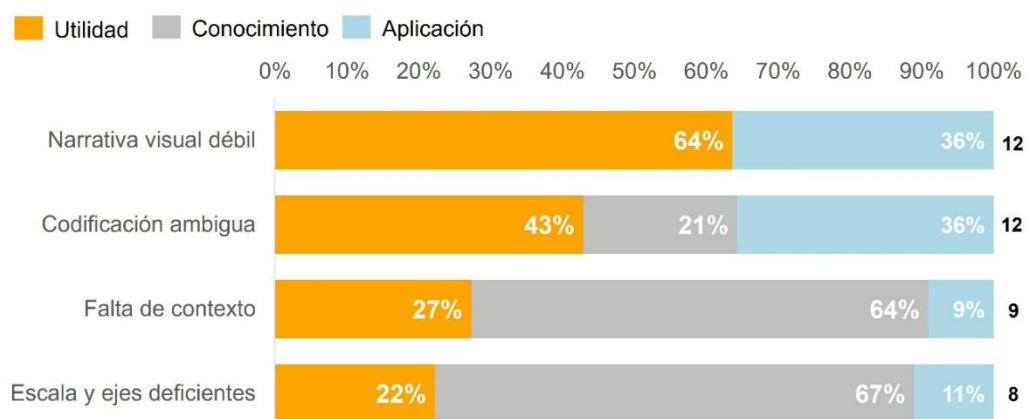


Fig. 29: Utilidad, conocimiento y aplicación de las recomendaciones según tipo de problema identificado. Los valores absolutos indican la frecuencia de cada error

Se observa que la categoría “Narrativa visual débil” concentra el mayor porcentaje de sugerencias consideradas útiles (64 %) y aplicadas (36 %), a pesar de que ninguna había

Doctorado en Informática

sido identificada previamente por el equipo (0 % de conocimiento). Este patrón sugiere que el modelo permitió identificar aspectos que no habían sido considerados previamente por el equipo. Las recomendaciones aportaron contenido nuevo y suficientemente concreto como para ser implementado en el rediseño. En contraste, las categorías “Falta de contexto” y “Escala y ejes deficientes” presentan altos niveles de conocimiento previo (64 % y 67 %, respectivamente), pero baja aplicación (9 % y 11 %). Esto sugiere que, si bien el equipo estaba familiarizado con las buenas prácticas asociadas a estos aspectos, dichas recomendaciones no fueron priorizadas o resultaron difíciles de incorporar en el rediseño.

Un ejemplo de recomendación considerada útil y aplicada es el caso de Elementos Asociados. La sugerencia generada por el sistema proponía: *“Utilice esquemas de color coincidentes para las categorías de tickets abiertos y cerrados en ambos gráficos para crear asociaciones visuales, asegurándose de que se utilice el mismo color para los tickets “abiertos” y “cerrados” en los gráficos de esfuerzo y de puntos de historia para mantener la coherencia.”* Esta mejora fue implementada en el rediseño para facilitar la interpretación visual del contenido.

En cambio, la recomendación referida a Atribución de Fuentes fue reconocida por todos los integrantes del equipo, pero no se implementó, al considerarse de bajo impacto para los objetivos del tablero y poco viable en el formato final.

Estos resultados indican que la utilidad percibida no se vincula necesariamente al conocimiento previo, y que la aplicabilidad está condicionada por la claridad de las sugerencias y el contexto de uso.

5.6.4.2. Impacto de las mejoras aplicadas

Comprensión e información. Para evaluar el impacto de las mejoras aplicadas, los participantes completaron el cuestionario presentado en la sección 5.6.3. Cada pregunta abordó una tarea específica asociada a la comprensión de la información visual, e incluyó una escala de confianza vinculada a la respuesta. El análisis se centró en comparar los

Doctorado en Informática

resultados obtenidos por grupo, considerando tanto la precisión como la confianza reportada.

El grupo control utilizó visualizaciones del proyecto original. El grupo experimental trabajó con versiones rediseñadas que incorporaban recomendaciones seleccionadas por el equipo. En línea con los resultados de la PI-1, la mayoría de las mejoras aplicadas no implicaron cambios estructurales, sino ajustes orientados a reforzar la narrativa o facilitar la interpretación. Estos ajustes incluyeron, por ejemplo, *tooltips* contextuales, resúmenes o elementos de ayuda integrados, aprovechando las características interactivas del tablero.

La Fig. 30 muestra un ejemplo de visualización original (panel izquierdo) y su versión mejorada (panel derecho).

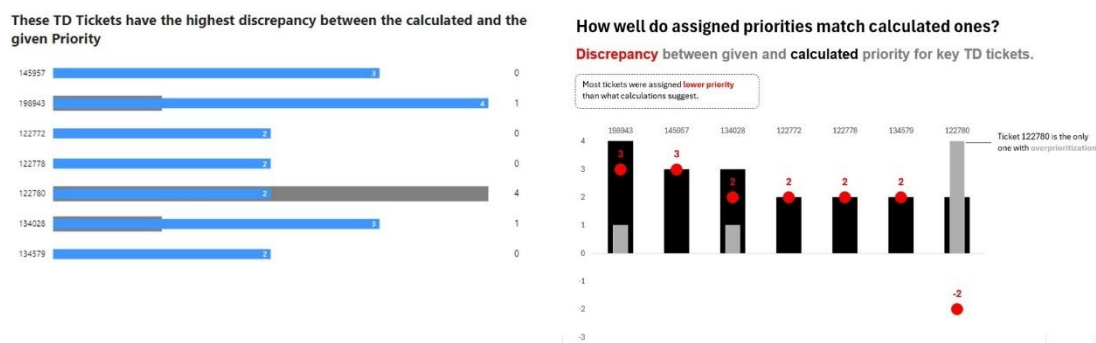


Fig. 30: Ejemplo de visualización original y rediseñada

Los resultados se resumen en la Fig. 31. El gráfico presenta, para cada pregunta del cuestionario (P1 a P5), la proporción de respuestas correctas (barras) y el nivel promedio de confianza (puntos) por grupo. La confianza fue normalizada a una escala de 0 a 1 para facilitar la comparación.

Doctorado en Informática

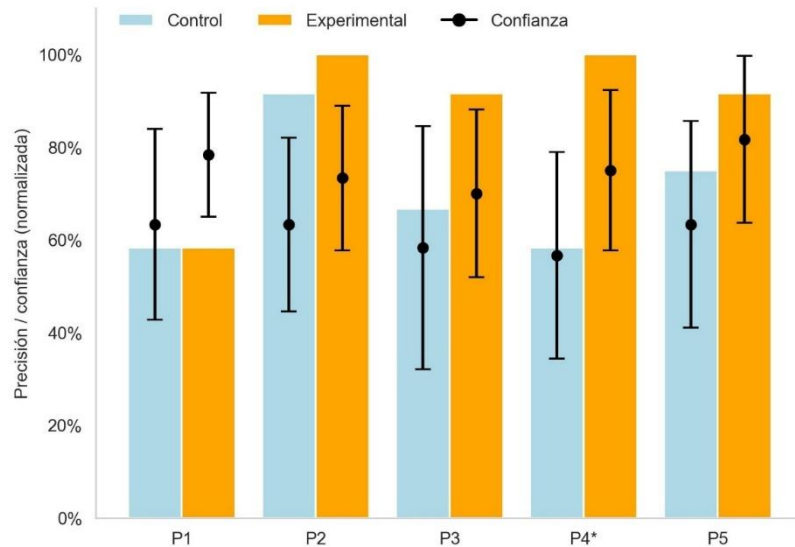


Fig. 31: Precisión y confianza normalizada por pregunta y grupo

En cuatro de las cinco preguntas, el grupo experimental alcanzó mayores niveles de precisión. La diferencia más destacada se registró en P4, relacionada con encontrar valores extremos. En ese caso, el 100 % de los participantes del grupo experimental respondió correctamente, frente al 58 % del grupo control. En la mayoría de los casos, los niveles de confianza fueron consistentes con la precisión observada. No se detectaron casos en los que un grupo manifestara mayor confianza sin haber obtenido mejores resultados.

Para analizar diferencias entre grupos, se aplicaron pruebas estadísticas no paramétricas, dado que las distribuciones no cumplían con los supuestos de normalidad. Se utilizó la prueba U de Mann–Whitney para comparar la cantidad total de respuestas correctas y el nivel de confianza promedio por participante. La proporción de respuestas correctas por pregunta se contrastó mediante la prueba exacta de Fisher (Nelson & Blauvelt, 2007) dada la naturaleza categórica de los datos y el tamaño reducido de la muestra.

Los resultados estadísticos fueron consistentes con los patrones descriptivos. El grupo experimental reportó niveles de confianza significativamente superiores ($U = 20.5$, $p = .003$). Si bien no se observaron diferencias significativas en la cantidad total de respuestas correctas ($U = 41.5$, $p = .068$), se identificó una tendencia favorable. A nivel individual,

Doctorado en Informática

solo la pregunta P4 mostró una diferencia estadísticamente significativa entre grupos ($p = .037$), denotada en el gráfico con *. Las demás preguntas no presentaron variaciones significativas.

En conjunto, estos resultados indican que las mejoras implementadas tuvieron un efecto positivo en la comprensión de tareas específicas, en particular aquellas vinculadas a la identificación de valores extremos. Esto respalda la utilidad del modelo como herramienta para identificar aspectos del diseño que inciden en la comprensión de los usuarios finales.

Usabilidad. La escala SUS permitió explorar cómo fue percibida la usabilidad del tablero por parte de los participantes luego de interactuar con las visualizaciones. Los puntajes obtenidos reflejan la evaluación subjetiva del sistema considerando aspectos como facilidad de uso, integración, claridad de los elementos y necesidad de apoyo externo.

La Tabla 19 muestra el puntaje total obtenido por cada grupo, incluyendo medidas de tendencia central y dispersión. El grupo experimental alcanzó un puntaje promedio significativamente mayor ($M = 83.5$, $DE = 3.9$) en comparación con el grupo control ($M = 63.1$, $DE = 6.2$). Esta diferencia, confirmada mediante la prueba U de Mann–Whitney ($U = 0.0$, $p < .001$), sugiere una percepción más positiva de usabilidad por parte de quienes utilizaron las visualizaciones rediseñadas.

Tabla 19: Media, mediana, rango y desviación estándar del SUS por grupo

Grupo	Media	DE	Min.	Max.	Mediana
Control	63.12	6.22	52.5	70	65
Experimental	83.54	3.91	77.5	90	82.5

Al clasificar los puntajes según los umbrales propuestos por Bangor et al. (2008) se observa que 10 participantes se ubicaron en el rango “Bueno”, 10 en “Aceptable” y 4 en “Excelente”. Ningún caso se ubicó en la categoría “Pobre”. Esta distribución indica una percepción favorable del sistema, aunque con mayor variabilidad dentro del grupo control.

Doctorado en Informática

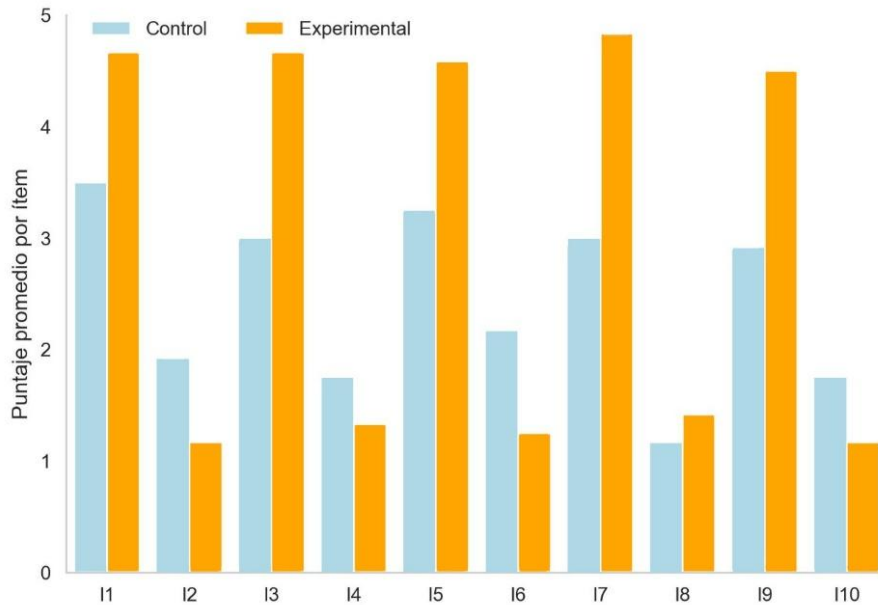


Fig. 32: Puntaje promedio por ítem y grupo

La Fig. 32 presenta el puntaje promedio por ítem (I) del cuestionario SUS, discriminado por grupo. Cada ítem fue calificado en una escala de 1 (muy en desacuerdo) a 5 (muy de acuerdo). El grupo experimental reportó mayores niveles de acuerdo en todos los ítems afirmativos (por ejemplo, “El sistema fue fácil de usar”), y puntuaciones más bajas en los ítems con formulación negativa (por ejemplo, “Necesitaría ayuda técnica para usarlo”), lo cual es coherente con una evaluación favorable.

En conjunto, los resultados del SUS indican que las mejoras aplicadas al diseño de las visualizaciones tuvieron un impacto positivo sobre la percepción de usabilidad. A diferencia de otros indicadores, este efecto se manifestó de forma clara en todos los participantes del grupo experimental y alcanzó niveles de significación estadística elevados. Estos hallazgos refuerzan la utilidad del modelo como herramienta para asistir en la revisión de visualizaciones desde una perspectiva centrada en la experiencia del usuario.

Doctorado en Informática**5.6.4.3. Percepción general del uso del modelo**

La percepción del equipo sobre el uso del modelo fue evaluada mediante una encuesta con tres preguntas abiertas. A continuación, se presenta una síntesis de las respuestas obtenidas.

Etapas de aplicación dentro del proceso de diseño. Todos los participantes coincidieron en que el modelo resulta útil una vez que se dispone de una primera versión de la visualización. Dos de ellos lo situaron en etapas finales, como paso de validación antes de publicar o entregar el material, mientras que otro propuso su uso en fases intermedias, cuando aún hay margen para realizar ajustes estructurales. En general, se percibe como una herramienta de revisión, con valor en momentos donde ya existe una base sobre la cual iterar.

Tareas o decisiones más asistidas por el modelo. Las respuestas indicaron que el modelo resulta especialmente útil para detectar problemas que suelen pasar desapercibidos, como omisiones en la narrativa, falta de atribuciones o deficiencias en la legibilidad. También se lo valoró como guía para decidir por dónde comenzar a mejorar una visualización, especialmente cuando el diseño ya parece correcto, pero genera dudas sobre su claridad o efectividad. Estas observaciones coinciden con los resultados previos en cuanto a utilidad y aplicación de las recomendaciones.

Obstáculos para su uso en escenarios reales. Entre las limitaciones señaladas se encuentran la posible sobrecarga en equipos con muchas visualizaciones, la extensión del cuestionario para usuarios no expertos, y la necesidad de integrar la herramienta al flujo de trabajo de forma natural. Frente a estos puntos, los participantes destacaron que la modalidad asistida y la integración temprana del modelo podrían reducir significativamente estas barreras.

Estas observaciones marcaron el cierre del proceso de validación en tres etapas. La Fig. 33 resume el recorrido realizado y los objetivos alcanzados en cada instancia. Los tres objetivos de uso en entorno real fueron cumplidos y se actualizó la aplicación web sumando el asistente basado en IA.

Doctorado en Informática

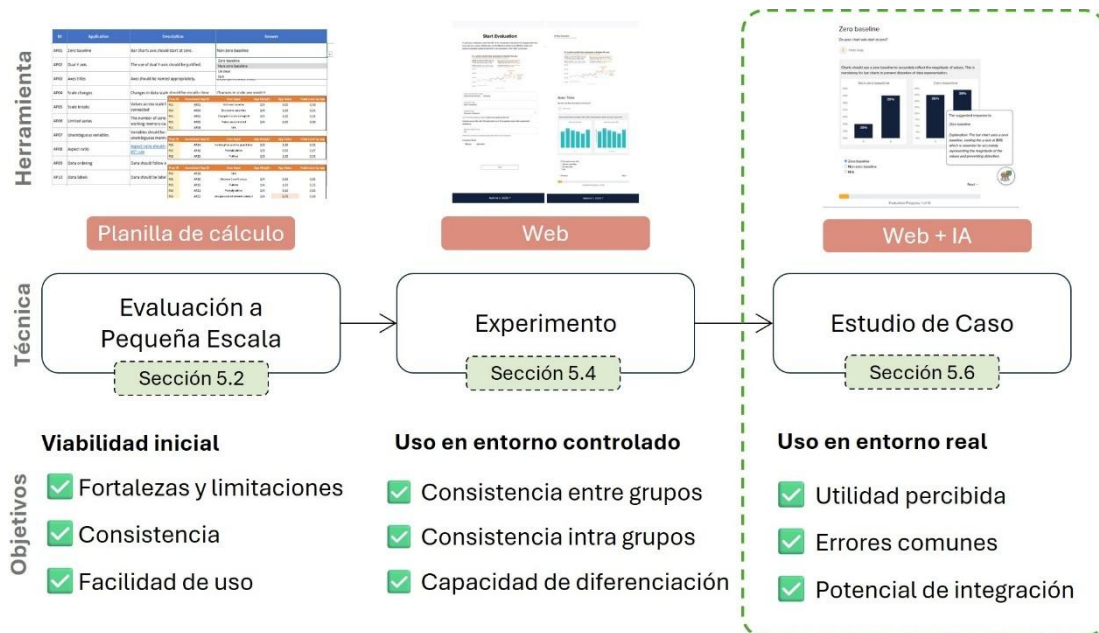


Fig. 33: Cumplimiento de objetivos en la tercera etapa de validación.

5.7. Conclusiones

La estrategia de validación desarrollada en este capítulo permitió analizar el modelo de calidad NaVi-Q desde distintos enfoques complementarios: una evaluación exploratoria, un experimento controlado y un estudio de caso aplicado. Cada instancia aportó evidencia específica sobre distintos aspectos del modelo, tales como su viabilidad inicial, su capacidad para generar estimaciones consistentes, su sensibilidad para discriminar niveles de calidad, y su utilidad en contextos de uso reales.

En conjunto, los resultados obtenidos permiten caracterizar fortalezas y limitaciones del modelo desde una perspectiva conceptual y práctica. Se observó que NaVi-Q permite detectar errores no triviales, guiar mejoras con impacto comprobado sobre la comprensión, la información y la usabilidad, y generar recomendaciones percibidas como útiles por los equipos de trabajo. Esto respalda su valor como herramienta de apoyo en procesos de diseño colaborativo, tanto en contextos controlados como aplicados.

Doctorado en Informática

En base a estos hallazgos, se consideraron los requisitos propuestos por Trendowicz & Kopczyńska (2014), quienes establecen una serie de condiciones que un método de evaluación debe cumplir para ser efectivo, comprensible y aplicable en diferentes contextos. La Tabla 20 resume estos requisitos junto con una breve apreciación sobre cómo NaVi-Q los satisface, considerando la estructura conceptual y la implementación práctica. Además, se incluye una columna que indica la sección de la tesis donde se aborda o ejemplifica cada requisito.

Tabla 20: Requisitos para métodos de evaluación de calidad

Requisito	Cumple	Justificación	Sección
R01. Soporta la toma de decisiones grupal y múltiples puntos de vista	✓	Permite crear perfiles de calidad personalizados mediante AHP-Express, integrando preferencias de distintos interesados.	4.7
R02. Comprensible para los interesados en la calidad	✓	Las evaluaciones mostraron buena comprensión del instrumento. Se mejoraron instrucciones y ayudas contextuales.	5.2.6.3; 5.4.5.7
R03. Maneja información incierta	✓	Se admiten respuestas subjetivas y “N/A”. Las sugerencias generadas por IA ayudan a resolver ambigüedades.	4.5.3; 5.5.2
R04. Abarca las preferencias de los interesados	✓	Las preferencias se expresan mediante pesos definidos por los usuarios en perfiles de calidad.	5.2.2; 5.6.2
R05. Aborda información incompleta.	✓	Las respuestas “N/A” no afectan la validez de la evaluación ni impiden el cálculo del puntaje.	4.5.3
R06. Maneja las interdependencias entre elementos del modelo de calidad	✓	Las propiedades se organizan jerárquicamente y su asociación con los criterios fue validada empíricamente.	4.3; 5.4.5.7
R07. Combina enfoques compensatorios y no compensatorios	✓	El modelo combina agregación ponderada con recomendaciones que destacan puntos críticos no compensables.	4.6
R08. Soportado por una herramienta automática	✓	Se implementó una herramienta web con modalidades de evaluación manual y asistida.	5.3.1

Doctorado en Informática

Requisito	Cumple	Justificación	Sección
R09. Permite incorporar evaluaciones subjetivas de expertos	✓	Las respuestas siempre son provistas por el usuario, y se integran al cálculo final en ambas modalidades.	4.5.3
R10. Admite la evaluación comparativa	✓	Permite comparar visualizaciones mediante puntajes globales y desglosados.	4.6.4
R11. Proporciona resultados repetibles	✓	Se observó alta consistencia entre evaluadores con distintos grados de experiencia.	5.4.5.3; 5.4.5.5
R12. Personalizable	✓	La estructura y los pesos de los perfiles de calidad pueden ajustarse según el dominio de aplicación.	4.7
R13. Admite un modelo de calidad jerárquico	✓	La estructura se organiza en tres niveles: criterio, propiedad y aplicación.	4.3
R14. Escalable y extensible	✓	Se puede extender o reducir el conjunto de propiedades sin alterar la lógica del modelo.	4.3; 5.35.5
R15. Admite evaluaciones intermedias	✓	Se generan puntajes a nivel global, por criterio y por propiedad.	4.6.4

La Fig. 34 resume las tareas del marco de DSR abordadas en este capítulo. Se observa que el proceso culmina con la obtención del artefacto definitivo. La articulación progresiva de las etapas ilustra el valor de adoptar un enfoque iterativo para construir artefactos que respondan de manera efectiva a necesidades reales.

Doctorado en Informática

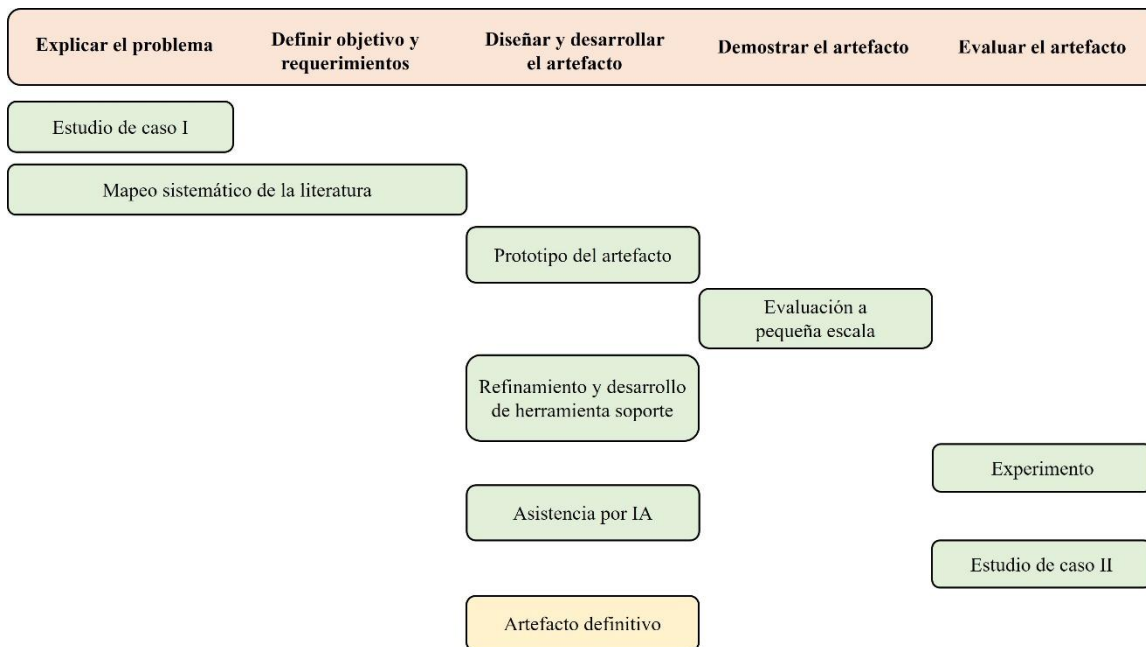


Fig. 34: Actividades abarcadas en el capítulo

Capítulo 6. Conclusiones

Este capítulo presenta las conclusiones de la investigación desarrollada en la tesis doctoral. La sección 6.1 resume las actividades realizadas y su relación con los objetivos y contribuciones definidos al inicio del trabajo. La sección 6.3 expone las principales limitaciones del estudio y las amenazas a su validez. La sección 6.4 identifica posibles líneas de trabajo futuro a partir de los resultados obtenidos, y la sección 6.5 reúne las publicaciones generadas para la difusión de los hallazgos.

6.1. Objetivos de la tesis abordados

El objetivo general fue desarrollar un modelo para evaluar la calidad y asistir en la mejora de visualizaciones narrativas utilizadas en las interfaces de usuario de sistemas software. Para alcanzar dicho objetivo, se adoptó el marco de DSR, que estructura la investigación en torno al diseño, desarrollo y evaluación de un artefacto orientado a resolver un problema práctico. A continuación, se describen las actividades realizadas en cada fase y su relación con los objetivos planteados en el Capítulo 1.

Explicar el problema. Para delimitar el problema y comprender sus causas, se realizaron dos tareas iniciales. En primer lugar, se condujo un estudio de caso exploratorio (Lezcano Airaldi et al., 2021), siguiendo las pautas de Yin (2017), que permitió identificar limitaciones en las prácticas actuales de evaluación de visualizaciones en entornos reales (sección 1.2).

En segundo lugar, se llevó a cabo un estudio de mapeo sistemático (Lezcano Airaldi, Irrazábal, & Diaz Pace, 2025) descrito en el Capítulo 1, siguiendo las guías metodológicas de Kitchenham & Charters (2007), con el propósito de establecer una base de

Doctorado en Informática

conocimiento sobre las buenas prácticas (sección 3.2) y los criterios existentes para evaluar visualizaciones narrativas (sección 3.3), así como también los métodos de evaluación existentes (sección 3.4).

Estas actividades permitieron precisar la necesidad de un enfoque de evaluación integral, identificando la falta de herramientas que contemplen tanto las buenas prácticas generales de visualización como los aspectos específicos de las visualizaciones narrativas. Este análisis contribuyó al cumplimiento del OE1, centrado en identificar buenas prácticas para el diseño de visualizaciones narrativas efectivas.

Definir objetivos y requerimientos. En primer lugar, se identificó que el tipo de artefacto más adecuado para abordar el problema planteado era un modelo prescriptivo, capaz de establecer una estructura para guiar la evaluación de visualizaciones narrativas (sección 2.1.1). A partir del análisis del contexto y del público objetivo, se establecieron tres requerimientos básicos: (1) que el modelo fuera útil para usuarios no expertos, (2) que pudiera adaptarse a distintos escenarios de uso, y (3) que generara resultados confiables y reproducibles (sección 2.1.2). Estos requisitos guiaron tanto el diseño conceptual del modelo (sección 4.3) como las decisiones técnicas adoptadas para su implementación práctica (sección 5.3.1).

Diseñar y desarrollar el artefacto. El modelo propuesto, denominado NaVi-Q (Capítulo 1), fue construido en seis pasos iterativos adaptados de Siebert et al. (2022), en base a los estándares ISO/IEC 25010 y ISO/IEC 33003 (sección 4.1). Se definió su alcance, centrado en la evaluación de visualizaciones narrativas tipo documento (estáticas o interactivas), con foco en propiedades de calidad externas observables (sección 4.2).

Se diseñó un metamodelo jerárquico compuesto por cinco criterios (comprensión, compromiso, información, memorabilidad y usabilidad), descompuestos en 17 propiedades y 55 aplicaciones evaluables mediante instrumentos específicos (sección 4.3). Cada elemento del modelo tiene un peso asociado que puede ajustarse según el contexto mediante perfiles de calidad (sección 4.7).

Doctorado en Informática

Para su operacionalización, se definieron reglas de evaluación siguiendo un enfoque formativo y compensatorio (sección 4.6). Se utilizaron funciones de medición ponderadas y técnicas de decisión multiatributo para calcular los puntajes de calidad (sección 4.6.2).

Estas actividades dieron cumplimiento al OE2, orientado a construir un modelo de evaluación con distintos niveles de abstracción y métricas.

Demostrar el artefacto. La viabilidad inicial del modelo fue explorada mediante una evaluación a pequeña escala (sección 5.2), orientada a observar su funcionamiento en una primera versión conceptual. Esta instancia permitió aplicar el modelo con un conjunto reducido de usuarios y visualizaciones, y recolectar datos tanto cuantitativos como cualitativos. Se analizaron tres aspectos principales: las fortalezas y limitaciones del modelo percibidas por los participantes, la consistencia entre evaluadores (estimada mediante correlación intraclase) y la dificultad reportada al utilizar el instrumento.

Los resultados permitieron confirmar que los evaluadores aplicaron el modelo de manera consistente, y señalaron ajustes necesarios para facilitar su aplicación. A partir de estos hallazgos, se realizó una primera iteración del artefacto que incluyó una reducción del número de aplicaciones (sección 5.3), mejoras en la redacción de preguntas y el desarrollo de la herramienta de software para operacionalizar el modelo y utilizarlo en etapas posteriores de validación (sección 5.3.1). Este desarrollo se corresponde con el OE3, que plantea el desarrollo de una herramienta de software como soporte tecnológico del modelo.

Evaluar el artefacto. La evaluación formal del modelo se llevó a cabo en dos etapas complementarias. En primer lugar, se realizó un experimento controlado con 40 participantes (sección 5.4), en el que se analizaron aspectos como la capacidad del modelo para discriminar niveles de calidad predefinidos, la consistencia de los puntajes generados por distintos perfiles de usuario y su correlación con evaluaciones de referencia. También se evaluó la claridad percibida de los criterios mediante una encuesta estructurada, y se exploró la experiencia de uso de la herramienta con escalas tipo Likert y análisis temático.

Doctorado en Informática

A partir de los resultados obtenidos, se llevó a cabo una segunda iteración de diseño del artefacto orientada a reducir la carga cognitiva del proceso y facilitar su aplicación (sección 5.5). En esta etapa se incorporaron verificaciones automáticas para determinadas aplicaciones del modelo, así como un sistema de evaluación asistida por IA basado en modelos de lenguaje multimodales.

En segundo lugar, se condujo un estudio de caso en un entorno aplicado (sección 5.6), que permitió observar el uso del modelo en un flujo de trabajo real, evaluar la utilidad percibida de las recomendaciones generadas, identificar patrones de mejora recurrentes, y medir el impacto de las visualizaciones mejoradas sobre la comprensión, la información percibida y la usabilidad, utilizando cuestionarios validados y encuestas cualitativas. Estas instancias permitieron abordar el OE4, centrado en validar el modelo a través de su implementación y uso en contextos controlados y aplicados.

Luego de completar las actividades propuestas en el marco de DSR, la contribución principal de la tesis consiste en un modelo de evaluación que permite diagnosticar y asistir en la mejora de visualizaciones narrativas a partir de características observables, aplicable en distintas etapas del desarrollo y en diversos contextos de uso. Este modelo, acompañado de su implementación como herramienta web interactiva, busca facilitar evaluaciones consistentes, promover el uso de buenas prácticas y brindar soporte a equipos de trabajo que buscan mejorar la forma en que comunican información a través de sus visualizaciones.

De esta manera, la hipótesis planteada en el Capítulo 1, según la cual “un modelo semiautomatizado basado en buenas prácticas puede incrementar el valor de las visualizaciones narrativas desde etapas tempranas de desarrollo” fue sustentada por la evidencia empírica obtenida a lo largo de esta investigación doctoral.

6.2. Contribuciones adicionales

Además de las actividades asociadas al desarrollo, implementación y validación del modelo de calidad, la tesis generó una serie de aportes complementarios que pueden resultar útiles en contextos similares o servir de base para desarrollos futuros.

Doctorado en Informática

Formalización de un vocabulario evaluativo adaptado al dominio visual: Se definió un conjunto estructurado de 36 aplicaciones evaluativas con escalas de respuesta específicas, derivadas de estándares, guías y estudios previos sobre buenas prácticas en visualización. Este vocabulario fue validado empíricamente y documentado de forma accesible para su reutilización (sección 4.5 y Apéndice F).

Automatización parcial de evaluaciones mediante análisis de imagen: Se implementaron verificaciones automáticas específicas para algunas aplicaciones evaluativas, como el contraste entre elementos y la accesibilidad para personas con daltonismo. Estas funciones se integraron a la interfaz de evaluación manual como apoyo para el usuario, basadas en técnicas de análisis perceptual y simulación de visión (sección 5.5.1).

Implementación de mecanismos de asistencia automatizada: Se integraron funcionalidades basadas en modelos de lenguaje multimodales para asistir al usuario durante la evaluación: (a) sugerencias contextuales en preguntas específicas, y (b) recomendaciones de mejora generadas a partir de los resultados obtenidos (sección 5.5.2). Esta implementación permite explorar el potencial de LLMs en tareas de evaluación sin supervisión directa.

Prácticas de diseño consistentes: Al definir criterios de calidad claros (sección 3.3) y ofrecer un enfoque sistemático de evaluación (sección 4.6), el modelo puede asistir a los creadores de visualizaciones mediante un conjunto de buenas prácticas. En organizaciones, esto podría favorecer la generación de resultados consistentes. El modelo también puede funcionar como una herramienta de formación para comprender los componentes fundamentales de las visualizaciones narrativas, facilitando la colaboración entre equipos técnicos y no técnicos. En el ámbito académico, proporciona un marco para estudiar y desarrollar estas prácticas de diseño en distintos campos y aplicaciones.

Marco comparativo de referencia: El modelo puede funcionar como un punto de referencia para contrastar diferentes enfoques de evaluación. En contextos industriales, podría contribuir a establecer un estándar para comparar resultados a lo largo del tiempo,

Doctorado en Informática

monitorear avances e identificar áreas de mejora. En el ámbito académico, puede resultar útil en estudios experimentales.

6.3. Amenazas a la validez

Si bien los estudios de validación permitieron evaluar el funcionamiento de NaVi-Q en entornos controlados y aplicados, los resultados deben interpretarse considerando ciertas amenazas a la validez. Estas amenazas no invalidan los hallazgos, pero sí delimitan el alcance de las conclusiones y señalan aspectos que deben ser explorados en investigaciones futuras. A continuación, se presentan las principales limitaciones organizadas en tres categorías: aquellas relacionadas con la generalización de los resultados y el contexto de aplicación; las vinculadas con el alcance conceptual del modelo y su capacidad para capturar decisiones de diseño dependientes del contexto; y las referidas al diseño de los estudios.

6.3.1. Validez externa

La validez externa de los estudios realizados se ve condicionada por ciertas limitaciones relacionadas con el contexto de aplicación, el perfil de los participantes involucrados y el dominio específico en el que se desarrollaron las evaluaciones. Estas condiciones pueden influir en la capacidad de generalizar los resultados obtenidos a otros entornos, poblaciones o ámbitos de uso.

Homogeneidad de los participantes. A lo largo de las distintas instancias de validación se involucró a participantes con perfiles diversos en cuanto a su experiencia técnica, pero pertenecientes mayoritariamente a entornos académicos o profesionales del ámbito informático. Esta relativa homogeneidad limita la posibilidad de generalizar los resultados a otros contextos organizacionales o poblaciones con menor familiaridad con visualizaciones de datos, especialmente aquellas no entrenadas en tareas analíticas o de lectura estructurada. Futuras validaciones podrían incluir participantes de distintos perfiles profesionales y culturales, a fin de evaluar la aplicabilidad del modelo en contextos más heterogéneos.

Doctorado en Informática

Influencia cultural en la interpretación narrativa. Si bien el modelo fue diseñado para aplicar principios generales de percepción visual y estructura narrativa, distintos estudios han demostrado que ciertos códigos gráficos no son universales. Según Ware (2020), los códigos arbitrarios —como los colores o símbolos usados en diagramas— se aprenden y mantienen dentro de convenciones culturales específicas. El autor señala que “los conjuntos de símbolos creados por diferentes culturas pueden conducir a interpretaciones opuestas de un mismo elemento visual”, y ofrece como ejemplo el uso del color rojo, que puede indicar advertencia en contextos occidentales pero buena fortuna en culturas orientales. En este sentido, si bien el modelo propone recomendaciones basadas en convenciones perceptivas ampliamente aceptadas, su aplicación podría generar ambigüedades en contextos multiculturales, especialmente cuando se evalúan narrativas visuales destinadas a audiencias heterogéneas. Una posible línea de trabajo futura consiste en adaptar los perfiles de calidad a distintas convenciones culturales o niveles de alfabetización visual, según las características del público objetivo

Tamaño de la muestra y dominio de aplicación. Si bien se pretendió incluir una variedad de tipos de visualización, la selección utilizada en los estudios no refleja necesariamente toda la diversidad de estilos narrativos disponibles en la actualidad. El conjunto de visualizaciones analizadas se centró en formatos para los cuales NaVi-Q fue originalmente concebido, tales como gráficos de líneas, barras, dispersión y sectores. Se incluyeron tanto visualizaciones estáticas como interactivas, y cada aplicación del modelo fue evaluada en todas sus opciones de respuesta posibles. No obstante, aunque el modelo contempla ciertos elementos interactivos, aún no ha sido probado con formatos altamente dinámicos o exploratorios, como visualizaciones con desplazamiento secuencial (*scroll-based*), tableros con datos en tiempo real o interacciones complejas. Esto se debe a que NaVi-Q fue diseñado principalmente para evaluar visualizaciones de uso cotidiano, comunes en entornos profesionales, donde predominan representaciones directas y estructuradas. Por otra parte, el estudio de caso se desarrolló en un proyecto vinculado a la ingeniería de software, lo cual restringe el análisis a un único dominio. La evaluación del modelo en otros campos, como la salud, la educación o el periodismo de datos, podría

Doctorado en Informática

aportar evidencia adicional sobre su aplicabilidad en contextos con diferentes prioridades narrativas e informativas.

6.3.2. Validez de constructo

La validez de constructo se refiere al grado en que un instrumento mide realmente lo que pretende medir. En este caso, si bien el modelo fue concebido como una herramienta sistemática basada en principios perceptivos y buenas prácticas de diseño, su aplicación práctica depende de decisiones situadas que pueden afectar la forma en que se interpretan sus recomendaciones o se utilizan sus resultados.

Tal como se discutió en la Sección 3.2, muchas de las buenas prácticas recopiladas en la literatura no son universalmente aplicables ni siempre convergentes. En determinadas situaciones, pueden entrar en conflicto entre sí o con los objetivos específicos del autor de la visualización. Por ejemplo, utilizar un color destacado para resaltar elementos importantes puede interferir con la codificación de categorías cuando ese mismo color ya se emplea con otro propósito. Este tipo de tensiones son comunes en problemas de diseño complejos, donde el espacio de soluciones óptimas se ve limitado por la cantidad y el tipo de información que debe representarse.

Ware (2020) y Desbarats (2023) advierten que, si bien es posible establecer guías de diseño basadas en principios perceptivos para situaciones simples, en contextos más complejos es el diseñador quien debe tomar decisiones informadas y administrar los recursos gráficos con criterio. En este sentido, aunque NaVi-Q ofrece una estructura sistemática para identificar aspectos problemáticos y orientar mejoras, no reemplaza el proceso de toma de decisiones propio del diseño. Su utilidad radica en servir como herramienta de apoyo que contribuye a hacer visibles ciertos aspectos de la calidad narrativa y facilitar la reflexión crítica, sin prescribir soluciones únicas ni cerradas.

6.3.3. Validez de conclusión

La validez de conclusión se refiere al grado de certeza con el que pueden interpretarse los resultados obtenidos. En este trabajo, dicha validez se ve condicionada por la ausencia de

Doctorado en Informática

una instancia específica diseñada para comparar el modelo propuesto con otros enfoques de evaluación disponibles.

Esta decisión se debe a que, actualmente, no existen modelos comparables que aborden de forma estructurada y parametrizable la calidad de visualizaciones narrativas desde una perspectiva orientada al usuario. La mayoría de los enfoques disponibles consisten en listas de verificación, heurísticas o marcos teóricos que no permiten una comparación directa ni reproducible en términos de cobertura, precisión o resultado.

Ante esta situación, se optó por incorporar evaluaciones de referencia en cada estudio, ya sea mediante análisis independientes, criterios consensuados por expertos o líneas base definidas ad hoc. Estas instancias permitieron interpretar los resultados del modelo y mitigar, en parte, la falta de un marco alternativo formal con el cual contrastarlo. Aun así, la comparación con otros enfoques constituye una línea futura relevante para consolidar el posicionamiento y alcance de NaVi-Q en el ecosistema de herramientas de evaluación.

6.4. Trabajos futuros

A partir de los resultados obtenidos y las limitaciones identificadas, se desprenden distintas líneas de trabajo que podrían contribuir a ampliar, refinar o complementar el modelo propuesto y su implementación práctica. A continuación, se presentan algunas direcciones relevantes para investigaciones futuras.

Ampliación y adaptación del modelo NaVi-Q: Una posible línea de trabajo consiste en validar el modelo en otros dominios de aplicación, como la salud, el periodismo o la educación. Estos campos presentan estructuras narrativas, objetivos comunicativos y audiencias que difieren de los analizados en esta tesis, por lo que su inclusión permitiría evaluar la flexibilidad del modelo ante distintos enfoques narrativos.

Asimismo, resulta pertinente explorar la ampliación del metamodelo para incluir nuevas propiedades y aplicaciones asociadas a formatos de visualización más dinámicos, como visualizaciones con desplazamiento secuencial, tableros interactivos en tiempo real, o gráficos con alto grado de exploración. Si bien NaVi-Q contempla ciertos elementos

Doctorado en Informática

interactivos, su estructura actual está orientada a visualizaciones más estructuradas y de uso general.

Avances en la integración con modelos de lenguaje. Dado que la herramienta de evaluación incorpora funcionalidades asistidas por IA, una línea de investigación posible consiste en comparar el desempeño de diferentes modelos de lenguaje (como GPT-4o, Claude o Gemini) en tareas de evaluación y recomendación. Esta comparación podría realizarse utilizando métricas cuantitativas como precisión y recuperación, evaluando hasta qué punto los modelos coinciden con respuestas de referencia o generadas por expertos humanos.

Además, resulta de interés experimentar con técnicas de *fine tuning* o ingeniería de prompts orientadas a mejorar la precisión y contextualización de las sugerencias generadas por el modelo. Este tipo de ajustes permitiría adaptar el comportamiento de los asistentes según el tipo de usuario, el dominio de aplicación o el nivel de detalle requerido.

Extensiones funcionales de la herramienta. Desde el punto de vista técnico, la herramienta desarrollada puede enriquecerse con funcionalidades adicionales que amplíen su alcance y utilidad. Una de ellas es la incorporación de un módulo de entrenamiento interactivo, orientado a usuarios que deseen aprender o reforzar buenas prácticas mediante casos ilustrativos y retroalimentación inmediata basada en el modelo.

Otra posible extensión consiste en ofrecer una API pública que permita integrar el motor de evaluación en otros sistemas, como editores gráficos o plataformas de diseño colaborativo. Esto permitiría incorporar evaluaciones de calidad directamente en los flujos de trabajo existentes.

Finalmente, podría desarrollarse un modo de revisión iterativa que registre las distintas versiones de una visualización evaluada, permitiendo observar su evolución a lo largo del tiempo y analizar el impacto de los cambios introducidos. Este tipo de historial podría facilitar procesos de mejora continua y diseño basado en evidencia.

6.5. Difusión de resultados

Los resultados de esta tesis fueron difundidos a través de distintas instancias de publicación, participación en eventos científicos y colaboraciones académicas. A continuación, se detallan las principales actividades realizadas.

6.5.1. Publicaciones derivadas de la tesis

- Lezcano Airaldi A, Diaz-Pace JA, Irrazábal E. (2021) Data-driven Storytelling to Support Decision Making in Crisis Settings: A Case Study. *Journal of Universal Computer Science*, 27(10): 1046-1068. <https://doi.org/10.3897/jucs.66714>

Estudio de caso I. Analiza el uso de visualizaciones narrativas en contextos de crisis, y permitió identificar los desafíos prácticos que motivaron la definición del problema de investigación.

- Lezcano Airaldi, A. (2022). Evaluation and Improvement of Narrative Visualizations. In *Anais do XXV Congresso Ibero-Americano em Engenharia de Software*, (pp. 368-375). Porto Alegre: SBC. [doi:10.5753/cibse.2022.20986](https://doi.org/10.5753/cibse.2022.20986)

Simposio doctoral. Presenta la propuesta inicial de la investigación, incluyendo su motivación, objetivos y marco metodológico.

- Lezcano Airaldi, A., Irrazábal, E., & Diaz Pace, A. (2025). Best practices and evaluation methods for narrative information visualizations: A systematic review. In *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)* (pp. 378–389). SciTePress. <https://doi.org/10.5220/0013202000003928>

Estudio de mapeo sistemático. Revisión de buenas prácticas y métodos de evaluación para visualizaciones narrativas, que fundamenta conceptualmente el diseño del modelo de calidad.

Doctorado en Informática

- Lezcano Airaldi, A., Carruthers, J. A., Irrazábal, E., & Diaz-Pace, A. (2025). Measuring narrative visualizations: A model-based evaluation framework and validation. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3570485>

Experimento de validación. Presenta el modelo propuesto, su estructura jerárquica y los resultados del experimento controlado con 40 participantes.

- **[En redacción]** Lezcano Airaldi, A., Wiese, M., Besier, A., & Irrazábal, E. Improving Narrative Visualizations in Technical Debt Dashboards: A Case Study.

Estudio de caso II. Presenta los resultados del estudio realizado junto al grupo SWK de la Universidad de Hamburgo. Analiza el impacto de las recomendaciones del modelo NaVi-Q en visualizaciones de deuda técnica, evaluando su utilidad, aplicabilidad y efecto sobre la comprensión, información y usabilidad.

6.5.2. Otras publicaciones

- **[En revisión]** Lezcano Airaldi, A., Romera, L., Maalej, W. *How well can AI generate backlogs from app mockups?*

Enviado a la conferencia Foundations of Software Engineering (FSE 2026). El artículo presenta un estudio sobre el uso de modelos de lenguaje multimodal para generar ítems de backlog a partir de maquetas de aplicaciones. El trabajo fue realizado en colaboración con el grupo MAST de la Universidad de Hamburgo.

- **[En redacción]** Lezcano Airaldi, A., Marcori, Joel & Irrazábal, E. Evaluación automatizada de visualizaciones narrativas: comparación de modelos de lenguaje de gran escala.

Resumen aceptado en el Congreso Brasileiro de Software (CBSOFT) 2025 para su presentación en formato póster. Manuscrito en redacción. El estudio compara el desempeño de distintos modelos de lenguaje multimodales (como GPT-4o, Gemini y Claude) en evaluación de visualizaciones narrativas, utilizando métricas como precisión, recuperación y F1.

Doctorado en Informática**6.5.3. Participación en eventos**

- Mentoría de ACM SIGSOFT (2021)

Participación en la primera edición de la *Software Engineering School for Latin America*, organizada por ACM SIGSOFT en el marco del XII Congresso Brasileiro de Software (CBSOFT 2021). En ese contexto, se accedió a un programa de mentoría con investigadores internacionales en el área de Ingeniería de Software, orientado al desarrollo de carrera académica y la planificación de proyectos doctorales.

- Escuela de Informática del CACIC (2022)

Dictado de curso sobre visualización narrativa en el marco de la Escuela de Informática del Congreso Argentino de Ciencias de la Computación (CACIC). Durante uno de los talleres se aplicó el método AHP-Express para establecer los pesos del perfil de calidad Estándar. Participaron 51 personas, cuya configuración de preferencias fue utilizada posteriormente en la evaluación a pequeña escala (sección 5.2.2).

- Corrientes Data – Semana del Gobierno Abierto (2023)

Tercer puesto en el concurso de visualización de datos “Corrientes Data – Visualizando mi ciudad”, organizado por la Municipalidad de Corrientes en el marco de la Semana del Gobierno Abierto. La participación permitió aplicar principios de diseño narrativo y evaluar el impacto comunicativo de distintas decisiones visuales.

- CBSOFT 2023 – Simposio doctoral y presentación de póster

Presentación de póster en el marco de la Escola Latino-Americana de Engenharia de Software, actividad académica realizada como parte del XIV Congresso Brasileiro de Software: Teoria e Prática (CBSOFT 2023), llevado a cabo en Campo Grande, Brasil. La participación fue posible gracias a una beca otorgada por ACM SIGSOFT, obtenida tras haber resultado seleccionada en un concurso para jóvenes investigadores.



Doctorado en Informática

6.5.4. Registro del artefacto

El modelo y la herramienta desarrollados en el marco de esta tesis, denominados NaVi-Q, han sido registrados y se encuentran protegidos por derechos de autor bajo el número de expediente EX-2025-60931702- -APN-DNDA#MJ, conforme a la legislación vigente en el país. Esta protección garantiza la autoría y preserva el uso del artefacto para fines académicos y de transferencia tecnológica.

Doctorado en Informática

Apéndice A. Estudios primarios del mapeo sistemático

Este apéndice presenta los 116 estudios primarios seleccionados en el mapeo sistemático descrito en el Capítulo 3. La Tabla A. 1 detalla el código asignado a cada estudio (de S1 a S116), junto con sus datos bibliográficos relevantes.

Tabla A. 1: Estudios primarios incluidos en el mapeo sistemático

Código	Año	Título	Autores
S01	2002	Graphical Encoding for Information Visualization: An Empirical Study	L. Nowell and R. Schulman and D. Hix
S02	2010	Narrative Visualization: Telling Stories with Data	Edward Segel and Jeffrey Heer
S03	2011	Visualization Rhetoric: Framing Effects in Narrative Visualization	Jessica Hullman and Nicholas Diakopoulos
S04	2016	Suggested Interactivity: Seeking Perceived Affordances for Information Visualization	Jeremy Boy and Louis Eveillard and Françoise Detienne and Jean Daniel Fekete
S05	2016	Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use	Annemarie Quispel and Alfons Maes and Joost Schilperoord
S06	2016	A Study on Designing Effective Introductory Materials for Information Visualization	Yuzuru Tanahashi and Nick Leaf and Kwan Liu Ma
S07	2017	The Attraction Effect in Information Visualization	Evanthia Dimara and Anastasia Bezerianos and Pierre Dragicevic
S08	2017	Extending Open Data Platforms with Storytelling Features	Brolcháin N., Porwol L., Ojo A., Wagner T., Lopez E., Karstens E.
S09	2000	Evaluating Visualizations Based on the Performed Task	Octavio Juarez and Chris Hendrickson and James H. Garrett
S10	2005	Improving 2D scatterplots effectiveness through sampling, displacement, and user perception	Enrico Bertini and Giuseppe Santucci
S11	2009	Investigating the Comprehension Support for Effective Visualization Tools – A Case Study	Harkirat Padda and Ahmed Seffah and Sudhir Mudur
S12	2010	Perceptual Guidelines for Creating Rectangular Treemaps	Nicholas Kong and Jeffrey Heer and Maneesh Agrawala
S13	2011	Whisper, Don't Scream: Grids and Transparency	Lyn Bartram and Maureen C. Stone

**Doctorado en Informática**

Código	Año	Título	Autores
S14	2011	Purposeful Visualization System	Xiaoyan Bai and David White and David Sundaram
S15	2011	A Study on Dual-Scale Data Charts	Petra Isenberg and Anastasia Bezerianos and Pierre Dragicevic and Jean Daniel Fekete
S16	2010	An Empirical Study on Using Visual Embellishments in Visualization	Borgo R., Abdul-Rahman A., Mohamed F., Grant P., Reppa I., Floridi L., Chen M.
S17	2013	Selecting the Aspect Ratio of a Scatter Plot Based on Its Delaunay Triangulation	Martin Fink and Jan Henrik Haunert and Joachim Spoerhase and Alexander Wolff
S18	2014	A Principled Way of Assessing Visualization Literacy	Jeremy Boy and Ronald A. Rensink and Enrico Bertini and Jean-Daniel Fekete
S19	2015	Clutter-Aware Label Layout	Yu Meng and Hui Zhang and Mengchen Liu and Shixia Liu
S20	2017	Towards Perceptual Optimization of the Visual Design of Scatterplots	Luana Micallef and Gregorio Palmas and Antti Oulasvirta and Tino Weinkauff
S21	2018	Evaluating Cartogram Effectiveness	Sabrina Nusrat and Md Jawaherul Alam and Stephen Kobourov
S22	2021	Rainbows Revisited: Modeling Effective Colormap Design for Graphical Inference	Khairi Reda and Danielle Albers Szafir
S23	2021	Smile or Scowl? Looking at Infographic Design Through the Affective Lens	Xingyu Lan and Yang Shi and Yueyao Zhang and Nan Cao
S24	2021	Visual Reasoning Strategies for Effect Size Judgments and Decisions	Alex Kale and Matthew Kay and Jessica Hullman
S25	1992	Improving the Visualization of Hierarchies with Treemaps: Design Issues and Experimentation	David Turo and Brian Johnson
S26	2006	Evaluating Visual Table Data Understanding	Nathalie Henry and Jean-Daniel Fekete
S27	2006	Effects of 2D Geometric Transformations on Visual Memory	Heidi Lam and Ronald A. Rensink and Tamara Munzner

Doctorado en Informática

Código	Año	Título	Autores
S28	2007	Complexity Analysis for Information Visualization Design and Evaluation	Ying Zhu and Xiaoyuan Suo and G. Scott Owen
S29	2007	The Effect of Aesthetic on the Usability of Data Visualization	Nick Cawthon and Andrew Vande Moere
S30	2008	Perceptual Dependencies in Information Visualization Assessed by Complex Visual Search	Ronald Van and Den Berg and Jos B.T.M. Roerdink and Fransw Cornelissen
S31	2008	Comprehension of Visualization Systems - Towards Quantitative Assessment	Harkirat Padda and Sudhir Mudur and Ahmed Seffah and Yojana Joshi
S32	2009	Evaluation of Symbol Contrast in Scatterplots	Jing Li and Jarke J. Van Wijk and Jean Bernard Martens
S33	2009	Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization	Simone Garlandini and Sara Irina Fabrikant
S34	2010	Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts	Bateman S., Mandryk R., Gutwin C., Genest A., McDine D., Brooks C.
S35	2010	Using Cognitive Fit Theory to Evaluate the Effectiveness of Information Visualizations: An Example Using Quality Assurance Data	Jay M. Teets and David P. Tegarden and Roberta S. Russell
S36	2010	An Heuristic Set for Evaluation in Information Visualization	An heuristic set for evaluation in information visualization
S37	2010	Graphical Perception of Multiple Time Series	Waqas Javed and Bryan McDonnell and Niklas Elmqvist
S38	2011	Eye tracking for visualization evaluation: Reading values on linear versus radial graphs	Joseph Goldberg and Jonathan Helfman
S39	2011	The Effect of Colour and Transparency on the Perception of Overlaid Grids	Lyn Bartram and Billy Cheung and Maureen Stone
S40	2012	Comparing Averages in Time Series Data	Michael Correll and Danielle Albers and Steve Franconeri and Michael Gleicher
S41	2012	Towards Adaptive Information Visualization: On the Influence of User Characteristics	Dereck Toker and Cristina Conati and Giuseppe Carenini and Mona Haraty
S42	2012	How Capacity Limits of Attention Influence Information Visualization Effectiveness	Steve Haroz and David Whitney
S43	2012	Evaluating the Effect of Style in Information Visualization	Vande Moere A., Tomitsch M., Wimmer C., Christoph B., Grechenig T.
S44	2013	Individual User Characteristics and Information Visualization: Connecting the Dots through Eye Tracking	Dereck Toker and Cristina Conati and Ben Steichen and Giuseppe Carenini

Doctorado en Informática

Código	Año	Título	Autores
S45	2013	Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting	Fuchs J., Fischer F., Mansmann F., Bertini E., Isenberg P.
S46	2013	Data Visualisation, User Experience and Context: A Case Study from Fantasy Sport	Rob Euman and José Abdelnour-Nocera
S47	2013	A Deeper Understanding of Sequence in Narrative Visualization	Hullman J., Drucker S., Riche N., Bongshin Lee, Fisher D., Adar E.
S48	2013	What Makes a Visualization Memorable?	Borkin M., Vo A., Bylinskii Z., Isola P., Sunkavalli S., Oliva A., Pfister H.
S49	2014	Sample-Oriented Task-Driven Visualizations: Allowing Users to Make Better, More Confident Decisions	Nivan Ferreira and Danyel Fisher and Arnd Christian König
S50	2014	Improving Information Perception of Graphical Displays – an Experimental Study on the Display of Column Graphs	Falschlunger Lisa and Eisl Christoph and Losbichler Heimo and Greil Andreas
S51	2014	Task-Driven Evaluation of Aggregation in Time Series Visualization	Danielle Albers and Michael Correll and Michael Gleicher
S52	2014	Would you prefer pie or cupcakes? Preferences for data visualization designs of professionals and laypeople in graphic design	Annemarie Quispel and Alfons Maes
S53	2014	Evaluation of information visualization techniques: analysing user experience with reaction cards	Tanja Merčun
S54	2014	Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error	Michael Correll and Michael Gleicher
S55	2014	The relation between visualization size, grouping, and user performance	Connor C. Gramazio and Karen B. Schloss and David H. Laidlaw
S56	2015	An Evaluation of the Impact of Visual Embellishments in Bar Charts	Drew Skau and Lane Harrison and Robert Kosara
S57	2016	Beyond Memorability: Visualization Recognition and Recall	Borkin M., Bylinskii Z., Kim N., Bainbridge C., Yeh C., Borkin D., Pfister H., Oliva A.
S58	2015	Towards an intelligent evaluation method of medical data visualizations	Saber Amri and Hela Ltifi and Mounir Ben Ayed
S59	2017	VLAT: Development of a Visualization Literacy Assessment Test	Sukwon Lee and Sung-Hee Kim and Bum Chul Kwon
S60	2017	Narratives in Crowdsourced Evaluation of Visualizations: A Double-Edged Sword?	Evanthia Dimara and Anastasia Bezerianos and Pierre Dragicevic
S61	2017	Assessing User Engagement in Information Visualization	Ya Hsin Hung and Paul Parsons

Doctorado en Informática

Código	Año	Título	Autores
S62	2017	Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences	S. McKenna and N. Henry Riche and B. Lee and J. Boy and M. Meyer
S63	2018	Aesthetic Experimental Study on Information Visualization Design Under the Background of Big Data	Tian Lei and Nan Ni and Qiumeng Zhu and Sijia Zhang
S64	2018	Modeling Color Difference for Visualization Design	Danielle Albers Szafir
S65	2018	Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support	Evanthia Dimara and Anastasia Bezerianos and Pierre Dragicevic
S66	2018	Assessing the Graphical Perception of Time and Speed on 2D+Time Trajectories	Charles Perin and Tiffany Wun and Richard Pusch and Sheelagh Carpendale
S67	2018	Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series	Yunhai Wang and Fubo Han and Lifeng Zhu and Oliver Deussen and Baoquan Chen
S68	2018	Evaluating Interactive Graphical Encodings for Data Visualization	Bahador Saket and Arjun Srinivasan and Eric D. Ragan and Alex Endert
S69	2018	What's the Difference? Evaluating Variants of Multi-Series Bar Charts for Visual Comparison Tasks	Arjun Srinivasan and Matthew Brehmer and Bongshin Lee and Steven M. Drucker
S70	2018	The Development of Heuristics for Evaluation of Dashboard Visualizations	Dawn Dowding and Jacqueline A Merrill
S71	2018	An Instrument for Evaluating the Quality of Data Visualizations	Raissa Barcellos and Jose Viterbo and Flavia Bernardini and Daniela Trevisan
S72	2018	Improving Perception Accuracy in Bar Charts with Internal Contrast and Framing Enhancements	Jose Diaz and Oscar Meruvia-Pastor and Pere-Pau Vazquez
S73	2019	A Heuristic Approach to Value-Driven Evaluation of Visualizations	Wall E., Agnihotri M., Matzen L., Divis K., Haass M., Endert A., Stasko J.
S74	2019	Mapping Color to Meaning in Colormap Data Visualizations	Schloss K., Gramazio C., Silverman A., Parker M., Wang A.
S75	2019	Face to Face: Evaluating Visual Comparison	Brian Ondov and Nicole Jardine and Niklas Elmqvist and Steven Franconeri
S76	2019	Understanding Visual Cues in Visualizations Accompanied by Audio Narrations	Ha Kyung Kong and Zhicheng Liu and Wenjie Zhu and Karrie Karahalios
S77	2019	Comparing the effectiveness of visualizations of different data distributions	Ariane M. B. Rodrigues and Gabriel D. J. Barbosa and Helio Lopes and Simone D. J. Barbosa
S78	2020	A Baseline Study of Emphasis Effects in Information Visualization	Aristides Mairena and Martin Dechant and Carl Gutwin and Andy Cockburn

Doctorado en Informática

Código	Año	Título	Autores
S79	2020	The Effect of Color Scales on Climate Scientists' Objective and Subjective Performance in Spatial Data Analysis Tasks	Dasgupta A., Poco J., Rogowitz B., Han K., Bertini E., Silva C.
S80	2020	Structure and Empathy in Visual Data Storytelling: Evaluating their Influence on Attitude	J. Liem and C. Perm and J. Wood
S81	2022	Declutter and Focus: Empirically Evaluating Design Guidelines for Effective Data Communication	Ajani K., Lee E., Xiong C., Knafllic C., Kemper W., Franconeri S.
S82	2021	Graph Design: The Data-ink Ratio and Expert Users	Kevin McGurgan and Elena Fedoroksaya and Tina Sutton and Andrew Herbert
S83	1994	Minimalism and the Syntax of Graphs	D. J. Gillan and E. H. Richman
S84	1984	Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods	William S. Cleveland and Robert McGill
S85	2008	Exploring the Role of Individual Differences in Information Visualization	Cristina Conati and Heather Maclaren
S86	2009	Visual Intelligence Density: Definition, Measurement, and Implementation	Xiaoyan Bai and David White and David Sundaram
S87	2019	Task-Based Effectiveness of Basic Visualizations	Bahador Saket and Alex Endert and Cagatay Demiralp
S88	1996	The use or misuse of three-dimensional graphs to represent lower-dimensional data	Michael Siegrist
S89	2014	Four Experiments on the Perception of Bar Charts	Justin Talbot and Vidya Setlur and Anushka Anand
S90	2016	How Colors in Business Dashboards Affect Users' Decision Making	Palash Bera
S91	2022	Improving Visualization Design for Effective Multi-Objective Decision Making	Bianchi Dy and Nazim Ibrahim and Ate Poorthuis and Sam Joyce
S92	2021	The Unmet Data Visualization Needs of Decision Makers within Organizations	Evanthia Dimara and Harry Zhang and Melanie Tory and Steven Franconeri
S93	2021	Remote Usability Assessment of Topic Visualization Interfaces with Public Participation Data: A Case Study	Ivania Yovanovic and Iñaki Goñi and Constanza Miranda
S94	2021	Visual Arrangements of Bar Charts Influence Comparisons in Viewer Takeaways	Xiong C., Setlur V., Bach B., Lin K., Koh E., Franconeri S.
S95	2021	Which Emphasis Technique to Use? Perception of emphasis techniques with varying distractors, backgrounds and visualization types	Aristides Mairena and Carl Gutwin and Andy Cockburn

Doctorado en Informática

Código	Año	Título	Autores
S96	2021	Characteristic traits of visualization for decision-making in the early stages of buildign design	Pil Brix Purup and Steffen Petersen
S97	2021	Once Upon A Time In Visualization Unde for decision-making in the early stages of buildity Causality	Choudhry A., Sharma M., Chundury P., Kapler T., Gray D., Ramakrishnan N., Elmqvist N.
S98	2022	Assessment Model for Information Visualization of Human-Computer Interface in the COVID-19 Pandemic	Liu J., Wahab A., Ang M., Chaw J., Ng K.
S99	2023	How the Preattentive Process is Exploi for decision-making in the early stages of buildiw Review	Luisa Barrera-Leon and Fulvio Corno and Luigi De Russis
S100	2022	Narrative medical visualization to communicate disease data	Meuschke M., Garrison L., Smit N., Bach B., Mittenentzwei S., Weiß V., Bruckner S., Lawonn K., Preim B.
S101	2022	VisRecall Quantifying Information Visu for decision-making in the early stages of building Answering	Yao Wang and Chuhan Jiao and Mihai Bace and Andreas Bulling
S102	2023	A Characterization of Interactive Visu for decision-making in the early stages of buildial Temporal	Benedikt Mayer and Nastasja Steinhauer and Bernhard Preim and Monique Meuschke
S103	2023	Breaking the Fourth Wall of Data Stori for decision-making in the early stages of buildion Interaction	Yang Shi and Tian Gao and Xiaohan Jiao and Nan Cao
S104	2023	Evaluating User Experience in Information Visualization Systems: UXIV an Evaluation Questionnaire	Eliane Zambon Victorelli and Julio Cesar dos Reis
S105	2023	Exploring Pie Charts and Part-To-Whole for decision-making in the early stages of buildich Approach	D. Bacic and A. Krbanjevic and N. Jukic
S106	2024	Adopting Human-data Interaction Guidelines and Participatory Practices for Supporting Inexperienced Designers in Information Visualization Applications	Eliane Zambon Victorelli and Julio Cesar Dos Reis
S107	2024	Affective Visualization Design Leverag for decision-making in the early stages of buildita Data	Xingyu Lan and Yanqiu Wu and Nan Cao
S108	2024	Design Paterns for Data-Driven News Articles	Shan Hao and Zezhong Wang and Benjamin Bach and Larissa Pschetz
S109	2024	Evaluating narrative visualization: a survey of practitioners	Nina Errey and Jie Liang and Tuck Wah Leong and Didar Zowghi
S110	2024	Exploring the Impact of Gestalt Laws of Human Perception in Business Information Visualization Context: An Eye-Tracking Study	Dinko Bačić



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

Código	Año	Título	Autores
S111	2024	How Do Low-Vision Individuals Experience Information Visualization?	Yanan Wang and Yuhang Zhao and Yea Seul Kim
S112	2024	Struggles and Strategies in Understanding Information Visualizations	Maryam Rezaie and Melanie Tory and Sheelagh Carpendale
S113	2024	Towards Visualization Thumbnail Designs That Entice Reading Data-Driven Articles	Kim H., Kim J., Han Y., Hong H., Kwon O., Park Y., Elmqvist N., Ko S., Kwon B.
S114	2024	VisCollage: Annotative Collages for Organizing Data Event Charts	Xiao Han Li and Yi Ting Hung and Jia Yu Pan and Wen Chieh Lin
S115	2024	Why Change My Design: Explaining Poorly Constructed Visualization Designs with Explorable Explanations	Leo Yu Ho Lo and Yifan Cao and Leni Yang and Huamin Qu
S116	2023	Attitudinal effects of data visualizations and illustrations in data stories	Manuela Garreton and Francesca Morini and Pablo Celhay and Marian Dork and Denis Parra

Doctorado en Informática

Apéndice B. Lista completa de aplicaciones

Este apéndice presenta la lista completa de las 55 aplicaciones iniciales definidas en el modelo NaVi-Q. La Tabla B. 1 detalla, para cada aplicación, su identificador (de AP01 a AP55), una breve descripción, el instrumento utilizado para su evaluación y la escala de valores correspondiente.

Tabla B. 1: Lista completa de aplicaciones

ID	Aplicación	Descripción	Opciones	Valor
AP01	Línea base en cero	El eje de los gráficos de barras debe comenzar en cero.	Línea base en cero	1
			Línea base no en cero	0
			No es claro	0
			N/A	
AP02	Doble eje Y	El uso de doble eje Y debe estar justificado.	No hay doble eje Y	1
			El doble eje Y es necesario	0.5
			El doble eje Y no está justificado	0
			N/A	
AP03	Títulos de ejes	Los ejes deben estar nombrados adecuadamente.	Títulos de ejes descriptivos	1
			Títulos genéricos	0.5
			Sin títulos	0
AP04	Cambios de escala	Los cambios en la escala de datos deben ser visualmente claros.	Los cambios son explícitos	1
			Los cambios no son explícitos	0
			N/A	
AP05	Cortes de escala	Los valores a través de cortes de escala no deben conectarse visualmente.	Los valores están conectados	0
			Los valores no están conectados	1
			N/A	
AP06	Series limitadas	El número de series debe limitarse a la capacidad de la memoria de trabajo.	Hasta 2 series	1
			Entre 3 y 5 series	0.5
			Más de 5 series	0
AP07	Variables no ambiguas	Las variables deben estar definidas de forma precisa y no ambigua.	Las variables son precisas	1
			Las variables son ambiguas	0
AP08	Relación de aspecto		Cumplida	1

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
		La relación de aspecto debe seguir la regla del banking a 45°.	Parcialmente cumplida	0.5
			No cumplida	0
			N/A	
AP09	Orden de los datos	Los datos deben seguir un orden específico.	Los datos están ordenados intencionalmente	1
			Los datos están ordenados automáticamente	0.5
			Los datos no están ordenados	0
AP10	Etiquetas de datos	Los datos deben estar etiquetados directamente.	Los datos están etiquetados directamente	1
			Los datos están etiquetados mediante leyendas	0.5
			Los datos están etiquetados de ambas formas	0.25
			Los datos no están etiquetados	0
AP11	Cambios en los datos	Los cambios en los datos deben ser visualmente explícitos.	Los cambios son explícitos	1
			Algunos cambios son explícitos	0.5
			Los cambios no son explícitos	0
			N/A	
AP12	Valores atípicos	Los valores atípicos o anomalías deben ser identificados claramente.	Los valores atípicos están identificados	1
			Los valores atípicos no están identificados	0
			N/A	
AP13	Datos proyectados	Los datos proyectados deben diferenciarse adecuadamente.	Los datos proyectados están diferenciados	1
			Los datos proyectados no están diferenciados	0
			N/A	
AP14	Efectividad del canal visual	Los canales visuales deben corresponder al tipo de dato representado.	Posición/longitud – datos cuantitativos	1
			Color/forma – datos cualitativos	1
			Ninguno de los anteriores	0
AP15	Codificación intuitiva	La codificación visual debe ser significativa y fácil de interpretar.	Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP16	Elementos emparejados	Las series y leyendas deben asociarse mediante color o tipografía.	Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0
			N/A	
AP17	Uniformidad del gráfico	El tipo de gráfico debe ser consistente en paneles múltiples.	Se usa un solo tipo de gráfico	1
			Se usan pocos tipos de gráfico	0.5
			Se usan muchos tipos de gráfico	0
			N/A	
AP18	Esquema de color	El esquema de color debe adaptarse al tipo de datos.	Colores secuenciales para datos ordenados o numéricos	1
			Colores divergentes para datos ordenados o numéricos	1
			Colores categóricos para datos nominales o no ordenados	1
			Ninguna de las anteriores	0
			N/A	
AP19	Colores limitados	La cantidad de colores debe estar dentro de los límites de la memoria de trabajo.	Hasta 2 colores	1
			Entre 3 y 5 colores	0.5
			Más de 5 colores	0
AP20	Color funcional	El color debe usarse para comunicar información, no solo para decoración.	Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0
			N/A	
AP21	Colores intuitivos	Los colores deben tener una resonancia semántica con los datos.	Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0
			N/A	
AP22	Colores accesibles	Los colores deben ser distinguibles por usuarios con discapacidades visuales.	Distinguibles para daltonismo	0.25
			Distinguibles en escala de grises	0.25
			Contraste suficiente entre colores	0.25
			No causan fatiga o incomodidad visual	0.25
			Todas las anteriores	1

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP23	Mensaje principal	El mensaje clave debe ser explícito en títulos o subtítulos.	Ninguna de las anteriores	0
			N/A	
			El mensaje principal aparece en el título	1
			El mensaje principal no aparece en el título	0
AP24	Títulos breves	Los títulos deben ser breves y descriptivos.	N/A	
			Entre 6 y 12 palabras	1
			Más de 12 palabras	0.5
			Menos de 6 palabras	0.5
AP25	Redacción	El texto debe estar optimizado para la comunicación.	No hay título	0
			Títulos y subtítulos en voz activa	0.33
			Títulos y subtítulos como oraciones completas	0.33
			Cada oración transmite una idea única	0.33
			Todas las anteriores	1
			Ninguna de las anteriores	0
AP26	Gramática y ortografía	El texto debe estar libre de errores gramaticales u ortográficos.	N/A	
			Sin errores	1
			Errores menores	0.75
			Errores importantes	0.50
			Múltiples errores	0.25
AP27	Tamaño de fuente	El tamaño de fuente debe ser jerárquico y legible.	Texto difícil de entender	0
			Cumplida	1
			Parcialmente cumplida	0.5
AP28	Alineación del texto	Los elementos de texto deben estar alineados para optimizar el espacio.	No cumplida	0
			Alineación a la izquierda	1
			Alineación a la derecha	0.5
AP29	Orientación del texto	La orientación del texto debe facilitar la legibilidad.	Alineación centrada	0.5
			Orientación horizontal	1
			Orientación vertical	0.5
			Orientación diagonal	0
			Orientaciones mixtas	0

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP30	Elementos paratextuales	Etiquetas, leyendas y anotaciones deben usarse con moderación.	Son apropiados	1
			Son excesivos o poco claros	0.5
			Son escasos o insuficientes	0.2
			No hay elementos paratextuales	0
AP31	Contraste	El texto debe ser legible sobre el fondo.	Hay suficiente contraste entre texto y fondo	1
			No hay suficiente contraste	0
AP32	Líneas de cuadrícula	Las cuadrículas deben ser útiles y no intrusivas.	No se usan líneas de cuadrícula	1
			Se usan con moderación	0.5
			Se usan moderadamente	0.2
			Se usan en exceso	0
			N/A	
AP33	Efectos 3D	Los efectos visuales 3D deben evitarse salvo que sean necesarios.	No hay efectos 3D	1
			Los efectos 3D son necesarios	0.5
			Los efectos 3D son ornamentales	0
AP34	Énfasis visual	Los datos importantes deben resaltarse visualmente.	Énfasis mediante color	1
			Énfasis mediante tamaño	1
			Énfasis mediante anotaciones	1
			Énfasis mediante grosor	1
			Énfasis mediante animaciones	1
			Énfasis mediante transparencia	1
			Sin énfasis particular	0
AP35	Marcadores de datos	Los marcadores de datos no esenciales deben minimizarse o eliminarse.	No hay marcadores	1
			Solo se usan los necesarios	1
			Están todos presentes	0
			Faltan pero son necesarios	0
			N/A	
AP36	Marcas y etiquetas de escala	Las marcas y etiquetas deben usarse solo si son necesarias.	No hay marcas ni etiquetas	1
			Solo se incluyen las necesarias	1
			Están todas presentes	0
			Faltan pero son necesarias	0
			N/A	

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP37	Fondo	El fondo debe ser discreto y no interferir con la visualización.	No se usó fondo	1
			El fondo es simple y no interfiere	0.5
			El fondo es excesivo o distrae	0
AP38	Líneas de ejes	Los ejes no deben incluir líneas innecesarias.	Los ejes no tienen líneas innecesarias	1
			Algunas líneas son innecesarias	0.5
			Todas las líneas son innecesarias	0
			N/A	
AP39	Bordes	La visualización no debe tener bordes innecesarios.	No hay bordes innecesarios	1
			Hay bordes innecesarios o excesivos	0
AP40	Atribución de fuentes	Las fuentes de datos deben estar citadas.	Las fuentes están citadas o enlazadas	1
			Las fuentes no están citadas ni enlazadas	0
			N/A	
AP41	Identificación del autor	Debe indicarse quién elaboró la visualización.	El autor está identificado	1
			No hay identificación del autor	0
			N/A	
AP42	Supuestos de conocimiento	La visualización no debe requerir conocimientos previos para comprenderse.	Se entiende sin conocimientos previos	1
			Requiere contexto, pero es accesible	0.5
			Requiere conocimientos previos, difícil de entender	0
			N/A	
AP43	Sobre-simplificación	No deben simplificarse en exceso fenómenos complejos.	Los datos están representados con precisión	1
			Los datos están levemente simplificados	0.5
			Los datos están muy simplificados	0.2
			Los datos están sobre-simplificados y llevan a error	0
			N/A	

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP44	Referencias adicionales	Deben incluirse referencias adicionales si es necesario.	Se incluyen referencias adicionales	1
			No se incluyen referencias adicionales	0
			N/A	
AP45	Audiencia	La visualización debe estar dirigida claramente a su audiencia.	El tipo de gráfico es lo suficientemente complejo	0.33
			Se consideran las necesidades de la audiencia	0.33
			Los títulos tienen orientación a la acción	0.33
			Todas las anteriores	1
			Ninguna de las anteriores	0
AP46	Mensajes y resúmenes	El texto debe resumir y explicar la información presentada.	Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0
			N/A	
AP47	Ilustraciones	Las ilustraciones, si están presentes, deben ser relevantes para el contenido.	Las ilustraciones están relacionadas con los conceptos	1
			Las ilustraciones son decorativas	0.5
			Las ilustraciones son irrelevantes o confusas	0
			N/A	
AP48	Interactividad sugerida	Los elementos interactivos deben estar claramente señalizados.	Todos los elementos interactivos están indicados/marcados	1
			Algunos elementos están indicados/marcados	0.5
			No hay indicaciones	0
			N/A	
AP49	Retroalimentación en la interacción	Debe proporcionarse retroalimentación al usuario durante la interacción.	Indicadores visuales	1
			Indicadores sonoros	1
			Tooltips	1
			Mensajes de error	1
			Animación	1
			Otro	1
			No hay retroalimentación	0

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
AP50	Visión general primero; detalles bajo demanda	Los datos generales deben preceder la información específica.	N/A	
			Cumplida	1
			Parcialmente cumplida	0.5
			No cumplida	0
AP51	Secuencia de eventos	Los eventos deben tener un orden intencional.	N/A	
			Orden lineal	1
			Orden condicional	1
			Orden híbrido	1
			Ninguno de los anteriores	0
AP52	Transiciones	Las transiciones deben guiar la navegación en visualizaciones con múltiples paneles.	N/A	
			Transiciones temporales	1
			Transiciones causales	1
			Transiciones discursivas	1
			Transiciones de granularidad	1
			Transiciones combinadas	1
			Otro tipo de transición	1
			No hay transiciones	0
AP53	Navegación	Deben estar disponibles mecanismos de navegación entre paneles múltiples.	N/A	
			Desplazamiento (scroll)	1
			Botones	1
			Barra deslizante	1
			Otro	1
			No hay navegación disponible	0
AP54	Progreso	El avance debe estar claramente indicado en visualizaciones con múltiples paneles.	N/A	
			Número de página o diapositiva	1
			Porcentaje completado	1
			Línea de tiempo o mapa de ruta	1
			Otro	1
			El progreso no está indicado	0
AP55	Accesibilidad para impresión		N/A	
			Tipografía	0.25
			Tamaño de fuente	0.25



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

ID	Aplicación	Descripción	Opciones	Valor
			Colores	0.25
			Contraste	0.25
		Todos los elementos deben estar optimizados para impresión.	Todos los elementos están optimizados	1
			Ninguno de los elementos está optimizado	0
			N/A	

Apéndice C. Materiales: evaluación a pequeña escala

Este apéndice reúne los materiales utilizados en el estudio de viabilidad descrito en la sección 5.1. La evaluación incluyó el uso del perfil de calidad Estándar, definido mediante el método AHP-Express, y se aplicó sobre un conjunto de 18 visualizaciones (12 estáticas y 6 interactivas). Se incluyen los siguientes materiales:

- La hoja de cálculo utilizada por los evaluadores, con criterios, propiedades, aplicaciones y fórmulas de puntuación.
- El script de Python empleado para seleccionar las visualizaciones estáticas mediante muestreo aleatorio.
- El formulario de encuesta posterior a la evaluación, diseñado para recoger percepciones sobre la experiencia de uso y dificultad del modelo.

C.1. Hoja de cálculo de evaluación

Para llevar a cabo la evaluación, cada participante completó una hoja de cálculo que incluía los criterios, propiedades y aplicaciones del modelo, junto con menús desplegables para seleccionar respuestas y fórmulas que calculaban automáticamente los puntajes parciales y totales. A continuación, se presenta un fragmento de dicha hoja (Fig. C. 1 y Fig. C. 2), con el fin de ilustrar su estructura y funcionamiento. El archivo completo, en formato editable, se encuentra disponible en el soporte digital complementario⁶.

⁶ <https://doi.org/10.17605/OSF.IO/KCUHA>

Doctorado en Informática

ID	Application	Description	Answer
AP01	Zero baseline	Bar charts axis should start at zero.	Non-zero baseline
AP02	Dual Y axis	The use of dual Y-axis should be justified.	Zero baseline Non-zero baseline Unclear
AP03	Axes titles	Axes should be named appropriately.	N/A
AP04	Scale changes	Changes in data scale should be visually clear	Changes in scale are explicit
AP05	Scale breaks	Values across scale breaks should not be connected	Values are connected
AP06	Limited series	The number of series should be limited to working memory capacity.	Between 3 and 5 series
AP07	Unambiguous variables	Variables should be defined in an accurate, unambiguous manner.	Variables are accurate
AP08	Aspect ratio	Aspect ratio should comply to the banking at 45° rule	N/A
AP09	Data ordering	Data should follow a specific order.	Data is ordered intentionally
AP10	Data labels	Data should be labeled directly.	Data is labeled by legends

Fig. C. 1: Tabla para respuestas

Prop ID	Associated App ID	User Input	App Weight	App Value	Total score by app
P01	AP01	Non-zero baseline	1/4	0,00	0,00
	AP03	Descriptive axes titles	1/4	1,00	0,25
	AP04	Changes in scale are explicit	1/4	1,00	0,25
	AP05	Values are connected	1/4	0,00	0,00
	AP08	N/A			
Prop ID	Associated App ID	User Input	App Weight	App Value	Total score by app
P02	AP14	Nonlength is used for quantitative	1/3	1,00	0,33
P02	AP16	Partially fulfilled	1/3	0,50	0,17
	AP20	Fulfilled	1/3	1,00	0,33
Prop ID	Associated App ID	User Input	App Weight	App Value	Total score by app
P03	AP18	N/A			
P03	AP19	Between 3 and 5 colors	1/4	0,50	0,13
P03	AP20	Fulfilled	1/4	1,00	0,25
P03	AP21	Partially fulfilled	1/4	0,50	0,13
P03	AP22	enough contrast between colors;	1/4	0,75	0,19
Criteria ID	Associated Prop ID	Prop Weight	Prop Value	Total Score by Prop	
C01	P01	1/9	0,50	0,06	
C01	P02	1/9	0,83	0,09	
C01	P03	1/9	0,69	0,08	
C01	P05	1/9	0,70	0,08	
C01	P06	1/9	0,75	0,08	
C01	P07	1/9	0,68	0,08	
C01	P11	1/9	0,70	0,08	
C01	P12	1/9	0,81	0,09	
C01	P13	1/9	0,70	0,08	
Criteria ID	Associated Prop ID	Prop Weight	Prop Value	Total Score by Prop	
C02	P03	1/8	0,69	0,09	
C02	P05	1/8	0,70	0,09	
C02	P11	1/8	0,70	0,09	
C02	P12	1/8	0,81	0,10	
C02	P13	1/8	0,70	0,09	
C02	P14	1/8	0,76	0,09	
C02	P16	1/8	0,75	0,09	
C02	P17	1/8	1,00	0,13	

Total Quality Score

0,73

Fig. C. 2: Cálculo de resultados

C.2. Script de selección aleatoria de visualizaciones

Para seleccionar de forma sistemática las visualizaciones estáticas empleadas en la evaluación, se utilizó un script en Python que aplica muestreo aleatorio simple sobre una carpeta con imágenes. El script recorre los archivos disponibles en el directorio de entrada, selecciona seis imágenes al azar y las copia en un directorio de salida. A continuación, se presenta el código completo utilizado en esta etapa.

Doctorado en Informática

```
import os
import random
import shutil

# Set the path to the folder containing the images
folder_path = ''

# Set the path to the folder where you want to save the selected images
output_folder_path = ''

# Create the output folder if it doesn't exist
if not os.path.exists(output_folder_path):
    os.makedirs(output_folder_path)

# Get a list of all the image files in the folder
image_files = [f for f in os.listdir(folder_path) if f.endswith('.jpg') or
                f.endswith('.png')]

# Select 6 random image files
selected_images = random.sample(image_files, k=6)

# Copy the selected images to the output folder
for image_file in selected_images:
    src_path = os.path.join(folder_path, image_file)
    dst_path = os.path.join(output_folder_path, image_file)
    shutil.copy(src_path, dst_path)

# Print the list of selected image files
print(selected_images)
```

C.3. Encuesta posterior a la evaluación

Como parte del estudio, se solicitó a los participantes que completaran una encuesta posterior a la aplicación del modelo NaVi-Q, con el objetivo de recoger percepciones sobre su experiencia, identificar fortalezas y limitaciones, y evaluar la adecuación de los distintos componentes del modelo (criterios, propiedades y aplicaciones). El formulario incluía preguntas cerradas, basadas en escalas de Likert de cinco puntos, y preguntas abiertas orientadas a captar comentarios cualitativos.

1. Evaluación general

- ¿Cómo calificaría su experiencia general utilizando el modelo de calidad para visualizaciones narrativas?
- (Respuesta obligatoria – escala: Muy difícil / Algo difícil / Ni fácil ni difícil / Algo fácil / Muy fácil)

2. Comentarios adicionales (opcional)

Por favor, indique cualquier comentario adicional que desee compartir.

Doctorado en Informática

3. Fortalezas

Describa las fortalezas que encontró en el modelo durante su evaluación.

4. Limitaciones

Describa las limitaciones que encontró al utilizar el modelo.

5. Adecuación de los criterios de calidad

- ¿Cómo calificaría la adecuación de los criterios de calidad para guiar su evaluación?
Nada adecuados – Poco adecuados – Medianamente adecuados – Muy adecuados
– Completamente adecuados
- ¿Hubo algún criterio que encontró particularmente relevante o que haya echado en falta?

6. Adecuación de las propiedades

- ¿Cómo calificaría la capacidad de las propiedades para captar los aspectos importantes de las visualizaciones narrativas? (Misma escala que la anterior)
- ¿Hubo alguna propiedad que encontró particularmente relevante o que haya echado en falta?

7. Adecuación de las aplicaciones

- ¿Cómo calificaría la adecuación de las aplicaciones para cuantificar las propiedades de visualización? (Misma escala que la anterior)
- ¿Hubo alguna aplicación que encontró particularmente relevante o que haya echado en falta?

8. Recomendaciones generales

En base a su experiencia, ¿cuáles son sus recomendaciones generales sobre el uso del modelo? Indique sugerencias de mejora u otras observaciones que considere relevantes.

Apéndice D. Materiales: experimento

Este apéndice reúne los materiales utilizados en el estudio experimental descrito en la sección 5.4. El estudio aplicó un diseño factorial con una muestra ampliada de participantes y visualizaciones, con el objetivo de analizar la capacidad del modelo para discriminar niveles de calidad y producir evaluaciones consistentes. Se incluyen los siguientes materiales:

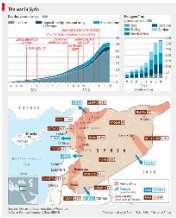
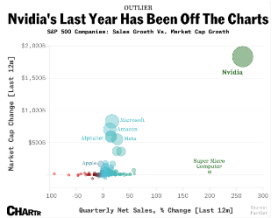
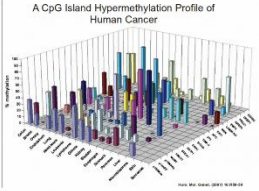
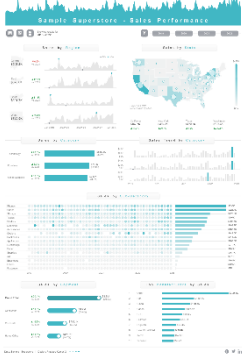
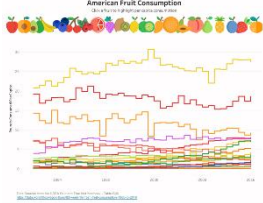
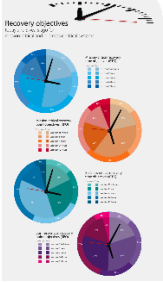
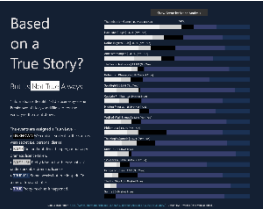
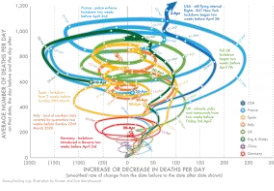
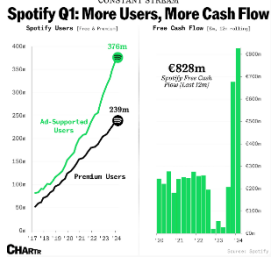
- Conjunto de visualizaciones utilizadas, seleccionadas a partir de fuentes heterogéneas (MASSVIS, Tableau, blogs y repositorios públicos).
- Formulario de asociación entre propiedades y criterios, utilizado para evaluar la validez de las propiedades del modelo.
- Encuesta de experiencia de uso, diseñada para recoger percepciones sobre la claridad del instrumento, la utilidad de las aplicaciones y la satisfacción general con el proceso de evaluación.

D.1. Visualizaciones utilizadas

La Tabla D. 1 presenta las 30 visualizaciones evaluadas. Cada una recibió un identificador único (de V01 a V30) y se organizaron en tres grupos, de acuerdo con sus puntajes de referencia: calidad baja (sombreado rojo), media (sombreado amarillo) y alta (sombreado verde). Las miniaturas incluidas a continuación permiten una referencia visual rápida; las versiones completas junto con las evaluaciones de referencia se encuentran disponibles en el soporte digital complementario.

Doctorado en Informática

Tabla D. 1: Visualizaciones evaluadas agrupadas por nivel de calidad

V01		V11		V21	
V02		V12		V22	
V03		V13		V23	
V04		V14		V24	
V05		V15		V25	

V06 Issues - Grouped by Assignee

Status	Nick Fury	Vincent Wong	Carol Danvers	Micah Runner	Van Helsing
Backlog	16 (30.8%)	8 (15.4%)	3 (5.8%)	17 (32.7%)	8 (15.4%)
In Progress	0	0	0	0	0
Blocked	0	0	0	0	0
On Hold	0	0	0	0	0
Closed	0	0	0	0	0

V16 Letting the Dogs Out

Bank of America's holding is expected to rise as mortgage REITs continue to rise. Mortgage REITs are expected to rise as mortgage REITs continue to rise.

V26 amazon: Bigger Than Ever At 30

Amazon's annual revenue has grown from \$1.2 billion in 1998 to \$30 billion in 2020. The chart shows a steady upward trend with a significant acceleration in the late 2010s.

V07 Pre-tax profits of the 1,000 largest banks

The charts show the distribution of pre-tax profits for the 1,000 largest banks. In 2007, the top 10 banks accounted for 48.7% of total profits. In 2011, the top 10 banks accounted for 46.2% of total profits.

V17 No Longer Gaining Ground

The chart shows unemployment rates for various states from 2009 to 2012. The rates for most states have declined since 2009, but they remain higher than the U.S. rate in June.

V27 THE GENDER GAP

The scatter plot shows the gender gap in various countries. The y-axis represents the gender gap score, and the x-axis represents the country. The chart shows a positive correlation between the gender gap score and the country's economic status.

V08 Profit vs. Month

The bar chart shows profit for each month from January to May. The profit starts at \$1.0 million in January, rises to \$1.5 million in February, and continues to rise to \$2.5 million in May.

V18 July 4th Often Leads To More Wildfires

The line chart shows the number of wildfires recorded by day of the year. The number of wildfires peaks in July, with a significant increase on July 4th.

V28 Real Estate Stocks Are Lagging

The line chart shows the performance of real estate stocks compared to the S&P 500. Real estate stocks have lagged behind the S&P 500 since 2007, with a significant decline in 2008.

V09 History of James Bond Movies

The bar chart shows the history of James Bond movies from 1962 to 2015. The chart shows a steady increase in the number of movies over time, with a peak in the 1970s and 1980s.

V19 Where Are Prices Still Rising?

The bar chart shows the percentage of households that have increased prices in various categories. The highest percentage of households have increased prices in the 'All Items' category, followed by 'Food away from home' and 'Gasoline (all types)'.

V29 It's a seller's market! Home sales boom in Hamilton this year

The line chart shows home sales in Hamilton from 2010 to 2015. The chart shows a significant increase in home sales in 2015, indicating a seller's market.

V10 BY THE NUMBERS

The bar chart shows the number of opiate deaths in 2014, 2015, and 2016. The number of deaths has increased significantly over the three-year period.

V20 Health Club

The image shows a man in a health club setting, likely a personal trainer or a member. The man is wearing a white shirt and is looking down at something in his hands.

V30 HEALTH SPENDING

The line chart shows health spending in the U.S. from 2000 to 2015. The chart shows a steady increase in health spending over the 15-year period.

Doctorado en Informática

D.2. Formulario de criterios y propiedades

Se solicitó a los participantes que indicaran el grado de asociación percibida entre criterio y sus propiedades. Para ello, se utilizó una matriz de evaluación basada en una escala de Likert de cinco puntos, donde 1 indicaba "sin asociación" y 5 "asociación muy fuerte".

¿Qué tan asociada considera que está cada propiedad con el criterio?

Para cada caso, indique el nivel de asociación que observa entre cada propiedad y el criterio correspondiente, utilizando la escala: 1 = Sin asociación, 5 = Asociación muy fuerte.

Criterio: comprensión

Propiedad	1	2	3	4	5
Precisión	☐	☐	☐	☐	☐
Saliencia	☐	☐	☐	☐	☐
Color	☐	☐	☐	☐	☐
Orden visual	☐	☐	☐	☐	☐
Enfoque	☐	☐	☐	☐	☐
Accesibilidad	☐	☐	☐	☐	☐
Claridad	☐	☐	☐	☐	☐
Contexto	☐	☐	☐	☐	☐
Narrativa	☐	☐	☐	☐	☐

Criterio: compromiso

Propiedad	1	2	3	4	5
Color	☐	☐	☐	☐	☐
Orden visual	☐	☐	☐	☐	☐
Claridad	☐	☐	☐	☐	☐
Contexto	☐	☐	☐	☐	☐
Narrativa	☐	☐	☐	☐	☐
Guía para el usuario	☐	☐	☐	☐	☐
Consistencia	☐	☐	☐	☐	☐
Interactividad	☐	☐	☐	☐	☐

Criterio: información

Doctorado en Informática

Propiedad	1	2	3	4	5
Precisión	☐	☐	☐	☐	☐
Disposición	☐	☐	☐	☐	☐
Enfoque	☐	☐	☐	☐	☐
Fiabilidad	☐	☐	☐	☐	☐
Redundancia	☐	☐	☐	☐	☐
Integridad	☐	☐	☐	☐	☐
Contexto	☐	☐	☐	☐	☐

Criterio: memorabilidad

Propiedad	1	2	3	4	5
Color	☐	☐	☐	☐	☐
Enfoque	☐	☐	☐	☐	☐
Redundancia	☐	☐	☐	☐	☐
Claridad	☐	☐	☐	☐	☐
Contexto	☐	☐	☐	☐	☐
Narrativa	☐	☐	☐	☐	☐
Guía para el usuario	☐	☐	☐	☐	☐
Metonimia	☐	☐	☐	☐	☐

Criterio: usabilidad

Propiedad	1	2	3	4	5
Saliencia	☐	☐	☐	☐	☐
Disposición	☐	☐	☐	☐	☐
Accesibilidad	☐	☐	☐	☐	☐
Fiabilidad	☐	☐	☐	☐	☐
Claridad	☐	☐	☐	☐	☐
Guía para el usuario	☐	☐	☐	☐	☐
Consistencia	☐	☐	☐	☐	☐
Interactividad	☐	☐	☐	☐	☐

D.3. Encuesta de experiencia de uso

Al finalizar el experimento, se solicitó a los participantes que completaran una encuesta orientada a evaluar su experiencia con la herramienta web. La encuesta incluyó cinco

Doctorado en Informática

preguntas cerradas con escala de Likert de cinco puntos, junto con una pregunta abierta destinada a recoger sugerencias de mejora.

Encuesta de experiencia de uso

1. Comprensión del modelo

¿Qué tan claras fueron las instrucciones y los enunciados dentro de la herramienta web para evaluar las visualizaciones?

Muy poco claras – Poco claras – Neutras – Claras – Muy claras

2. Apoyo durante la evaluación

¿Considera que la ayuda o el apoyo proporcionado en cada pregunta fue suficiente para completar tu evaluación?

Nada suficiente – Algo insuficiente – Neutro – Suficiente – Muy suficiente

3. Utilidad de las aplicaciones

¿Qué tan útiles te resultaron las aplicaciones para evaluar las visualizaciones?

Nada útiles – Poco útiles – Neutras – Útiles – Muy útiles

4. Claridad de las relaciones

¿Qué tan clara fue la conexión entre los criterios/propiedades y sus aplicaciones específicas durante la evaluación?

Muy poco clara – Poco clara – Neutra – Clara – Muy clara

5. Satisfacción general

¿Qué tan satisfecho/a estás, en general, con tu experiencia utilizando la herramienta web?

Muy insatisfecho/a – Insatisfecho/a – Neutro/a – Satisfecho/a – Muy satisfecho/a

6. Sugerencias de mejora

¿Qué mejoras sugerís para facilitar el uso de la herramienta web?
(Respuesta abierta, 1–2 oraciones)

Doctorado en Informática

Apéndice E. Materiales: estudio de caso

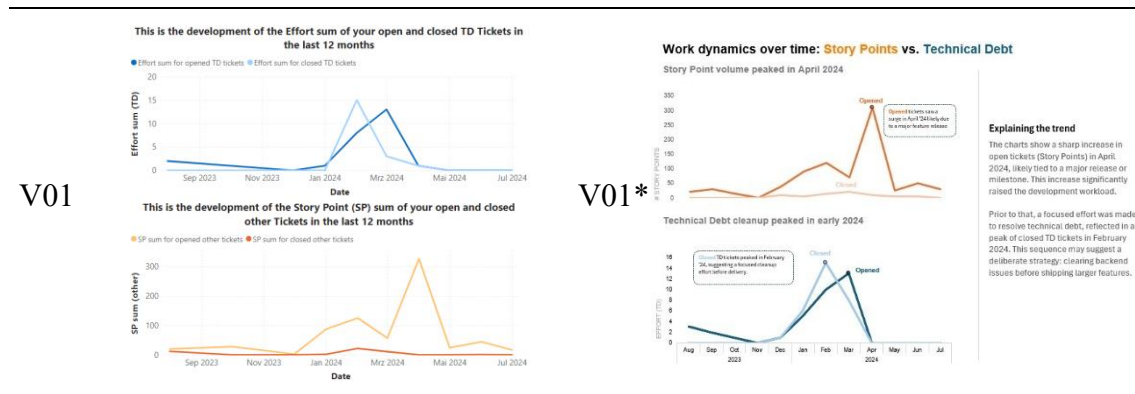
Este apéndice reúne los materiales utilizados durante el estudio de caso presentado en la sección 5.6, orientado a analizar el funcionamiento del modelo NaVi-Q en un entorno aplicado. Se incluyen los siguientes materiales:

- Visualizaciones evaluadas: conjunto de visualizaciones originales utilizadas en el proyecto, así como sus versiones mejoradas luego de aplicar las recomendaciones generadas por el modelo.
- Detalle del perfil de calidad personalizado: incluye los pesos de criterios y propiedades asignados a mediante el método AHP-Express.
- Cuestionario de comprensión e información completo: instrumento aplicado a los grupos control y experimental para evaluar el impacto de las mejoras implementadas.
- Formulario SUS: cuestionario utilizado para estimar la usabilidad percibida de las visualizaciones.

E.1. Visualizaciones evaluadas

La tabla a continuación presenta las visualizaciones utilizadas y su correspondiente versión rediseñada (identificada con un asterisco).

Tabla E. 1: Visualizaciones evaluadas (originales y rediseñadas)



Doctorado en Informática



Nota: V04 incluye vistas alternativas que fueron tratadas como visualizaciones distintas (V05-V07).

E.2. Perfil de calidad personalizado

Este perfil fue construido a partir de la priorización realizada por los miembros del equipo mediante el método AHP-Express, quienes asignaron importancia relativa a los distintos criterios y propiedades del modelo. A continuación, se presentan las tablas completas con los pesos finales de criterios y propiedades.

Tabla E. 2: Criterios de calidad

Criterio	Peso
Comprensión	0,30
Compromiso	0,10
Información	0,29
Memorabilidad	0,11

Doctorado en Informática

Criterio	Peso
Usabilidad	0,20

Tabla E. 3: Pesos por propiedad para el criterio Comprensión

Propiedad	Peso
Precisión	0,23
Saliencia	0,07
Color	0,04
Orden visual	0,14
Enfoque	0,10
Accesibilidad	0,10
Claridad	0,17
Contexto	0,09
Narrativa	0,06

Tabla E. 4: Pesos por propiedad para el criterio Compromiso

Propiedad	Peso
Color	0,07
Claridad	0,28
Contexto	0,14
Narrativa	0,11
Guía para el usuario	0,28
Interactividad	0,13

Tabla E. 5: Pesos por propiedad para el criterio Información

Propiedad	Peso
Precisión	0,26
Integridad	0,31
Fiabilidad	0,25
Contexto	0,18

Tabla E. 6: Pesos por propiedad para el criterio Memorabilidad

Propiedad	Peso
Color	0,10
Enfoque	0,19
Redundancia	0,10
Claridad	0,31
Contexto	0,18
Narrativa	0,12

Doctorado en Informática

Tabla E. 7: Pesos por propiedad para el criterio Usabilidad

Propiedad	Peso
Saliencia	0,05
Disposición	0,11
Accesibilidad	0,13
Fiabilidad	0,18
Claridad	0,15
Guía para el usuario	0,10
Interactividad	0,13
Consistencia	0,15

E.3. Cuestionario de comprensión e información

Pregunta 1: ¿En qué mes se registró el mayor esfuerzo para cerrar tickets de DT?

- ☐ A) Enero 2024
- ☐ B) Febrero 2024
- ☐ C) Marzo 2024
- ☐ D) Abril 2024

Nivel de confianza: [1] [2] [3] [4] [5]

Pregunta 2: Según el gráfico de oportunidades de mejora de bajo esfuerzo, ¿qué combinación de prioridad y esfuerzo caracteriza a estos ítems?

- ☐ A) Baja prioridad y alto esfuerzo
- ☐ B) Alta prioridad y bajo esfuerzo
- ☐ C) Alta prioridad y alto esfuerzo
- ☐ D) Baja prioridad y bajo esfuerzo

Nivel de confianza: [1] [2] [3] [4] [5]

Pregunta 3: ¿Cuál es el nivel de discrepancia más alto entre prioridad dada y calculada en los tickets de DT?

- ☐ A) 3
- ☐ B) 4

Doctorado en Informática

☐ C) 5

☐ D) 2

Nivel de confianza: [1] [2] [3] [4] [5]

Pregunta 4: ¿Cuál es la tendencia general del esfuerzo acumulado en los tickets de DT abiertos en los últimos 12 meses?

☐ A) Aumento sostenido

☐ B) Disminución marcada

☐ C) Estabilidad con una caída puntual

☐ D) Crecimiento exponencial

Nivel de confianza: [1] [2] [3] [4] [5]

Pregunta 5: ¿Qué visualización permite detectar qué tickets tienen un retorno de inversión más temprano?

☐ A) Gráfico de esfuerzo en el tiempo

☐ B) Diagrama de barras de discrepancias

☐ C) Tabla de ROI

☐ D) Diagrama de renvíos

Nivel de confianza: [1] [2] [3] [4] [5]

E.4. Formulario SUS

A continuación, encontrará una serie de afirmaciones sobre su experiencia al utilizar el tablero. Por favor, indique su nivel de acuerdo con cada enunciado, utilizando la siguiente escala:

- 1 = Muy en desacuerdo
- 2 = En desacuerdo
- 3 = Neutro
- 4 = De acuerdo

**Doctorado en Informática**

- 5 = Muy de acuerdo

Marque una única opción por fila.

Ítem	Enunciado	1	2	3	4	5
1	Me parece que usar el tablero fue sencillo.	☐	☐	☐	☐	☐
2	Considero que necesitaría ayuda técnica para poder usar este tablero.	☐	☐	☐	☐	☐
3	Creo que las diferentes partes del tablero están bien integradas.	☐	☐	☐	☐	☐
4	Me pareció que había demasiada complejidad en el uso del tablero.	☐	☐	☐	☐	☐
5	Me sentí seguro/a usando el tablero.	☐	☐	☐	☐	☐
6	Considero que las visualizaciones eran consistentes entre sí.	☐	☐	☐	☐	☐
7	Me gustaría usar un sistema como este con frecuencia.	☐	☐	☐	☐	☐
8	Encontré que las visualizaciones eran confusas.	☐	☐	☐	☐	☐
9	Pude aprender a usar el sistema rápidamente.	☐	☐	☐	☐	☐
10	Necesitaría aprender muchas cosas nuevas antes de poder usar el tablero.	☐	☐	☐	☐	☐

Doctorado en Informática

Apéndice F. Modelo completo NaVi-Q

Este apéndice presenta la versión final del modelo de calidad NaVi-Q. Se incluyen los elementos componentes (criterios, propiedades y aplicaciones) y las relaciones entre ellos en cada nivel de la jerarquía.

F.1. Criterios

La Tabla F. 1 presenta los criterios de evaluación definidos en el modelo, incluyendo su identificador, nombre y una breve descripción.

Tabla F. 1: Criterios de calidad

ID	Criterio	Definición
C01	Comprensión	Capacidad de facilitar la extracción de conocimiento y la interpretación del mensaje.
C02	Compromiso	Grado en que se mantiene el interés, atención o participación activa del usuario.
C03	Información	Calidad, procedencia y transparencia de los datos representados.
C04	Memorabilidad	Grado en que se retiene la información relevante o el mensaje principal.
C05	Usabilidad	Facilidad para alcanzar objetivos de manera efectiva y adaptada al usuario.

F.2. Propiedades

La Tabla F. 2 detalla las propiedades asociadas a cada criterio y su función evaluativa.

Tabla F. 2: Propiedades de calidad

ID	Propiedad	Descripción
P01	Precisión	Uso de una escala adecuada para representar los datos, evitando distorsiones o alteraciones que puedan inducir a error o confusión.
P02	Saliencia	Grado en que la visualización asocia la información relevante con elementos visuales prominentes.
P03	Color	Uso estratégico del color, considerando la cantidad de colores y su función dentro del gráfico.
P04	Disposición	Organización de los elementos visuales (proporciones, contraste, espacios en blanco) que influyen en la interpretación.

Doctorado en Informática

ID	Propiedad	Descripción
P05	Ruido visual	Presencia de elementos innecesarios que no aportan información nueva y pueden generar confusión (por ejemplo, “chart junk”).
P06	Enfoque visual	Uso de recursos visuales (como color o contraste) para resaltar patrones o información importante.
P07	Accesibilidad	Diseño orientado a la comprensión por parte de usuarios diversos, incluyendo distintas capacidades y niveles de conocimiento.
P08	Fiabilidad	Grado de transparencia y documentación respecto a la fuente y calidad de los datos utilizados.
P09	Redundancia	Presentación de la misma información por múltiples vías, facilitando la comprensión sin añadir ruido visual.
P10	Integridad	Representación fiel y completa de la información, sin omisiones ni distorsiones.
P11	Claridad	Capacidad para comunicar ideas complejas de manera clara y concisa, evitando elementos superfluos.
P12	Contexto	Inclusión de información contextual relevante que facilite la comprensión del usuario.
P13	Narrativa	Uso de técnicas narrativas para construir un mensaje estructurado y coherente.
P14	Guía al usuario	Claridad en la navegación e interpretación, especialmente en visualizaciones complejas o con múltiples paneles.
P15	Consistencia	Uniformidad en elementos visuales como paletas de color, tipografías o íconos.
P16	Interactividad	Posibilidad de que el usuario interactúe con los datos (por ejemplo, hacer zoom, desplazarse o aplicar filtros).

F.3. Aplicaciones

La Tabla F. 3 presenta la lista completa de aplicaciones junto con sus instrumentos de evaluación y respectivas opciones de respuesta posibles.

Tabla F. 3: Aplicaciones e instrumentos

ID	Aplicación	Descripción	Instrumento	Valor
AP01	Línea base en cero	Los ejes deben estar nombrados adecuadamente.	Línea base en cero / Línea base no en cero / N/A	1 / 0
AP02	Títulos de ejes	El número de colores debe estar dentro de los límites de la memoria de trabajo.	Títulos de ejes descriptivos / genéricos / sin títulos / N/A	1 / 0.5 / 0
AP03	Cambios en los datos	Los cambios en los datos deben ser visualmente explícitos.	Los cambios son explícitos / no son explícitos / N/A	1 / 0 / N/A

Doctorado en Informática

ID	Aplicación	Descripción	Instrumento	Valor
AP04	Series limitadas	El número de series debe limitarse a la capacidad de la memoria de trabajo.	Hasta 2 series / Entre 3 y 5 / Más de 5	1 / 0.5 / 0
AP05	Orden de los datos	Los datos deben seguir un orden específico.	Datos ordenados intencionalmente / Sin orden discernible	1 / 0
AP06	Etiquetas de series	Las series deben estar etiquetadas directamente.	Etiquetado directo / Por leyenda / Ambas formas / Sin etiquetado	1 / 0.5 / 0.25 / 0
AP07	Mapeo gráfico intuitivo	La codificación visual de la información debe ser significativa y clara.	Cumplido / Parcialmente cumplido / No cumplido	1 / 0.5 / 0
AP08	Elementos pareados	Los patrones visualizados deben asociarse mediante color o tipografía.	Cumplido / Parcialmente cumplido / No cumplido / N/A	1 / 0.5 / 0
AP09	Uniformidad en los gráficos	El tipo de gráfico debe ser consistente entre paneles múltiples.	Se usa un tipo / Se usan pocos tipos / Se usan muchos tipos / N/A	1 / 0.5 / 0
AP10	Esquema de color	El esquema de color debe estar alineado con el tipo de dato.	Valores numéricos / Datos divergentes / Datos categóricos / No coincide / N/A	1 / 1 / 1 / 0
AP11	Colores limitados	El número de colores debe estar dentro de los límites de la memoria de trabajo.	Hasta 2 colores / Entre 3 y 5 / Más de 5	1 / 0.5 / 0
AP12	Color funcional	El color debe usarse para transmitir información, no solo como adorno.	Tienen propósito / No están optimizados / Sin propósito / N/A	1 / 0.5 / 0
AP13	Colores intuitivos	Los colores deben estar en resonancia semántica con los datos.	Intuitivos / Parcialmente intuitivos / Contradictorios o confusos / N/A	1 / 0.5 / 0
AP14	Colores accesibles	Los colores deben distinguirse para personas con discapacidades visuales.	Diferenciables para daltonismo / No diferenciables	1 / 0
AP15	Mensaje principal	Los mensajes clave deben estar explícitos en títulos y subtítulos.	Claramente indicado / No indicado	1 / 0
AP16	Títulos breves	Los títulos deben ser breves y descriptivos.	Cortos y concisos / Largos o poco claros / Sin título	1 / 0 / 0
AP17	Escritura	El texto debe estar optimizado para la comunicación.	Voz activa / Frases completas / Ideas únicas / Ninguna / N/A	0.33 c/u / 0
AP18	Tamaño de fuente	El tamaño de fuente debe ser jerárquico y legible.	Tamaños jerárquicos / Uniformes pero legibles / Inconsistentes o inapropiados	1 / 0.5 / 0
AP19	Alineación del texto	Los elementos textuales deben estar alineados para optimizar el espacio.	Alineado a la izquierda / Derecha / Centrado	1 / 0.5 / 0

Doctorado en Informática

ID	Aplicación	Descripción	Instrumento	Valor
AP20	Orientación del texto	La orientación del texto debe mejorar la legibilidad.	Horizontal / Vertical / Diagonal / Mixta	1 / 0.5 / 0
AP21	Anotaciones	Las etiquetas, leyendas y anotaciones deben usarse con moderación.	Apropiadas / Excesivas / Escasas	1 / 0.5 / 0
AP22	Contraste	El texto debe ser legible sobre el fondo.	Suficiente / Insuficiente	1 / 0
AP23	Líneas de cuadrícula	Las cuadrículas deben ser visibles y no intrusivas.	Apropiadas / Excesivas / N/A	1 / 0 / -
AP24	Efectos 3D	Se deben evitar los efectos visuales 3D salvo que sean necesarios.	Sin efectos 3D / Con fines decorativos	1 / 0
AP25	Resaltado visual	Los puntos y patrones importantes deben destacarse.	Énfasis efectivo / Poco claro / Sin énfasis	1 / 0.5 / 0
AP26	Marcadores de datos	Los marcadores de datos no esenciales deben minimizarse o eliminarse.	Apropiados / Excesivos / Escasos / N/A	1 / 0.5 / 0
AP27	Ticks y etiquetas	Las marcas y etiquetas deben estar solo si son necesarias.	Apropiadas / Excesivas / N/A	1 / 0 / -
AP28	Fondo y bordes	El fondo y los bordes deben ser discretos.	No se usaron / Son neutros / Son distractores	1 / 0.5 / 0
AP29	Fuente de datos	Se deben citar las fuentes de los datos.	Citadas / No citadas / N/A	1 / 0 / -
AP30	Identificación del autor	Debe indicarse la autoría.	Incluida / No incluida / N/A	1 / 0 / -
AP31	Audiencia	Se debe considerar a la audiencia en toda la visualización.	Complejidad adecuada / Títulos orientados a la acción / Ninguna	0.5 / 0.5 / 0
AP32	Mensajes y resúmenes	Se deben usar elementos de texto para resumir y explicar la información.	Claros y útiles / Poco claros o ausentes / N/A	1 / 0 / -
AP33	Ilustraciones	Las ilustraciones, si las hay, deben estar relacionadas con el tema.	Relacionadas / Decorativas o irrelevantes / N/A	1 / 0 / -
AP34	Interactividad sugerida	Los elementos interactivos deben estar señalizados.	Todos señalizados / Algunos / Ninguno / N/A	1 / 0.5 / 0
AP35	Estructura y flujo	La información general debe preceder a la específica.	Clara y lógica / Poco clara o confusa / N/A	1 / 0 / -
AP36	Exploración de datos	La navegación debe permitir al usuario explorar los datos.	Exploración completa / Limitada / Sin exploración / N/A	1 / 0.5 / 0

Doctorado en Informática

F.4. Estructura jerárquica del modelo

Esta sección presenta las relaciones entre los distintos niveles del modelo. En primer lugar, en la Tabla F. 4 se detalla qué propiedades se asocian a cada criterio de evaluación. Luego, la Tabla F. 5 indica las aplicaciones vinculadas a cada propiedad.

Tabla F. 4: Relación entre criterios y propiedades de calidad

ID Criterio	Propiedades asociadas
C01	P01; P02; P03; P05; P06; P07; P11; P12; P13
C02	P03; P05; P11; P12; P13; P14; P15; P16
C03	P01; P04; P06; P08; P09; P10; P12
C04	P03; P06; P09; P11; P12; P13; P14;
C05	P02; P04; P07; P08; P11; P14; P15; P16

Tabla F. 5: Relación entre propiedades y aplicaciones

ID Propiedad	Aplicaciones asociadas
P01	AP01; AP02; AP03
P02	AP08; AP12
P03	AP10; AP11; AP12; AP13; AP14
P04	AP19; AP20; AP26; AP27; AP28
P05	AP04; AP11; AP19; AP20; AP23; AP24; AP25; AP26; AP27; AP28
P06	AP08; AP12; AP25
P07	AP06; AP07; AP13; AP14; AP18; AP20; AP22; AP26; AP27; AP28
P08	AP05; AP06; AP29; AP30
P09	AP02; AP06; AP07; AP08; AP13; AP15; AP21; AP32; AP33
P10	AP24; AP29
P11	AP04; AP09; AP15; AP16; AP26; AP27; AP28; AP32; AP33



Universidad Nacional del Nordeste

Universidad Nacional de Misiones

Universidad Tecnológica Nacional

Doctorado en Informática

ID Propiedad	Aplicaciones asociadas
P12	AP02; AP15; AP17; AP21; AP32; AP33; AP35
P13	AP08; AP13; AP15; AP17; AP21; AP31; AP32; AP33; AP35
P14	AP21; AP25; AP31; AP32; AP34; AP35
P15	AP08; AP09; AP10; A11
P16	AP34; A36

Apéndice G. Manual de usuario NaVi-Q

Este apéndice presenta una guía práctica para el uso de la herramienta NaVi-Q, orientada a usuarios que deseen evaluar la calidad de visualizaciones narrativas. La plataforma permite realizar evaluaciones manuales o automáticas, generar informes interactivos y recibir sugerencias de mejora basadas en criterios de calidad predefinidos.

G.1. Acceso a la plataforma

Para comenzar a utilizar NaVi-Q, es necesario contar con un usuario registrado.

1. Acceder a la página principal de la herramienta: <https://www.naviq.com.ar> (ver Fig. G. 1).

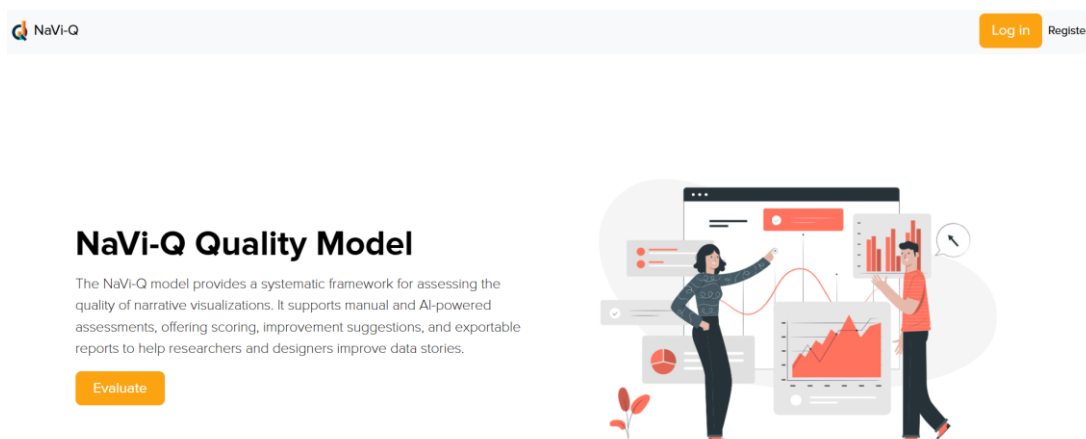


Fig. G. 1: Pantalla de inicio de la herramienta

2. Hacer clic en “Registrarse” y completar el formulario con nombre, correo electrónico y contraseña (Fig. G. 2).
3. Una vez registrado, iniciar sesión desde la opción “Ingresar”.

Doctorado en Informática

Register

Register

Already have an account? Log in.

Log In

☐ Remember me

Log in

Don't have an account? Register here

Fig. G. 2: Formularios de registro y acceso

G.2. Crear una nueva evaluación

Para iniciar una evaluación:

1. Acceder a la sección "Nueva evaluación" (ver Fig. G. 3) haciendo clic en cualquiera de los botones "Evaluar" de la página de inicio.
2. Cargar una imagen de la visualización a evaluar (formato PNG o JPG).
3. Seleccionar un perfil de calidad disponible.
4. Establecer un "puntaje mínimo viable" si se desea utilizarlo como referencia.
5. Elegir el tipo de evaluación:
 - **Manual:** el usuario responde una serie de preguntas guiadas.
 - **Automática:** el sistema evalúa la visualización de forma asistida por IA.

Una vez completados estos pasos, se habilita el botón para comenzar.

Doctorado en Informática**Start Evaluation**

To start your evaluation, enter the URL of the visualization and select the quality profile that best suits your needs. Additionally, set the Minimum Viable Score (MVS) to define the lowest acceptable quality threshold for your evaluation. Click 'Start' to proceed.

Upload Visualization Image
Seleccionar archivo Sin archivos seleccionados

Evaluation Title
New Evaluation

Quality Profile

Can't find the profile you need? [Create a new quality profile here.](#)

Minimum Viable Score
0,5

Define the minimum acceptable quality threshold for the evaluation.

Evaluation Mode
☒ Manual ☐ Automatic

You will answer each question manually based on your interpretation of the visualization. If you're unsure about any question, you can get AI help for individual questions.

Start

Fig. G. 3: Iniciar nueva evaluación

G.3. Evaluación manual

En este modo, el sistema presenta cada aplicación (pregunta) de forma secuencial (ver Fig. G. 4). Para cada una:

- Se muestra una descripción y contenido de ayuda.
- Si corresponde, se visualizan resultados automáticos (por ejemplo, chequeos de contraste o accesibilidad).
- El usuario debe seleccionar una o más respuestas según corresponda.
- Es posible solicitar una sugerencia generada por IA para mejorar ese aspecto de la visualización.

Doctorado en Informática

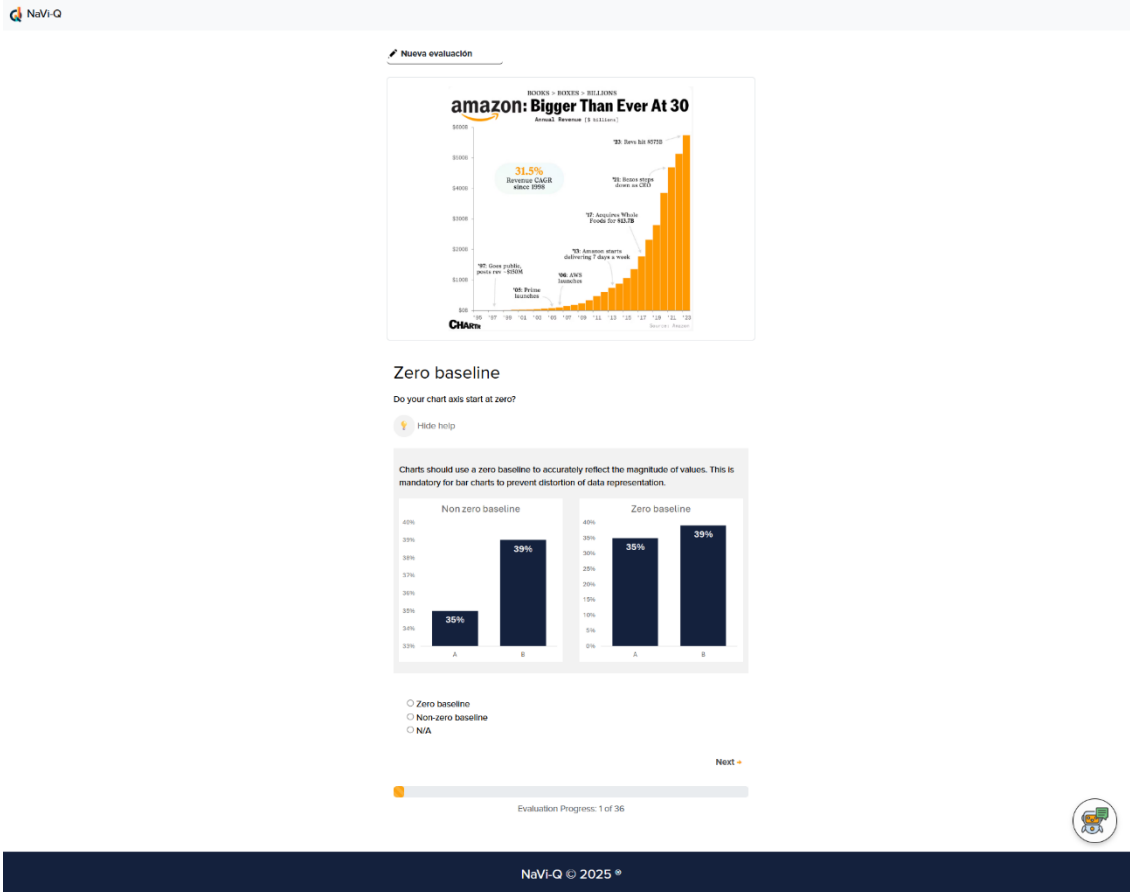


Fig. G. 4: Vista de pregunta individual en evaluación manual

G.4. Evaluación automática

Este modo permite obtener un diagnóstico sin intervención manual. El sistema aplica:

- Chequeos automáticos (por ejemplo, contraste, colores accesibles).
- Evaluación mediante IA: analiza la imagen y su descripción para responder cada pregunta del perfil.

G.5. Reporte de resultados

Una vez finalizada la evaluación, se accede a un reporte interactivo (ver Fig. G. 5). El mismo contiene:

- Puntaje total representado gráficamente.

Doctorado en Informática

- Desglose por criterios y propiedades.
- Tabla con todas las aplicaciones evaluadas y las respuestas seleccionadas.
- Recomendaciones de mejora basadas en el impacto potencial de cada aplicación.
- Sugerencias específicas para cada recomendación mediante IA.
- Botones para exportar los resultados en formatos estándar.

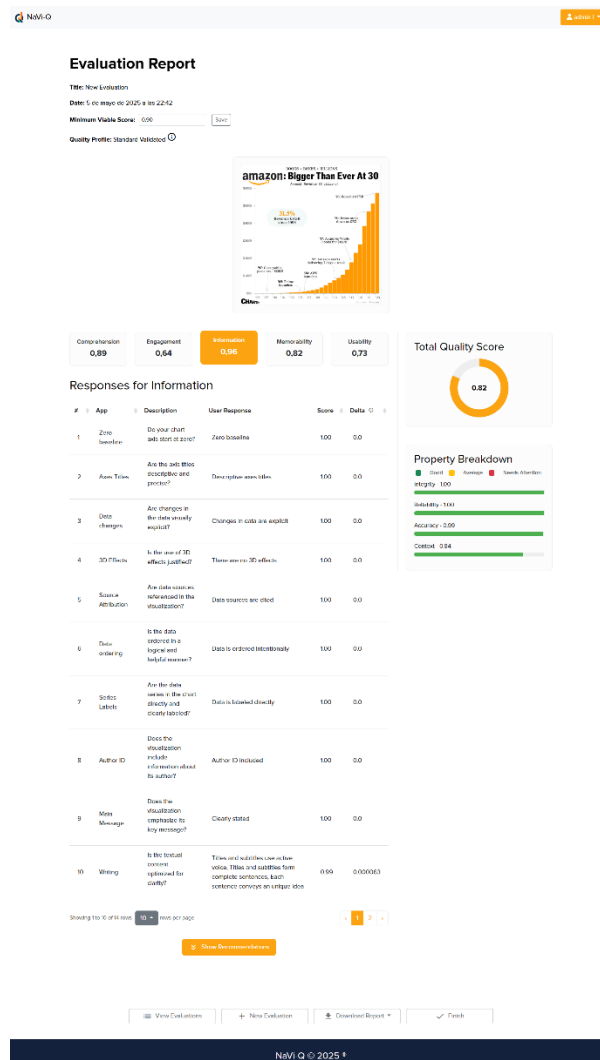


Fig. G. 5: Reporte de resultados

Doctorado en Informática

G.6. Generación de recomendaciones

Tras completar una evaluación, la herramienta identifica qué aplicaciones tienen mayor potencial de mejora, calculando para cada una un valor **delta**, que representa la diferencia entre el puntaje alcanzado y su contribución máxima esperada, considerando su peso dentro del perfil.

Las recomendaciones se ordenan según este delta, priorizando aquellas cuya mejora tendría mayor impacto sobre el puntaje total. Además de presentar estas recomendaciones de manera general, la herramienta permite, para cada una, generar una sugerencia específica de implementación mediante inteligencia artificial. Esta sugerencia es contextual y está orientada a orientar al equipo de diseño sobre cómo abordar concretamente la mejora sugerida.

Road to your Minimum Viable Score

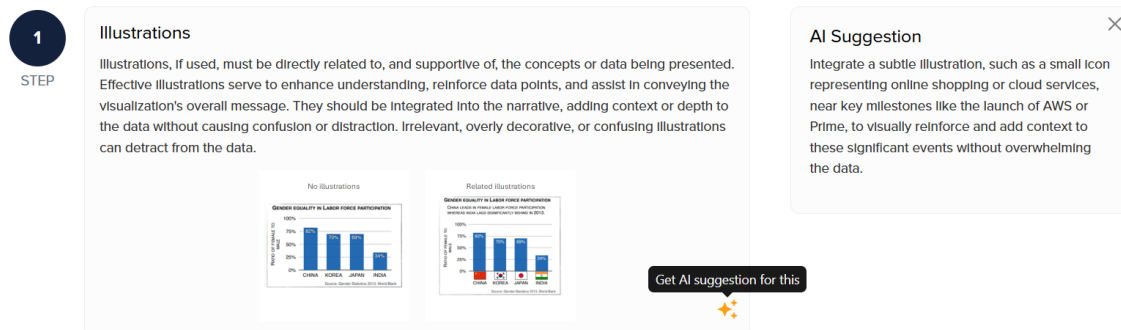


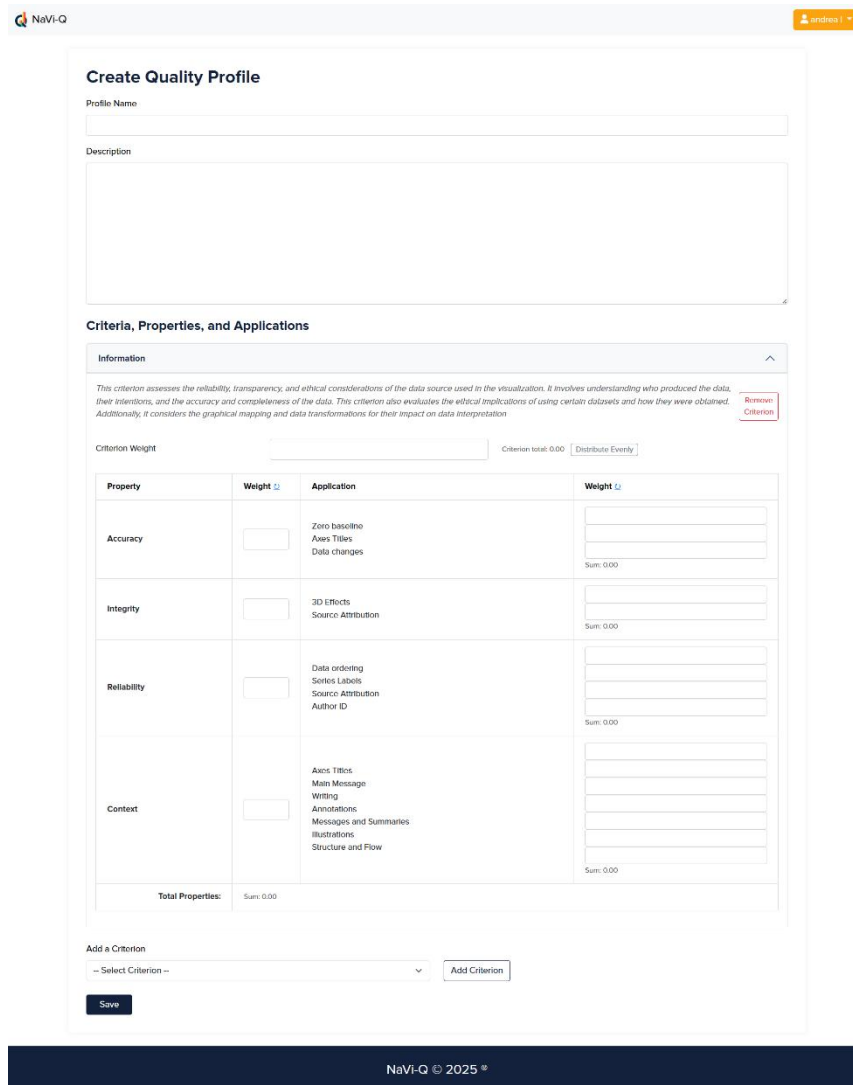
Fig. G. 6: Sugerencia de mejora general (modelo) y específica (mediante IA)

G.7. Gestión de perfiles

La herramienta permite configurar diferentes perfiles de calidad, que definen cómo se estructura una evaluación (ver Fig. G. 7). Cada perfil contiene:

- Criterios de calidad.
- Propiedades asociadas a cada criterio.
- Aplicaciones

Doctorado en Informática



NaVi-Q

Profile Name

Description

Criteria, Properties, and Applications

Information

This criterion assesses the reliability, transparency, and ethical considerations of the data source used in the visualization. It involves understanding who produced the data, their intentions, and the accuracy and completeness of the data. This criterion also evaluates the ethical implications of using certain datasets and how they were obtained. Additionally, it considers the graphical mapping and data transformations for their impact on data interpretation.

Criterion Weight

Criterion total: 0.00

Distribute Evenly

Property	Weight	Application	Weight
Accuracy		Zero baseline Axes Titles Data changes	Sum: 0.00
Integrity		3D Effects Source Attribution	Sum: 0.00
Reliability		Data ordering Series Labels Source Attribution Author ID	Sum: 0.00
Context		Axis Titles Main Message Writing Annotations Messages and Summaries Illustrations Structure and Flow	Sum: 0.00
Total Properties:	Sum: 0.00		

Add a Criterion

-- Select Criterion --

Add Criterion

Save

NaVi-Q © 2025

Fig. G. 7: Gestión de perfiles de calidad

Para crear un nuevo perfil, el usuario debe:

1. Seleccionar los criterios que formarán parte del perfil.
2. Una vez seleccionados, el sistema carga automáticamente las propiedades asociadas a cada criterio, según la estructura establecida por el modelo.
3. Ajustar los pesos de cada nivel jerárquico:
 - Peso de cada criterio respecto al perfil.
 - Peso de cada propiedad respecto al criterio.

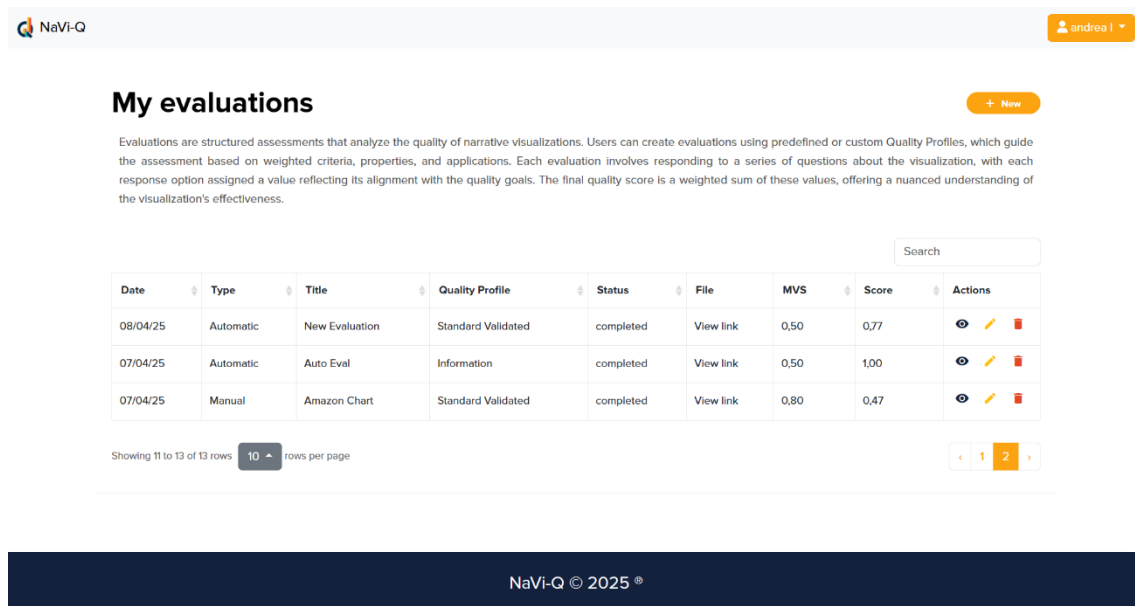
Doctorado en Informática

- Peso de cada aplicación respecto a la propiedad.

Importante: Las asociaciones entre criterios, propiedades y aplicaciones están definidas por el modelo base y no pueden modificarse al crear un perfil. El usuario solo puede ajustar los pesos relativos de cada elemento.

G.8. Listado de evaluaciones y de perfiles de calidad

Desde el menú principal, el usuario puede acceder a una vista general de las evaluaciones realizadas y de los perfiles de calidad disponibles (ver Fig. G. 8 y Fig. G. 9).



My evaluations + New

Evaluations are structured assessments that analyze the quality of narrative visualizations. Users can create evaluations using predefined or custom Quality Profiles, which guide the assessment based on weighted criteria, properties, and applications. Each evaluation involves responding to a series of questions about the visualization, with each response option assigned a value reflecting its alignment with the quality goals. The final quality score is a weighted sum of these values, offering a nuanced understanding of the visualization's effectiveness.

Date	Type	Title	Quality Profile	Status	File	MVS	Score	Actions
08/04/25	Automatic	New Evaluation	Standard Validated	completed	View link	0,50	0,77	  
07/04/25	Automatic	Auto Eval	Information	completed	View link	0,50	1,00	  
07/04/25	Manual	Amazon Chart	Standard Validated	completed	View link	0,80	0,47	  

Showing 11 to 13 of 13 rows 10 rows per page 1 2

NaVi-Q © 2025

Fig. G. 8: Listado de evaluaciones

Evaluaciones

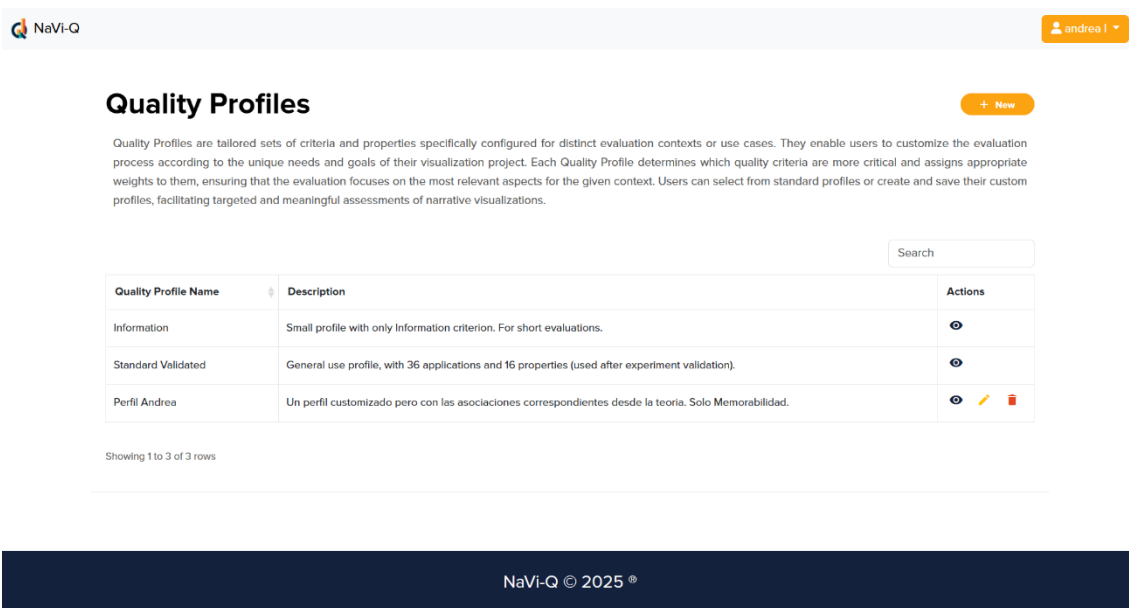
En la sección "Mis evaluaciones", se listan todas las evaluaciones realizadas por el usuario, con la siguiente información:

- Título asignado a la evaluación.
- Fecha de creación.
- Tipo de evaluación (manual o automática).

Doctorado en Informática

- Perfil de calidad utilizado.
- Puntaje total obtenido.
- Estado (completa, borrador, etc.).

Cada entrada del listado permite acceder al reporte correspondiente o continuar una evaluación en curso.



Quality Profiles + New

Quality Profiles are tailored sets of criteria and properties specifically configured for distinct evaluation contexts or use cases. They enable users to customize the evaluation process according to the unique needs and goals of their visualization project. Each Quality Profile determines which quality criteria are more critical and assigns appropriate weights to them, ensuring that the evaluation focuses on the most relevant aspects for the given context. Users can select from standard profiles or create and save their custom profiles, facilitating targeted and meaningful assessments of narrative visualizations.

Quality Profile Name	Description	Actions
Information	Small profile with only Information criterion. For short evaluations.	
Standard Validated	General use profile, with 36 applications and 16 properties (used after experiment validation).	
Perfil Andrea	Un perfil customizado pero con las asociaciones correspondientes desde la teoría. Solo Memorabilidad.	

Showing 1 to 3 of 3 rows

NaVi-Q © 2025

Fig. G. 9: Listado de perfiles de calidad

Perfiles de calidad

En la sección "Perfiles de calidad", se listan tanto los perfiles predeterminados como los perfiles personalizados creados por el usuario. Para cada perfil se muestra:

- Nombre del perfil.
- Descripción.
- Acciones disponibles: ver, editar (si es personalizado) o eliminar.

Referencias

- Ajani, K., Lee, E., Xiong, C., Knafllic, C. N., Kemper, W., & Franconeri, S. (2022). Declutter and Focus: Empirically Evaluating Design Guidelines for Effective Data Communication. *IEEE Transactions on Visualization and Computer Graphics*, 28(10), 3351–3364. <https://doi.org/10.1109/TVCG.2021.3068337>
- Aliyev, R., Temizkan, H., & Aliyev, R. (2020). Fuzzy Analytic Hierarchy Process-Based Multi-Criteria Decision Making for Universities Ranking. *Symmetry* 2020, Vol. 12, Page 1351, 12(8), 1351. <https://doi.org/10.3390/SYM12081351>
- Amini, F. (2020). *Data Videos: Turning Data into Engaging Narratives*. 173.
- Bach, B., Stefaner, M., Boy, J., Drucker, S., Bartram, L., Wood, J., Ciuccarelli, P., Engelhardt, Y., Köppen, U., & Tversky, B. (2018). Narrative Design Patterns for Data-Driven Storytelling. In *Data-Driven Storytelling* (pp. 107–133). CRC Press (Taylor & Francis). <https://doi.org/10.1201/9781315281575-5>
- Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly: Management Information Systems*, 35(2), 261–292. <https://doi.org/10.2307/23044044>
- Bai, X., White, D., & Sundaram, D. (2009). Visual intelligence density. *Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction - CHINZ '09*, 93–100. <https://doi.org/10.1145/1577782.1577799>
- Bai, X., White, D., & Sundaram, D. (2010). Purposeful Visualization System. *2010 Second World Congress on Software Engineering*, 1, 241–244. <https://doi.org/10.1109/WCSE.2010.100>



Doctorado en Informática

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *Intl. Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bansiya, J., & Davis, C. G. (2002). A Hierarchical Model for Object-Oriented Design Quality Assessment. *IEEE Transactions on Software Engineering*, 28(1), 4–17. <https://doi.org/10.1109/32.979986>
- Barcellos, R., Viterbo, J., Bernardini, F., & Trevisan, D. (2018). An Instrument for Evaluating the Quality of Data Visualizations. *2018 22nd International Conference Information Visualisation (IV)*, 169–174. <https://doi.org/10.1109/iV.2018.00038>
- Barrera-Leon, L., Corno, F., & De Russis, L. (2023). How the Preattentive Process is Exploited in Practical Information Visualization Design: A Review. *International Journal of Human-Computer Interaction*, 39(4), 707–720. <https://doi.org/10.1080/10447318.2022.2049137>
- Battle, L., Duan, P., Miranda, Z., Mukusheva, D., Chang, R., & Stonebraker, M. (2017). Beagle: Automated Extraction and Interpretation of Visualizations from the Web. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018-April*, 1–8. <https://doi.org/10.1145/3173574.3174168>
- Bera, P. (2016). How colors in business dashboards affect users' decision making. *Communications of the ACM*, 59(4), 50–57. <https://doi.org/10.1145/2818993>
- Bertin, J. (2010). *Semiology of Graphics: Diagrams, Networks, Maps* (1st ed.). Esri Press.
- Blei, D. M., & Lafferty, J. D. (2009). *Topic Models*.
- Boehm, et al. B. W. (1978). *Characteristics of Software Quality*. North-Holland Publishing Company. https://books.google.com/books/about/Characteristics_of_Software_Quality.html?hl=es&id=Cdm0AAAAIAAJ
- Bollen, K. A., & Ting, K. F. (2000). A tetrad test for causal indicators. *Psychological Methods*, 5(1), 3–22. <https://doi.org/10.1037/1082-989X.5.1.3>



Doctorado en Informática

- Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., Pfister, H., & Oliva, A. (2016). Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 519–528. <https://doi.org/10.1109/TVCG.2015.2467732>
- Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2306–2315. <https://doi.org/10.1109/TVCG.2013.234>
- Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3), 198–213. <https://doi.org/10.1177/1473871615594652>
- Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963–1972. <https://doi.org/10.1109/TVCG.2014.2346984>
- Brooke, J. (1996). SUS: A “Quick and Dirty” Usability Scale. In *Usability Evaluation In Industry* (1st ed., pp. 207–212). CRC Press. <https://doi.org/10.1201/9781498710411-35>
- Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology*, 29(3), 461–473. <https://doi.org/10.1080/14640747708400622>
- Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364–367. <https://doi.org/10.1080/01621459.1974.10482955>
- Card, S., Mackinlay, J., & Shneiderman, B. (1999). Readings in Information Visualisation. Using Vision to Think.: Using Vision to Think. In *Morgan Kaufmann* (Issue January). Morgan Kaufmann Publishers.



Doctorado en Informática

- Carpendale, S. (2008). Evaluating Information Visualizations. In *Information Visualization: Vol. 4950 LNCS* (Issue January 1970, pp. 19–45). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70956-5_2
- Carswell, C. M. (1992). Choosing Specifiers: An Evaluation of the Basic Tasks Model of Graphical Perception. *https://Doi.Org/10.1177/001872089203400503*, 34(5), 535–554. <https://doi.org/10.1177/001872089203400503>
- Cawthon, N., & Moere, A. Vande. (2007). The effect of aesthetic on the usability of data visualization. *Proceedings of the International Conference on Information Visualisation*, 637–645. <https://doi.org/10.1109/IV.2007.147>
- Cherubini, F., & Nielsen, R. K. (2016). Editorial Analytics: How News Media are Developing and Using Audience Data and Metrics. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2739328>
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554. <https://doi.org/10.1080/01621459.1984.10478080>
- Desbarats, N. P. (2023). *Practical charts: the essential guide to creating clear, compelling charts for reports and presentations*. Practical Reporting, Inc.
- Diakopoulos, N. (2018). Ethics in Data-Driven Visual Storytelling. In *Data-Driven Storytelling* (pp. 233–248). A K Peters/CRC Press. <https://doi.org/10.1201/9781315281575-10>
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218. <https://doi.org/10.1016/J.JBUSRES.2008.01.009>
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development.

Doctorado en Informática

<https://doi.org/10.1509/Jmkr.38.2.269.18845>, 38(2), 269–277.
<https://doi.org/10.1509/JMKR.38.2.269.18845>

Dibia, V. (2023). *LIDA: A Tool for Automatic Generation of Grammar-Agnostic Visualizations and Infographics using Large Language Models*. iv.

Dimara, E., Bezerianos, A., & Dragicevic, P. (2017a). Narratives in crowdsourced evaluation of visualizations: A double-edged sword? *Conference on Human Factors in Computing Systems - Proceedings, 2017-May*, 5475–5484.
<https://doi.org/10.1145/3025453.3025870>

Dimara, E., Bezerianos, A., & Dragicevic, P. (2017b). The Attraction Effect in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 471–480. <https://doi.org/10.1109/TVCG.2016.2598594>

Dimara, E., & Stasko, J. (2022). A Critical Reflection on Visualization Research: Where Do Decision Making Tasks Hide? *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 1128–1138.
<https://doi.org/10.1109/TVCG.2021.3114813>

Dowding, D., & Merrill, J. A. (2018). The Development of Heuristics for Evaluation of Dashboard Visualizations. *Applied Clinical Informatics*, 9(3), 511–518.
<https://doi.org/10.1055/s-0038-1666842>

Dykes, B. (2019). *Effective Data Storytelling: How to Drive Change with Data, Narrative and Visuals*. Wiley.

Edmond, C., & Bednarz, T. (2021). Three trajectories for narrative visualisation. *Visual Informatics*, 5(2), 26–40. <https://doi.org/10.1016/j.visinf.2021.04.001>

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174.
<https://doi.org/10.1037/1082-989X.5.2.155>

Eisner, W. (2008). *Graphic Storytelling & Visual Narrative* (W. W. Norton, Ed.).



Doctorado en Informática

- Elmqvist, N., & Yi, J. S. (2015). Patterns for visualization evaluation. *Information Visualization*, 14(3), 250–269. <https://doi.org/10.1177/1473871613513228>
- Errey, N., Liang, J., Leong, T. W., & Zowghi, D. (2024). Evaluating narrative visualization: a survey of practitioners. *International Journal of Data Science and Analytics*, 18(1), 19–34. <https://doi.org/10.1007/s41060-023-00394-9>
- Errey, N., Zowghi, D., Leong, T. W., Wu, Y., Yuan, X., Huang, W., & Liang, C. J. (2025). Heuristics for evaluating narrative visualization: a validation study. *Journal of Visualization*. <https://doi.org/10.1007/s12650-025-01051-y>
- Evergreen, S. D. H. (2012). Death by boredom: The role of visual processing theory in written evaluation communication. [Western Michigan University]. In *Dissertation Abstracts International Section A: Humanities and Social Sciences*. <https://scholarworks.wmich.edu/dissertations/403>
- Fansher, M., Walls, L., Hao, C., Subramonyam, H., Boduroglu, A., Shah, P., & Witt, J. K. (2025). Narrative visualizations: Depicting accumulating risks and increasing trust in data. *Cognitive Research: Principles and Implications*, 10(1), 1–27. <https://doi.org/10.1186/S41235-025-00613-W/FIGURES/15>
- Fayers, P. M., & Hand, D. J. (2002). Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2), 233–253. <https://doi.org/10.1111/1467-985X.02020>
- Feng, M., Peck, E., & Harrison, L. (2019). Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 501–511. <https://doi.org/10.1109/TVCG.2018.2865117>
- Field, A. P. (2005). Kendall's Coefficient of Concordance. *Encyclopedia of Statistics in Behavioral Science*. <https://doi.org/10.1002/0470013192.BSA327>



Doctorado en Informática

- Figueiras, A. (2014). How to tell stories using visualization. *Proceedings of the International Conference on Information Visualisation*, 18–26. <https://doi.org/10.1109/IV.2014.78>
- Forsell, C., & Johansson, J. (2010). An heuristic set for evaluation in information visualization. *Proceedings of the International Conference on Advanced Visual Interfaces - AVI '10*, 199. <https://doi.org/10.1145/1842993.1843029>
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, 32(2), 124. <https://doi.org/10.2307/749671>
- Fu, X., Wang, Y., Dong, H., Cui, W., & Zhang, H. (2019). Visualization Assessment: A Machine Learning Approach. *2019 IEEE Visualization Conference, VIS 2019*, 126–130. <https://doi.org/10.1109/VISUAL.2019.8933570>
- Garreton, M., Morini, F., Celhay, P., Dork, M., & Parra, D. (2023). Attitudinal effects of data visualizations and illustrations in data stories. *IEEE Transactions on Visualization and Computer Graphics*, 30(7), 4039–4054. <https://doi.org/10.1109/TVCG.2023.3248319>
- Gottschall, J. (2012). *The Storytelling Animal: How Stories Make Us Human* (Houghton Mifflin Harcourt, Ed.; 1st ed.). Mariner Books.
- Haase, H. (1998). Evaluating the quality of scientific visualizations: the Q-VIS reference model. In R. F. Erbacher & A. Pang (Eds.), *Visual Data Exploration and Analysis V* (Vol. 3298, Issue January, pp. 123–131). <https://doi.org/10.1117/12.309534>
- Hanusz, Z., Tarasinska, J., & Zielinski, W. (2016). Shapiro–Wilk Test with Known Mean. *REVSTAT-Statistical Journal*, 14(1), 89–100–189–100. <https://doi.org/10.57805/REVSTAT.V14I1.180>
- Haroz, S., & Whitney, D. (2012). How Capacity Limits of Attention Influence Information Visualization Effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2402–2410. <https://doi.org/10.1109/TVCG.2012.233>

Doctorado en Informática

- He, Y., Cao, S., Shi, Y., Chen, Q., Xu, K., & Cao, N. (2024). *Leveraging Large Models for Crafting Narrative Visualization: A Survey*. 1–20. <http://arxiv.org/abs/2401.14010>
- Healey, C. G., Booth, K. S., & Enns, J. T. (1993). Harnessing preattentive processes for multivariate data visualization. *Proceedings - Graphics Interface*, 107–117.
- Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using mechanical Turk to assess visualization design. *Conference on Human Factors in Computing Systems - Proceedings, 1*, 203–212. <https://doi.org/10.1145/1753326.1753357>
- Helmy, S. E., Eladl, G. H., & Eisa, M. (2021). Fuzzy Analytical Hierarchy Process (FAHP) Using Geometric Mean Method to Select Best Processing Framework Adequate to Big Data. *Journal of Theoretical and Applied Information Technology*, 15(1). www.jatit.org
- Henry Riche, N., Hurter, C., Diakopoulos, N., & Carpendale, S. (2018). *Data-Driven Storytelling* (1st ed.). A K Peters/CRC Press.
- Heradio, R., Fernandez-Amoros, D., Galindo, J. A., Benavides, D., & Batory, D. (2022). Uniform and scalable sampling of highly configurable systems. *Empirical Software Engineering*, 27(2), 1–34. <https://doi.org/10.1007/S10664-021-10102-5/TABLES/7>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2231–2240. <https://doi.org/10.1109/TVCG.2011.255>
- Hullman, J., Drucker, S., Riche, N. H., Bongshin Lee, Fisher, D., & Adar, E. (2013). A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2406–2415. <https://doi.org/10.1109/TVCG.2013.119>



Doctorado en Informática

- Hung, Y. H., & Parsons, P. (2017). Assessing user engagement in information visualization. *Conference on Human Factors in Computing Systems - Proceedings, Part F1276*, 1708–1717. <https://doi.org/10.1145/3027063.3053113>
- Hwang, C.-L., & Yoon, K. (1981). *Multiple Attribute Decision Making Methods and Applications A State-of-the-Art Survey* (1st ed., Vol. 186). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-48318-9>
- ISO 9241-11. (2018). *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- ISO/IEC 25010. (2023). *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model*. <https://www.iso.org/standard/78176.html>
- ISO/IEC 33003. (2015). *Information technology — Process assessment — Requirements for process measurement frameworks*. <https://www.iso.org/standard/54177.html>
- Johannesson, P., & Perjons, E. (2014). An introduction to design science. In *An Introduction to Design Science* (Vol. 9783319106). <https://doi.org/10.1007/978-3-319-10632-8>
- Johansson, J., & Forsell, C. (2016). Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 579–588. <https://doi.org/10.1109/TVCG.2015.2466992>
- Jung, H. W. (2013). Investigating measurement scales and aggregation methods in SPICE assessment method. *Information and Software Technology*, 55(8), 1450–1461. <https://doi.org/10.1016/j.infsof.2013.02.004>
- Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and*

Doctorado en Informática

- Lecture Notes in Bioinformatics*), 4950 LNCS(May 2014), 154–175.
https://doi.org/10.1007/978-3-540-70956-5_7
- Kitchenham, B., & Charters, S. M. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/J.JCM.2016.02.012>
- Kosara, R., & MacKinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, 46(5), 44–50. <https://doi.org/10.1109/MC.2013.36>
- Kosslyn, S. (2006). *Graph Design for the Eye and Mind*. Oxford University Press.
- Lahrache, S., El Ouazzani, R., & El Qadi, A. (2018). Visualizations memorability through visual attention and image features. *Procedia Computer Science*, 127, 328–335. <https://doi.org/10.1016/j.procs.2018.01.129>
- Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9), 1520–1536. <https://doi.org/10.1109/TVCG.2011.279>
- Lan, X., Shi, Y., Zhang, Y., & Cao, N. (2021). Smile or Scowl? Looking at Infographic Design Through the Affective Lens. *IEEE Transactions on Visualization and Computer Graphics*, 27(6), 2796–2807. <https://doi.org/10.1109/TVCG.2021.3074582>
- Lan, X., Wu, Y., & Cao, N. (2024). Affective Visualization Design: Leveraging the Emotional Impact of Data. *IEEE Transactions on Visualization and Computer Graphics*, 30(1), 1–11. <https://doi.org/10.1109/TVCG.2023.3327385>
- Leal, J. E. (2020). AHP-express: A simplified version of the analytical hierarchy process method. *MethodsX*, 7, 100748. <https://doi.org/10.1016/j.mex.2019.11.021>



Doctorado en Informática

- Lee, B., Riche, N. H., Isenberg, P., & Carpendale, S. (2015). More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications*, 35(5), 84–90. <https://doi.org/10.1109/MCG.2015.99>
- Lee, S., Kim, S.-H., & Kwon, B. C. (2017). VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- Lei, T., Ni, N., Zhu, Q., & Zhang, S. (2018). Aesthetic experimental study on information visualization design under the background of big data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10919 LNCS, 218–226. https://doi.org/10.1007/978-3-319-91803-7_16
- Lezcano Airaldi, A., Diaz-Pace, J. A., & Irrazábal, E. (2021). Data-driven Storytelling to Support Decision Making in Crisis Settings: A Case Study. *JUCS - Journal of Universal Computer Science*, 27(10), 1046–1068. <https://doi.org/10.3897/jucs.66714>
- Lezcano Airaldi, A., Irrazábal, E., & Diaz Pace, A. (2025). Best Practices and Evaluation Methods for Narrative Information Visualizations: A Systematic Review. *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering*, 378–389. <https://doi.org/10.5220/0013202000003928>
- Lezcano Airaldi, A., Irrazábal, E., & Diaz-Pace, J. A. (2025). Best Practices and Evaluation Methods for Narrative Information Visualizations : A Systematic Review Evaluation in Information. In M. Mannion, T. Mannisto, & L. Maciaszek (Eds.), *Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering* (pp. 378–389). SCITEPRESS - Science and Technology Publications. <https://doi.org/10.5220/0000189300003928>
- Lichtenthäler, R., & Wirtz, G. (2022). *A Review of Approaches for Quality Model Validations in the Context of Cloud-native Applications*. February, 24–25.



Doctorado en Informática

- Liu, F. H. F., & Hai, H. L. (2005). The voting analytic hierarchy process method for selecting supplier. *International Journal of Production Economics*, 97(3), 308–317. <https://doi.org/10.1016/J.IJPE.2004.09.005>
- Lo, L. Y.-H., & Qu, H. (2024). *How Good (Or Bad) Are LLMs at Detecting Misleading Visualizations?* <http://arxiv.org/abs/2407.17291>
- Matzen, L. E., Haass, M. J., Divis, K. M., Wang, Z., & Wilson, A. T. (2018). Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 563–573. <https://doi.org/10.1109/TVCG.2017.2743939>
- McKenna, S., Henry Riche, N., Lee, B., Boy, J., & Meyer, M. (2017). Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum*, 36(3), 377–387. <https://doi.org/10.1111/CGF.13195>
- McKnight, P. E., & Najab, J. (2010). Kruskal-Wallis Test. *The Corsini Encyclopedia of Psychology*, 1–1. <https://doi.org/10.1002/9780470479216.CORPSY0491>
- McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *The Corsini Encyclopedia of Psychology*, 1–1. <https://doi.org/10.1002/9780470479216.CORPSY0524>
- Morisio, M., Stamelos, I., & Tsoukias, A. (2002). A new method to evaluate software artifacts against predefined profiles. *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering*, 27, 811–818. <https://doi.org/10.1145/568760.568899>
- Munzner, T. (2014). *Visualization Analysis and Design* (1st ed.). A K Peters/CRC Press.
- Nelson, F. R. T., & Blauvelt, C. T. (2007). Fisher Exact Test. In *A Manual of Orthopaedic Terminology* (pp. 353–369). Elsevier. <https://doi.org/10.1016/B978-0-323-04503-2.10013-1>
- Nielsen, J. (1994). *Usability Inspection Methods*. Wiley. <https://www.nngroup.com/books/usability-inspection-methods/>



Doctorado en Informática

- Norman, D. A. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things* (1st ed.).
- Nowell, L., Schulman, R., & Hix, D. (2002). Graphical encoding for information visualization: an empirical study. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, 2002-Janua, 43–50.
<https://doi.org/10.1109/INFVIS.2002.1173146>
- Padda, H., Mudur, S., Seffah, A., & Joshi, Y. (2008). Comprehension of Visualization Systems - Towards Quantitative Assessment. *First International Conference on Advances in Computer-Human Interaction*, 83–88.
<https://doi.org/10.1109/ACHI.2008.19>
- Padda, H., Seffah, A., & Mudur, S. (2007). Visualization Patterns: A Context-Sensitive Tool to Evaluate Visualization Techniques. *2007 4th IEEE International Workshop on Visualizing Software for Understanding and Analysis*, 88–91.
<https://doi.org/10.1109/VISSOF.2007.4290705>
- Patton, M. Q. (2014). Qualitative research and evaluation methods: Integrating Theory and Practice. In *SAGE Publications, Inc.* (4th ed.). Sage publications.
- Perin, C., Wun, T., Pusch, R., & Carpendale, S. (2018). Assessing the Graphical Perception of Time and Speed on 2D+Time Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 698–708.
<https://doi.org/10.1109/TVCG.2017.2743918>
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. <https://doi.org/10.1016/J.INFSOF.2015.03.007>
- Plaisant, C. (2004). The challenge of information visualization evaluation. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, 109–116.
<https://doi.org/10.1145/989863.989880>



Doctorado en Informática

- Post, F. H., Nielson, G. M., & Bonneau, G.-P. (2003). *Data Visualization: The State of the Art*. Springer US. <https://doi.org/10.1007/978-1-4615-1177-9>
- Quispel, A., Maes, A., & Schilperoord, J. (2016). Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use. *Information Visualization*, 15(3), 238–252. <https://doi.org/10.1177/1473871615606478>
- Rezaie, M., Tory, M., & Carpendale, S. (2024). Struggles and Strategies in Understanding Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 30(6), 3035–3048. <https://doi.org/10.1109/TVCG.2024.3388560>
- Robson, C. (2017). *Small-Scale Evaluation: Principles and Practice*. Sage Publications.
- Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2), 131–164. <https://doi.org/10.1007/s10664-008-9102-8>
- Saaty, T. L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill.
- Sakarang, S. F., Asli, M. F., & Hamzah, M. (2025). Narrative visualization: Taxonomy, methodologies, challenges, and opportunities. *Information Visualization*, 24(2), 199–223. <https://doi.org/10.1177/14738716251315916>
- Saket, B., Srinivasan, A., Ragan, E. D., & Endert, A. (2018). Evaluating Interactive Graphical Encodings for Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(3), 1316–1330. <https://doi.org/10.1109/TVCG.2017.2680452>
- Santos, B. S., Ferreira, B. Q., & Dias, P. (2015). Heuristic Evaluation in Information Visualization Using Three Sets of Heuristics: An Exploratory Study. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9169, pp. 259–270). https://doi.org/10.1007/978-3-319-20901-2_24

Doctorado en Informática

- Schloss, K. B., Gramazio, C. C., Silverman, A. T., Parker, M. L., & Wang, A. S. (2019). Mapping Color to Meaning in Colormap Data Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 810–819. <https://doi.org/10.1109/TVCG.2018.2865147>
- Sedgwick, P. (2014). Spearman's rank correlation coefficient. *BMJ*, 349. <https://doi.org/10.1136/BMJ.G7327>
- Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. <https://doi.org/10.1109/TVCG.2010.179>
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47–69. <https://doi.org/10.1023/A:1013180410169>
- Shao, H., Martinez-Maldonado, R., Echeverria, V., Yan, L., & Gašević, D. (2024). Data Storytelling in Data Visualisation: Does it Enhance the Efficiency and Effectiveness of Information Retrieval and Insights Comprehension? *Conference on Human Factors in Computing Systems - Proceedings*, 21. https://doi.org/10.1145/3613904.3643022/SUPPL_FILE/3613904.3643022-TALK-VIDEO.VTT
- Shin, S., Hong, S., & Elmqvist, N. (2024). *Visualizationary: Automating Design Feedback for Visualization Designers using LLMs*. 1–13. <http://arxiv.org/abs/2409.13109>
- Siebert, J., Joeckel, L., Heidrich, J., Trendowicz, A., Nakamichi, K., Ohashi, K., Namba, I., Yamamoto, R., & Aoyama, M. (2022). Construction of a quality model for machine learning systems. *Software Quality Journal*, 30(2), 307–335. <https://doi.org/10.1007/s11219-021-09557-y>
- Sless, D. (1981). *Learning and Visual Communication* (Croom Helm, Ed.; 1st ed.). Taylor & Francis Group. <https://doi.org/10.4324/9780429021909>



Doctorado en Informática

- Stasko, J. (2014). Value-driven evaluation of visualizations. *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization, 10-Novembe*, 46–53. <https://doi.org/10.1145/2669557.2669579>
- Stevens, S. S. (1951). *Mathematics, Measurement, and Psychophysics*. John Wiley & Sons, Inc.
- Tavana, M., Soltanifar, M., & Santos-Arteaga, F. J. (2021). Analytical hierarchy process: revolution and evolution. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04432-2>
- Tory, M., & Moller, T. (2005). Evaluating visualizations: do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5), 8–11. <https://doi.org/10.1109/MCG.2005.102>
- Trendowicz, A., & Kopczyńska, S. (2014). Adapting Multi-Criteria Decision Analysis for Assessing the Quality of Software Products. Current Approaches and Future Perspectives. In *Advances in Computers* (Vol. 93, pp. 153–226). Elsevier. <https://doi.org/10.1016/B978-0-12-800162-2.00004-X>
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press. <https://www.jstor.org/stable/530384?origin=crossref>
- Viégas, F. B., & Wattenberg, M. (2006). Technical forum: Communication-minded visualization: A call to action. *IBM Systems Journal*, 45(4), 801–812. <https://doi.org/10.1147/sj.454.0801>
- Wagner, S., Goeb, A., Heinemann, L., Kläs, M., Lampasona, C., Lochmann, K., Mayr, A., Plösch, R., Seidl, A., Streit, J., & Trendowicz, A. (2015). Operationalised product quality models and assessment: The Quamoco approach. *Information and Software Technology*, 62(1), 101–123. <https://doi.org/10.1016/j.infsof.2015.02.009>
- Wagner, S., Lochmann, K., Winter, S., Goeb, A., & Klaes, M. (2009). Quality models in practice: A preliminary analysis. *2009 3rd International Symposium on Empirical*

Doctorado en Informática

- Software Engineering and Measurement, ESEM 2009*, 464–467.
<https://doi.org/10.1109/ESEM.2009.5316003>
- Wainer, H. (1992). Understanding Graphs and Tables. *Http://Dx.Doi.Org/10.3102/0013189X021001014*, 21(1), 14–23.
<https://doi.org/10.3102/0013189X021001014>
- Waldner, M., Diehl, A., Gracanin, D., Splechtna, R., Delrieux, C., & Matkovic, K. (2019). A Comparison of Radial and Linear Charts for Visualizing Daily Patterns. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 1–1.
<https://doi.org/10.1109/TVCG.2019.2934784>
- Wall, E., Agnihotri, M., Matzen, L., Divis, K., Haass, M., Endert, A., & Stasko, J. (2019). A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 491–500.
<https://doi.org/10.1109/TVCG.2018.2865146>
- Wang, Y., Han, F., Zhu, L., Deussen, O., & Chen, B. (2018). Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series. *IEEE Transactions on Visualization and Computer Graphics*, 24(2), 1141–1154.
<https://doi.org/10.1109/TVCG.2017.2653106>
- Wang, Y., Zhao, Y., & Kim, Y. S. (2024). How Do Low-Vision Individuals Experience Information Visualization? *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3613904.3642188>
- Ware, C. (2020). *Information Visualization: Perception for Design (Interactive Technologies)* (4th ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-812875-6.01001-X>
- Wieringa, R. J. (2014). Design science methodology: For information systems and software engineering. In *Design Science Methodology: For Information Systems and Software Engineering*. <https://doi.org/10.1007/978-3-662-43839-8>

Doctorado en Informática

- Willett, W., Heer, J., Hellerstein, J. M., & Agrawala, M. (2011). CommentSpace: Structured support for collaborative visual analysis. *Conference on Human Factors in Computing Systems - Proceedings*, 3131–3140. <https://doi.org/10.1145/1978942.1979407>
- Williams, R., Scholtz, J., Blaha, L. M., Franklin, L., & Huang, Z. (2018). Evaluation of Visualization Heuristics. In M. Kurosu (Ed.), *Human-Computer Interaction. Theories, Methods, and Human Issues* (Vol. 10901, Issue January, pp. 208–224). Springer International Publishing. https://doi.org/10.1007/978-3-319-91238-7_18
- Wohlin, C., & Rainer, A. (2022). Is it a case study?—A critical analysis and guidance. *Journal of Systems and Software*, 192, 111395. <https://doi.org/10.1016/j.jss.2022.111395>
- Woolson, R. F. (2005). Wilcoxon Signed-Rank Test. *Encyclopedia of Biostatistics*. <https://doi.org/10.1002/0470011815.B2A15177>
- Wu, A., Wang, Y., Shu, X., Moritz, D., Cui, W., & Zhang, H. (2018). *AI4VIS : Survey on Artificial Intelligence Approaches for Data Visualization. XX(Xx)*.
- Yi, J. S. (2010, April). Implications of Individual Differences on Evaluating Information Visualization Techniques. *Proceedings of beyond Time and Errors: Novel Evaluation Methods for Visualization*.
- Yi, J. S., Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). *Toward a Deeper Understanding of the Role of Interaction in Information Visualization. October*.
- Yin, R. K. (2017). Case Study Research Design and Applications. In *The Canadian Journal of Program Evaluation*. Sage Publications Ltd.
- Yoon, K., & Hwang, C.-L. (2011). Multiple Attribute Decision Making: An Introduction. In *Multiple Attribute Decision Making* (Vol. 10). SAGE Publications, Inc. <https://doi.org/10.4135/9781412985161>
- Zdanovic, D., Lembcke, T. J., & Bogers, T. (2022). The influence of data storytelling on the ability to recall. *CHIIR 2022 - Proceedings of the 2022 Conference on Human*



Doctorado en Informática

Information Interaction and Retrieval, 67–77.
<https://doi.org/10.1145/3498366.3505755>

Zeleny, M. (1982). *Multiple Criteria Decision Making*. McGraw-Hill.

Zhao, J., Xu, S., Chandrasegaran, S., Bryan, C. J., Du, F., Mishra, A., Qian, X., Li, Y., & Ma, K. L. (2021). ChartStory: Automated Partitioning, Layout, and Captioning of Charts into Comic-Style Narratives. *IEEE Transactions on Visualization and Computer Graphics*, 1–15. <https://doi.org/10.1109/TVCG.2021.3114211>

Zhao, Z., & Elmqvist, N. (2023). The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization. *IEEE Computer Graphics and Applications*, 43(04), 97–110. <https://doi.org/10.1109/MCG.2023.3269850>

Zhu, Y. (2007). Measuring Effective Data Visualization. In *Advances in Visual Computing* (pp. 652–661). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76856-2_64

Zhu, Y., Suo, X., & Owen, G. S. (2007). Complexity Analysis for Information Visualization Design and Evaluation. In *Advances in Visual Computing: Vol. 4841 LNCS* (Issue PART 1, pp. 576–585). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76858-6_56