TU WIEN Informatics

# Visuelle Narrative gegen Irreführende Visualisierungen im Gesundheitswesen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieurin**

im Rahmen des Studiums

**Biomedical Engineering**

eingereicht von

**Anna Shilo, BSc**
Matrikelnummer 12045102

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Assistant Prof. Dr. Renata Georgia Raidou

Wien, 4. September 2023

Anna Shilo

Renata Georgia Raidou

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU Wien Informatics

# Visual narratives against misleading visualizations in health care

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

**Diplom-Ingenieurin**

in

**Biomedical Engineering**

by

**Anna Shilo, BSc**
Registration Number 12045102

to the Faculty of Informatics

at the TU Wien

Advisor: Assistant Prof. Dr. Renata Georgia Raidou

Vienna, 4th September, 2023

Anna Shilo | Renata Georgia Raidou

# Erklärung zur Verfassung der Arbeit

Anna Shilo, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 4. September 2023

Anna Shilo

# Danksagung

Ich bin meiner Betreuerin, Assistenzprofessorin Dr. Renata Georgia Raidou, die über ein Füllhorn an Ideen zu verfügen scheint, sehr dankbar. Ihre Betreuung war während der gesamten Arbeit von Ermutigung, ständiger Unterstützung und wertvollem Feedback geprägt. Ihr umfangreiches Wissen, ihre Führung durch das Dickicht des theoretischen Hintergrunds der Arbeit, ihre Geduld und ihre positive Einstellung habe ich stets sehr zu schätzen gewusst. Außerdem möchte ich mich bei Dipl.-Ing. Elisabeth Salomon, Ph.D. vom Zentrum für Medizinische Physik und Biomedizinische Technik der Medizinischen Universität Wien, Abteilung Nuklearmedizin, für die Durchführung des CT-Scans eines Supermarktfisches herzlich bedanken.

Ariadna Cherit Hernández, B.Sc., spielte eine entscheidende Rolle dabei, mich zu diesem Forschungsthema zu ermutigen, indem sie ihre wertvollen Erfahrungen aus der Arbeit in einem Krankenhaus und als Medizinstudentin weitergab. Ich schätzte ebenfalls sehr die Hilfe von Dipl.-Ing. Hanna Helene Steiner bei den Übersetzungen ins Deutsche und ihre hilfreichen Ratschläge, wie man am besten damit umgeht, wenn manch eine Situation zu überwältigend erscheint. Ein besonderer Dank geht an Simon Mariacher, der die Musik für das Spiel komponiert und bei der Erstellung von Animationen für letzteres geholfen hat. Darüber hinaus bin ich Felix Mariacher, B.Sc., sehr dankbar, welcher immer tiefstes Vertrauen in meine Arbeit hatte und mich ständig emotional sowie mental unterstützt hat. Ebenfalls danken möchte ich den freiwillen Testpersonen, welche es ermöglicht haben das Spiel zu testen und zu verifizieren. Schließlich möchte ich mich bei meiner Familie für ihre Unterstützung trotz der großen Entfernung bedanken.

# Acknowledgements

I'm extremely grateful to my supervisor Assistant Prof. Dr. Renata Georgia Raidou, who seems to possess a horn of plenty filled with ideas. Her supervision provided me with encouragement, constant support, and valuable feedback throughout the thesis. Her extensive knowledge, guidance through the thickets of the thesis theoretical background, patience, and positive attitude have been very much appreciated. Also, I would like to extend my sincere thanks to Dipl.-Ing. Elisabeth Salomon, Ph.D. from the Center for Medical Physics and Biomedical Engineering of the Medical University of Vienna, Nuclear Medicine department, for conducting the CT scan of a supermarket fish.

Ariadna Cherit Hernández, B.Sc., played a decisive role in encouraging me to pick this research topic by sharing her valuable experience from working in a hospital and being a medical student. I very much appreciate Dipl.-Ing. Hanna Helene Steiner's help with the translations to German and helpful advice on how to proceed when the situation seems too overwhelming. Special thanks to Simon Mariacher, who wrote the music and helped with animations for the game. Furthermore, I am extremely grateful to Felix Mariacher, B.Sc., who had a profound belief in my work and provided constant emotional support. I would like to thank the volunteer user study participants who made it possible to test and verify the game. Finally, I wish to thank my family for their support despite the long distance.

# Kurzfassung

Diese Diplomarbeit bietet einen Lösungsansatz gegen irreführende Visualisierungen im Gesundheitsbereich, die unzutreffende Erkenntnisse daraus vermitteln. Irreführende Aspekte solcher Visualisierungen ergeben sich aus Unsicherheiten, die in den einzelnen Schritten der medizinischen Visualisierungspipeline auftreten. Wir untersuchen die Gebiete des Storytellings und der Gamification, um das breite Publikum dabei zu unterstützen, irreführende Visualisierungen im Gesundheitswesen zu erkennen und zu bekämpfen. Unsere Forschungsfragen lauten: "*Welche Arten von Unsicherheitsfaktoren treten in der medizinischen Visualisierungspipeline auf und verbirgt sich dahinter eine gewisse Absicht?*" und "*Wie können wir die allgemeine Bevölkerung über die Existenz von Unsicherheitsfaktoren bei der Visualisierung aufklären?*"

Zur Beantwortung der Forschungsfragen haben wir eine Klassifizierung der Typen von Unsicherheitsfaktoren in der medizinischen Visualisierungspipeline entwickelt und das Lernspiel "DeteCATive" entworfen und umgesetzt, um der breiten Öffentlichkeit diese Konzepte auf ansprechende Weise zu vermitteln. Das Spiel umfasst acht Aufgaben, die amüsante fiktive Geschichten mit absichtlich irreführenden Visualisierungen von medizinischen Daten enthalten. Jede Geschichte beinhaltet eine eigene Auswahl an Annahmen. Der Spieler muss anhand der Geschichte bestimmen, ob eine Annahme richtig oder falsch ist, um Punkte und Belohnungen zu erhalten. Diese Punkte lassen sich am Ende des Spiels verwenden, um das Spielziel zu erreichen.

Um den pädagogischen Wert des Spiels zu bewerten, führten wir eine Nutzerstudie mit 21 Testpersonen durch. Diese Studie liefert uns wesentliche Einblicke sowie Erkenntnisse. Bestimmte irreführende Tricks bei der Visualisierung konnten von den Testpersonen nur schwer erkannt werden. Das Spiel wurde von den Testpersonen in Bezug auf Einprägsamkeit, Motivation und Interaktion positiv bewertet. Falsch beurteilte Annahmen beanspruchten mehr Zeit als richtig eingestufte, was auf die Bereitschaft der Testpersonen hinweist, mehr zu lernen. Zu weiterführenden Forschungsansätzen gehört die Untersuchung einer möglichen Korrelation zwischen den Unsicherheitsfaktoren und deren Nachweisbarkeit sowie die Untersuchung von weiteren Absichten in diesem Zusammenhang.

# Abstract

This thesis proposes a solution against misleading visualizations in health care, which convey inaccurate insights. Misleading elements of such visualizations originate from uncertainties emerging across the steps of the medical visualization pipeline. We investigate the field of storytelling and gamification to support the general audience in recognizing and addressing misleading visualizations in health care. Our research questions are: "*Which types of uncertainty arise in the medical visualization pipeline and is there any intent behind those?*" and "*How can we inform the general population about the existence of visualization uncertainty?*"

To answer the research questions, we created a taxonomy of uncertainty types in the medical visualization pipeline and designed and developed the educational game "DeteCATive" to convey these concepts to the general public in an engaging way. The game includes eight tasks that contain amusing fictional stories with misleading visualizations created with intent and based on medical data. Every story comes with its own set of assumptions. A player should define whether an assumption is correct or false based on the story to gain points and rewards. Then, these points can be spent at the end of the game to fulfill the game objective.

To assess the educational value of the game, we conducted a user study with 21 participants. This study provided us with significant insights. Certain misleading visualization tricks were hard to recognize by the participants. The game obtained positive participants feedback from the participants regarding memorability, reinforcement, and engagement. Incorrectly assessed assumptions required more time as opposed to correctly assessed ones, indicating the willingness of participants to learn more. Further research directions include the investigation of a potential correlation between uncertainty types and detectability or investigating further intents.

# Contents

**Kurzfassung** **xi**

**Abstract** **xiii**

**Contents** **xv**

**1 Introduction** **1**

**2 Related Work** **7**
2.1 Misleading Visualizations . . . . . . . . . . . . . . . . . . . . . . . . . 7
2.2 Uncertainty in the Medical Visualization Pipeline . . . . . . . . . . . . 9
2.3 Visualization for Educational Purposes . . . . . . . . . . . . . . . . . . 12

**3 Methodology** **17**
3.1 Uncertainty in Medical Data . . . . . . . . . . . . . . . . . . . . . . . . 17
3.2 Cases of Non-imaging Data Uncertainties . . . . . . . . . . . . . . . . 22
3.3 Imaging Data Uncertainties Cases . . . . . . . . . . . . . . . . . . . . 36
3.4 Intentional Misleading Elements in Medical Visualization . . . . . . . 52
3.5 Combating Misleading Visualizations through Storytelling and Gamification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

**4 Implementation** **69**
4.1 Creating Stories and Tasks . . . . . . . . . . . . . . . . . . . . . . . . 69
4.2 Unity Game Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77

**5 Results and Discussion** **83**
5.1 User Study . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
5.2 Analysis of User Study Outcomes . . . . . . . . . . . . . . . . . . . . . 86
5.3 Limitations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99

**6 Conclusion** **101**

**List of Figures** **105**

**List of Tables** **113**

CHAPTER 1

# Introduction

Data visualization is the scientific field that takes advantage of a user's perception and cognition to support the derivation of insights from complex data [Tel14]. Basic data visualization approaches include individual or multiple linked plots, dashboards, infographics, renderings, animations, simulations, and so on. These approaches provide improved data accessibility and tend to be persuasive, resulting in insights. However, the effectiveness of persuasion depends on the initial user's attitude [PMN+14]. Visuals enhance the delivery of messages derived from data, but they can also conceal a message, providing little or wrong insight [LGS+22].

Daily, we generate and have access to massive amounts of data from medicine, biology, and life sciences. For example, just one computed tomography (CT) scan produces a data set of several gigabytes (GBs). It is not only a complex issue to organize data storage, but also to structure and interpret data to obtain insights. In health care, data are very specialized, heterogeneous, and big [PRSL23]. Each patient's medical history usually includes diagnoses, prescriptions, data from laboratory examinations and imaging modalities, and other medical data. Presently, electronic health records (EHRs) aim to contain most of a patient's data structured in a pre-defined manner and accessible to various health professionals [DSSK19]. Redundant medical examinations were reduced when introducing EHRs, a rich source of big and complex data.

In health care, visualizations sometimes target the general population in the delivery of a message [NN21]. A famous historical example dates back to London in 1854 when the British doctor John Snow created a map of the cholera outbreak, based on which he could track the source of cholera: a contaminated town well (Figure 1.1). However, John Snow experienced mistrust from his colleagues, who supported a popular belief that breathing vapors rather than drinking contaminated water leads to cholera [Tut]. In modern medicine, visualized data derived from an imaging modality serve as diagnosis

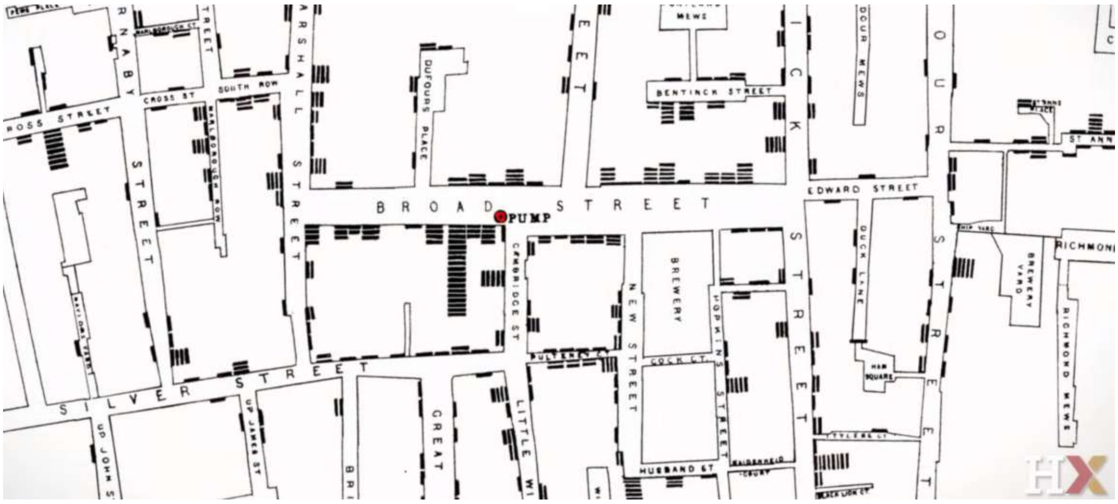



Figure 1.1: The piece of Dr. John Snow's map of London, where the hash marks represent the number of deaths at a specific address. The source of cholera, a contaminated pump, is highlighted in red. The screenshot originates from "John Snow and the 1854 Broad Street cholera outbreak" YouTube teaser video to the Harvard University online course: "PredictionX: John Snow and the Cholera Epidemic of 1854".

confirmation or provide reasons for treatment strategy changes [PRSL23].

When medical data analysis or clinical trials reveal new insights, the message should be conveyed to the general population to support informed health-related decisions. There are several factors influencing how well such messages are conveyed, understood, and translated into actual insights [MGS+21]. Within this work, we will focus on one aspect: uncertainties in the medical visualization pipeline and their potential influence on message delivery and insight derivation.

Primarily, a visualization extracted from erroneous data is doomed to be faulty [LGS+22]. However, even visualizations based on correct data can provide misleading premises due to alterations in the data processing [RPHL14, GSWS21]. Finally, even when the data are correct and the data processing is performed correctly, wrong visual encodings may have been selected to represent the data, thus leading a user to wrong conclusions [LPLK22]. This is particularly important, as the user may have insufficient knowledge about the data and the underlying processes. All these uncertainties may have a significant impact on the outcomes of the analytical process — especially, if these are made with intent to deceive.

Currently, diverse content creators can freely process publicly available non-imaging medical data and introduce a bias [ZSP+21]. Visualizations, mistakenly or intentionally created as misleading, *support misinformation propagation over social media.* Often they are created with easy-to-use editing and data analytic tools [KSO+23]. Such visualizations can provoke mistrust of the general public towards authorities and even underestimation of a critical situation [ZSP+21]. An event that recently happened in China clearly illustrates misinformation propagation. In late 2020 Xinhua News Agency and the People's Daily newspaper published news claiming that a new medicine can inhibit the novel coronavirus infection. Rapidly the news about a possible medicament against coronavirus spread across the Chinese social media Weibo, causing panic buying [HS20]. Several days later, the People's Daily account published a new post on Weibo, urging people not to rush with medicament purchases, since the research was only a preliminary study and the new medicament still needs to pass clinical trials [Peo].

Misleading visualizations *may also be created with intent*, i.e., a visual representation can be intentionally tweaked to support a narrative and push a deceiving idea to the audience (lobbyism). For example, in an article by Lisnic et al., a Twitter user used two charts to prove that vaccination against COVID-19 is ineffective [LPLK22]. From the charts, two premises could be obtained: vaccination rates in Iceland are higher than in Nigeria (Figure 1.2: right), but the number of confirmed cases in Nigeria is lower (Figure 1.2: left). Therefore, the general conclusion supports the idea of vaccine inefficiency and a user obtains "an illusion of data-driven insight" [LPLK22].

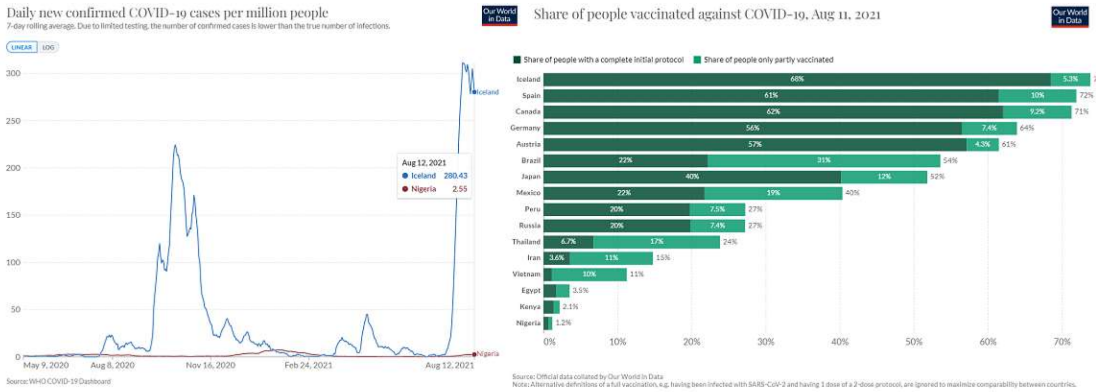



Figure 1.2: Charts, which were used in a tweet about the inefficiency of COVID-19 vaccination. Original figures from [LPLK22], reconstructed in Our World in Data to achieve better image quality.

Modern medical imaging software is even more demanding, as it requires a lot of *background knowledge to operate and — above all — to interpret*. For instance, radiology education without an internship takes about 6 years in Europe [RZR+17]. Here, a fundamental way of creating misleading data representations is, for instance, by forging medical images. Mirsky et al. summarised the possible motivation to forge 3D medical images and presented a sophisticated way of altering medical images using neural networks to falsify a medical condition [MMSE19]. In a more recent article, the framework "Jekyll" by Mangaokar et al. successfully attacked an X-ray and retinal fundus images to inject disease conditions [MPB+20].

In this work, we are particularly interested in deciphering ways that medical data visualizations can be misleading and their potential impact on our society's well-being, as discussed in the examples above. We aim to explore (intentionally and unintentionally) misleading visualizations in the medical field and propose a solution that improves the visual literacy of the general public.

There is already work on understanding and classifying misleading visualizations [LGS+22], on proposing educational approaches that improve visualization literacy [AGR21], and on classifying uncertainty visualization in medical imaging [GSWS21]. Our work aims to combine knowledge from all these three research fields, to suggest an approach that supports general audiences in *identifying and resolving health care-related misleading visualizations resulting from different uncertainty sources within the medical visualization pipeline*.

In this thesis, we will tackle two *research questions*:

1. Which types of uncertainty arise in the medical visualization pipeline and is there any intent behind those?
2. How can we inform the general population about the existence of visualization uncertainty?

To begin with, we study different uncertainty types by collecting vivid cases of visualizations with misleading elements appearing due to uncertainties in the visualization pipeline via an open-coding approach. Moreover, we elucidate the effect of uncertainties on the visualization pipeline and investigate whether these uncertainties have been intentionally introduced. Also, we combine storytelling and gamification in an educational approach that enables the general population to detect misleading visualizations in the health care field. Thus, we target improving visualization literacy about given types of uncertainty in the medical visualization pipeline.

We implement our storytelling approach for the creation of an educational tool to communicate uncertainty types in misleading visualizations in health care. This approach is accessible to the general population due to the use of accessible terminology rather than scientific and medical terms, engaging strategies from journalism and films, and simplified visualizations [SH10]. Gamification is a proven way of enriching the learning experience with typical elements used in gaming, thus supporting the education process in an engaging way [LACA18]. Employing gamification is proven to be beneficial not only in educational fields like engineering [YA21] or preclinical training [KSG+22], but it also has supported post-traumatic stress disorder (PTSD) treatment [DOIG19].

At the individual steps of the thesis, we address the following questions:

1) Open-coding approach:

- What kinds of misleading elements appear throughout the steps of the medical visualization pipeline?
- Where do these misleading elements originate from?
- Is there an intent behind the introduction of these misleading elements in the medical visualization pipeline?

2) Educational approach, combining storytelling and gamification, to combat misleading visualizations:

- How to recognize misleading elements in medical visualization?
- What is the impact of the misleading elements in medical visualization on the analytical capabilities of the intended audience?

The contributions of our work revolve around the creation of a taxonomy of uncertainty types in the medical visualization pipeline along with their potential intents and a corresponding educational tool to communicate those to the general population. As opposed to the first story-based game that supports visualization literacy by Huynh et al. [HNGC20], our game focuses on *intentionally misleading* stories with *medical data visualizations*, which include different *uncertainty types*. Our work aims to provide new insights to researchers of misleading visualization in the medical field and is beneficial for platforms communicating science to the general population.

More details on the state-of-the-art of the relevant research are included in Chapter 2. In Chapter 3, the taxonomy of uncertainty types intentionally introduced in the medical visualization pipeline and the corresponding approaches to combat these uncertainties through the development of an educational game are presented. Chapter 4 includes the implementation details of the educational game. The results of the user study used for the game evaluation and the limitations are discussed in Chapter 5. In Chapter 6, the thesis summary and future work conclude the thesis.

CHAPTER 2

# Related Work

Investigating the design of misleading visualizations and educating intended users against them are engaging topics in the visualization community. At the IEEE VIS conference, there is an annual event called VisLies, where the conference members discuss confusing and sloppy examples of misleading visualizations collected over a year. Since our work aims to explore misleading visualizations in the medical field and propose a solution that improves the visual literacy of the general public, we conduct below an analysis of fields relevant to our work. This chapter begins with scientific papers about misleading visualization. Then we focus more on uncertainty in the medical visualization pipeline, as this is often a source of misleading elements in the field. Afterward, we conclude with research on visualization for educational purposes, including storytelling and gamification.

## 2.1 Misleading Visualizations

A recent research proposing a taxonomy of misleading elements in visualization was conducted by Lo et al. [LGS+22]. Their research aims to explore misleading visualizations and improve visual literacy. They implemented the grounded theory method (GTM) [Mul14], which includes several stages. First, they collect relevant static images through two types of sources: search engines and social media platforms by using keywords. Second, they open-code the collected data along with constant revising of the developing theory until gathering more data does not yield any further insights. The limitations of their method include the majority of data found being in English, only static images being considered, data collection based only on reported cases, and overlaps between different sources and samples. Finally, they contribute with a taxonomy of misleading elements containing 74 types of issues distributed among 5 stages of the visual analytic (VA) process: input, visualization design, plotting, perceptions, and interpretation.

The VA process used by Lo et al. is the modified pipeline from McNutt et al., who


Figure 2.1: Each step of the VA pipeline creates an opportunity for errors caused by visualization creator choices to emerge, thus giving way to visualization mirages [MKC20].

investigated visualization mirages across the VA pipeline [MKC20]. As opposed to Lo et al. work, they focus more on the VA pipeline and the downstream effects of choices made at each pipeline step, which in the end influence the message a human perceives (Figure 2.1). Additionally, McNutt et al. propose the term visualization mirage. This includes any visualization that at a glance supports a specific message, however under closer examination it brings doubt, due to the choices made along the pipeline [MKC20].

Another important work by Lisnic et al. aims to investigate how charts mislead the general population in practice [LPLK22]. They used the official streaming endpoint to collect COVID-19-related posts with visualizations shared on Twitter. From the collected data a relevant sample was used for qualitative coding. The limitations of their method include the majority of data found being in English and only obvious examples being considered. Finally, they contribute with a typology of misleading post features: source of visualization, text polarity, visualization design violations, and reasoning errors. Their work showed that visualization design violations are not the actual way of spreading misinformation, thus confirming the earlier findings of Lee et al. that charts supporting misinformation arguments usually follow the design guidelines [LYI+21]. In the majority of cases, salient features of visualizations are used to weakly support a user's idea, however, the facts that counteract the idea are purposely omitted. Also, they explain how inductive reasoning supports the misinformation arguments (Figure 2.2) and propose visualization design guidelines.

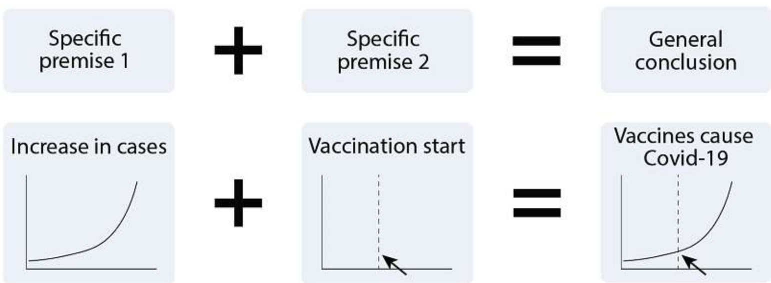



Figure 2.2: Example of inductive reasoning provided by Lisnic et al. [LPLK22]. In this case, two premises weakly support the general conclusion and provide no formal logical fallacies. To invalidate this conclusion, one needs to search for omitted facts.

These findings (including what elements make a visualization misleading [LGS+22], where they arise from [MKC20], and how they mislead [LPLK22]) are relevant to our work. However, the research on misleading visualization is general and the literature lacks investigations within the context of the medical field. Therefore, we need to dig deeper into the source of misleading elements originating across the medical visualization pipeline to explore further misleading visualizations in the medical field.

## 2.2 Uncertainty in the Medical Visualization Pipeline

Within the medical imaging pipeline, uncertainties are sources of misleading visualizations [RPHL14, GSWS21]. Medical pipeline processes differ for non-imaging and imaging data. Regarding non-imaging data, prior studies communicate the existence of uncertainties. These comprise, for example, missing data [RAS21] and incorrect values [WL22] in EHRs. Among others, Gschwandtner et al. proposed a categorized summary of existing dirty data taxonomies causing data quality problems shown in Figure 2.3 [GGAM12]. This work is relevant for the discussion about the quality of non-imaging medical data, especially, the uncertainties in EHRs, which can be potentially used for misleading visualizations.

| **Single source** |
|---|
| Missing data |
| Missing value |
| Missing tuple |
| Semi-empty tuple |
| Dummy entry (e.g., -999) |
| Syntax violation / wrong data type |
| Duplicates |
| Inconsistent duplicates / Contradicting records |
| Approximate duplicates |
| Unique value violation |
| Incorrect values |
| Misspellings |
| Domain violation (outside domain range) |
| Violation of functional dependency (e.g., age-birth) |
| Circularity in a self-relationship |
| Incorrect derived values (error in computing data) |
| Unexpected low/high values |
| Misfielded values |
| Invalid substring / Embedded values |
| Ambiguous data; imprecise, cryptic values, abbreviations |
| Outdated temporal data |
| Inconsistent spatial data (e.g., incomplete shape) |

| **Multiple sources** |
|---|
| References |
| Referential integrity violation / dangling data |
| Incorrect references |
| Heterogeneity of representations |
| Naming conflicts |
| Synonyms |
| Homonyms |
| Heterogeneity of syntaxes |
| Different word orderings |
| Uses of special characters |
| Heterogeneity of semantics |
| Heterogeneity of measure units (EUR vs. $) |
| Heterogeneity of aggregation/abstraction |
| Information refers to different points in time |
| Heterogeneity of encoding formats (ASCII, EBCDIC, etc.) |

Figure 2.3: Five general taxonomies of dirty data combined into one by Gschwandtner et al. [GGAM12]. Multiple sources refer to a situation when multiple data sources have to be integrated/joined.

Other papers communicate uncertainties at the data processing step of the medical visualization pipeline, for example, uncertainties arising from the parametrization of infectious disease outbreak simulations [RRGT+13]. Regarding the representation step of the medical visualization pipeline, Keizer et al. propose a visual dictionary for the representation of biomedical data in the fields of antimicrobial stewardship and infection control [KLS+21], including "DOs", "DON'Ts" and which visualization types should be used (bar, line charts, or dendrogram) for specific content. Additionally, using an icon array instead of a bar chart can lead to an improved understanding of mammography test outcomes [SPS11]. Also, Visschers et al. discussed the influence of presentation data format on risk perception by humans [VMPDV09] via framing effect. However, these facts are never assembled along the medical non-imaging data visualization pipeline.

The first step to a taxonomy of uncertainties in the medical imaging data visualization pipeline was made by Ristovski et al. [RPHL14]. They identified sources of uncertainties in the medical visualization pipeline (Figure 2.4) and described uncertainty types with a mathematical description of uncertainty types using random fields (RF) theory. The final pipeline only accounts for in-time most common medical applications. Uncertainty types are categorized by Ristovski et al. depending on spatial location (discrete or continuous), dimensionality (1D, 2D, 3D, nD), type of events (numerical, categorical, binary, volumetric), and their sources [RPHL14]. For example, the uncertainty of the segmentation outcome is discrete 3D RF with binary events because the imaging data segmented is discrete, usually volumetric, and the segmentation outcome is encoded as a binary event: the voxel belongs to the specific area or not. Moreover, depending on the uncertainty type, they proposed solutions for uncertainty-aware visualization.

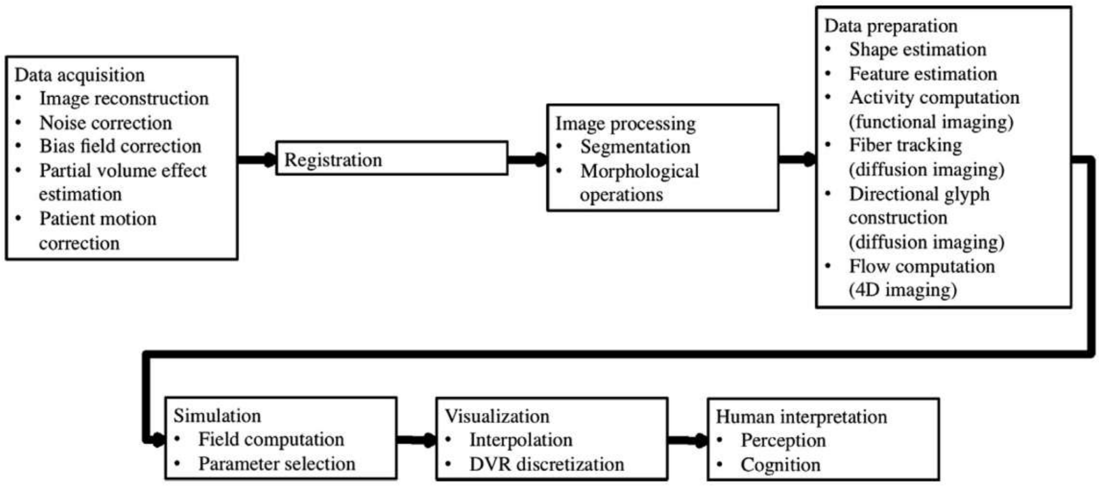



Figure 2.4: Medical visualization pipeline according to Ristovski et al. [RPHL14]. Each box represents a step (or substep), where uncertainty can appear.

Modern techniques of uncertainty-aware visualization and open issues of the field were summarized by Gillmann et al. [GSWS21], who additionally improved the taxonomy of uncertainty in the medical visualization pipeline created by Ristovski et al. [RPHL14]. To the former taxonomy, they added separation of each source of uncertainty into the steps of medical imaging: image acquisition, transformation, and visualization (Figure 2.5). Moreover, they added a new category to the categorization by Ristovski et al. [RPHL14] allocating sources between aleatoric and epistemic uncertainty (Figure 2.6). Aleatoric uncertainty refers to the statistical (random) nature of uncertainty, while epistemic refers to a systematic error [HW21].

Also, there is an example of uncertainty across a more specialized medical visualization pipeline: the tumor radiation treatment (RT) planning pipeline [Rai18]. In this work, visualization strategies to address sources of uncertainties specific to each step of the RT pipeline are present. To conclude, it is hard to find the one medical visualization taxonomy overarching all taxonomies. While none of the previous works investigates the potential intent behind these uncertainties, our work aims to fill this literature gap.

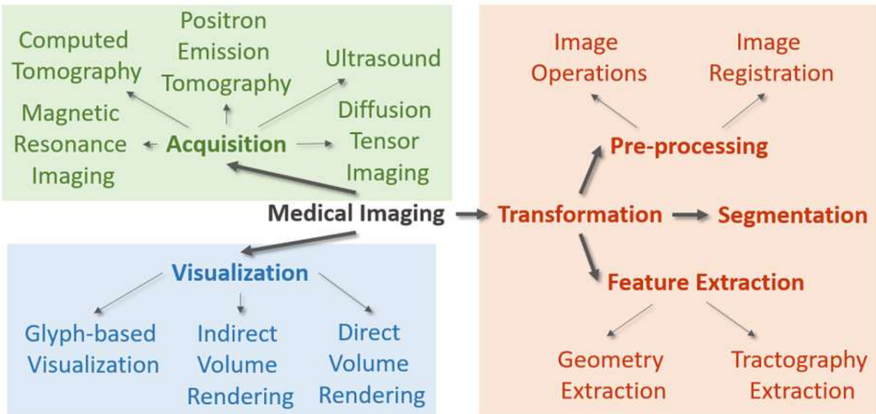



Figure 2.5: Steps within medical imaging, including image acquisition, transformation, or visualization. An uncertainty-aware visualization may be required to assess the uncertainties introduced at these steps and to communicate them to a clinician [GSWS21].

| Sources of Uncertainty | Type | Dimensionality of Event | Category | Description of Event |
|---|---|---|---|---|
| Positional uncertainty | $a$ | 3D | numerical | discrete |
| Pixel/voxel value uncertainty | $a$ | nD | numerical | discrete |
| Incompleteness of Data | $a$ | nD | numerical | discrete |
| Model inaccuracy | $e$ | 3D | spatial/volumetric/numeric | discrete/continuous |
| Model incompleteness | $e$ | 3D | spatial/volumetric/numeric | discrete/continuous |
| Parameter/boundary condition uncertainty | $a/e$ | nD | numerical | discrete |
| Rasterization uncertainty | $e$ | 2D/3D | numerical | continuous |
| Perceptual and cognitive uncertainty | $e/a$ | 3D | binary | continuous |
| Decision making bias | $e/a$ | 3D | binary | continuous |

Figure 2.6: Taxonomy of uncertainty sources in medical imaging [GSWS21]. The green colored lines coincide with the image acquisition step, while the orange and blue lines refer to the image transformation and visualization steps, respectively.

## 2.3 Visualization for Educational Purposes

**Storytelling**. Most visualization approaches targeting large audiences include digital storytelling. The development of a virtual story includes the collaboration of storytelling and computer graphics, used for processes like 3D modeling, animations, and rendering [TZ09]. Narrative visualizations can be found at art exhibitions or museums (Figure 2.7 a)) because it is an effective way to convey messages and build memorable experiences [CCN17]. Also, they can be enriched with virtual reality (VR) or augmented reality (AR) technologies to increase audience engagement as shown in Figure 2.7 b).

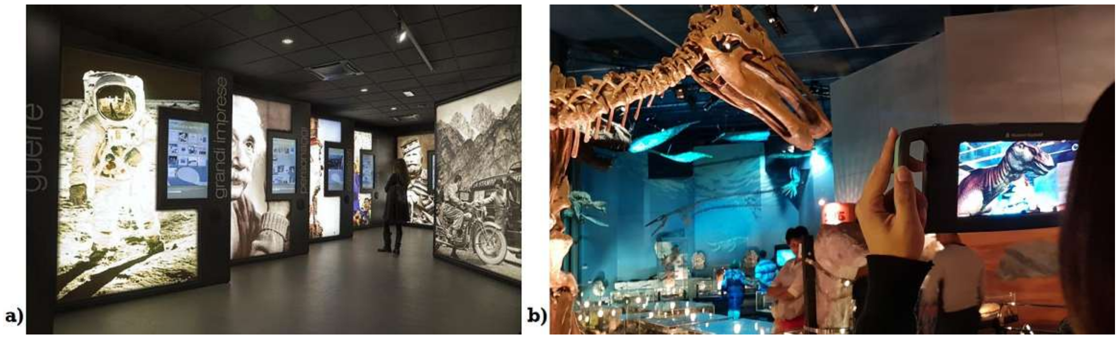


Figure 2.7: a) Spazio La Stampa exhibition in Turin, b) ExplorAR tablet with augmented reality application visualizing a dinosaur from its skeleton at National Museum Cardiff in Wales.

According to Bach et al. there are only three requirements when using narrative/storytelling design patterns: having a story, knowing your audience, and the effect the story should have on them [BSB+18]. When it comes to complex concepts, effectively communicating stories to a large audience involves simplification of the message, emphasizing only key concepts, and omitting minor details [Böt20]. In the medical field, storytelling is used to enhance the communication between a patient and a physician, for example, to explain the diagnostic or therapeutic procedure. For instance, in the work of Rozmovits et al., patients with prostate cancer feel less fear and are better prepared for brachytherapy, an internal radiation therapy, when the information delivered is not patchy and overwhelming [RZ04].

Other approaches aim to cope with insufficient information provided to the patients before the coronary angiography procedure, Brand et al. developed a graphic-based information brochure [BGH+19]. They indicated that standard written informed consent (IC) may not guarantee patient comprehension of procedural specifics and potential risks. Factors such as patient anxiety, literacy level, and clinicians' communication skills may contribute to this. Their study demonstrated that supplementing written IC with graphical information improves patient comprehension, reduces anxiety, and increases satisfaction. Two pages from the graphic-based information brochure are shown in Figure 2.8.

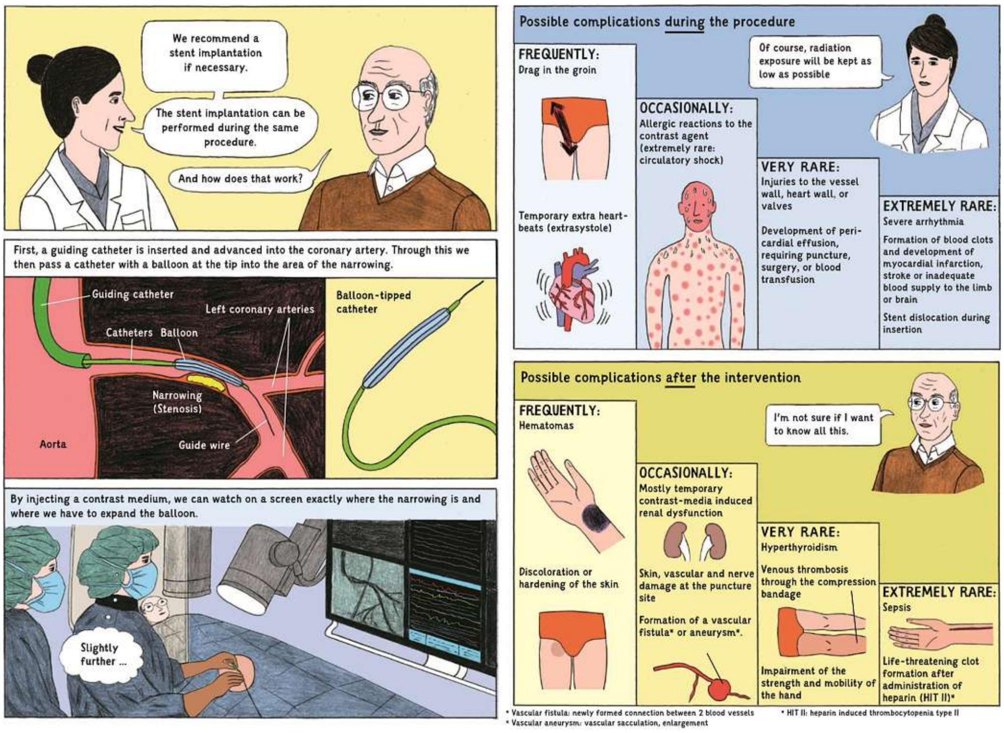



Figure 2.8: Fragments from the graphic-based information brochure about coronary angiography procedure [BGH+19]. The fragments include pieces of information about the procedure details and possible complications during and after the intervention. The full version of the brochure is available on Annals of Internal Medicine.

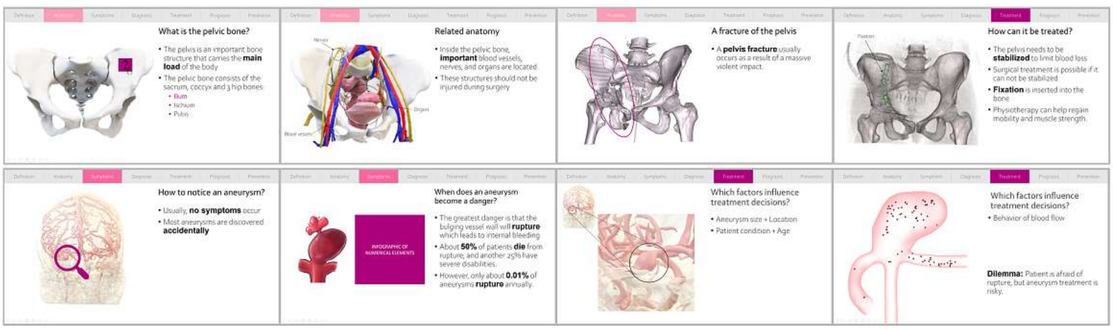



Figure 2.9: Fragments from top: pelvic fracture story, bottom: aneurysm storyboard. The stories include detailed anatomy with 3D models, condition-associated risks, and aspects of treatment [MGS+21].

In the field of narrative medical visualization, Meuschke et al. provided the first-ever implementation of data-driven narrative techniques to inform the general population about pelvic fractures (Figure 2.9 top), brain aneurysms (Figure 2.9 bottom), and liver cancer [MGS+21]. Additionally, they defined a seven-stage template to structure narrative

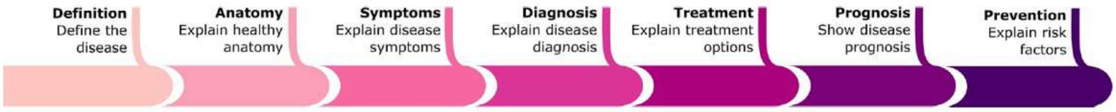



Figure 2.10: Template proposed by Meuschke et al. to communicate disease data using a seven-step template in narrative medical visualization [MGS+21].

medical visualization communicating disease data shown in Figure 2.10. This template is a generalized structure of 30 web blogs communicating diseases to the general audience [MGS+21]. These blogs belong to credible sources like university hospitals, scientific institutes, and online encyclopedias. In a later work by Meuschke et al., the authors improved the medical story design of the liver cancer story and summarized their insights into a research agenda for narrative medical visualization [MGS+22]. The improved liver story is shown in Figure 2.11. It contains an additional slide about the outcome of the patient's story using positive framing: showing a concrete treatment method and active prevention methods to convey that liver cancer is not a hopeless condition. Recently a case study for narrative visualization in medicine was published by Kleinau et al. [KSM+22]. This work dealt with explaining to the general population the impact of vortexes in an aortic arc using narratives enriched with data-driven visualizations.

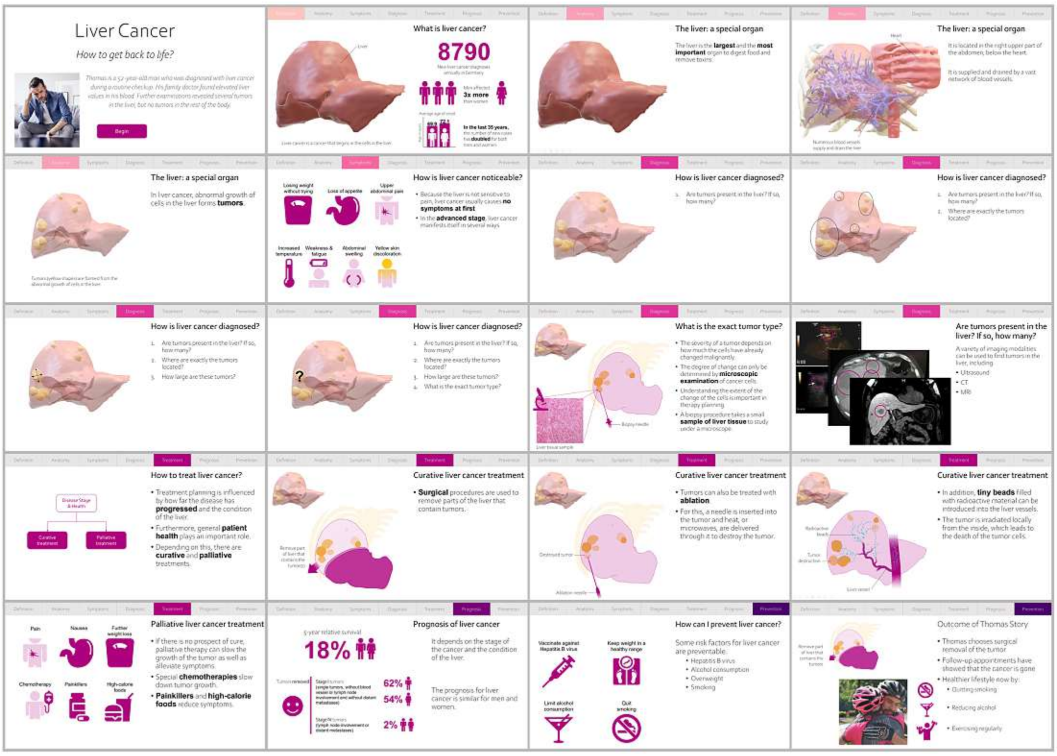



Figure 2.11: Complete liver cancer story [MGS+22].

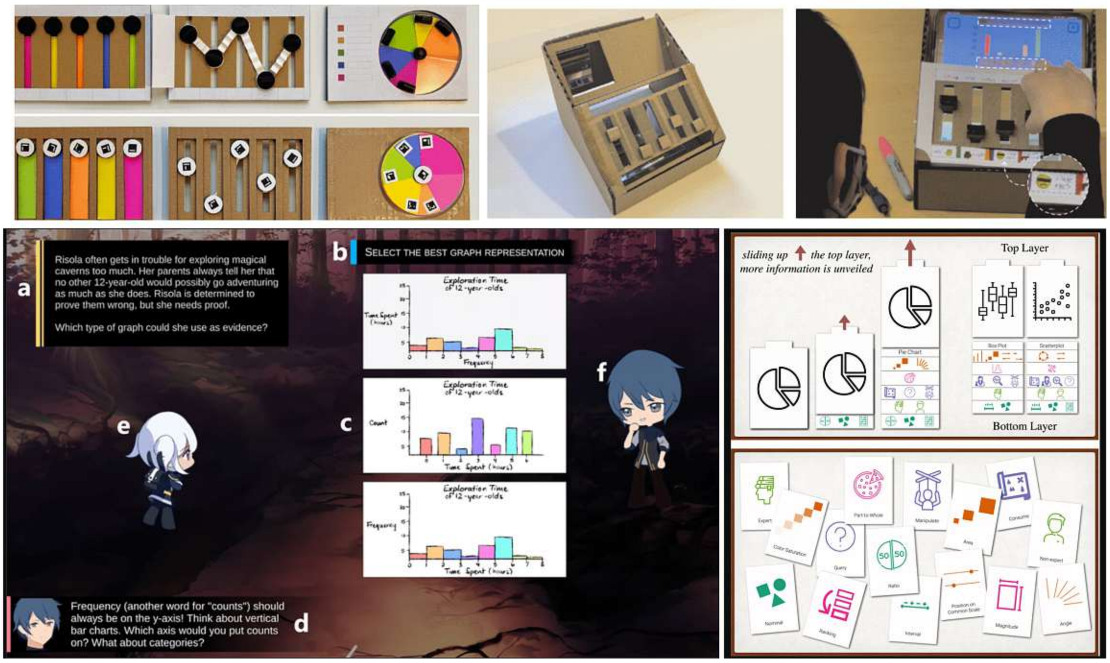



Figure 2.12: Games promoting visualization literacy. Top: do-it-yourself crafting game for learning through creation about visualizations [BVY+22], bottom left: role-playing game for learning how to read visualizations [HNGC20], bottom right: two card games for visualization design education [AGR21].

**Gamification**. Many educational approaches are enriched with gamification techniques because they increase engagement in the study process and provide challenges, which can motivate active participation [LACA18]. In the medical field, gamification can be used for education or support treatment. For example, an application Mindfulness Coach for mobile phones uses some gamification techniques to teach individuals struggling with PTSD to reduce their stress by making practices. The application included several "difficulty" levels enriched with progress tracking visualized as a growing tree and a rewarding system with badges.

Moreover, the gamification method is used to promote visualization literacy, for example, through a do-it-yourself (DIY) crafting game [BVY+22]. During the game, kids construct the "Data is Yours" toolkit from common low-cost materials (Figure 2.12 top middle). The age of the kids that participated in the user study was between 6–11 years old. This toolkit can be personalized to three types of interactive physical visualizations: bar, line, and pie (Figure 2.12 top left). The process of hand-crafting a visualization is supported with a digital component accessed via a mobile phone (Figure 2.12 top right). Finally, this toolkit encouraged kinds to actively interact with visualizations, demonstrating the potential of the DIY approach in improving visualization literacy.

Another example is a (role-playing game) RPG-style visualization game presented in a study by Huynh et al. [HNGC20]. In this game, a player's character, a magic school student, has to help other in-game characters by answering multiple-choice questions, shown Figure 2.12 (bottom left). The questions revolve around determining the appropriate visualization to solve a character's problem (proving a fact to the character's parents to not get in trouble) and visualization reading skills, thus launching the educational process. As the previous game, it addresses the gap in visualization literacy, but the target audience is older children (between 11 and 13). Overall, this approach increased engagement but did not show a significant difference in visualization reading skills [HNGC20].

Also, a card game [AGR21] approach is used to cope with visualization literacy. In this game, one player picks a visualization with specific components, while another player has to guess the visualization by asking questions about its components. However, it uses basic and not-field-specific data visualizations. Moreover, Schindler et al. generated templates for nested papercraft of biological and anatomical interest, which the audience can print and assemble (Figure 2.13) to learn more about the included structures [SKRW22]. To our knowledge, no articles related to gamification to improve medical data visualization literacy of the general audience have been published yet, thus giving additional value to our work.

Figure 2.13: Assembled nested papercraft of a human head (left) that reveals the different anatomical substructures, when color filters are used (right). Edited from [SKRW22].

# CHAPTER 3

# Methodology

Investigating, understanding, correcting, and raising awareness of misleading visualizations is a continuous effort of the visualization community. New categories of misleading elements are being identified and organized into taxonomies [LGS+22, GSWS21]. Novel methods to aid visualization design [DWQW22] and to detect and counteract misleading visualizations are under development [ZM22]. Typically, research papers are expert-oriented and can be challenging for the average visualization consumer to perceive. Furthermore, medical visualization could contain information that might not be accessible to someone lacking medical or clinical knowledge. Our goal is to raise awareness regarding misleading medical visualizations among the general population.

To achieve this objective, we design and develop an educational game that offers an approachable, understandable, and enjoyable way to learn about misleading visualizations in health care. This chapter outlines the three main components of our game creation process: the *theoretical background of uncertainty* being the source of misleading elements of visualizations in health care, *storytelling* as a way to capture interest in complex topics, and the integration of *gamification*.

In Section 3.1, we cover the basis of uncertainty in medical data. The chapter continues with cases of uncertainties within both non-imaging and imaging medical data in Section 3.2 and Section 3.3, respectively. Additionally, Section 3.4 addresses the intentional introduction of uncertainties. Section 3.5 provides applied concepts of storytelling and gamification.

## 3.1 Uncertainty in Medical Data

In this section, we elucidate how uncertainties in the medical data visualization pipeline may be the source of misleading visualizations. First, we categorize medical data into two groups: *imaging* and *non-imaging* data, and introduce some basic concepts of medical imaging techniques. Then we present a taxonomy of the uncertainty types that emerge throughout the visualization pipeline [RPHL14]. For better comprehension, we provide indicative cases for each uncertainty type. Furthermore, we classify the intention behind the introduction of any kind of uncertainty.

### 3.1.1 Medical Data

Medical data are heterogeneous since they are obtained from different sources and stored in various formats. Medical data include administrative data tied to billing and insurance, real-time health monitoring data (e.g., heart rate monitor standing at an operating theater), health surveys, patient-generated data, imaging data, and more. Storage and management of such complex data are challenging [GBR21].

A way to integrate patient data is EHRs. They contain laboratory test results, medical history, diagnoses, treatments, medications, and other relevant data. However, the overall adoption of EHRs is problematic due to challenges mainly related to standardization, security, and privacy [PAA+22]. Oppositely, imaging medical data are stored in a standardized way in digital imaging and communications in medicine (DICOM) format. A DICOM file includes not only the raw image data but also the technical details of the scan and accompanying patient and hospital data.

Medical imaging data are obtained using medical imaging modalities. These include CT, magnetic resonance imaging (MRI), X-rays, ultrasound, positron emission tomography (PET), single-photon emission computed tomography (SPECT), and others listed in Table 3.1. These imaging data sets can be represented as two- and three-dimensional visualizations, depending on the modality. For instance, CT and MRI images are three-dimensional volumes, but typically these volumes are sliced perpendicular to a given axis and shown slice-by-slice.

In our work, we consider two medical data types: imaging and non-imaging (Figure 3.1).

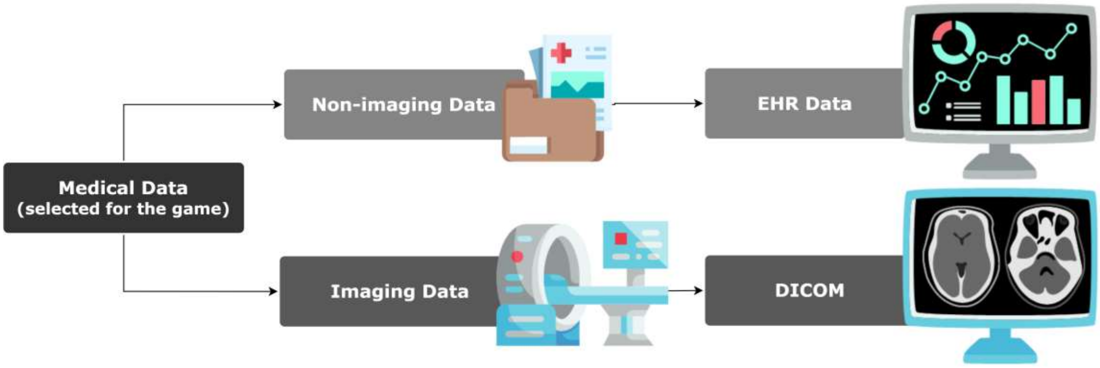



Figure 3.1: Medical data structures selected for the purposes of this work.

Under non-imaging medical data, we considered only tabular data coming from EHRs, clinical trial data, genomic data, and public health data, including cohort and population studies, and epidemiological or demographical data. Their representation differs from the visualization of imaging data. In this case, we commonly encounter statistical visualizations. These can be, for example, a bar chart indicating the population suffering from a specific type of cancer or a graph with the annual number of fatalities attributed to cardiovascular diseases.

### 3.1.2 Medical Imaging

Medical imaging techniques offer a sophisticated, noninvasive way of accessing both the anatomical and physiological information of a patient. For example, nuclear medicine techniques like PET and SPECT illuminate cellular metabolic activity and some physiological processes [Liv12]. These techniques show the traveling and accumulation of a radiotracer (radioactive substance) inside the body. Among these, in PET glucose analog [$^{18}$F]-Fluorodeoxyglucose (FDG) radiotracer [Gro] aids in detecting cancer or even arthritis in preclinical studies (Figure 3.2). While these methods alone have a good temporal resolution, they have insufficient spatial resolution and a lack of anatomical landmarks [Liv12]. In other words, it is hard to determine the precise location of a radiotracer inside the body in the visualization.

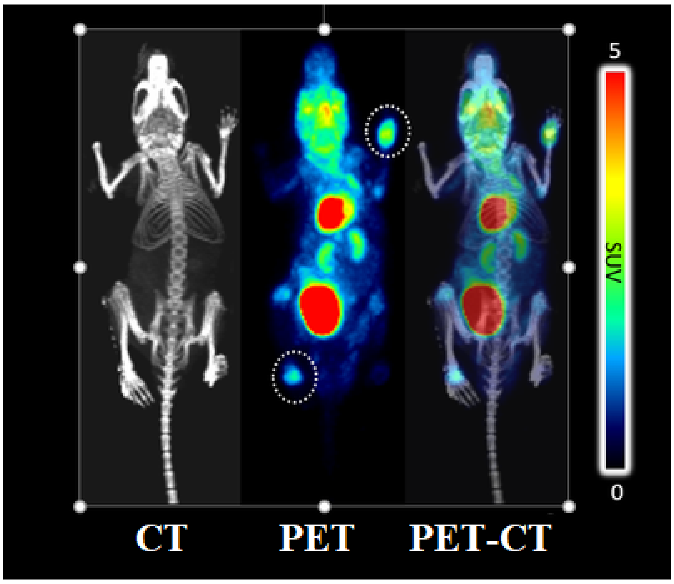



Figure 3.2: Fused PET and CT data used for arthritis inflammation assessment in a mouse. The CT scan delivers the anatomical information for the PET scan, while the SUV – standardized uptake value of [$^{18}$F]-FDG that shows metabolic activity. Edited from [Gro].

High spatial information can be delivered by combining nuclear medicine information with diagnostics information, delivered by MRI and CT (Figure 3.2). These two techniques use completely different methods of capturing data. Inside an MRI machine, there is a high magnetic field generated by superconducting coils [BKN+08]. The magnetic field forces atoms, usually hydrogen, inside a patient's body to align their spinning moment along the field. During the scan, the computer-generated radio waves abruptly kick the hydrogen atoms out of the aligned position. While still in the high magnetic field, the hydrogen atoms return to their aligned state and they emit a signal used to reconstruct a 3D model of the patient's anatomy. Hydrogen atoms are abundant in water- and fat-containing tissues and their complex interaction with surroundings influences the signal encoding relaxation times. Therefore, MRI is frequently used to visualize the soft tissue contents [BKN+08].

By contrast, a CT machine uses X-rays to gather data. A rotating gantry around the patient has an emitter and a detector placed diametrically. Emitted X-rays travel through the body, undergoing attenuation by different tissues before striking the detector. Employing Radon transformation, a computer program reconstructs the body of a patient by volumetric pixels known as voxels that contain X-ray attenuation data stored in Hounsfield Units (HUs) [SSHW05]. Finally, a CT scan provides 3D information with distinct bone-to-soft-tissue contrast. However, the main disadvantage of a CT scan is the patient's exposure to ionizing radiation [MPB+09].

| Source | Principle | Conditions Detected (examples) |
|---|---|---|
| X-rays | attenuation of X-rays | pneumonia, bone fractures |
| CT | attenuation of X-rays in 3D | internal bleeding, tumors |
| Mammography | attenuation of soft X-ray spectra | breast cancer, enlarged axillary lymph node |
| Angiography | injecting contrast and visualizing the vessels with real-time X-ray | blood vessel blockage, aneurysm |
| Fluoroscopy | drinking contrast medium and visualizing with real-time X-ray | digestive tract dysfunction |
| MRI | magnetization of water molecules | brain disorders, cancer |
| Ultrasound | real-time sonography (collecting echos from planar ultrasonic waves) | developing fetuses during pregnancies, kidney abnormalities, gynecological conditions |
| PET | detection of radiotracer emission | cancer stage, metabolic condition |
| SPECT | detection of radiotracer emission | neurological conditions, infections, and stroke |

Table 3.1: Summary of medical imaging modalities, indicating their working principle and examples of target conditions.

### 3.1.3 Uncertainty Taxonomy within the Medical Visualization Pipeline

To establish the theoretical basis for our work, we selected relevant articles about misleading visualizations and uncertainty visualization in health care. Then, we gather and analyze several misleading and uncertainty visualization taxonomies. Afterward, we gather representative cases of visualizations of both imaging and non-imaging data to observe these fallacies happening.

Having selected cases of misleading elements, we categorize them under different uncertainty types through an open-coding approach to create a comprehensive taxonomy. We follow the categorization of uncertainties provided by Griethe et al. [GS05] into the following categories: error, imprecision, non-specificity, and subjectivity. We augment this categorization with an additional uncertainty type: incompleteness (otherwise known as missingness) [ANI+20]. The final taxonomy includes misleading elements sorted in relation to uncertainty types, represented by colored diagram blocks within the remainder of this work (Figure 3.3):

- *Error*, indicating outlier or deviation from a true value;
- *Imprecision*, when the resolution of a value differs from the needed resolution;
- *Non-specificity*, meaning a lack of distinction for objects;
- *Incompleteness*, including missing attribute(s);
- *Subjectivity*, relating to the degree of subjective influence in the data.

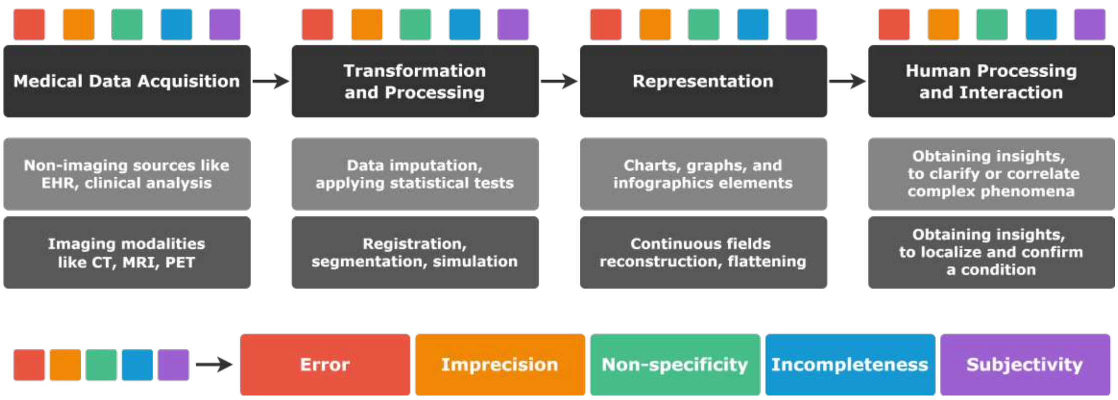



Figure 3.3: Medical data visualization pipeline, inspired and adapted from [RPHL14, SPBR20, GSWS21, LGS+22]. Black blocks: non-imaging/imaging data visualization pipeline steps. Grey blocks: processes or attributes describing each step of the visualization pipeline. Colored blocks: uncertainty types categories used to sort misleading elements appearing at each pipeline step.

Between raw data and its final visualization, medical data go through multiple steps that all involve uncertainties that can add misleading elements. These processes differ depending on the nature of the medical data. For example, at the transformation and processing step, segmentation, which is the isolation of a specific region of interest, is applied to imaging data, but not to non-imaging data. The final taxonomy entails two visualization pipelines, as processes that utilize non-imaging and imaging data are different and may entail different uncertainties and/or implications for the final outcome and the analytical process (Figure 3.3). In the upcoming sections, we will provide detailed information about these processes and misleading elements that they can introduce.

## 3.2 Cases of Non-imaging Data Uncertainties

This section provides cases of uncertainty types for every step of the visualization pipeline of non-imaging medical data, with figures that can be found at the end of each pipeline step description. Each description includes also a table with a brief overview. Many cases of misleading visualizations in this section have been retrieved or inspired from the Github Gallery with a collection of “bad" visualizations created by Lo et al. [LGS+22].

### Uncertainties in Data Acquisition

The most common source of uncertainties during non-imaging data acquisition is the presence of dirty data [GGAM12]. Table 3.2 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Data Acquisition | |
|---|---|
| ■ Error | Misspellings, unique value violation, contradicting records |
| ■ Imprecision | Unexpected low and high values, observational error |
| ■ Non-specificity | Using medical tests with different specificity and sensitivity |
| ■ Incompleteness | Missing data |
| ■ Subjectivity | Heterogeneity of representation |

Table 3.2: Uncertainty type cases in the data acquisition step of the non-imaging medical data visualization pipeline.

■ **Error**. Dirty data contain errors such as misspellings and duplicates. Duplicates may cause unique value violations and even lead to contradicting records [GGAM12]. For instance, EHRs are filled in by humans, thus there can be an inaccurate birth date entry, mistakenly assigning treatment plans to the wrong patient, or a diagnosis with a misspelled name [TP03].

■ **Imprecision**. When using medical instrumentation like a thermometer, the displayed measured value represents the sum of the true value and the observational error. The observational error equals systematic plus random errors, thus leading to imprecision of the measured value. Additionally, unexpected low and high values bring imprecision to the data set. Moreover, there is data variation, when medical data are obtained from different software, raters, technicians, and machines [MEB+23].

**Non-specificity**. Data uncertainty can be the result of laboratory tests with different specificity and sensitivity [BPDCRG+17]. In other words, some tests are not as “good” at determining a disease or a healthy condition in comparison to other tests (Figure 3.4). Sensitivity determines the ratio of sick patients detected as being sick (true positive rate), and specificity is the ratio of healthy patients detected as being healthy (true negative rate). Unfortunately, both measures counteract each other: the higher the specificity, the lower the sensitivity, and vice versa (Figure 3.5).

**Incompleteness**. Within a data set, a missing entry leads to the appearance of the uncertainty type known as incompleteness. The appearance reasons include the unavailability of certain information at the time of recording (e.g., due to drop-outs) or errors in data entry (e.g., a measurement was involuntarily omitted) [PL20]. The absence of an entry can potentially influence subsequent data analysis [GGAM12].

**Subjectivity**. Subjective uncertainty arises when individuals record acquired data heterogeneously. For example, they use different representations of measurement units or use synonyms for the same value [GGAM12]. This can introduce mismatched fields inside a merged database.

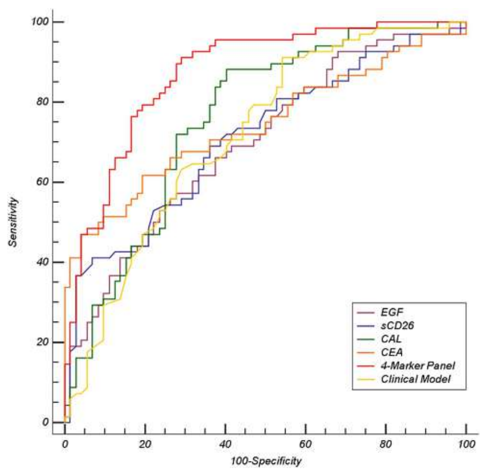



Figure 3.4: Receiver operating characteristic (ROC) curves for each individual marker showing different performance for lung cancer prediction [BPDCRG+17].

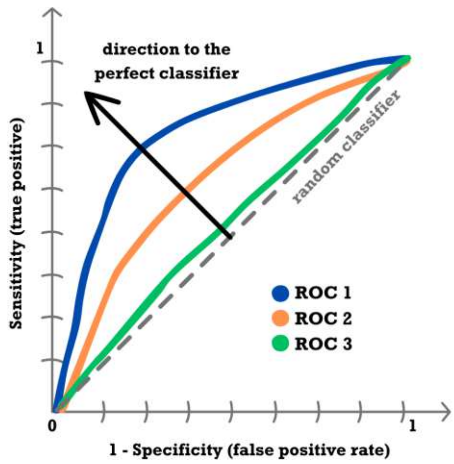



Figure 3.5: Hints to reading ROC curves.

### Uncertainties in Transformation and Processing

The next visualization pipeline step entails mathematical manipulations with data and preparation thereof before being visualized, for instance, in the form of a chart. Table 3.3 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Transformation and Processing | |
|---|---|
| Error | Calculation, imputation errors |
| Imprecision | Too few data points, interpolation and extrapolation of data |
| Non-specificity | "Lie with statistics", performing a statistical test on the wrong type of data |
| Incompleteness | Incomplete, missing calculations (such as normalization) |
| Subjectivity | Selective data, known as cherry-picking |

Table 3.3: Uncertainty type cases in the transformation and processing step of the non-imaging medical data visualization pipeline.

**Error**. Inaccurate mathematical manipulations can corrupt the data, result in overfitting models representing a data set, or propose questionable prediction [LGS+22]. Sometimes, missing data have to be replaced with substituted values, e.g. from imputations. However, this imputation can bring some uncertainties, as the imputed value might not be accurate. Feng et al. artificially introduced missing values into COVID-19 data from the Democratic Republic of the Congo [FHG21]. Then, they measured the crude bias (a measure of deviation) between the imputed values and true values, taking into account the percentage of missing data (Figure 3.6). Depending on the method of imputation, the grade of deviation of the new imputed data from the "true" data increased when the percentage of missing values rose.

**Imprecision**. Uncertainty can arise, for instance, from processing samples with too few data points. Such samples are not representative and pose limitations in statistical analysis, as it is hard to draw meaningful conclusions. There is a way to hide small samples, for example, inside a bar chart (Figure 3.7).

**Non-specificity**. In this case, we consider uncertainties that bring a lack of distinction. It happens, for example, when building the trend line on random data because it can misinform people about an existing trend [LGS+22].

Another case is transforming categories to values and performing, e.g., a statistical metric on it, as if working with numerical data. Categorical data have two subgroups: nominal and ordinal data. Ordinal data require the encoding that keeps the relationship (hierarchy) between the variables. Examples are the stages of liver disease (inflammation, fibrosis, cirrhosis, and liver failure), pain level, and education level [MMAF14]. In nominal data, no variable relationship exists and they can be listed in any order. Such variables can be gender, blood type (A, B, AB, O), eye color, or surgical outcome (dead or alive) [MMAF14].

Ordinal data do not have an equal separation between levels and often violate requirements for numerical metrics models, for example, being normally distributed. Treating ordinal data as if they were numeric leads to misinterpretations as in a non-medical example by Liddell et al. [LK18] (Figure 3.8). In this case, star rating distributions of two movies were analyzed using two tests: ordered probit (suitable for ordinal data) and metric (not suitable). As a result, the differences in means between these two movies were in opposite directions when using different statistical tests [LK18].

■ **Incompleteness**. Here we include missing calculations or steps of data processing. For example, missing normalization makes comparisons within a visualization meaningless [LGS+22]. Visualization with absolute numbers is misleading due to the hidden influence of some factors, which are removed when the data are normalized. Such factors can be the population or age structure as shown in Figures 3.9 and 3.10.

■ **Subjectivity**. The selection of a part of data that fits a claim and excluding the other part introduces bias. This phenomenon is known as cherry-picking. For example, picking selectively countries, specific ages or genders may facilitate drawing misleading conclusions out of the data. As an example, in Figure 3.11, a conclusion about the success of the Swedish COVID-19 policy was drawn without a comparison against any other country [LPLK22].

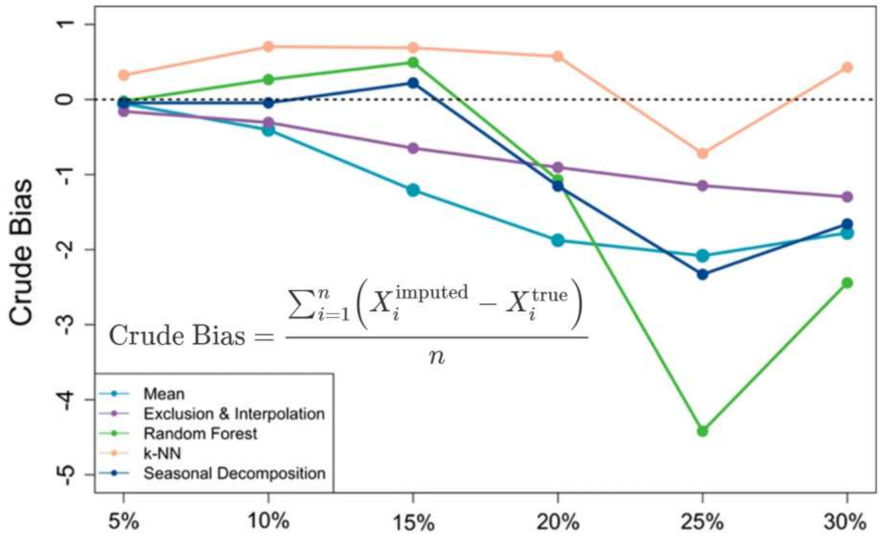



Figure 3.6: ■ Crude bias (a measure of divergence of the data with imputed values from the true data) plotted against the percentage of missing data in the true data set for 5 different imputation methods. Edited from [FHG21].

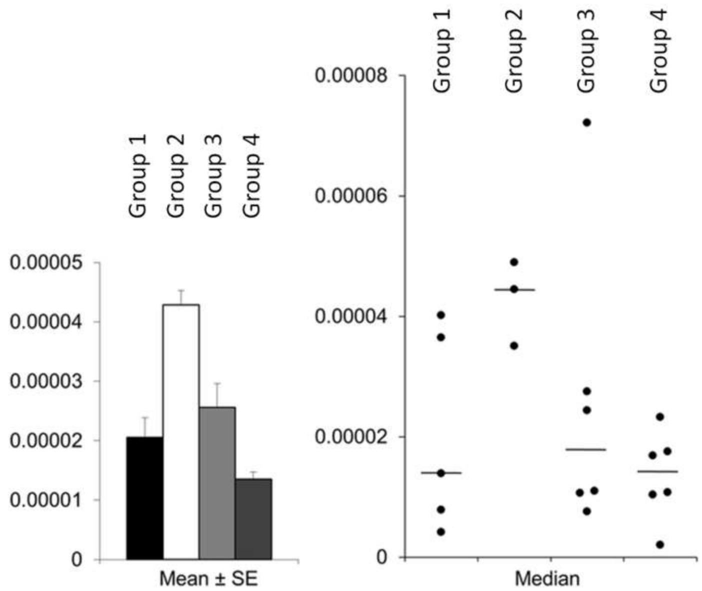



Figure 3.7: Different sample distributions can lose features when represented as a bar chart. The bar chart on the left obscures that each sample size is small. Observe, for example, that there is a large variance in group 1 and an outlier in group 3. These features are clearly visible in the scatter plot on the right. Edited from [WMWG14].

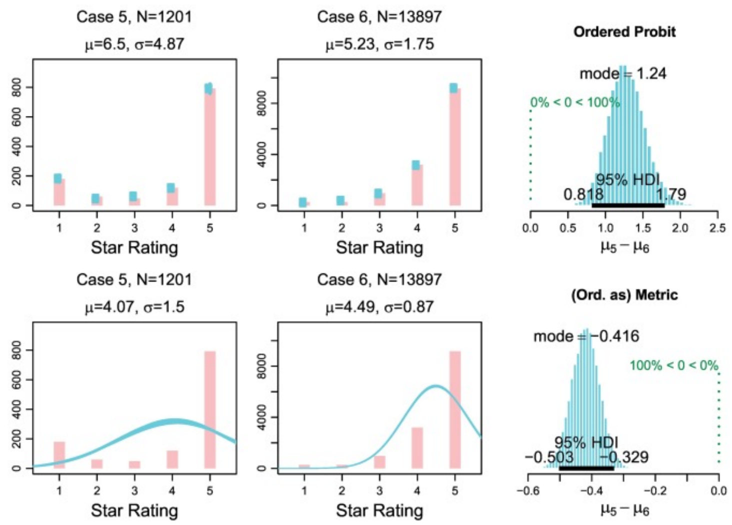



Figure 3.8: Comparison of ranking star rating of two movies (cases 5 and 6) by different statistical tests. The ordered probit model (top) describes the ordinal data more accurately. When applying the ordered probit test, the resulting difference in mean between the two movies is in the opposite direction to the outcome of the metric test (bottom) [LK18].

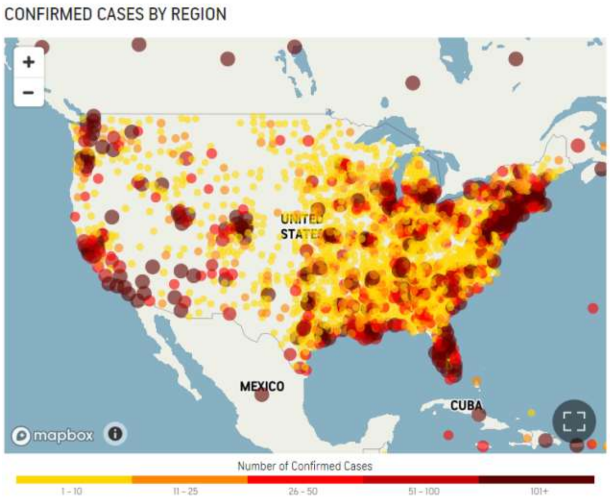



Figure 3.9: ■ Map of the United States of America with confirmed cases of an unknown disease. While data are not normalized on the population of the states, the hotspots are located in highly populated areas [LGS+b].

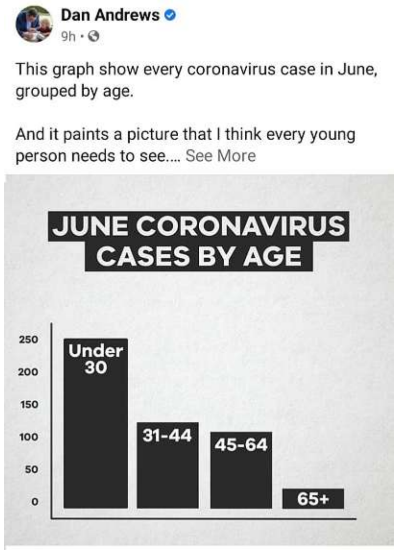



Figure 3.10: ■ Comparison of COVID-19 cases incidence between age groups, not accounting for demographical distribution (there are more young people than elderly people) [LGS+c].

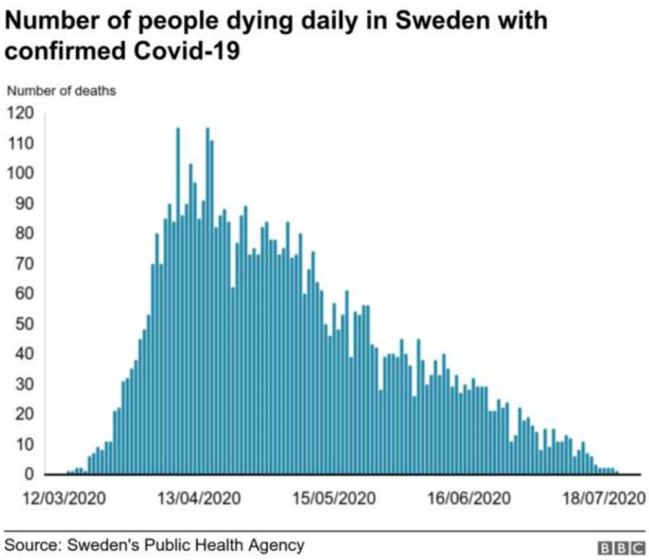



Figure 3.11: ■ The COVID-19 death curve in Sweden, where the COVID-19 restrictions were relatively mild. The author of this visualization was arguing against COVID-19 restrictions by providing a single piece of evidence. A comparable neighbor with fewer deaths or with soft measures but more deaths is missing [LPLK22].

### Uncertainties in Representation

At this pipeline step transformed and processed data are visualized in a form such as a map or graph. Table 3.4 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Representation | |
|---|---|
| Error | Color violation, wrong usage of scales, misrepresentation of data values |
| Imprecision | Uneven binning, inconsistent tick intervals, low axis resolution |
| Non-specificity | Heterogeneous data adaptation, data of different magnitude |
| Incompleteness | Incomplete chart, truncated axis, limiting data range, missing axis, units, labels, legend |
| Subjectivity | Choice of the chart, color mess, setting arbitrary thresholds, dual axis, inappropriate axis range, chaotic canvas |

Table 3.4: Uncertainty type cases in representation step of the non-imaging medical data visualization pipeline.

**Error**. There is a large group of representation fallacies, called color violation. An example of the wrong usage of color scale is using a gradient hue scale for categorical data [NJD20]. The gradient would create an illusion of a relationship among independent categories. Also, using the rainbow color scale for the temperature field is counterintuitive because the color change along the rainbow is not uniform [Sza18]. Moreover, discretizing the color map of the continuous variable creates abrupt borders between close values [LGS+22]. An example of the wrong usage of color is shown in Figure 3.12.

Within presenting fallacies, a case of misrepresentation is drawing bars disproportionately to the values that they contain [LGS+22]. Also, there is the wrong usage of scales — linear against logarithmic scales. A logarithmic scale, when used instead of a linear one, flattens the curve, i.e., makes the rate of change look smaller than it actually is [LGS+22]. Additionally, log scale graphs provoke less accurate understanding by the public, thus causing altering of the public policy [RSDG20].

**Imprecision**. This kind of uncertainty arises due to the uneven split of data into bins or intervals [LGS+22]. Inconsistent trick intervals hide the changing scale as shown in Figure 3.13. Such irregularity complicates the extraction of values from the bar plot. For instance, in Figure 3.13 extracting the percentage of the U.S. adult population with obesity in 1990 is impossible due to the year range being assigned to each observation.

**Non-specificity**. Presenting data of different magnitudes can cause problems with analyzing trends of smaller variables like in Figure 3.14. To address this issue, one approach is to employ a dual axis [LGS+22]. However, using a dual axis may generate uncertainty. For instance, in Figure 3.15 the axes are dual and truncated. This choice results in the counties with mask mandates appearing below the blue line representing the no-mask mandate counties. However, upon closer examination of the data values, the orange line representing the mask counties corresponds to a higher number of cases.

**Incompleteness**. An incomplete chart makes it hard to extract information from the chart. The missing elements include axis, axis ticks, units, value labels, title, axis title, and legend [LGS+22]. As an example, in Figure 3.16 no information that would at least imply what the visualization is about is present. Also, visualizing data over a chosen time range can blur the overall picture, as it is the visualization of an incomplete data set, known also as propagated selective data. A similar case, but regarding the y-axis, is the truncated axis [LGS+22].

**Subjectivity**. Within this type of uncertainty, we have many different cases including a choice of charts, which can cause using a confusing chart type, using indistinguishable colors like in Figure 3.12. Alongside this, reasoning errors like setting an arbitrary threshold, which serves for judging a phenomenon also cause misinterpretations [LPLK22]. A choice of using visual embellishments can lead to misleading elements like confusing legends, cluttering, and difficult-to-read text [LGS+22]. For example, we refer the reader to Figure 3.17.

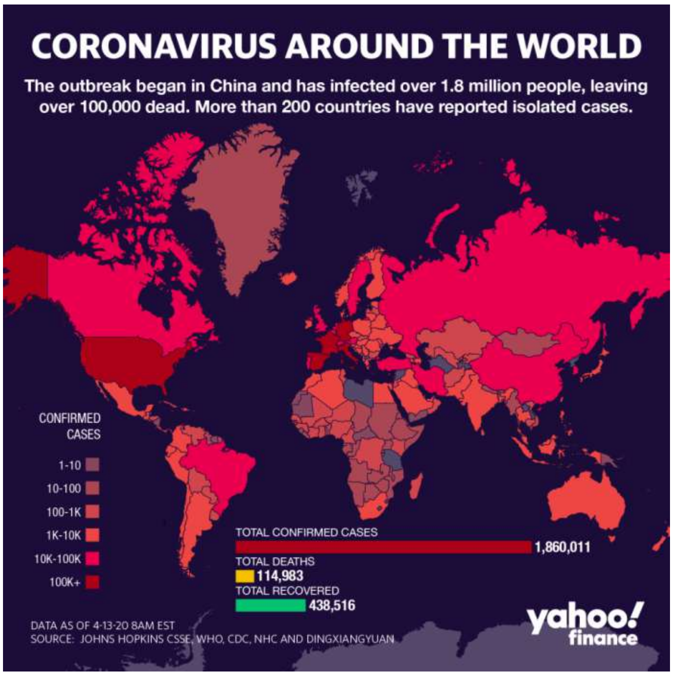



Figure 3.12: Visualization of COVID-19 statistics where a continuous variable (number of confirmed cases) has been discretized and hardly differentiable colors were used [LGS+a].

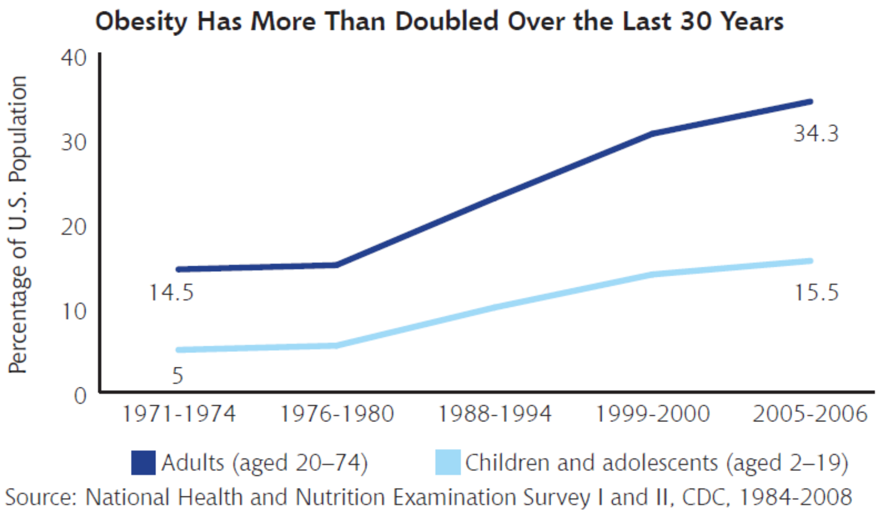



Figure 3.13: ■ Visualization with inconsistent tick intervals (random time ranges) and low time resolution, resulting in imprecision of extracted values for a specific year [LGS+f].

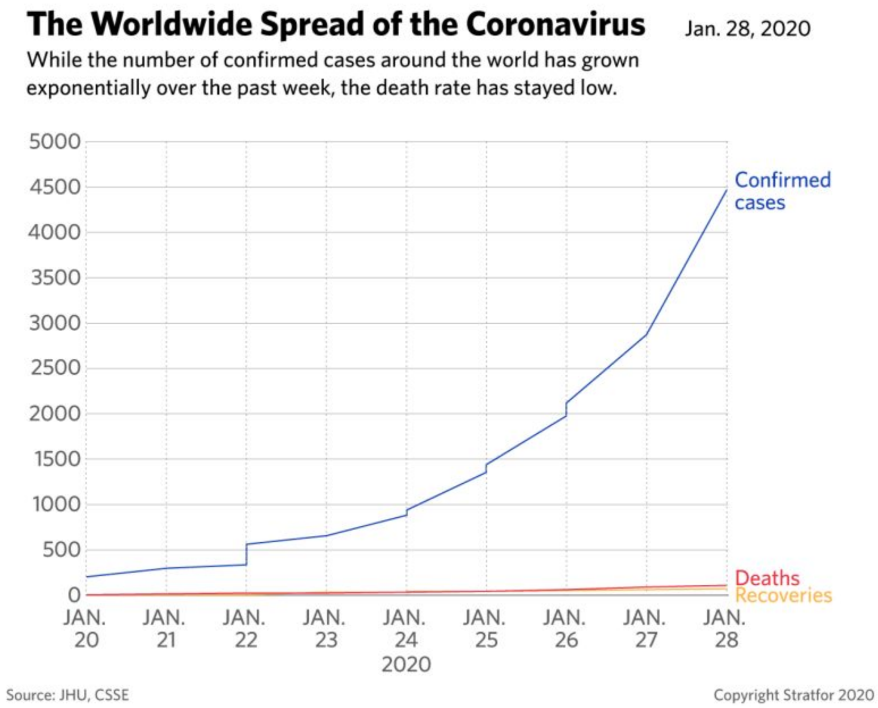



Figure 3.14: ■ Visualization with data of different magnitudes, leading to a low resolution of data with a smaller magnitude (deaths and recoveries) [LGS+d].

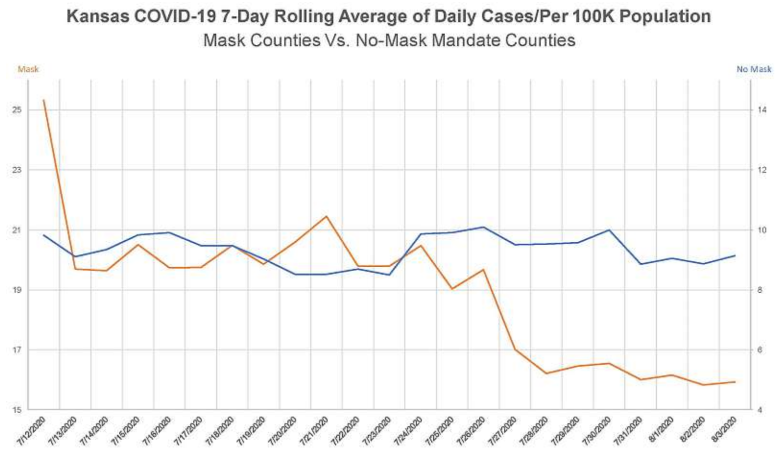



Figure 3.15: ■ Visualization with dual and truncated axes that make the orange graph look below the blue one, even though the orange line values are higher [LPLK22].

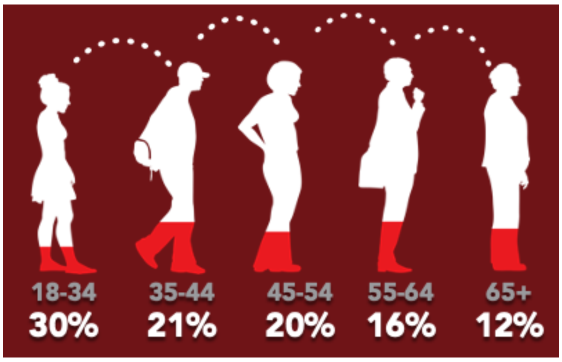



Figure 3.16: ■ Visualization with missing title, labels, and annotations. It is impossible to know what the infographic is about [LGS+22].

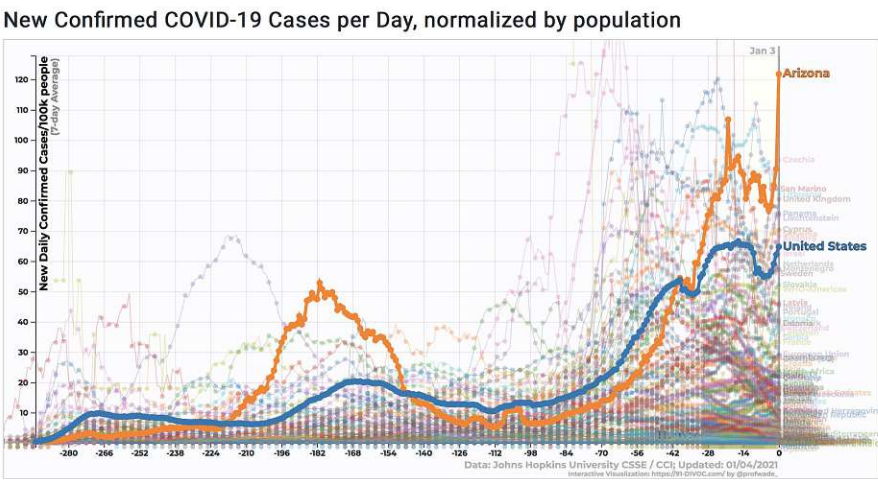



Figure 3.17: ■ Visualization with cluttering due to line overplotting [LGS+e].

### Uncertainties in Human Processing and Interaction

The next visualization pipeline step entails human interpretation of the ready visualization. Table 3.5 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Human Processing and Interaction | |
|---|---|
| ■ Error | Not color-blind friendly choices |
| ■ Imprecision | Low resolution of scales |
| ■ Non-specificity | Projection distortion and abstraction of a distribution |
| ■ Incompleteness | Lack of user guidance, lack of visual hierarchy, no communication about uncertainties |
| ■ Subjectivity | Visual illusion, false linkage, power of words, personal/domain bias |

Table 3.5: Uncertainty type cases in human processing and interaction step of the non-imaging medical data visualization pipeline.

■ **Error**. In this case, the uncertainty of a visualization perception can be caused by not taking into consideration the visual abilities of the user, for example, color blindness [MA14]. Also, the choice of a color map influences the incremental data variation perception. For example, Figure 3.18 shows the difference between the disproportional (c) and sequential ordering (d) of colors.

■ **Imprecision**. In this case, the low resolution of the color map leads to the imprecision of values perceived from a visualization. In Figure 3.19, it is hard to extract values. For example, the conduct disorder (sky blue) disability-adjusted life-years (DALYs) for the 10-14 age category are approximately 2. In this figure, extracting the value is problematic also due to the stacked data representation.

■ **Non-specificity**. Some visualization types, like bar charts, can cause oversimplification. In Figure 3.20, four sample groups with different features are presented: symmetric distribution, with an outlier, bimodal distribution, and with unequal sample sizes. When these samples are shown in a bar chart in Figure 3.20 (left) these features are obscured, leading to a lack of distinction for the sample distribution inside the bar chart.

■ **Incompleteness**. This uncertainty can be triggered by a lack of user guidance and visual hierarchy (Figure 3.21). User guidance is closely related to a knowledge gap, which is the difference between what is required to know to analyze a visualization and the actual knowledge of a user [CAA+20]. Moreover, missing information often results in uncertainty. This may involve, for example, hidden premises [LPLK22], trends and correlations presence [DWQW22].

■ **Subjectivity**. Visual illusions like three-dimensional pie charts or area encoding can cause uncertainty in the interpretation of the visualization [LGS+22]. For example, in Figure 3.22, the numerical values do not correspond to the visual representations. Also, this type of uncertainty includes false linkage (pattern-seeking, assigning meaning to unexplained salient features, invalid comparison), and power of words (misleading title, annotations, emotional features) [LGS+22].

Moreover, there are biases influencing the visualization perception, for instance, personal bias (previous knowledge), application domain (government representative or nurses), and context (technical specifics, physical environment) [MA14]. These types of bias provoke reasoning errors like incorrect reading of charts and misinterpretation of scientific studies [LPLK22].

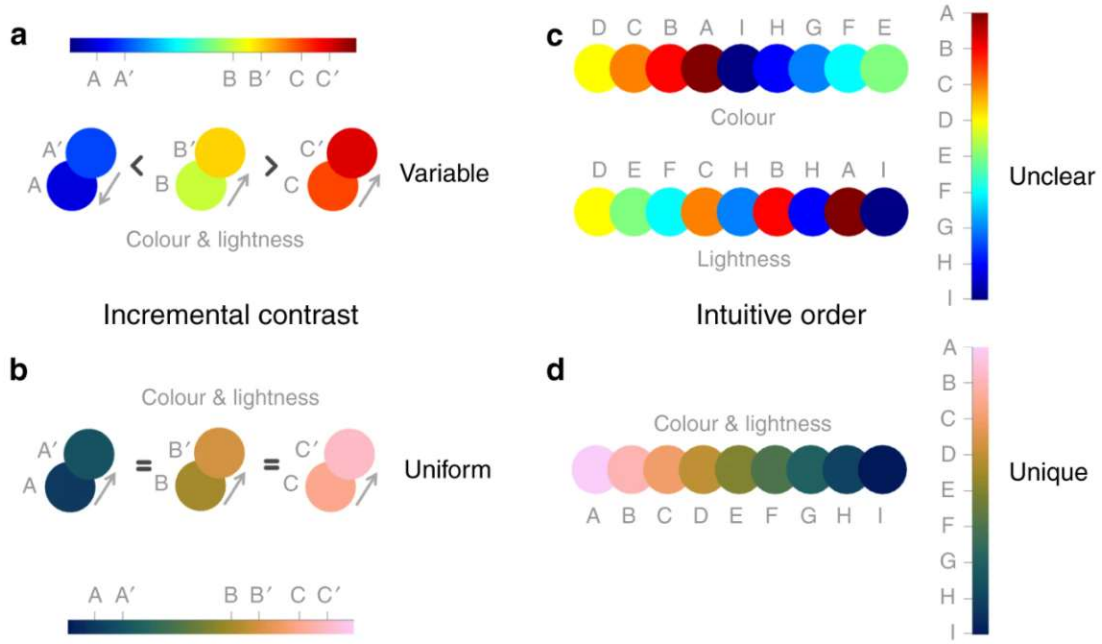



Figure 3.18: ■ Challenges when creating a color map with perceptual uniformity, when using a rainbow color map. Comparison of (a) "unscientific" color map (rainbow) and (b) "scientific" color map (batlow) shows over- or underestimation of incremental contrast of the rainbow color map. A sequential ordering of the color maps is shown in (c) and (d) [CSH20].

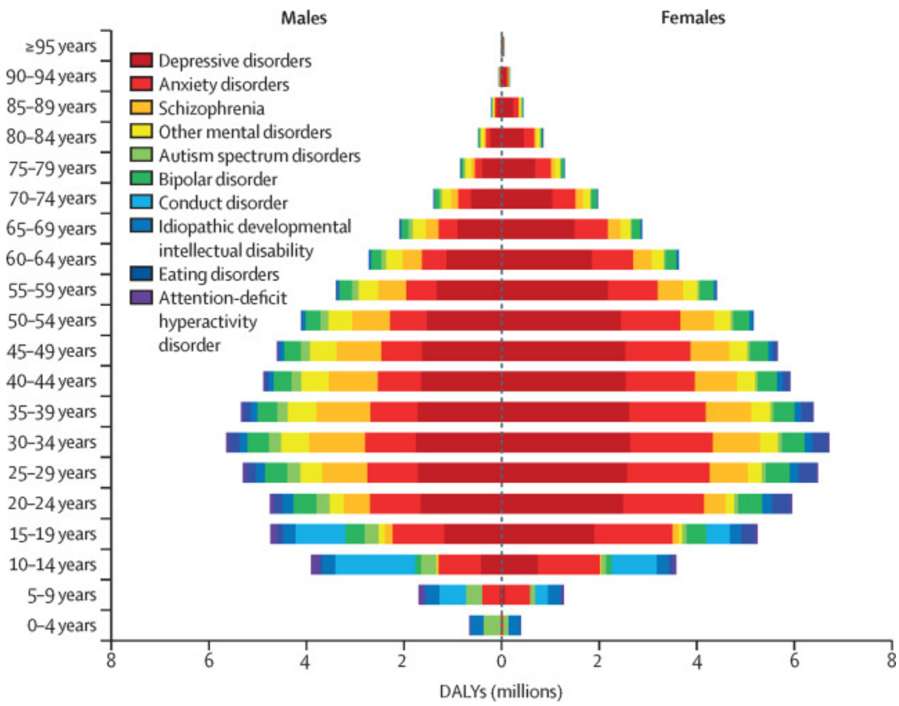



Figure 3.19: Visualization of disability-adjusted life-years (DALYs) with ineffective representation that leads to imprecision when trying to obtain scalar values [C+22].

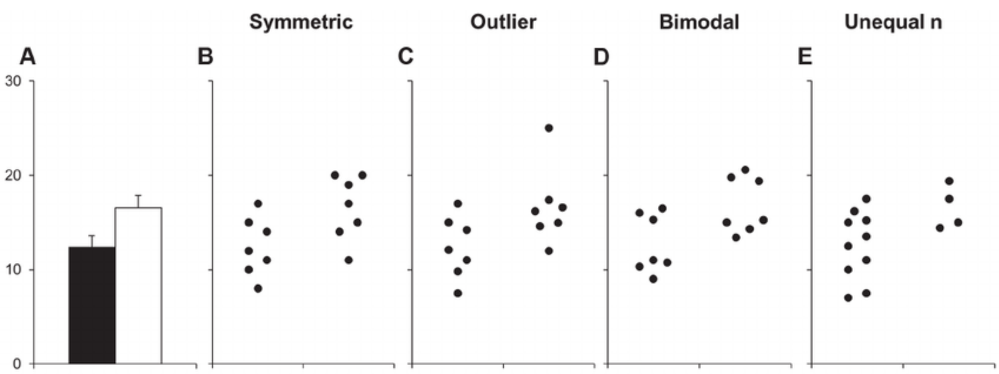



Figure 3.20: Bar chart visualization that hides the underlying data distributions causing a lack of data sample distinction. Edited from [WMWG14].

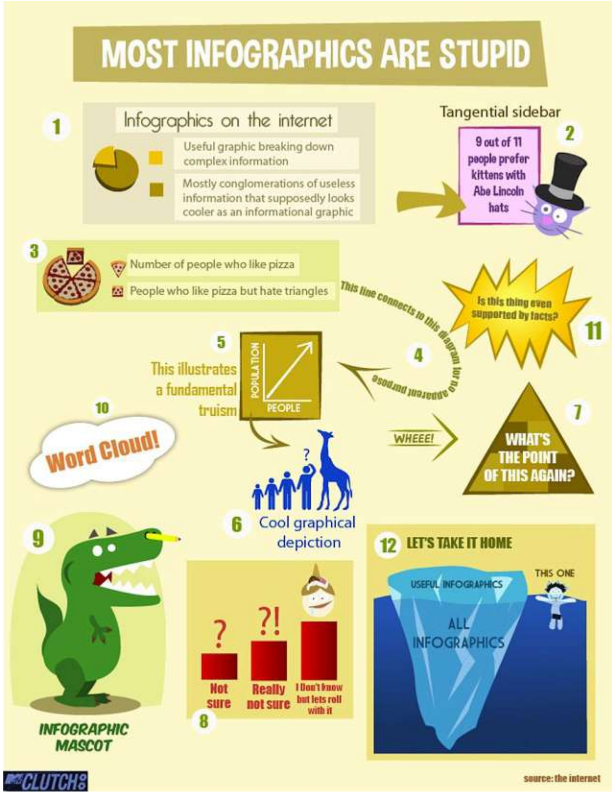



Figure 3.21: A humorous visualization pointing out the lack of visual hierarchy in infographics, where the enumerated consequential information units are ordered chaotically [MTV].

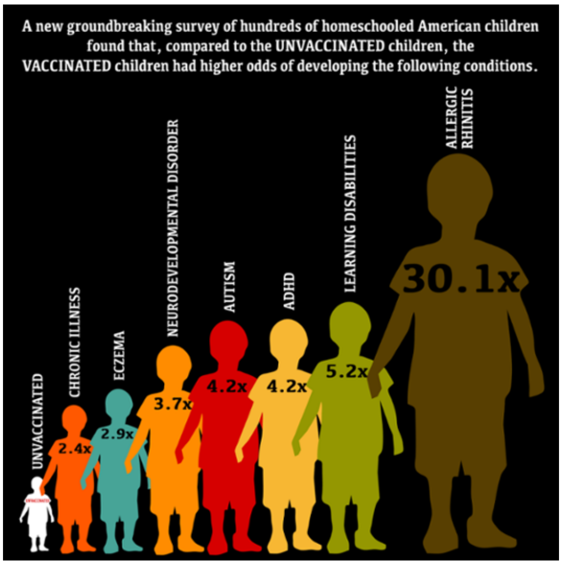



Figure 3.22: Visualization with pictorial area encoding. The height of the silhouette with 30.1x (brown) looks only two times higher than the 5.2x (green), but the area should be 6 times larger [LGS+g].

## 3.3 Imaging Data Uncertainties Cases

This section contains cases of uncertainty types for every step of the medical imaging data visualization pipeline, supported with figures that can be found at the end of each pipeline step description. Each description starts with a table, which contains a brief overview. Many cases have been inspired by the reviews of Ristovski et al. [RPHL14] and Gillmann et al. [GSWS21]. Table 3.6 summarizes the cases discussed in this subsection.

### Uncertainties in Data Acquisition

This medical imaging data visualization pipeline step coincides with the first step on the pipeline proposed by Ristovski et al. [RPHL14]. The step includes signal capture, image reconstruction, and different types of corrections. Table 3.6 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Data Acquisition | |
|---|---|
| ■ Error | Medical imaging artifacts |
| ■ Imprecision | Positional, temporal, and resolution uncertainty |
| ■ Non-specificity | Physics of imaging modalities |
| ■ Incompleteness | Limited field of view |
| ■ Subjectivity | Variations of the protocol |

Table 3.6: Uncertainty type cases in data acquisition step of the medical imaging data visualization pipeline.

■ **Error**. This uncertainty type originates from medical image artifacts. Artifacts are misrepresentations of the ground truth (real situation) due to not idealized conditions during medical imaging examinations [BF12]. For instance, an artifact may be a random noise, which obscures low-contrast borders. Also, during acquisition, a patient can accidentally move, while natural organ movements cannot be deliberately stopped. Therefore, scans contain motion artifacts [Rai18].

There are many different types of artifacts and they differ among the modalities. For example, shadowing artifacts appear in sonography, when an ultrasound wave cannot pass through a tissue boundary with a high impedance difference, resulting in signal loss [HPB13]. By contrast, CT scans may contain different artifacts and some of them are shown in Figure 3.23. Ring artifacts originate from miscalibration or failure of an X-ray detector element. Being used in hip replacements or dental fillings, metal can cause metal artifacts. Also, polychromatic beams and the energy-dependent attenuation of X-rays inside the body produce beam hardening artifacts.

■ **Imprecision**. An imprecision in medical imaging data may arise from different sources. For instance, the transducer is hand-held during the ultrasound scan, thus causing the positional uncertainty [GSWS21]. In radiotherapy, positional uncertainty due to an impression in the patient's position measurement can lead to the shift of the treatment area and lead to major consequences [TOI+20]. Also, variations in radiotracer emission start time (post-injection) introduce imprecision in PET signals [S+15]. Moreover, the partial volume effect, when different tissue types are enclosed in one voxel (Figure 3.24), can cause resolution uncertainty [PB07]. Also, in medical image reconstruction, a process of converting electric signals to an image is conducted under certain assumptions, so that the image is not reconstructed perfectly, but with a certain uncertainty [RPHL14].

■ **Non-specificity**. This uncertainty arises from the process of capturing signals by medical imaging modalities [GSWS21]. The process is dependent on the physics of imaging modality operation (as discussed in Section 1.1.2), and certain body parts structures are more emphasized than others. For example, Figure 3.25 shows how different modalities highlight different structures inside a turtle: hard materials such as eggshells are more prominent in the CT scan or the vascular system in the MR scan. The work by Lauridsen et al. contains one more exciting scan [LHW+11]: a contact medium-enhanced MR scan of the gastrointestinal tract of a spider shown in Figure 3.26.

Moreover, for the same reason, the sensitivity and specificity of different modalities toward detecting certain conditions vary. For instance, having high sensitivity, MR can show poor specificity, when detecting tumor [Rai18]. As an example, for breast cancer detection MR sensitivity and specificity range of 85.7 to 100 percent and 25 to 100 percent, respectively [APT+22]. Higher specificity of breast cancer detection can be achieved by combining mammography with ultrasound obtaining the range of 76.5%–82.5% [APT+22].

■ **Incompleteness**. This uncertainty type relates to information loss when employing the wrong field of view (FOV), an area of the body that is examined during the scan. If a pathological condition is located outside the scanned area or at the edges of the FOV, the examination of the scan is prone to a location-related error within the diagnostic error classification developed by Kim and Mansfield [OYA+21]. Additionally, missing data occur in longitudinal studies [OEI+22]. For example, with every decade of the longitudinal study in the elderly, there is a 25% higher risk of dropout rates [CBM05]. The reasons include physical and cognitive decline, health problems, and death [HAS09].

■ **Subjectivity**. Imaging protocols provide instructions on how to adjust the acquisition machine to access a particular body area and/or a pathological condition. They are used by technicians in medical imaging areas like MR, CT, or nuclear medicine. Unfortunately, protocols lack standardization: the individual preference of the radiologist, vendor-specific sequences, or tweaking of imaging protocols during the scan are common sources of a variety of data [SHMB17]. As an example, different MR pulse sequences produce various signals. As shown in Figure 3.27, the same structures are highlighted differently in scans obtained via different MR pulse sequences resulting in different contrasts.

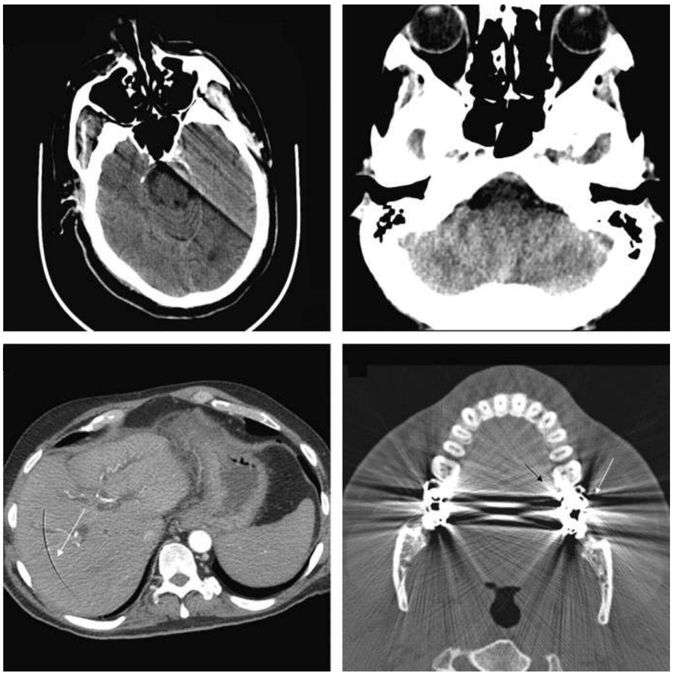


Figure 3.23: Variety of examples of CT artifacts. Top left: motion, top right: beam hardening, bottom left: ring (pointed by an arrow), bottom right: metal artifacts. Edited from [PSHK08].

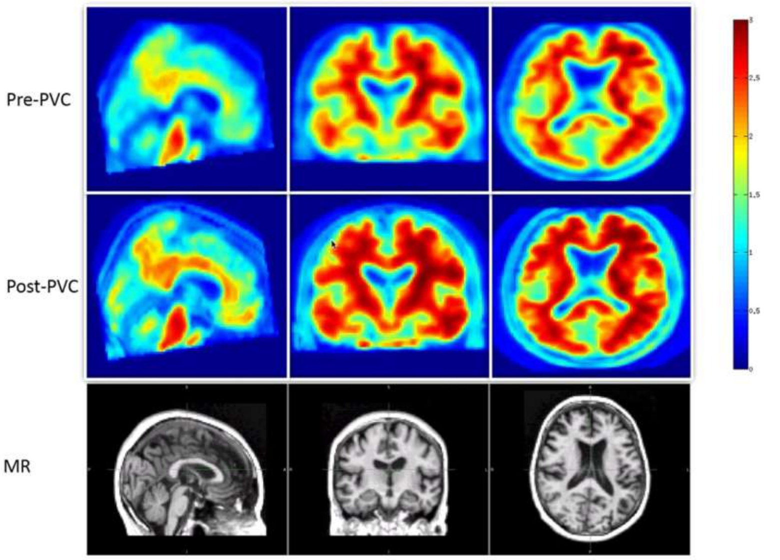



Figure 3.24: Partial volume correction in PET scans. Partial volume correction reduces the imprecision of values in voxels. Top: original scan, middle: after partial volume correction, bottom: T1-weighted MR scan used in partial volume correction. The color legend represents regional standard uptake value ratios [YHG+17].

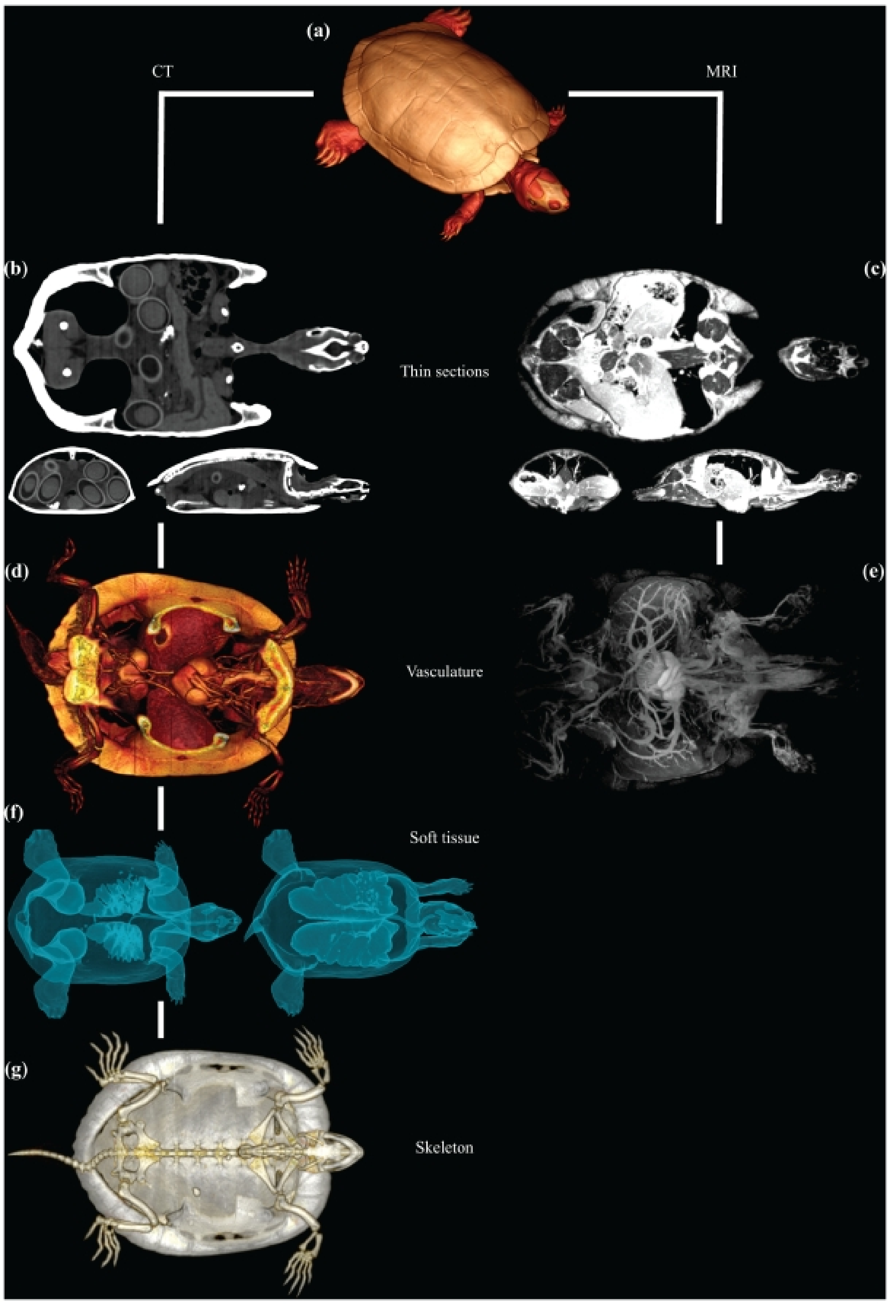



Figure 3.25: The red-eared slider examined via CT (left) and MR (right) techniques. The skeleton is captured in a CT scan, a slice of which is depicted in (b), and finally rendered in (g). Due to the injection of contrast substances, the vascular system can be visualized by CT (d) and MR (e). The soft tissues are visible in the MR scan (c), while the CT scan contains the lungs due to air-tissue radiodensity difference [LHW+11].

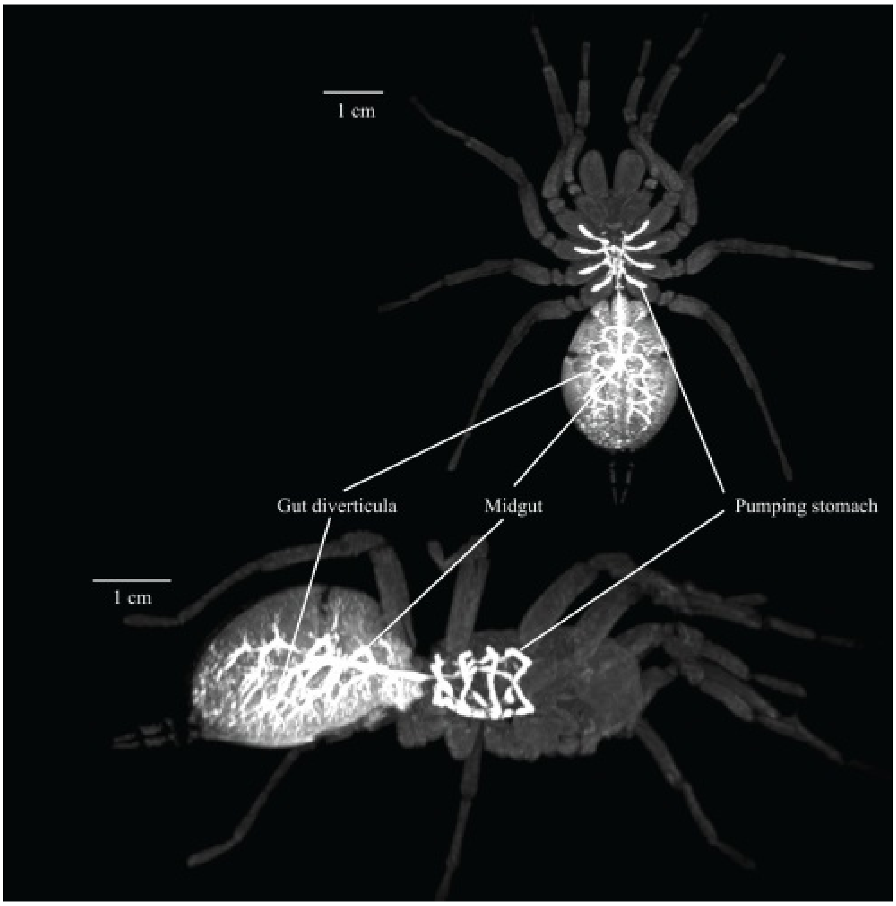



Figure 3.26: ■ MR scan of a whiteknee tarantula gastrointestinal tract. The structures are enhanced due to feeding the tarantula a contrast medium-filled cockroach [LHW+11].

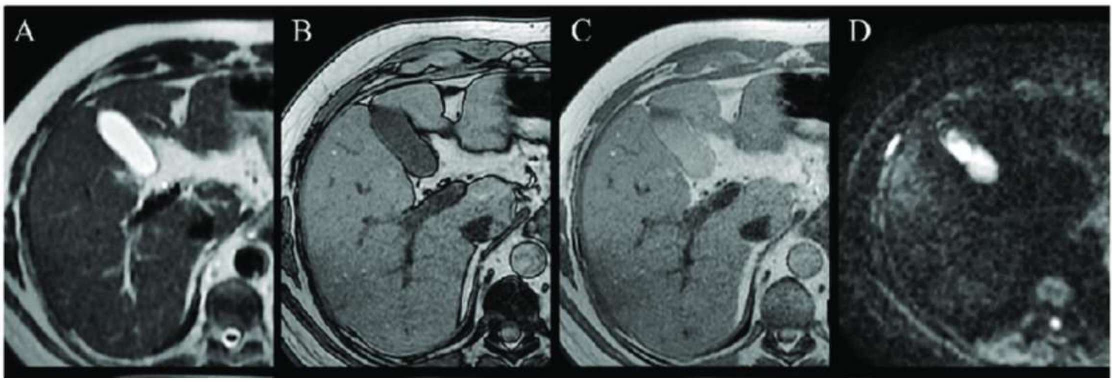



Figure 3.27: ■ MR scans of the abdomen using different pulse sequences: (A) T2-weighted, (B) T1-weighted out-of-phase, (C) T1-weighted in-phase, (D) Diffusion-weighted imaging. The structure shown on the left is the liver because the CT image was captured in the direction from feet to head. Edited from [RCF+21].

### Uncertainties in Transformation and Processing

This pipeline step includes registration, image processing, data preparation including shape estimation, flow computation, and others, and simulation steps from the pipeline proposed by Ristovski et al. [RPHL14]. Registration aims to align coordinate systems of multimodal scans, follow-up scans (recorded at different times), and scans of different patients [KSM+09]. Then the aligned images can be fused (Figure 3.2). Segmentation aims to classify each voxel among segments in a binary way, for example, “prostate” vs. “not prostate”. Other segmentation methods estimate the probability of a voxel belonging to each segment [RPHL14]. Table 3.7 summarizes the cases discussed in this subsection.

| **Types of Uncertainty in Transformation and Processing** | |
|---|---|
| Error | Algorithm failures |
| Imprecision | Imprecision of calculations executed by algorithms |
| Non-specificity | Overspecializations of algorithm |
| Incompleteness | Partial algorithm convergence |
| Subjectivity | Parametrization of algorithms |

Table 3.7: Uncertainty type cases in transformation and processing step of the medical imaging data visualization pipeline.

**Error**. At the transformation processing step, different mathematical models prepare the imaging data for visualization. These models support image registration, segmentation, and other complex computations and perform simulations. Errors in any of those mathematical models produce uncertainty. For example, a chosen segmentation algorithm can be incorrect and lead to over- or undersegmentation of an organ of interest [SMH10] (Figure 3.28).

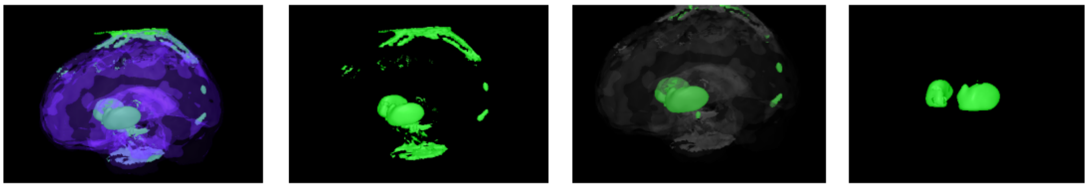
Similarly, a registration model can fail to align coordinate systems, thus leading to incorrect fusion of images from different modalities. Even though registration compensates for patient motion, severe motion artifacts may cause misregistration (propagated uncertainty) [BGC22]. Errors can appear in statistical methods of calculation of neuron activity combining images from different modalities inside a voxel (functional MR), reconstruction methods of fibers in diffusion imaging (diffusion-weighted MR), or statistical methods of simulation of treatment outcomes [RPHL14].

**Imprecision**. Registration imprecision originates from different spatial resolutions of images from different modalities (Figure 3.29). For instance, PET has a limited spatial resolution (4–5 mm) in comparison to MR spatial resolution (1 mm) [S+17]. Therefore, such registration delivers uncertainties into the fused image. Another example is the uncertainty of organ or tumor size computation during segmentation [GSWS21]. Additionally, the calculation of wall shear stress at the aneurysm in blood flow 4D MR image can be imprecise [RPHL14].

■ **Non-specificity**. Some mathematical models are so specialized that they cannot be used for all scans in general and require assigned images originating from specific medical imaging devices and protocols. Anatomical and pathological variability may pose a threat to the correct or accurate operation of a mathematical model [LZXP20] (Figure 3.30). Also, some models are better at the detection of some abnormalities than others, due to the abnormalities' diverse nature, and some models cannot account for organ motion. To reduce cardiac motion artifacts, each CT image slice must be acquired at a specific time of the heart cycle, therefore, the volumetric reconstruction contains slices from consecutive cycles [SSHW05].

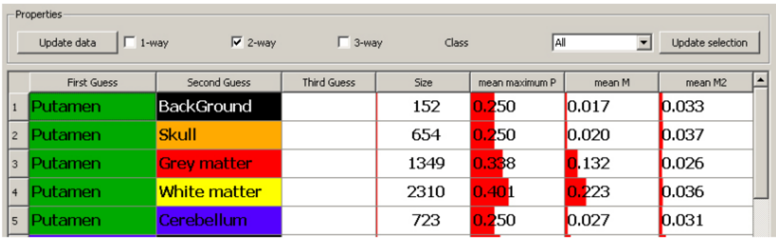
■ **Incompleteness**. This uncertainty type is related to the situation when a model algorithm does not converge to a result (partial convergence). For example, accurate segmentation of the prostate and surrounding organs are risk is essential for radiotherapy. However, the segmentation model, shown in Figure 3.31, provided an incomplete segmentation and left some areas undetected [RMB+16].

■ **Subjectivity**. Uncertainties coming from the parametrization of models belong to this type. For example, image operations including contrast enhancement, edge detection, and color correction can have different outcomes depending on a subjective choice of parameters [GSWS21]. Similarly, the choice of border conditions and parameters in a mathematical model can bring uncertainties in the outcome (Figure 3.32).



(a) (b) (c) (e)

(f)




| | First Guess | Second Guess | Third Guess | Size | mean maximum P | mean M | mean M2 |
|---|---|---|---|---|---|---|---|
| 1 | Putamen | BackGround | | 152 | 0.250 | 0.017 | 0.033 |
| 2 | Putamen | Skull | | 654 | 0.250 | 0.020 | 0.037 |
| 3 | Putamen | Grey matter | | 1349 | 0.338 | 0.132 | 0.026 |
| 4 | Putamen | White matter | | 2310 | 0.401 | 0.223 | 0.036 |
| 5 | Putamen | Cerebellum | | 723 | 0.250 | 0.027 | 0.031 |

Figure 3.28: ■ Oversegmented putamen treated by segmentation editing. (a) 3D PET image, (b) oversegmented putamen, (c) voxels with the white matter as the second best choice in the segmentation editing widget, (d) classification after two editing iterations, (f) segmentation editing widget. Edited from [SMH10].

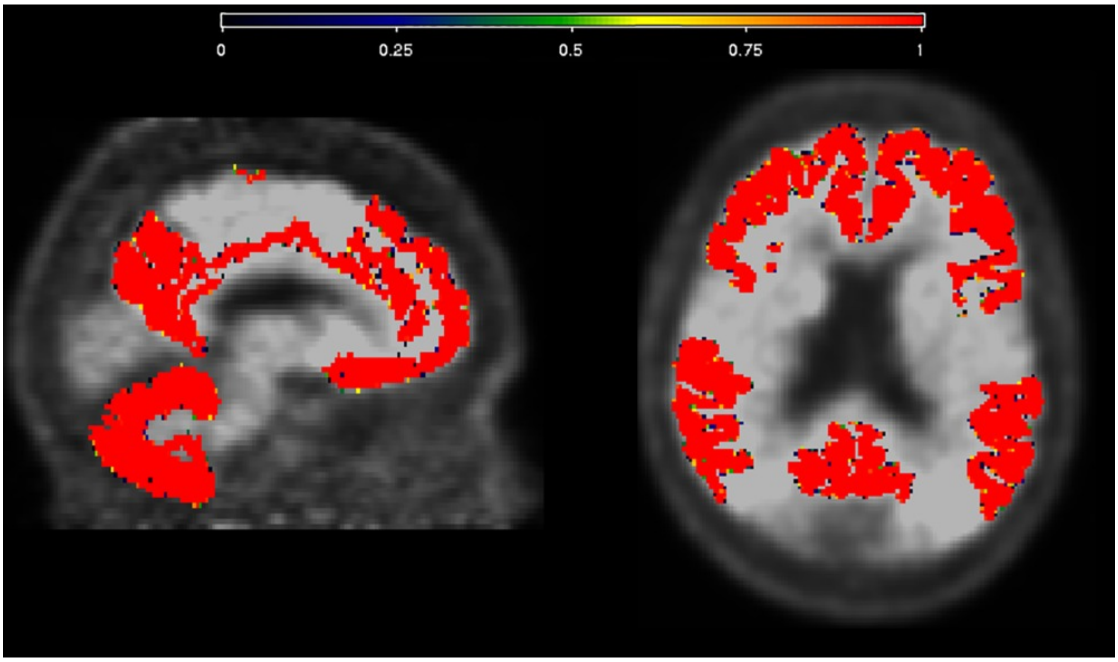



Figure 3.29: Heat map of voxels registration outcomes. Red voxels (corresponding to 1 in the color legend) were correctly located in 50 perturbed PET-MRI registrations. Voxels with other colors varied their position or were absent across perturbations, indicating PET-MRI rigid registration imprecision [S+17].

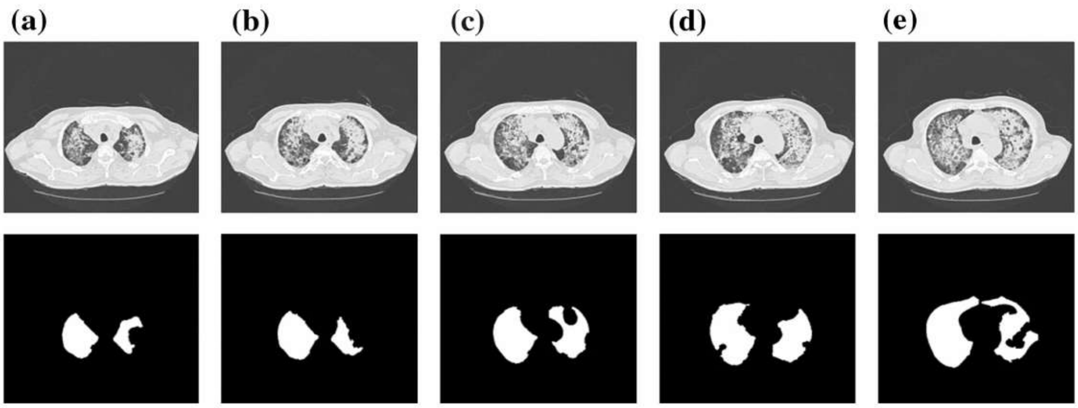



Figure 3.30: Results of lung segmentation that does not account for pathological variability. Above: CT scan slices, bottom: segmentation results [LZXP20].

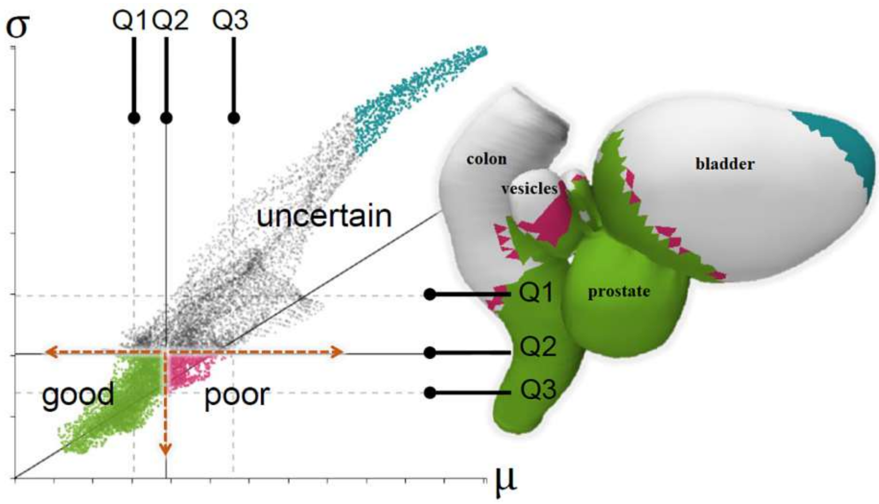



Figure 3.31: ■ Confidence scatter plot of prostate, bladder, vesicles, and rectum segmentation outcome. The scatter plot is based on mean error $\mu$ and the standard deviation $\sigma$. Segmentation outcomes with high standard deviation correspond to areas left undetected (cyan). Edited from [RMB+16].

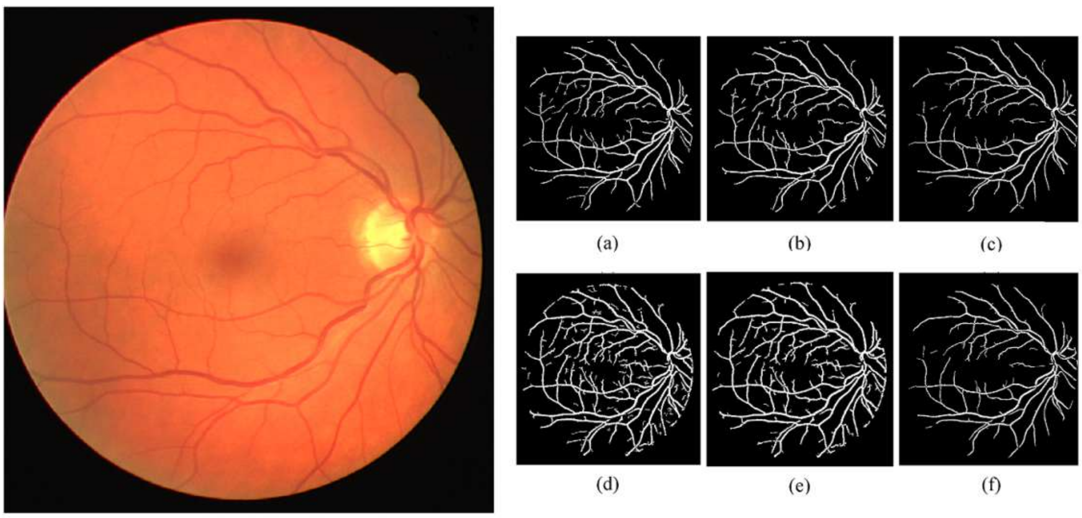



Figure 3.32: ■ Results of different segmentation techniques (a–f) for retinal vasculature indicating the influence of parametrization on the segmentation accuracy. Edited from [BAKS16].

## Uncertainty in Representation

This medical imaging data visualization pipeline step coincides with the visualization step from the pipeline proposed by Ristovski et al. [RPHL14]. It includes interpolation and direct volume rendering. Moreover, some cases are taken from flattening techniques. Table 3.8 summarizes the cases discussed in this subsection.

| Types of Uncertainty in Representation | |
|---|---|
| Error | Rendering artifacts |
| Imprecision | Reconstruction imprecision |
| Non-specificity | Specialization of illumination, flattening techniques |
| Incompleteness | Incomplete reconstruction |
| Subjectivity | Adjusting parametrization of representation (opacity, color, depth) |

Table 3.8: Uncertainty type cases in representation step of the medical imaging data visualization pipeline.

**Error**. The idea of direct volume rendering is to get a 3D representation of data. The calculations are based on the following rays through the data and accumulating optical properties. The chosen interpolation model in direct volume rendering might bring discontinuity artifacts [RPHL14]. Discretization of the volume rendering integral introduces a sampling artifact looking like "onion rings" (Figure 3.33). Also, illumination, shading, and scattering models can introduce rendering artifacts [JSYR14].

**Imprecision**. In Figure 3.34, the volume rendering steps are shown. The volume rendering equation integrates over the ray path, considering the sample contribution based on material properties. The sampling (selecting positions along the ray) method used in the integral influences the time and accuracy of rendering [ZEP11]. Volumetric data used for rendering is encoded in voxels, so the data are discrete. Interpolation of the discrete data is used to reconstruct the information between slices, thus creating a smooth continuous field [Dri19]. The low spatial resolution brings more uncertainties to the borders between voxels, where data are interpolated [RPHL14]. Also, exact geodesic distance (shortest-path distance) calculation is important for medical flattening techniques. Imprecision of geodesic distance calculation can result in error-prone mapping with large distortions [KMM+18].

**Non-specificity**. Some visualization techniques are highly specialized, which can cause a lack of distinction for some body structures. For example, some illumination models can support material properties and provide much realism to the visualization, while others use simple illumination techniques [JSYR14]. Different flattening techniques are originally designed for data from specific medical modalities and for different targets, including the circulation system (aneurysm maps, vessel flattening), the brain, the colon, and others (Figure 3.35). However, some techniques can be transferable to other organs, but most of the techniques lack generalizability [KMM+18].

■ **Incompleteness**. During rendering, issues can lead to incomplete reconstruction (Figure 3.36). For example, incomplete reconstruction can happen while using the early-ray termination method. When a ray accumulates sufficient opacity values, samples further on the ray are neglected and do not contribute to the rendering. This method reduces the time of rendering. However, with an insufficient threshold, the rendering can lose quality and terminate too early, missing out on some information [LR06].

■ **Subjectivity**. Subjective preferences for a rendered model appearance give rise to this type of uncertainty. For example, they can influence the parametrization of representation techniques, leading to subjective adjustments of opacity, transfer function (color-mapping), and depth-simulating methods that can make the visualization appealing to the visualization creator. In Figure 3.37, several options for creating "shadows" in the rendering of intestines are shown.

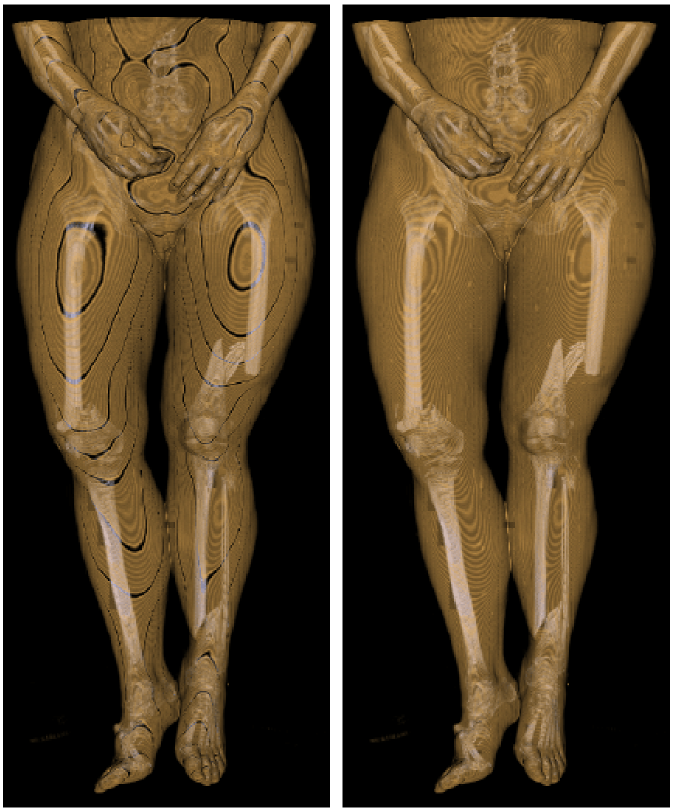


Figure 3.33: ■ Comparison of interpolation methods. The left one (no interblock interpolation) contains block boundary artifacts, removed by the maximum distance method on the left. Sampling artifacts present in both images [LLY06].

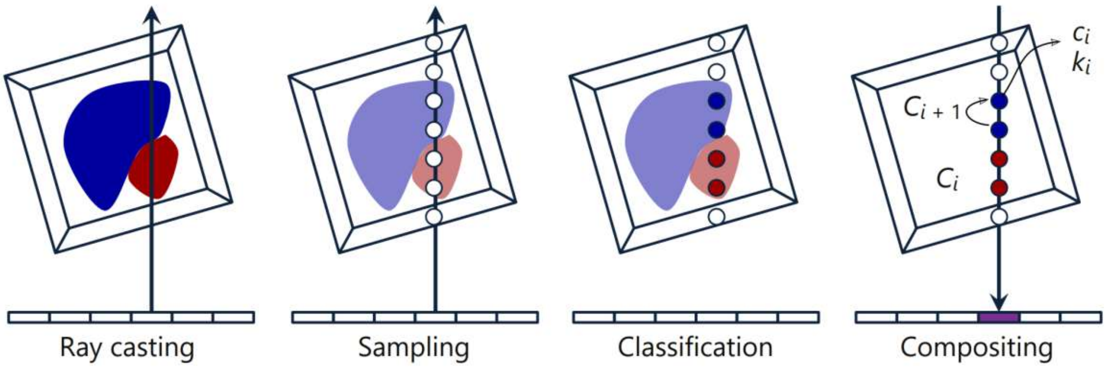



Figure 3.34: The volume rendering pipeline contains four steps: ray casting, sampling, classification, and composing. Samples along the ray contribute to the composed value depending on their material properties. The interpolation method (type and order) of all sample values along the ray results in uncertainties. Edited from [SPBR20].

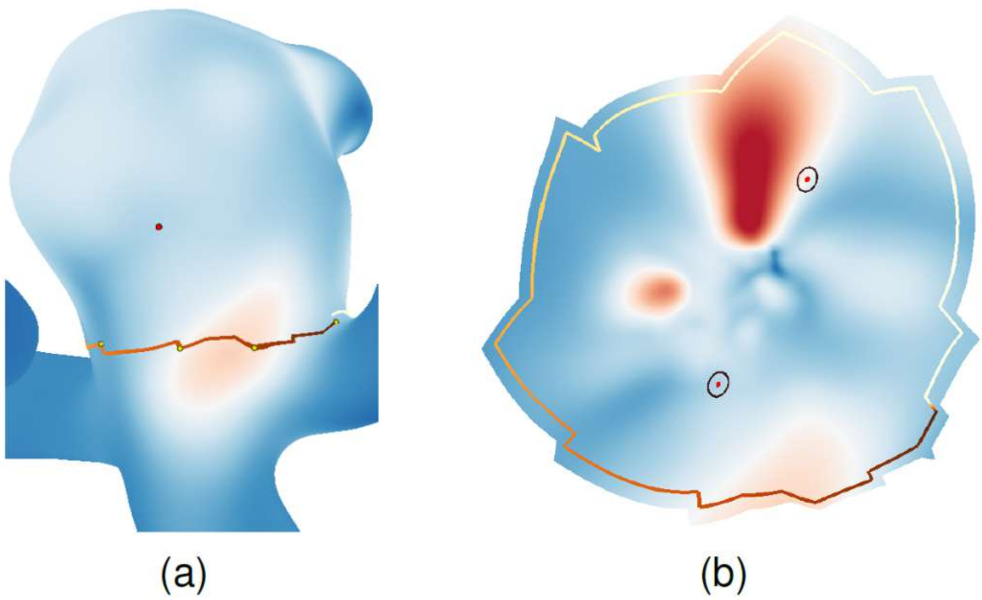



Figure 3.35: Flattening technique used for aneurysm visualization. The aneurysm dome is cut at the aneurysm's neck (ostium) region (a). The surface of the dome is unfolded (b). The red color represents the areas with a high risk of rupture (wall thickness dependent) [MVB+16].

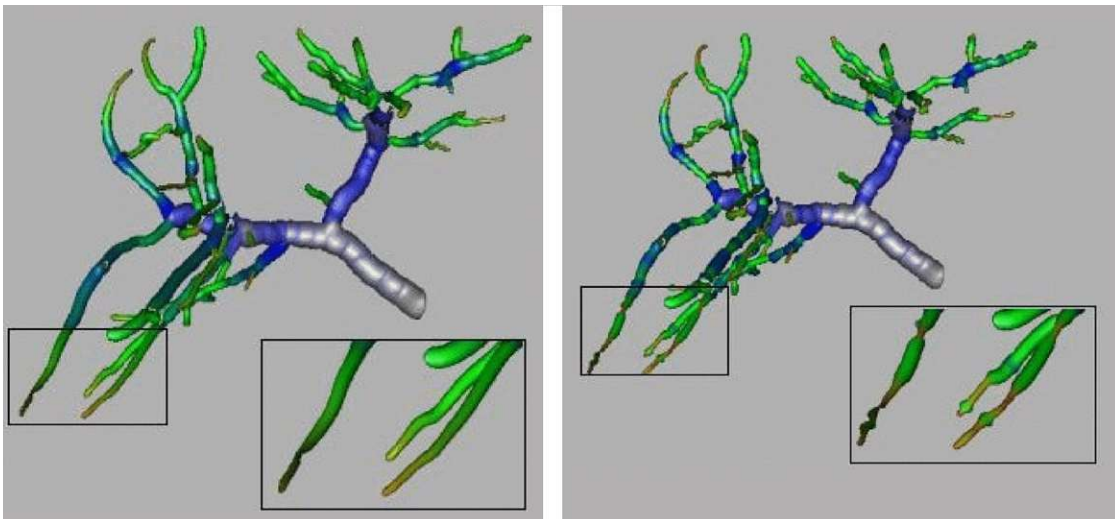


Figure 3.36: The smoothing of vessels (left). Due to the discontinuous nature of radiological data, the rendering contains a disrupted vessel representation without smoothing (right). Color coding reflects the diameter of a vessel [HPSP01].

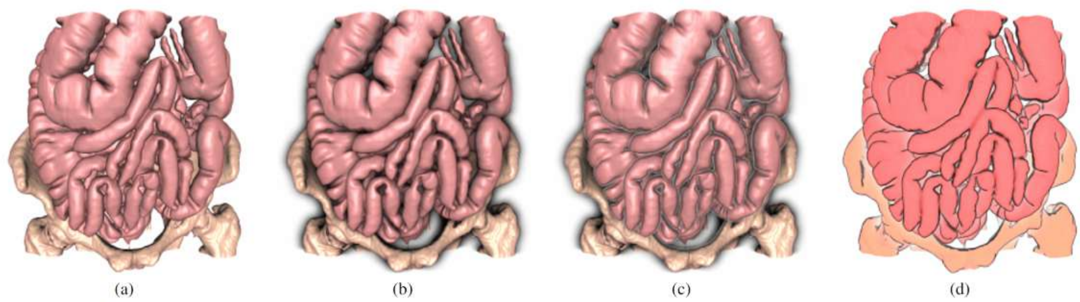



Figure 3.37: Volume renderings of intestines: (a) without halos, (b) with shadow-like and (c) with semi-transparent halos, (d) without shading, but with a smooth white and black contour [BG07]. Shape and depth perception might be affected by different representations.

### Uncertainties in Human Processing and Interaction

This medical imaging data visualization pipeline step coincides with the human interpretation step from the pipeline proposed by Ristovski et al. [RPHL14]. These uncertainty types are related to how the final visualization is perceived by a user at the end of the visualization pipeline. Table 3.9 summarizes the cases discussed in this subsection.

| **Types of Uncertainty in Human Processing and Interaction** | |
|---|---|
| Error | Non-invertibility of color maps |
| Imprecision | Inverting a color in a color legend to a scalar value |
| Non-specificity | Uncertainty of depth, material perception, projection distortion |
| Incompleteness | Lack of guidance, knowledge gap |
| Subjectivity | Decision-making bias |

Table 3.9: Uncertainty type cases in human processing and interaction step of the medical imaging data visualization pipeline.

**Error**. Non-invertibility of color legends can bring perceptional uncertainties to people with conditions like color-vision deficiency and color blindness. These conditions impair the ability to perceive some color legends. For example, Figure 3.38 compares the color legends that are suitable for color-blind people and people with color-vision deficiency with the rainbow (jet) color legend. The rainbow legend is non-invertible because the same values can correspond to colors that look the same for a color-blind person.

**Imprecision**. The low resolution of scales may bring a perceptional uncertainty when inverting a color in a color map to a scalar value. For example, in Figure 3.39 colors encode the cortical thickness of the hip joins. It could be hard to identify the exact value, represented by a color legend. For instance, the orange color may encode a value between 1 mm and 1.3 mm.

**Non-specificity** Some subjective uncertainties might be triggered also by a lack of specificity. The depth and material characteristics are defined by the capabilities of the rendering method (Figure 3.40). Their perception may vary between different users. Also, projection distortion appears when a 3D object is projected onto a 2D map. This happens not only in flattening techniques. For instance, an X-ray scan is a 2D projection of an object, which causes loss of depth information [RPHL14].

**Incompleteness**. Lack of user guidance in a visualization may cause uncertainties. The guidance serves to fill a knowledge gap that shows the difference between what should be known to analyze a visualization and what a user actually knows [CAA+20]. The guidance can also highlight areas to explore and get insights from. It is possible to actively make a user focus on a feature using attention guidance techniques, which include selective camera angles, highlighting, adjusting contrast, and using blur to reduce the noise of context objects [EMW23].

■ **Subjectivity**. Interpretation of a visualization begins with obtaining information from a visualization (usually quantitative data) and ends with a decision [RPHL14]. Interpretation fallacies give rise to this type of uncertainty and decision-making bias in clinical routine. Dealing with medical image data on a daily basis or very rarely can affect the data perception [GSWS21]. For instance, there are measures of variability in measurements of the same user and between users.

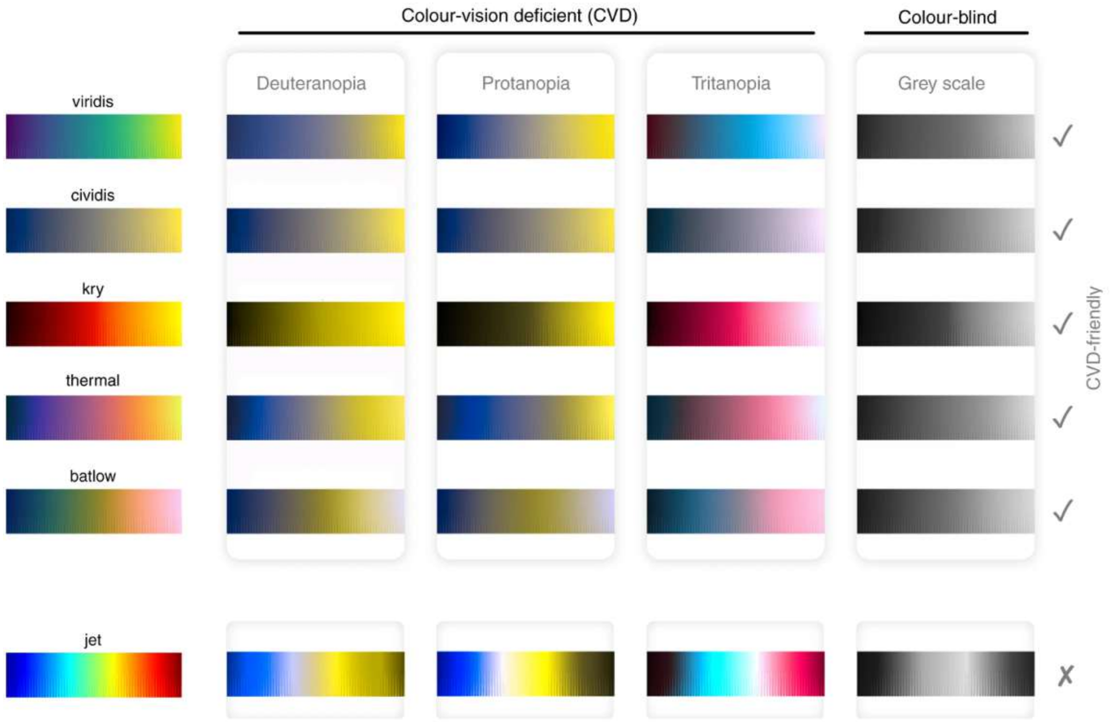



Figure 3.38: ■ Comparison of color legends and the way they are seen by color-blind people and people with color vision deficiency [CSH20].

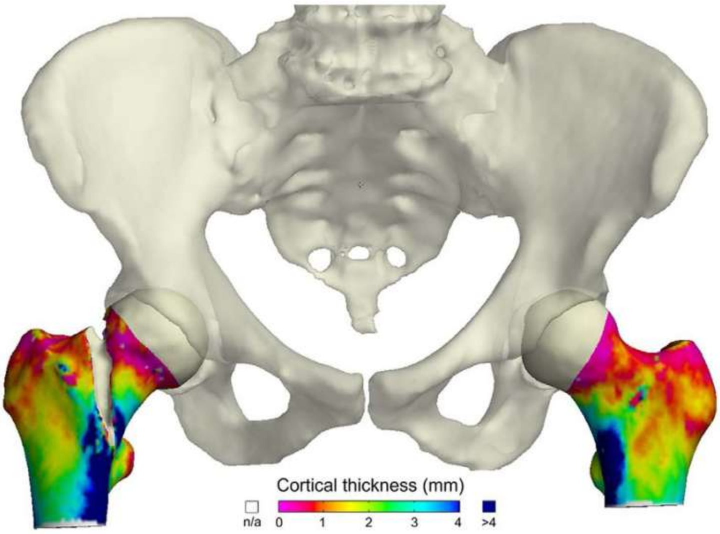



Figure 3.39: ■ Pelvis and femora CT scan rendering of a patient with osteoporosis that caused a hip fracture. The color legend encodes the cortical thickness and poses perceptional uncertainty when trying to obtain an exact value for a specific color [PTM+12].

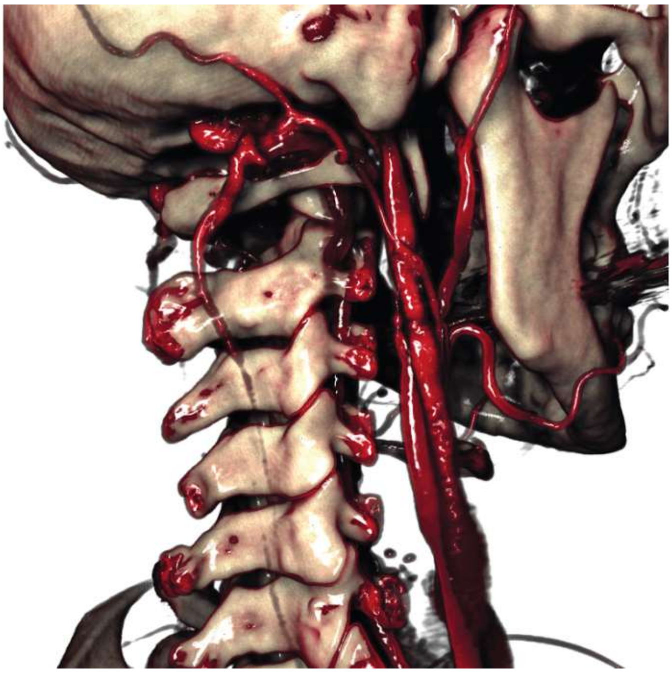


Figure 3.40: ■ CT scan rendering with specular reflection of blood vessels and isotropic phase function used for bones to minimize the uncertainty of material perception [JKRY12].

## 3.4 Intentional Misleading Elements in Medical Visualization

Our work aims to raise awareness about the aforementioned types of uncertainties and the potential intentions behind them. The introduction of uncertainty can be intentional or accidental. In this subsection, we assess the probability of uncertainty type being intentionally introduced, summarized in Tables 3.10 and 3.11 that refer to non-imaging and imaging data visualization pipelines, respectively.

| NON-IMAGING DATA | ■ Err. | ■ Impr. | ■ Non-sp. | ■ Incom. | ■ Subj. |
|---|---|---|---|---|---|
| Data Acquisition | ✓ | ✓ | ○ | ✓ | ○ |
| Transform. & Processing | ✓ | ✓ | ✓ | ✓ | ✓ |
| Representation | ✓ | ✓ | ✓ | ✓ | ✓ |
| Human Process. & Interaction | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 3.10: Introduction of uncertainties to non-imaging medical data at different visualization pipeline steps. Legend: ✓: intentional, ○: not known intention.

When acquiring non-imaging medical data, we make an assumption that the data are not fabricated and consider the source of uncertainties to be due to erroneous self-reporting. Patients may lie to appear in a good light, to avoid embarrassment or negative consequences, and to achieve secondary benefits, like special medication or payments [PS09]. For simplicity, let's take a patient who smokes and has to fill in a survey as a patient self-reporting. Maybe due to being ashamed of smoking, this patient would lie. For example, the patient can write that he does not smoke (■ error). Also, he can write that he smokes two cigarettes a day, but he smokes a pack per day (■ imprecise). Finally, he can just leave the space in the survey next to the question about smoking blank (■ incompleteness). Cases of non-specificity and subjectivity uncertainty types are not known.

During the transformation and processing of data, it is possible to make the data look as if it supports an idea. Using a more appealing imputation function (■ error) can add some artifacts and the wrong statistical test can prove a hypothesis, which can be used for falsifying research [KAK22](■ non-specificity). Incorrect extrapolation methods (■ imprecision) can be used, for example, in low-level mutagenic risks estimations [Bro90]. Also, normalization can be missing, thus making comparisons invalid, (■ incompleteness), and the data can be selective, showing only the "supporting the theory" part of data (■ subjectivity).

At the representation step, all four types of uncertainties can be intentionally introduced. A normal scale instead of logarithmic (■ error) in a graph may increase the persuasiveness of the idea. Inconsistent binning size (■ imprecision) may give more emphasis to a specific year. Data of a smaller magnitude can be obscured by larger values of another sample that is being simultaneously presented (■ non-specificity). The same way works with the truncated axis (■ incompleteness) and setting an arbitrary threshold (■ subjectivity).

Regarding human processing, Lo et al. propose that the creators of intentionally misleading visualization, draw attention and assign meaning to unexplained salient features in well-designed charts [LGS+22]. Additionally, design violations can influence the perception, for example, actively mislead by introducing artifacts by color legends (■ error) [BI07] with low resolution (■ imprecision). Additionally, to falsify results, it is possible to hide distributions with a bar chart (■ non-specificity). Hidden premises (■ incompleteness) and text polarity (■ subjectivity) also can serve intentional purposes [LPLK22].

| IMAGING DATA | ■ Err. | ■ Impr. | ■ Non-sp. | ■ Incom. | ■ Subj. |
|---|---|---|---|---|---|
| Data Acquisition | × | × | × | × | × |
| Transform. & Processing | ✓ | ✓ | ○ | ✓ | ✓ |
| Representation | ○ | ○ | ○ | ○ | ○ |
| Human Process. & Interaction | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 3.11: Introduction of uncertainties to medical imaging data at different visualization pipeline steps. Legend: ✓- intentional, ○ - not known intention, × - not intentional.

We assume that during medical imaging data acquisitions, it is almost impossible to intentionally introduce uncertainties. Being located in hospitals or medical centers due to their size and costs, the imaging machines use approved protocols and medical experts follow medical ethics and work for the benefit of the patient's health. However, medical imaging data can be vulnerable when transferred over the internet, due to increased risks of image manipulation especially with the rise of artificial intelligence [DD22].

Manipulations in the transformation and processing visualization pipeline step, include copy-move forgery (■ error), blurry effects (■ imprecision), removing an area (■ incompleteness), adjusting contrast, sharpening, and brightness (■ subjectivity) [DD22]. While some of these cases are shown in Figure 3.41, cases of non-specificity are not known. More advanced methods use generative adversarial networks (GANs). For example, such methods could be employed to mimic histopathological images [AZP+20] and inject specific pathological conditions in chest X-rays and retinal images [MPB+20]. The resulting forged image can lead to the wrong diagnosis of a patient's disease. *Why would somebody do it?* The attackers' motivation can be political, ideological, attention-, revenge- or money-related when asking for ransom for the original medical data as proposed by Mirsky et al. [MMSE19] (Figure 3.42).

In the representation step, creators of medical imaging data visualizations follow qual-

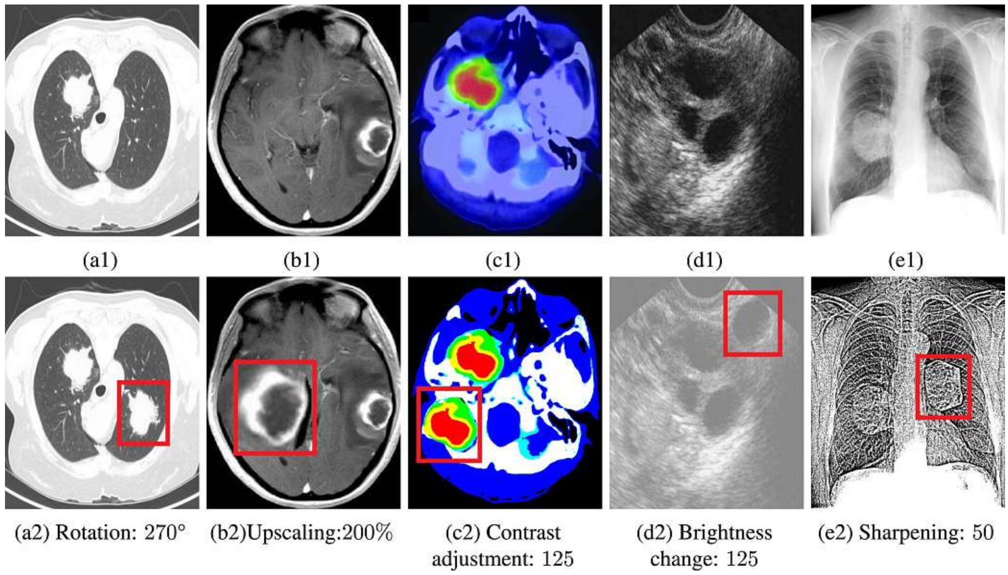



Figure 3.41: Copy-move forgery in medical images. Top: original CT, MRI, PET, ultrasound, and digital X-ray scans. Bottom: forged images with different types of attacks. Red colored rectangles indicate tampered regions [DD22].

+ : Add Evidence
− : Remove Evidence
± : Either

● : Target Effect
○ : Side Effect

| | | | Goal: Steal Job Position | Affect Elections | Remove Leader | Sabotage Research | Falsify Research | Hold Data Hostage | Insurance Fraud | Murder | Terrorize |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Motivation | | Ideological | | | + | | | | | | ± |
| | | Political | | + | + | | | | | − | |
| | | Money | + | | + | + | − | ± | + | | |
| | | Fame/Attn. | + | | + | | ± | | | | |
| | | Revenge | + | | + | | | | | − | + |
| Effect | Physical | Injury | ○ | ○ | ○ | | ○ | ○ | | ○ | ● |
| | | Death | | | | | | ○ | | ● | ● |
| | Mental | Trauma | ○ | ○ | ○ | | | ○ | | ○ | ● |
| | | Life Course | ● | ● | ● | ○ | | | ○ | ○ | ● |
| | Monetary | Cause Loss | ○ | | ○ | ● | | ○ | | ○ | ● |
| | | Payout | ● | | | ○ | ● | ● | ● | | |
| Attack Type | | Untargeted | | | | X | | X | | | X |
| | | Targeted | X | X | X | X | X | | X | X | |

Figure 3.42: The motivation behind attacking medical imaging data by manipulation. Targeted attacks aim for a specific patient, while untargeted attack has no specific target patient [MMSE19].

ity criteria: expressiveness, effectiveness, and appropriateness when representing data [MA14]. Intentionally a visualization can be tweaked and appear misleading, but it would be a violation of effectiveness (showing exactly the information contained in the data). A misleading representation can originate from a visualization designer's dilemma [LGS+22], but this is not necessarily the result of a bad intention. Rather it could be the result of, for example, a compromise for distortion during the flattening of an organ [KMM+18].

For the human processing and interaction step, we consider the following cases. Intentional use of improper color legend (■ error) can actively mislead users by introducing artifacts [BI07]. Also, the low resolution of the color legend can be used to conceal the exact values (■ imprecision), due to publication bias indicating that research with positive results has a higher chance of being published [KAK22]. Employing projection distortion can influence the perception of distances, like the diameter of vascular structures or tumor's shape (■ non-specificity) [KMM+18]. Moreover, not communicating the data uncertainty (■ incompleteness) is used to present a visualization as accurate and true at all conditions [Hul19]. As the context influences the interpretation of the accompanying visualization, it can serve for bad intentions (■ subjectivity) [KLK18].

## 3.5 Combating Misleading Visualizations through Storytelling and Gamification

We investigate a solution against misleading representations (due to intentionally or unintentionally introduced uncertainties) through a narrative approach for educational purposes. We anticipate that a narrative approach that integrates storytelling and gamification aspects may appeal to the general population by presenting data in a simple, engaging, and clear way [MGS+21]. Moreover, interactivity within the narrative approach enhances memorability [MGS+22] and can be enriched with gamification techniques to support learning [HNGC20].

By combining storytelling with gamification, we build an educational game that supports visualization audiences to recognize and avoid misleading visualizations in health care. During the game creation, we aim to make misleading elements accessible to the general population by using laypeople terminology instead of scientific and medical terms, defining a clear objective, developing an intuitive and appealing user interface (UI), and using rewarding strategies for building engagement.

### 3.5.1 Storytelling

In our approach, we use eight simple examples as proof-of-concept: four imaging and four non-imaging stories including visualizations with artificially introduced fallacies. They are commonly occurring and indicative of misleading visualizations in health care. The stories contain a misleading visualization and accompanying text, to make them look like

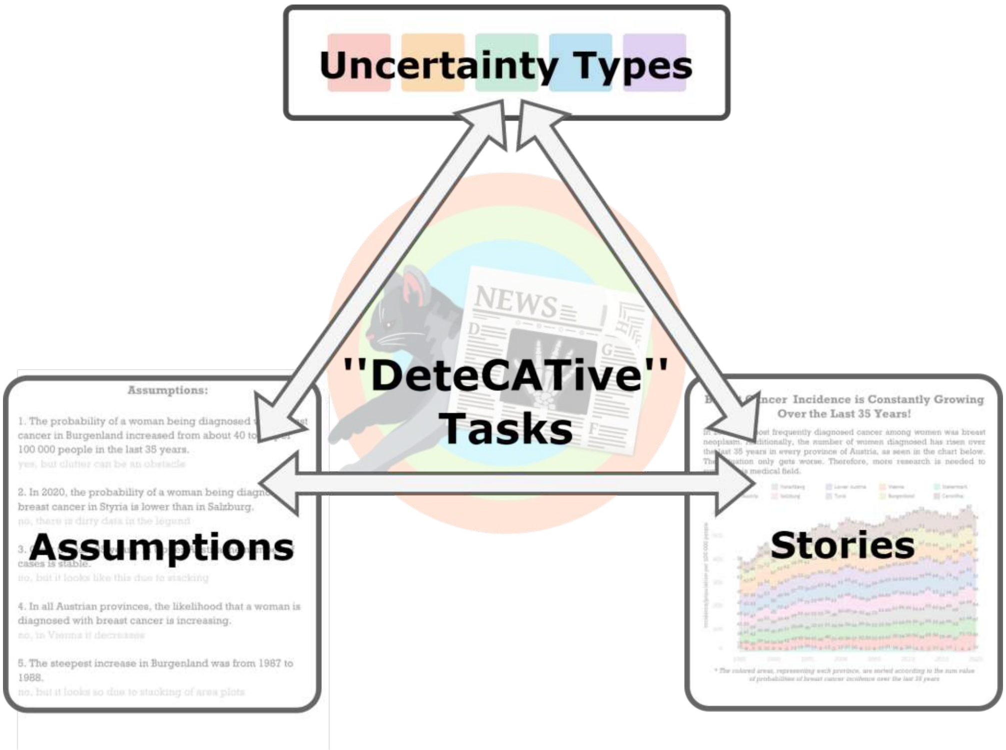



Figure 3.43: The cycle of interconnection between stories, uncertainty types, and assumptions used within our proposed approach.

a newspaper snippet or an advertisement found on the street. For each story, we include a list of assumptions and explanations assembling them into tasks (Figure 3.43). The learning should happen while assessing these assumptions, whether they are true or false, thus a misleading element in the story can come to light by reading the explanations.

Behind the creation of every misleading visualization lies an exhausting brainstorming, aimed to generate the main trick and motive, referred to as “intention”. The ideas for non-imaging visualizations were based on the data found in publicly available sources. However, creating medical imaging visualization that intentionally misleads proved to be complicated. For example, the idea for one task came from various inspiration sources. At the beginning of the visualization design, we used MeVisLab to generate some noisy areas in a brain MR scan to imitate cancer. Then we remembered a 16th-century portrait of Gregor Baci located in Schloss Ambras in Innsbruck (Warning: the portrait is not for the faint-hearted since it depicts a facial piercing injury with a lance), and an accident of Phineas Gage, 19th-century railroad constructor in the USA, who survived when an iron rod impaled his head at work.

Another indicative source of inspiration was the summer practice in the High Field MR Centre located on the Medical University Campus at the General Hospital of Vienna in 2022. The summer practice was under the supervision of Prof. Dipl.-Phys. Dr. Andreas Berg, who told real-life gruesome stories about not following the MR safety rules. Finally, all the inspiration sources mentioned above combined into the idea of a construction worker impaled with a magnetic metal rod and scanned with the rod in an MR machine, which is, of course, impossible.

After creating visualizations based on ideas of tricks and intentions, we composed the accompanying text, thus creating the stories. Within the text, our aim was to hide (similarly to a real-life scenario) between the lines the intent, provide context, and give some hints or details about the functionality of a medical imaging machine. For each story, multiple assumptions were written, each relating to misleading elements within a visualization. The explanations for each assumption serve an educational purpose and include visual and text components. The textual explanation requirements are short, precise, and not ambiguous. The visual component supplements the textual component. To evaluate the tasks, a small focus group of three TU Wien students and two students from other Austrian universities was asked for a pilot trial. Their feedback was valuable, leading to the reformulation of some assumptions to increase clarity. Moreover, we got confirmation that some misleading tricks are hard to recognize.

The following paragraphs present the tasks used in the game. The task description includes the idea, data, intention, and uncertainty types. These descriptions are accompanied by figures containing the stories, assumptions with short answers, and types of uncertainties across the visualization pipeline. In the game, we explain the origin of the misleading elements and deliver the potential motivation for making such visualizations. Due to the game being oriented toward the general population, we did not include information about the uncertainty types.

**Task 1** uses a cumulative representation of COVID-19 data in Austria: deaths, recoveries, and new cases. When the cumulative nature of the representation is not noticed, the bar chart may look like the numbers never stop growing. The story shown in Figure 3.44 (right) aims to support a public health initiative to continue wearing masks in public transport. Assumptions to the story in Figure 3.44 (left) refer to the intentionally introduced uncertainty types: error (mistyped year), imprecision (large binning interval), non-specificity (data of different magnitude), and subjectivity (wrong choice of chart and misleading title) tracked over the visualization pipeline in Figure 3.45.

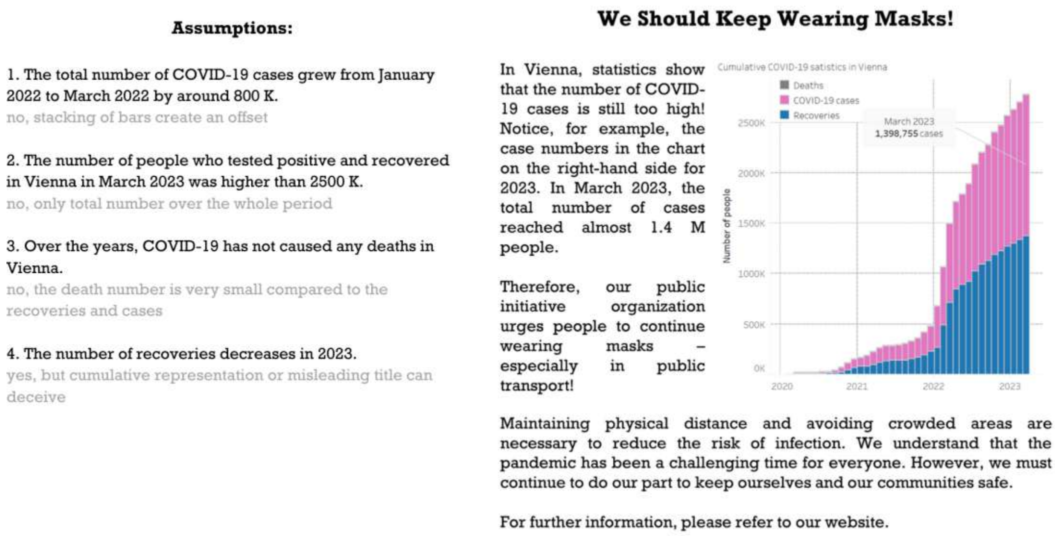



Figure 3.44: Task 1 story with COVID-19 statistics used to support public health initiatives. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | ■ Error: mistyped year of the data row [A4]<br><br>* invisible when data are represented as bar chart |
| **Transformation/ Processing** | no |
| **Representation** | ■ Imprecision: large binning interval (month) [A4]<br><br>■ Non-specificity: data of different magnitude (deaths)[A3]<br><br>■ Subjectivity: wrong choice of chart: cumulative representation [A2] and stacked bar chart [A1] |
| **Human Processing** | ■ Subjectivity: power of words (misleading story title that cases keep rising) [A4] |

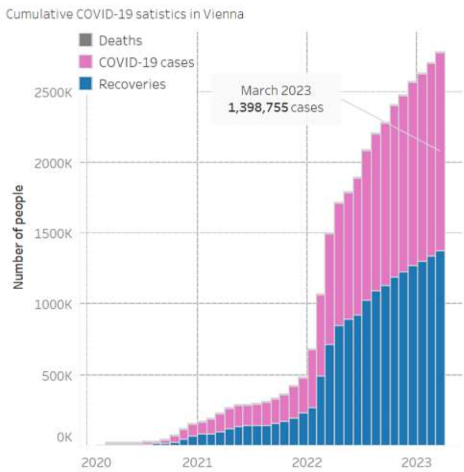



Figure 3.45: Uncertainty types in the misleading visualization of task 1.

**Task 2** uses a volume rendering of a CT scan of a human abdomen. An undersegmented and separately colored liver imitates a severely damaged fatty liver condition. The story shown in Figure 3.46 (right) aims to create a warning against alcoholism prompting the viewer to seek help. Assumptions to the story in Figure 3.46 (left) refer to the uncertainty types: error (under segmentation), non-specificity (physics of the imaging modality, object orientation), incompleteness (knowledge gap, lack of guidance), and subjectivity (choice of transfer function) tracked over the visualization pipeline in Figure 3.47.


Figure 3.46: Task 2 story with undersegmented liver used to create a health warning. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | Non-specificity: physics of the imaging modality [A4] |
| **Transformation/ Processing** | Error: under segmentation [A5] |
| **Representation** | Non-specificity: orientation of the object [A1]<br>Subjectivity: choice of transfer function [A3] |
| **Human Processing** | Incompleteness: knowledge gap [A1,4]<br>Incompleteness: lack of guidance [A2] |

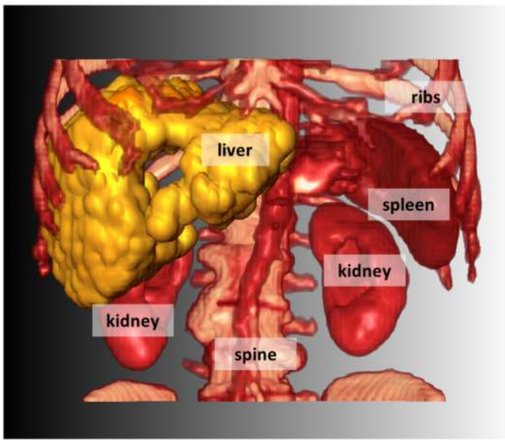



Figure 3.47: Uncertainty types in the misleading visualization of task 2.

**Task 3** uses a color map showing melanoma (malignant skin cancer) cases across Austrian provinces. A tweaked color legend and values not normalized to the population of provinces bring forward Salzburg as the healthiest province. The story shown in Figure 3.48 (right) aims to persuade people to buy real estate in Salzburg. Assumptions in Figure 3.48 (left) refer to the uncertainty types: error (color violation), imprecision (low scale resolution), incompleteness (missing data and normalization), and subjectivity (inappropriate axis range) tracked over the visualization pipeline in Figure 3.49.

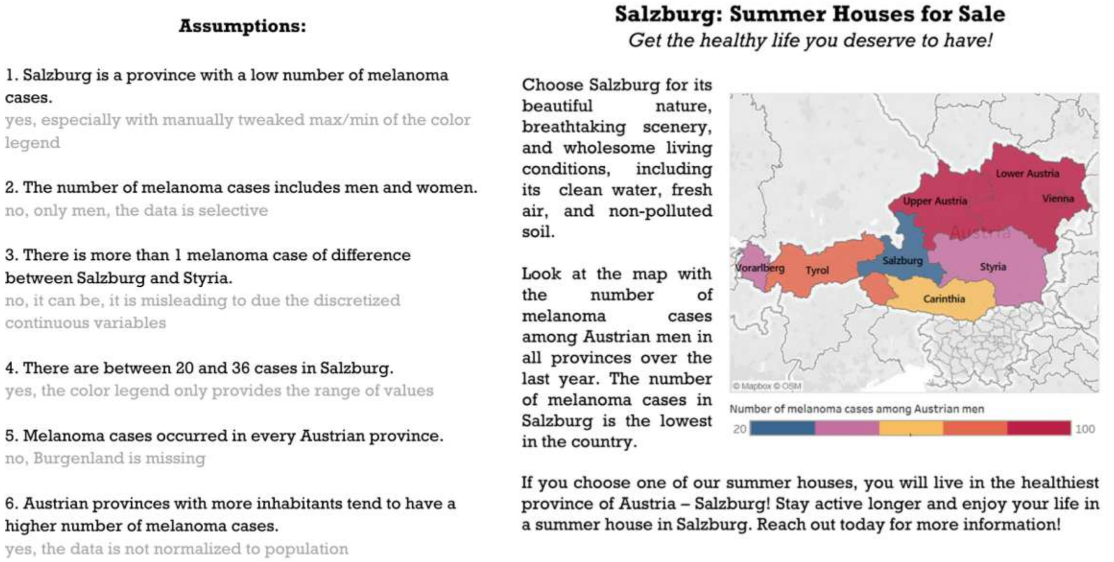



Figure 3.48: Task 3 story with melanoma incidence in Austrian provinces used to persuade people to buy real estate. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | ■ Incompleteness: missing data (Burgenland) [A5] |
| **Transformation/ Processing** | ■ Incompleteness: missing normalization [A6]<br>■ Subjectivity: selective data, only male data are used [A2] |
| **Representation** | ■ Error: color violation, discretized continuous value [A3] and wrong choice of color scale [A6]<br>■ Subjectivity: inappropriate axis range in the legend [A1] (changed max and min values) |
| **Human Processing** | ■ Imprecision: low resolution of scales (what is the number of cases?) [A4] |

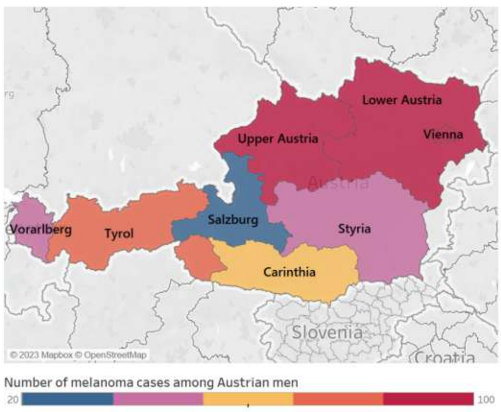



Figure 3.49: Uncertainty types in the misleading visualization of task 3.

**Task 4** uses a slice from an MRI scan and the respective volume rendering from the entire volume of a human head. An artificially introduced structure imitates a magnetic metal rod that impaled the human. The story shown in Figure 3.50 (right) aims to create sensation and draw attention to safety measures at work. The knowledge gap regarding the physics of MR scan acquisition and MR safety measures could make people believe in the plausibility of the story. Assumptions in Figure 3.50 (left) refer to the uncertainty types: error (artificially introduced noise), non-specificity (physics of the imaging modality, visibility of tissues), and subjectivity (knowledge gap) tracked over the visualization pipeline in Figure 3.51.

**Assumptions:**

1. A magnetic metal rod in an MR scan slice appears as a white circle inside a human brain.
no, magnetic items cannot brought into the MR room because of the strong magnetic field

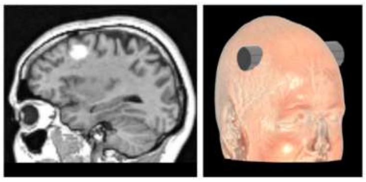
2. The eye of the patient is seen in the scan because it absorbed more X-rays than the surrounding tissues.
no, there are no X-rays in an MR machine

3. The white circumference (simply put, the outer shell) of the head represents the skull.
no, it is skin, skull is below the surface of the skin and barely visible in an MR scan

4. In the scan, it is possible to differentiate between the white and grey matter of the brain.
yes, soft tissues are visible

**Construction Worker Survived a Terrible Accident!**

Last Friday, a terrible accident happened at the construction site of the city bridge. A construction worker got impaled with a metal rod through his head. Fortunately, the man survived because the rod did not damage the vital brain areas.



Left: An MR brain slice Right: MR scan 3D rendering

Currently, the unfortunate worker is staying at the main hospital getting all the necessary care. Upon hospital admission, a magnetic resonance (MR) scan was made to check the degree of brain damage.

MR scans are used to emphasize and visualize soft tissues inside a body. An MR imaging machine does not use radiation. In this case, a very high magnetic field plays an essential role in the sophisticated process of capturing tissue information.

Our journalistic team managed to get the MR scan of the worker (see above), and it looks terrifying! At last, we want to warn you to take care and follow safety rules at work.

Figure 3.50: Task 4 story with noise in an MR scan used to create sensation and draw attention to work safety measures. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | ■ Non-specificity: physics of the imaging modality [A3] |
| **Transformation/ Processing** | ■ Error: artificially introduced noise [A1] |
| **Representation** | ■ Non-specificity: visibility of tissues [A2,3] |
| **Human Processing** | ■ Incompleteness: knowledge gap [A1,2,4] |

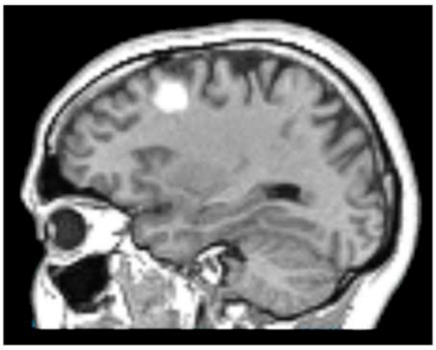


Figure 3.51: Uncertainty types in the misleading visualization of task 4.

**Task 5** uses a stacked area chart to show breast cancer incidence among women over the years across Austrian provinces. A stacked area chart complexities perceiving the overall picture, thus necessitating correct reading of values. The story shown in Figure 3.52 (right) aims to generate media attention toward the rising number of breast cancer cases in all Austrian provinces, omitting that breast cancer incidence in Vienna actually decreases. Assumptions in Figure 3.52 (left) refer to the uncertainty types: subjectivity (inconsistencies of representation, wrong choice of chart, clutter, area encoding) tracked over the visualization pipeline in Figure 3.53.

**Assumptions:**

1. The probability of a woman being diagnosed with breast cancer in Burgenland increased from about 40 to 65 per 100 000 people in the last 35 years.
yes, but clutter can be an obstacle

2. In 2020, the probability of a woman being diagnosed with breast cancer in Styria is lower than in Salzburg.
no, there is dirty data in the legend

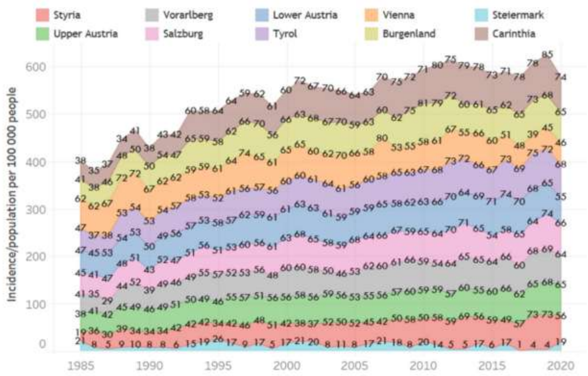
3. Over the last 35 years, in Upper Austria the number of cases is stable.
no, but it looks like this due to stacking

4. In all Austrian provinces, the likelihood that a woman is diagnosed with breast cancer is increasing.
no, in Vienna it decreases

5. The steepest increase in Burgenland was from 1987 to 1988.
no, but it looks so due to stacking of area plots

**Breast Cancer Incidence is Constantly Growing Over the Last 35 Years!**

In 2020, the most frequently diagnosed cancer among women was breast neoplasm. Additionally, the number of women diagnosed has risen over the last 35 years in every province of Austria, as seen in the chart below. The situation only gets worse. Therefore, more research is needed to support this medical field.




** The colored areas, representing each province, are sorted according to the sum value of probabilities of breast cancer incidence over the last 35 years*

Figure 3.52: Task 5 story with breast cancer statistics used to generate attention toward the rise of breast cancer cases. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | ■ Subjectivity: inconsistencies of representation (e.g., Styria vs Steiermark) [A2] |
| **Transformation/ Processing** | no |
| **Representation** | ■ Subjectivity: wrong choice of chart, area chart is problematic for human eye to estimate values [all A] and stacked [A3,5]<br><br>■ Subjectivity: clutter [A1] |
| **Human Processing** | ■ Subjectivity: area encoding (decrease is not clearly visible) [A4] |

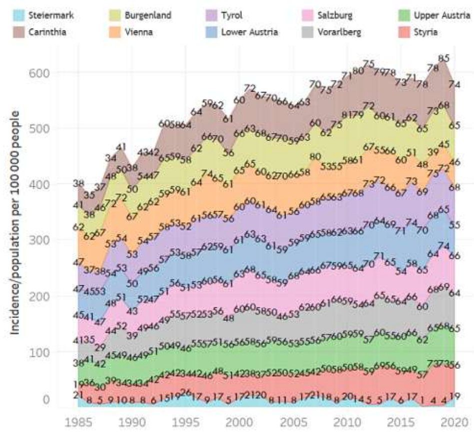



Figure 3.53: Uncertainty types in the misleading visualization of task 5.

**Task 6** uses a volume rendering of a human breast CT scan. The volume is sliced in a specific place and rotated by 180 degrees to imitate dextrocardia, a rare condition when the heart is located on the opposite side. The story shown in Figure 3.54 (right) aims to raise the selling rates of the newspaper by “clickbaiting”, using a sensational title to attract readers. Additionally, the mirrored slice of a CT scan was inspired by reflecting medical errors during bilateral procedures [SPB21]. Assumptions in Figure 3.54 (left) refer to the uncertainty types: non-specificity (physics of the imaging modality, visibility of tissues), incompleteness (knowledge gap), and subjectivity (choice of slice location and orientation) tracked over the visualization pipeline in Figure 3.55.

**Assumptions:**

1. The heart of the sportswoman is located between the lungs and above the liver and spleen.
yes, liver and sleep locates in abdomen, which is below the heart, and the heart locates between the lungs

2. The lungs can be differentiated from the surrounding tissues because the patient inhaled a special contrast medium before the CT scan.
no, the contrast originate from the air density

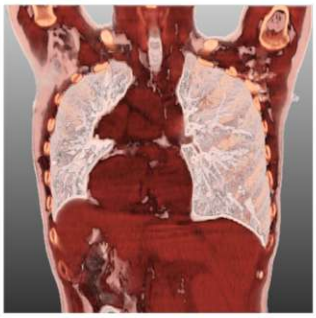
3. The border between spleen and liver is as well-visible as the borders of these organs with the lungs.
no, the contrast is too low

4. Bones can be differentiated from the surrounding soft tissue in a CT scan.
yes, bones are clearly visible in a CT scan

5. The liver of this sportswoman is located on the opposite side in comparison to an average patient.
no, the location is normal, the scan is 180 degrees rotated (mirrored along the coronal and sagittal planes)

**CT Scan Reveals Rare Heart Placement in Famous Sportswoman!**

Last month the most famous sportswoman in the world underwent a CT scan as part of a routine health checkup.



During a CT scan, a part of a patient's body is irradiated with a rotating X-ray beam, and a signal is detected. In the end, the detected signal is transformed into a 3D model showing the inner structures of a patient's body. As X-rays are absorbed more by dense tissues such as bone, these are distinguishable in the acquired images of the human body.

To everyone's astonishment, the chest scan of the most famous sportswoman in the world revealed that her heart is positioned on the right side of her chest rather than the left. Could it be the reason for her success in sports? Can it lead to some health problems? Will it cause the end of the sport carrier? To get the answers, buy our next week's edition with an interview with the most famous sportswoman in the world!

Figure 3.54: Task 6 story imitating dextrocardia in a human breast CT slice used to raise the selling rates. The short answers to assumptions are in grey.

| | |
|---|---|
| **Data Acquisition** | ■ Non-specificity: physics of the imaging modality [A3] |
| **Transformation/ Processing** | no |
| **Representation** | ■ Non-specificity: visibility of tissues [A3,4]<br><br>■ Subjectivity: choice of slice location and 180 degrees rotation of the slice [A5] |
| **Human Processing** | ■ Incompleteness: knowledge gap [A1,2,5] |

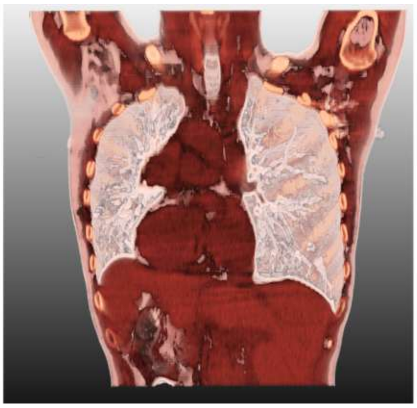


Figure 3.55: Uncertainty types in the misleading visualization of task 6.

**Task 7** uses a stacked bar chart with COVID-19 data in Austria: deaths, recoveries, and new cases. Data aggregated annually obscures that the time spans until March 2023. The story shown in Figure 3.56 (right) aims to support a narrative for political gain stating that, due to the party effort, the number of cases in 2023 is much lower than in 2022. This is not legitimate because 2023 has not ended yet. Assumptions in Figure 3.56 (left) refer to the uncertainty types: imprecision (large binning interval), non-specificity (data of different magnitude), incompleteness (missing data and normalization), and subjectivity (invalid comparison, wrong choice of chart) tracked over the visualization pipeline in Figure 3.57.

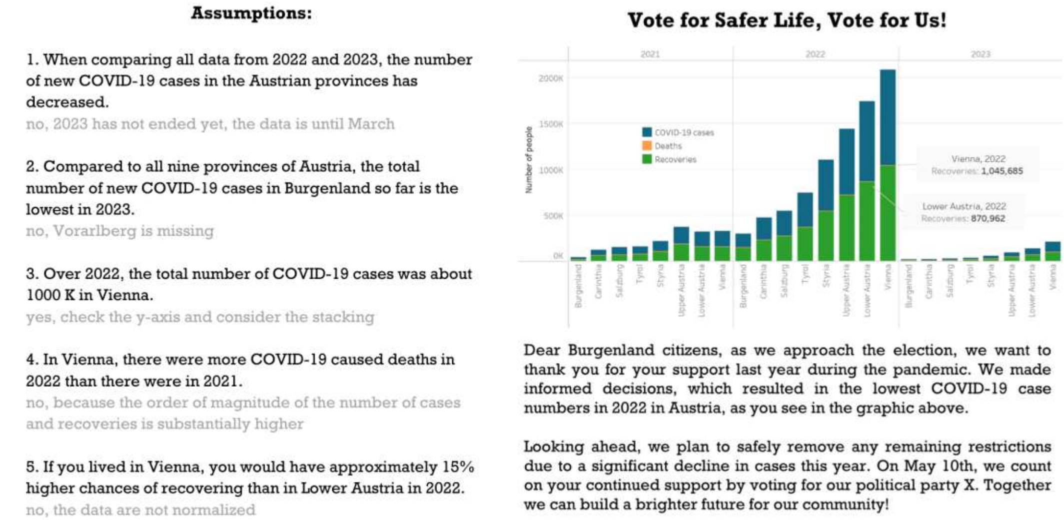



Figure 3.56: Task 7 story with COVID-19 statistics used to support a narrative for political gain. The short answers to assumptions are in grey.

| Data Acquisition | Transformation/ Processing | Representation | Human Processing |
|---|---|---|---|
| Incompleteness: missing data (Vorarlberg) [A2] | Incompleteness: missing normalization [A5]<br>Subjectivity: selective data (no data for the rest of 2023) [A1] | Imprecision: large binning interval (years) [A1]<br>Non-specificity: data of different magnitude [A4]<br>Subjectivity: wrong choice of chart [A3] | Subjectivity: invalid comparison of the full year data (2022) with a part of 2023 [A1] |

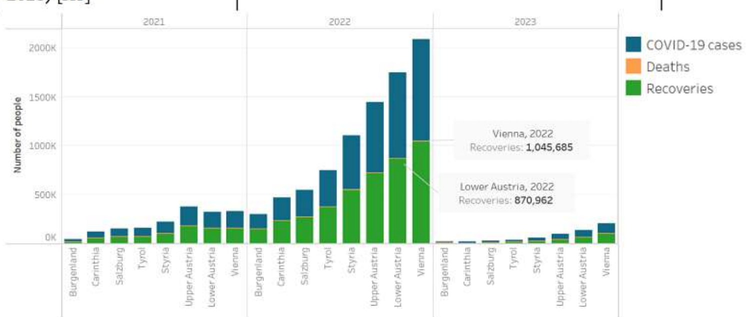



Figure 3.57: Uncertainty types in the misleading visualization of task 7.

**Task 8** uses a 2D topogram of a fish. Objects, including staples, were placed on top of the fish body during topogram acquisition. Due to the loss of spatial information in 2D images, the objects seem to be inside the fish. The story shown in Figure 3.58 (right) aims to raise awareness about lake pollution by showing a fish filled with garbage. Assumptions in Figure 3.58 (left) refer to the uncertainty types: error (noise), non-specificity (physics of the imaging modality, visibility of tissues), incompleteness (loss of spatial information), and subjectivity (2D projection perception) tracked over the visualization pipeline in Figure 3.59.

**Assumptions:**

1. The fish spine and skull should not be visible in an X-ray image.
yes, bones are clearly visible in an X-ray scan

2. Staples look white in the image because they absorbed more radiation, being dense materials.
yes, it is the principle of an X-ray scan

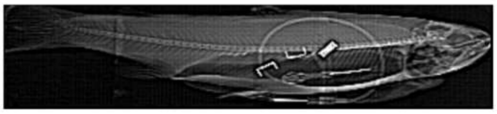
3. The lower part of the fish spine towards the tail was damaged.
no, it is the noise

4. The staples are inside the belly of the fish.
no, they are above the surface of the fish

**Contaminated Fish Pose Threat to Public Health!**

Environmental experts issue a grave warning to address lake pollution! The lake fish, feeding on waste in the polluted waters, may end up in local markets, posing potential health hazards to consumers.

Researchers have unveiled evidence of widespread pollution in the beloved lake. The harmful waste accumulated there impacts the ecosystem. Fish accidentally ingest garbage, like staples or probes (see an X-ray image below, where denser materials such as metal or bone appear whiter as they absorb more radiation).



Consuming contaminated fish can lead to serious health issues, including gastrointestinal problems and long-term illnesses. Immediate action is necessary to tackle the pollution crisis, implementing proper waste management systems and raising awareness about the risks to public health.

Preserving the health of our lake ecosystem is paramount. Let us unite, adopt responsible environmental practices, and safeguard our environment and the community well-being!

Figure 3.58: Task 7 story with a 2D topogram of a fish X-ray scan used to draw attention to lake pollution. The short answers to assumptions are in grey.

| Data Acquisition | Transformation/ Processing | Representation | Human Processing |
|---|---|---|---|
| Error: noise [A3]<br>Non-specificity: physics of imaging modality [A2] | no | Non-specificity: visibility of tissues [A1]<br>Incompleteness: loss of spatial information [A4] | Subjectivity: 2D projection perception [A4] |

Figure 3.59: Uncertainty types in the misleading visualization of task 8.

### 3.5.2 Gamification

While working on the theoretical part behind the methodology, and reading related scientific papers, we imagined a private detective investigating newspaper snippets. Hence, we decided to create an educational game, where a player can enjoy a similar atmosphere and be engaged upon finding misleading elements. Additionally, solving riddles or puzzles activates problem-solving skills and can increase motivation, thus improving the learning process, which is an essential aspect of an educational game [Jel17]. This educational game aims to provide support on how to recognize and avoid misleading visuals for educational purposes through gamification.

Gamification means the augmentation of the learning process with some game elements [LACA18]. Such games encourage participation more than high-score performance, shorten the feedback cycle when solving a game problem, and promote active exploration [LH11]. Our educational game contains some key elements of gamification [Mau19]. The player is exposed to the main objective here, rescuing stray cats (Figure 3.60). Storytelling is used to create the game narrative and introduce different "levels" (tasks). In these levels, the player can achieve points and get immediate feedback and even rewards. The objective of cat rescue and the detective atmosphere gave the name to the game "DeteCATive" which combines both aspects.

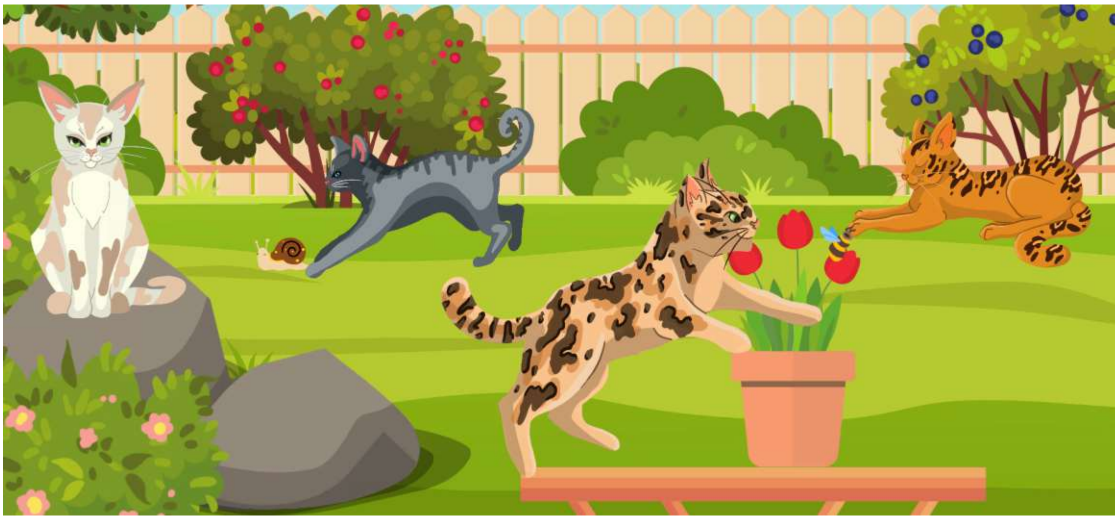


Figure 3.60: Stray cats that can become foster cats at the end of "DeteCATive" are used as the main objective to engage and reward the audience.

Generally, the gameplay core of "DeteCATive" is solving riddles. However, it can be not enough to sustain the motivation of players through the game. Positive reinforcement strategies, like a rewarding system, are used to sustain the motivation of a gamer. Wang et al. propose a list of rewarding systems such as score, experience points, item or resource granting, achievement, feedback, cutscene, and unlocking mechanics systems [WS11]. In "DetCATive", we used two rewarding systems: score and achievement due to the relative simplicity of implementation by C# codes in Unity. The score system includes points that the player achieves by correctly assessing assumptions. One correctly assessed assumption equals one point, and one incorrectly assessed assumption provides zero points (scores). These points are used at the end of the game to rescue the stray cats.

The achievements also contain three versions of badges shown in Figure 3.61. When the player finds all the missing provinces of Austria, the reward is the "Eagle-eye" badge. To get the "Scholar" badge, the player should get points for assessing each assumption about the anatomical position of an organ. The hardest badge is "Critical thinker" (Figures 3.61 (middle) and 3.62). To obtain it, all main tricks with medical imaging data should be identified.


Figure 3.61: Badges rewarding system.

Below, the conditions for each badge are listed. The number following the assumption represents the number of the task (before the dot) and the number of an assumption in the task (after the dot).

"Eagle-eye" badge conditions:

- Assumption 3.5, missing Burgenland;
- Assumption 7.2, missing Vorarlberg;

"Critical thinker" badge conditions:

- Assumption 2.5, under-segmentation of the liver;
- Assumption 4.1, faked MR scan;
- Assumption 6.5, rotation of the CT scan slice;
- Assumption 8.4, staples positioned not inside the fish;

"Scholar badge" conditions:

- Assumption 2.1, kidney anatomical location;
- Assumption 6.1, heart anatomical location.

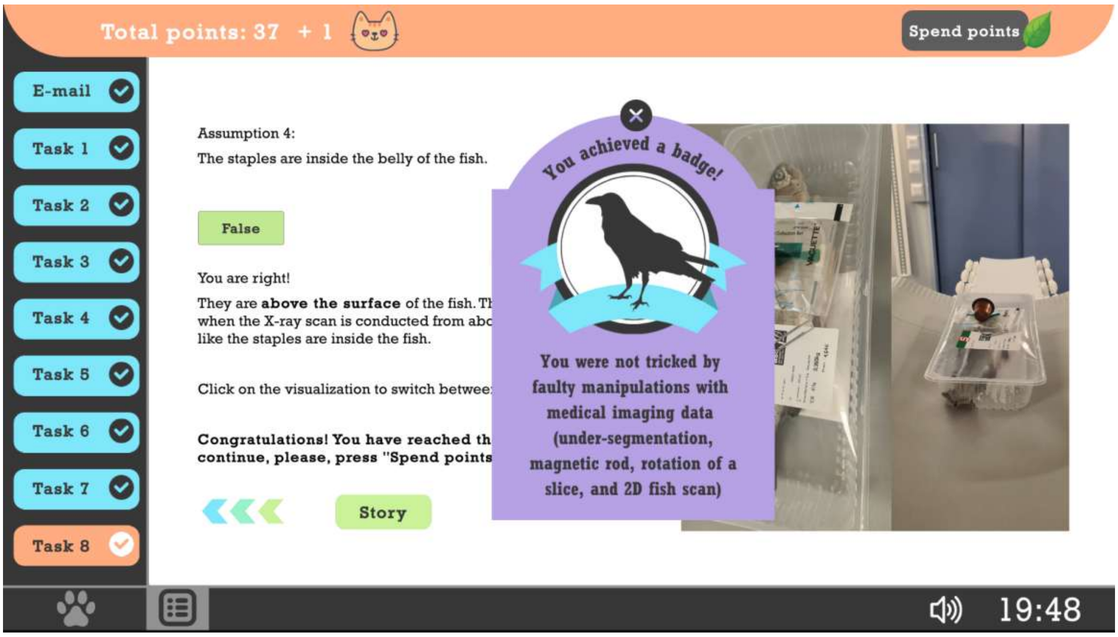



Figure 3.62: Screenshot of a view when a player achieves a badge.

CHAPTER 4

# Implementation

In this chapter, we dive into the practical aspects of the thesis. Section 4.1 outlines the creation of stories and tasks. The process began with visualization ideas development, which determined the specific data requirements. Once the suitable data was found, the visualizations were created in Tableau and MeVisLab, shaping the ideas into a visual form. Subsequently, each visualization was enriched with text creating a misleading story. The creation of tasks involved formulating assumptions referring to misleading elements in stories. Each assumption had an accompanying explanation that clarified why it was true or false. Supporting visualizations were used to enhance understanding of explanations. In Section 4.2, we share the details about using the Unity engine to integrate all tasks into an educational game.

## 4.1 Creating Stories and Tasks

To create a visualization for each game task, we needed to gather suitable data from open sources. Specifically, the non-imaging visualizations originate from the following open-source data sets:

- COVID-19 data in Austria (26.02.2020-25.03.2023):
  Source: BMSGPK, Österreichisches COVID-19 Open Data Informationsportal
  Name: COVID-19: Zeitliche Darstellung von Daten zu Covid19-Fällen je Bundesland, accessed in March 2023
- Cancer statistics in Austria (1983-2020):
  Source: Statistik Austria
  Name: Cancer statistics, accessed in March 2023
- Population statistics in Austria (1983 - 2020):
  Source: STATcube – Statistical Database of Statistik Austria

Name: STATcube: Population at the beginning of the year since 1982, accessed in March 2023

- Cancer classification for 2016:
  Source: WHO
  Name: International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version, accessed in March 2023

Medical data used for the medical imaging visualizations also originate from open sources, with the exception of one data set obtained via a collaboration with the General Hospital of Vienna. Specifically, we obtained imaging data from the following sources:

- Human abdomen CT scan
  Source: MeVisLab demo data path
  (Installation Directory)Packages/MeVisLab/Resources/DemoData
  Name: Liver1_CT_venous.small.dcm, accessed in March 2023
  Image size: 93x93x61
  Voxel size: 3.9858x3.9858x3.9858

- Human head MR scan
  Source: MeVisLab demo data path
  (Installation Directory)Packages/MeVisLab/Resources/DemoData
  Name: BrainMultiModal/ProbandT1.tif, accessed in March 2023
  Image size: 109x91x80
  Voxel size: 1.9531x1.9531x2

- Human breast CT scan
  Source: The Cancer Imaging Archive
  Collection: NSCLC-Radiomics
  Subject ID: LUNG1-056, accessed in March 2023
  Study date: 05-23-2008
  Image size: 512x512x134
  Voxel size: 0.9766x0.9766x3

- Supermarket fish CT scan
  Source: collaboration with Dipl.-Ing. Elisabeth Salomon, PhD
  from the Center for Medical Physics and Biomedical Engineering of the Medical University of Vienna, Nuclear Medicine Department
  arranged in May 2023
  Image size: 512x512x179
  Voxel size: 0.1621x0.1621x2

Following the data collection, we use Tableau, data analytics software, to visualize the non-imaging data. Tasks 1 and 7 are based on the COVID-19 data in Austria. In task 1 we used quantitative data including the population of a province, cumulative numbers of deaths, recoveries, and COVID-19 cases, along with qualitative data: reporting date (d/m/y) aggregated in months and provinces of Austria (Figure 4.1). Similarly, in task 7, we used the same data, though with the reporting date aggregated yearly and absolute numbers of deaths, recoveries, and COVID-19 cases (Figure 4.2).


Figure 4.1: Tableau interface with the visualization for task 1. Cumulative data from Vienna is used for this bar chart.

On the other hand, tasks 3 and 5 centered around the cancer statistics in Austria (Figures 4.3 and 4.4). The data set contained qualitative data: the reporting years, ICD-10 codes of cancer type, sex of patients, province of residence, and quantitative data representing the number of records per reporting year. However, due to the absence of quantitative data normalization to the province population, the population statistics for each Austrian province were joined with the cancer statistics. Population statistics in Austria contained the population number (quantitative data) per year and province (qualitative data). Also, the type of cancer in the cancer statistics was encoded with the ICD-10 code. Therefore, we used the WHO classification data (qualitative) to assign the name of a cancer type to the ICD-10 codes. Additionally, in task 3, the breast cancer incidence normalized to the population for Styria was manually split into two groups: Styria and Steiermark, to introduce dirty data (see the legend in Figure 4.4).


Figure 4.2: Tableau interface with the visualization for task 7. COVID-19 data are summed up over three years: 2021, 2022, and 2023. Vorarlberg is excluded from the visualization.

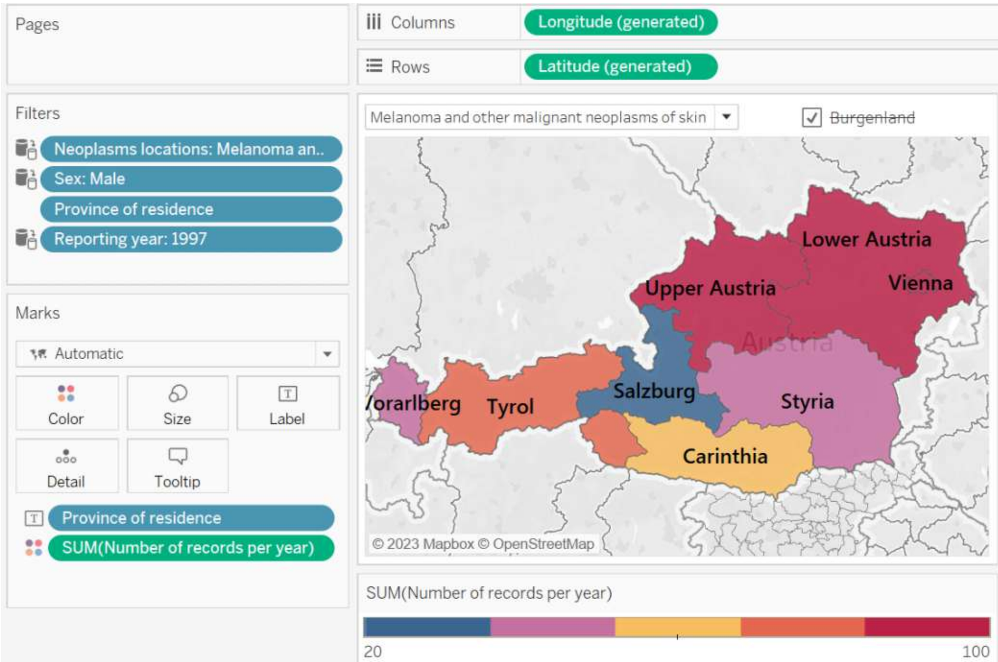



Figure 4.3: Tableau interface with the visualization for task 3. Four filters are applied to the data: neoplasm location, sex, province of residence, and reporting year. Burgenland is excluded from the provinces of residence list.

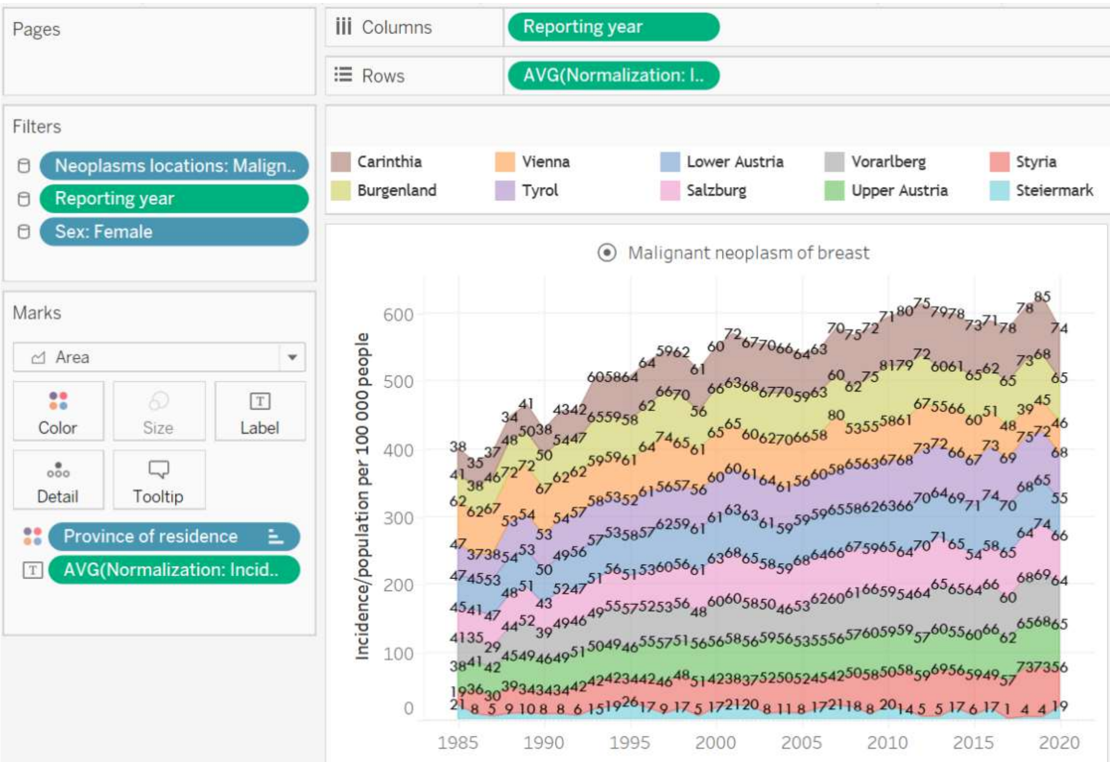



Figure 4.4: Tableau interface with the visualization for task 5. Filters are applied to the data – only malignant neoplasm of breast data for women is shown. The data of Styria is divided between two variables: Steiermark and Styria to create dirty data.

To visualize the imaging data, we used MeVisLab, a medical image processing software. Each imaging task was based on its own data set. Task 2 utilized the human abdomen CT scan (Figure 4.5), in which the liver was undersegmented due to a higher lower threshold in RegionGrowingMacro and larger sizes of y and z kernels in CloseGap. In task 4, we used the human head MR scan with an artificially introduced source of Gaussian noise by using vtkImageGaussianSource shown in Figure 4.6 (left) that imitated a magnetic metal rod that impaled a fictional construction worker.

For task 6, we used a CT scan of a human breast (Figure 4.7). A SoClipBox was used to cut a slice from the volume. Then the slice was rotated in the SoExaminerViewer to imitate dextrocardia. To create a visualization for task 8, a 2D X-ray fish scan acquired during the CT scan (topogramm), was utilized to create the visualization shown in Figure 4.8 (left). Also, the fish volume rendering was rotated in the SoExaminerViewer to make an illusion that staples are inside the fish. The fish volume rendering under a revealing angle is shown in Figure 4.8 (right). Several adjustments were made to some MeVisLab networks to produce a visualization for explanations of assumptions. For instance, in Figures 4.6 and 4.8 the most left networks provided the explanatory visualizations.

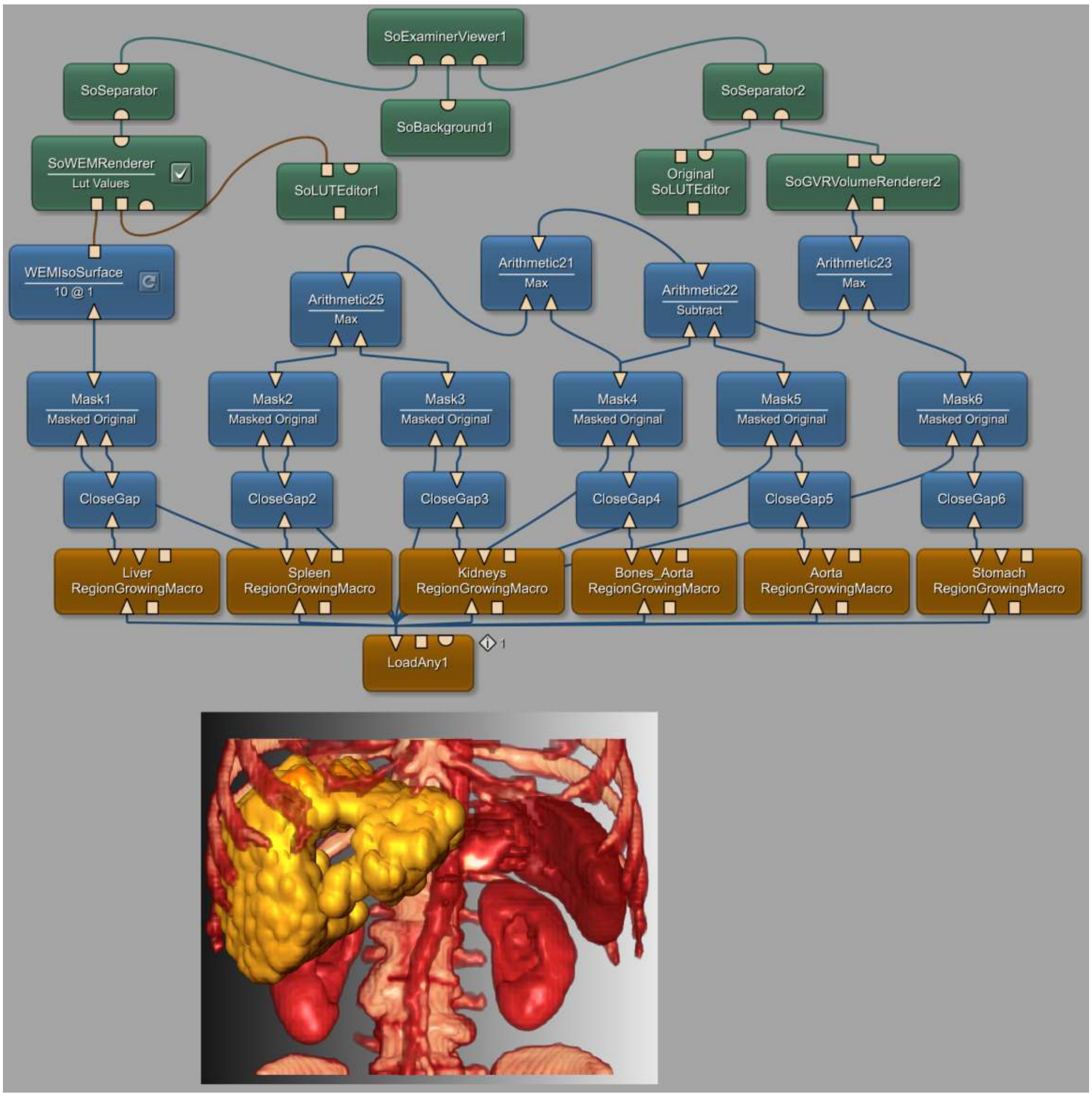



Figure 4.5: MeVisLab network for task 2. Organs are extracted from the original data via region growing and close gap approaches. Then indirect volume rendering is employed for the organs while the liver is surface rendered to hide the undersegmentation.

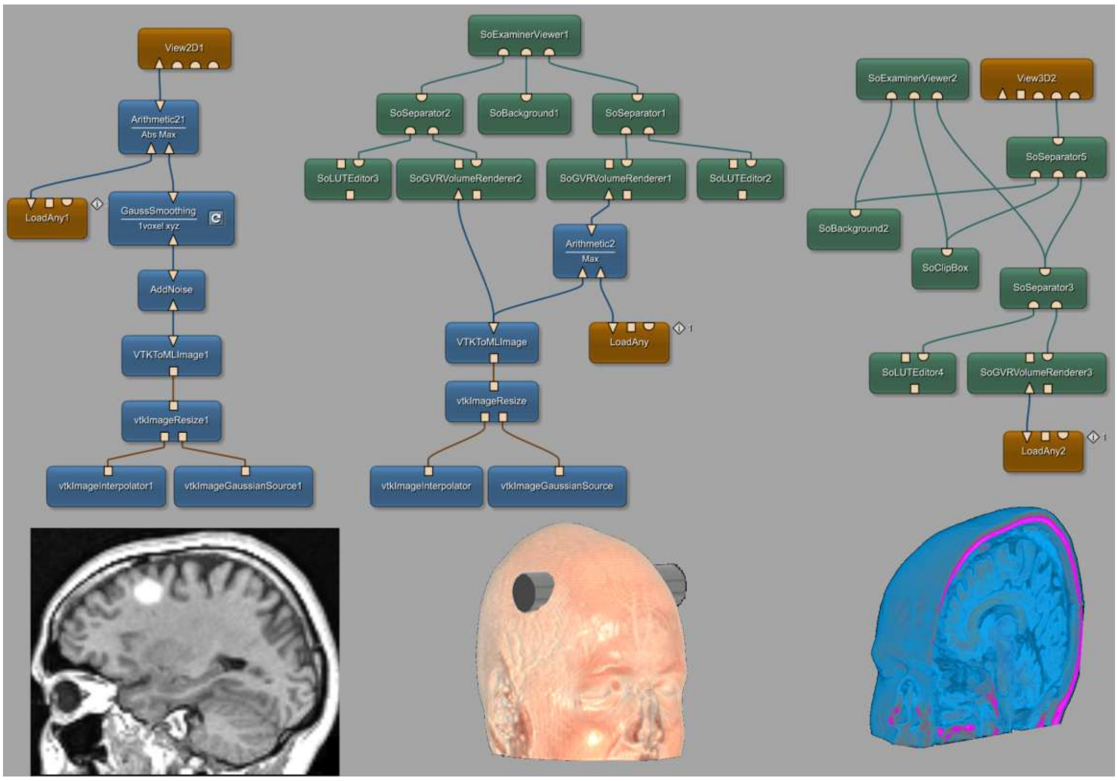



Figure 4.6: MeVisLab networks for task 4. The left-most network was used to create the noise inside the MR scan. The middle network was used to create a 3D volume rendering of a head impaled with a rod. For the explanatory visualizations, the right-most network was used.

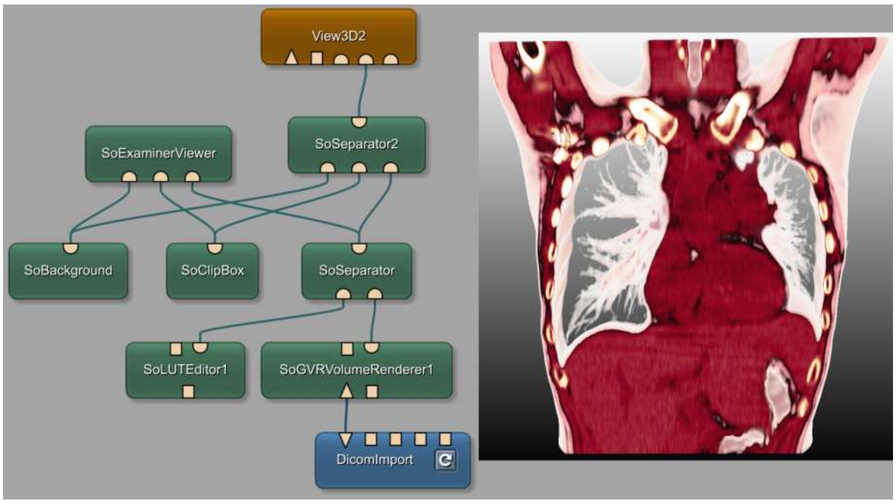



Figure 4.7: MeVisLab network for task 6. A SoClipBox allowed the creation of a slice out of the volume model, which then was rotated 180 degrees in the SoExaminerViewer to simulate a fake dextrocardia case.

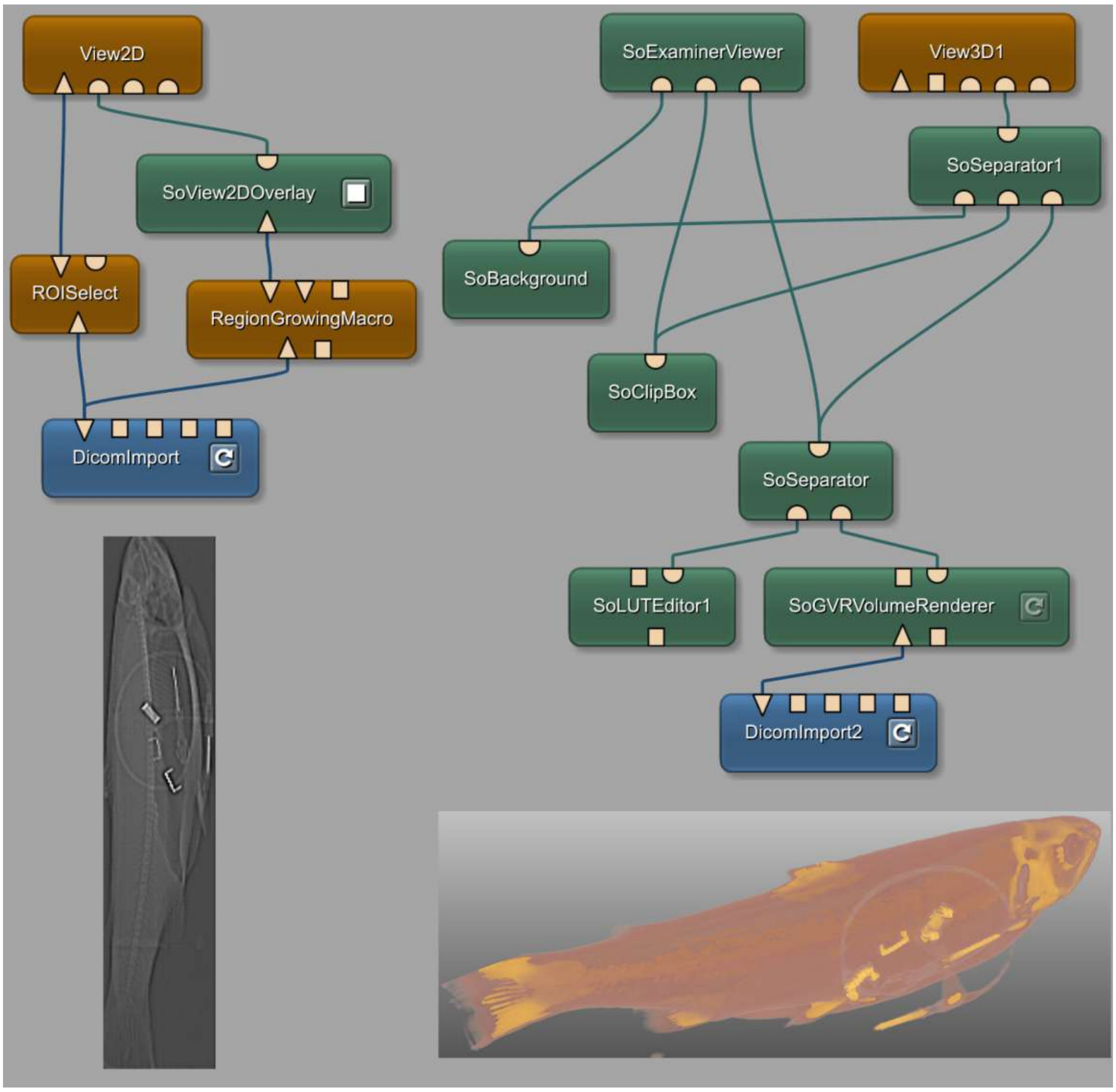



Figure 4.8: MeVisLab networks for task 8. The left-most network used CT data to create misleading visualization. For the explanatory visualizations, the right network used the 3D CT scan data.

## 4.2 Unity Game Design

To present misleading visualizations in health care in an accessible and educational form, gamification was used. There are three main concepts that we followed while developing the educational game "DeteCATive". First, the game revolves around rescuing stray cats. Second, the gameplay core is solving riddles. Lastly, the reward system should sustain player motivation during the game. These concepts were integrated during the game development using Unity, a game development platform.

The game structure contains three main scenes: start, main, and final. These scenes have several views inside them. Sometimes the player can move back and forth among the views within a scene. In the start scene, the player inputs the name or Participant ID used for the game activity file (Figure 4.9) and submerges into the game atmosphere. The game atmosphere components include medical imaging scans, statistical visualizations, detectives, and cats (Figure 4.10).


Figure 4.9: Screenshot of welcome view of the start scene of the educational game "DeteCATive".

In the main scene, players are informed that rescuing stray cats necessitates their participation in a test. The test assesses the ability to identify misleading elements within visualizations used in health care. The better the scores on the test, the more cats can be rescued. Once players have read how this test works, they can begin with task 1. The next task is available only upon the completion of the previous one. Some screenshots of the main scene views are shown in Figure 4.11.


Figure 4.10: Screenshots of a detective table view of the start scene of the educational game “DeteCATive”.

The main scene view shown in Figure 4.12 (above) presents an assumption, a misleading visualization, and a dropdown menu with true and false options. After selecting an option, the player can submit the answer. After clicking the submit button, the assumption explanation appears as shown in Figure 4.12 (below). Then players have options to move forward to the next assumption by clicking on the triple arrow, or they can revisit the story or a previous task view to refer to explanations for already solved assumptions.

In the final scene, the player can spend the point on the stray cats as shown in Figure 4.13 (above). Also, there is a view with the test result, indicating the accuracy of the test and the badges achieved during the game as shown in Figure 4.13 (below). On the last page of the game, the player can observe the rescued cats playing in the garden and exit the game.

To ensure that the UI is accessible and intuitive for users, we followed some UI development guidelines. They include consistency in typography, color usage, symmetry and hierarchy of layout, and intuitiveness of UI objects appearance [ARB04]. Also, the scene layout should be simple and symmetrical, and the objects within the layout should follow the hierarchy representing the level of importance.


Figure 4.11: Screenshots of the main scene of the education game "DeteCATive": overall instruction view (above) and task 2 (below).

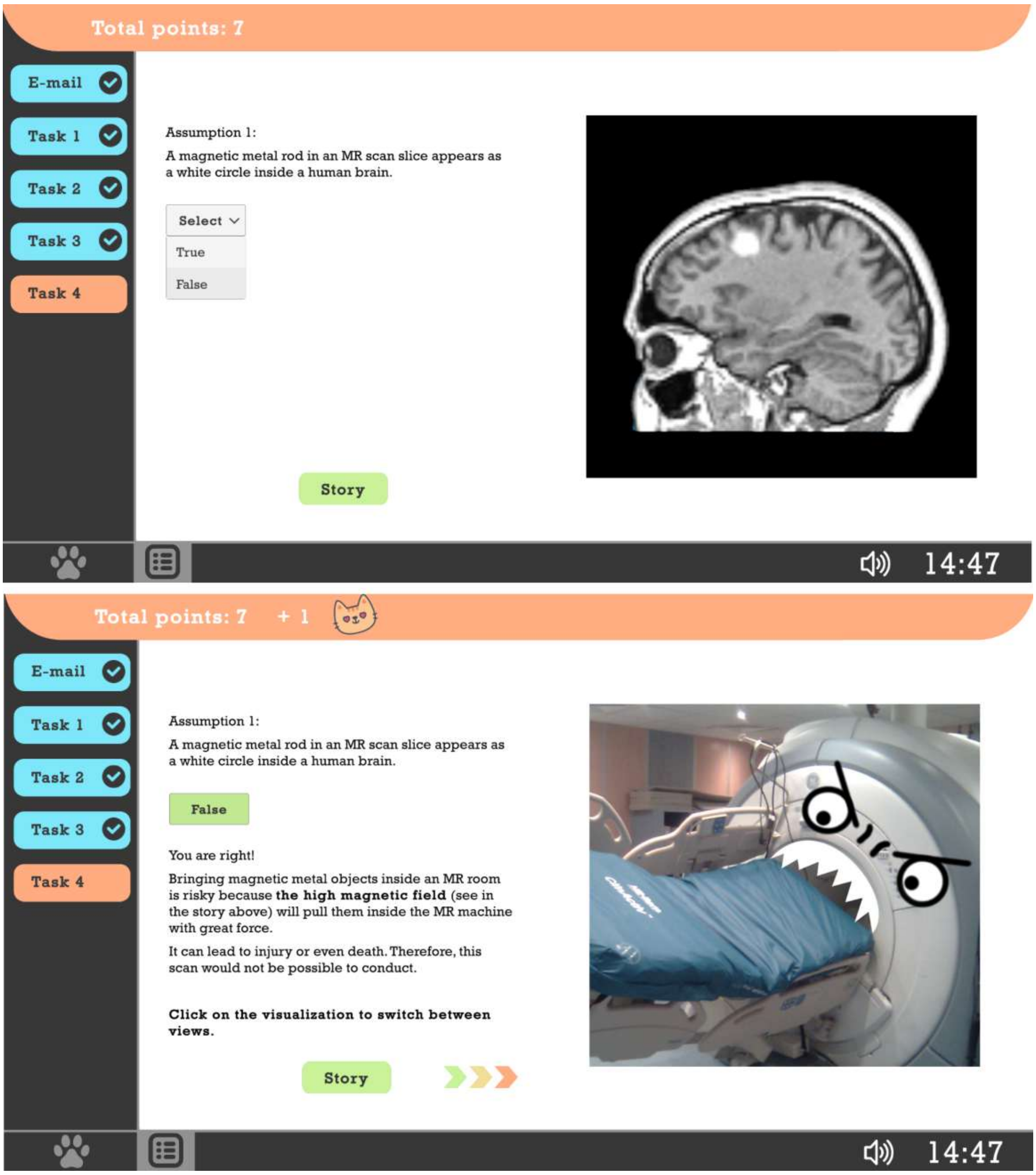



Figure 4.12: Screenshots of the main scene of the education game “DeteCATive”: task 4, assumption 1 (above) and the explanation to it (below).

The visual appeal of the "DeteCATive" UI is achieved by the consistency of fonts, color palette, and layout. Also, smooth animations were added to UI elements: buttons, sliders, dropdowns, and input fields. The color palette was chosen with the help of pallet generator Coolors. Smooth animations of UI elements were coded with C# and the garden animations (final scene) of cats, plants, and a bee were made in Blender.

The game ambience was complemented by background music, written by Simon Mariacher. Also, when a box with a cat opens, a sound effect plays. The sound effect Cute-cat-meow-sound.mp3 originates from Orange Free Sounds, a stock audio collection. It is permitted for non-commercial use under the license "Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

All images in the sprite collection of the game are made in Canva, a design tool. However, there are several exceptions. The angry MR machine in Figure 4.12 (below) was found on the website of Kryptonite Solutions company. The liver damage stages are taken from the provider of stock images Shutterstock as shown in Figure 4.11 (below). Each image in comics complementing stories is generated in online AI art generators like tools like Hotpot, Gencraft, Canva AI image generator, and others. For example, in Figure 4.11 (below) the comic shows a place, where one could find a snippet about "Alcoholism leads to liver damage!". Moreover, we hand-drew stray cats that can be rescued by the player on an iPad with an Apple pencil borrowed from a friend.

To record the users' data, we wrote a C# that recorded game activity data into a .txt file at each start of the game. The file includes the Participant ID input at the beginning of the game (Figure 4.9) and the time when the player started the game. Additionally, the file contains information about which buttons were clicked and the time difference between clicks. When the "Spend points" button is clicked (the button indicates the test exit), the click timeline terminates. And the total time required to solve the test is written in the file. The final part of the game activity file includes the accuracy of the test and the list of achievements. All these data processing outcomes are discussed in the Chapter 5 Results.


Figure 4.13: Screenshots of final scenes of the education game "DeteCATive": choosing foster cats (above) and final results and awards views (below). In this case, a player collected 26 points and achieved two badges: "Eagle-eye" and "Scholar". Cats can be picked by spending collected points. The green squares under each task represent a correctly assessed assumption.

CHAPTER 5

# Results and Discussion

This chapter contains the evaluation of the educational game through a user study. The user study objective is to assess the educational value of "DeteCATive". In Section 5.1, we present the detailed design of the user study, along with the data we were interested in collecting and sources used for the inspiration for questionnaire questions. Then we present the results of the user study and the statistical tests that were used for data analysis in Section 5.2. Finally, we discuss the limitations of the game in Section 5.3.

## 5.1 User Study

In the user study, we ask the participants to play and evaluate "DeteCATive", an educational game that copes with misleading visualizations in health care. The game contains fictional stories with medical data visualizations, which have been presented in Section 3.5. The participants, while playing our game, read the assumptions and decide if these are true or false, based on the stories. For the correct answers, they collect points and even badges.

Within the user study, we asked the participants to sign the consent form for the anonymous processing of their data. Then they were asked to fill in their information using Google Forms. This form included the participant ID, age group, prior knowledge in medicine/biology, and experience in visualization. Participants could pick an ID of their choice. It could be any name or number, but not their real name. Moreover, they were informed in the consent form that the participant ID serves to revoke participants' data. Also, the participant ID connects the data from the questionnaire and game activity file. After a briefing regarding the content of the study, they were asked to play the game. The game on average took about 40 min to play. After the game, they were asked to complete a questionnaire.

We recorded the data using two approaches: a questionnaire in Google Forms and a game activity file recorded during the playing of the game. The latter was used to collect the speed of task performance and rate of errors which can be used to assess the usability [LBI+11]. To this end, a C# code was written for the Unity-developed game, which records the game activity. The data recorded in the game activity file are the following:

1. Story reading time;
2. Number of referencing back to the story;
3. Accuracy of answers;
4. Assumption reading time;
5. Explanation reading time;
6. Position of game exit.

Each task contains a story (text and visualization) and assumptions with explanations/clarifications to them. The *story reading time* indicates the time between the first opening of a task and the first switching to the assumptions. When a participant returns to the story while analyzing an assumption or a clarification of it, we count it as a *reference back to the story*. *Accuracy of answers* relates to whether a participant correctly or incorrectly assessed an assumption during the game. This is measured through tracking points gathered during the game.

The *assumption reading time* is the time between reading an assumption and submitting an answer. The *explanation reading time* is the time between submitting an answer and moving to the next assumption or task. *Position of game exit* refers to the point during the game after which task a participant presses the button "Spend points" and exits the game. This measurement can relate to the willingness to learn more.

The questionnaire was another way to collect the participants' data to evaluate game design decisions and users' likes and dislikes regarding the design [LBI+11]. The questionnaire included open, multiple-choice questions, and 7-point Likert scales (Strongly disagree /.../ Strongly agree). In the questionnaire, we used some aspects of the evaluation of narrative visualization summarized by Meuschke et al. [MGS+22], like memorability, aesthetics, and cognitive involvement. To check whether the reinforcement strategy worked, we included also this aspect. We also took inspiration from the VisEngage questionnaire, used for the evaluation of user experience through engagement [HP17]. This work separates the engagement into several aspects, including aesthetics, captivation, challenge, and others, and proposes examples of questionnaire questions to measure each aspect.

In the final version of the questionnaire, we combined all relevant evaluation aspects discussed above into the following questionnaire:

1. Memorability [MGS+22];
2. Reinforcement;
3. Engagement [HP17]:
    - Aesthetics;
    - Cognitive involvement;
    - Captivation.
4. Subjective likes and dislikes [LBI+11];

*Memorability* includes "short-time" and "long-time" memory questions. "Short-time" questions relate to the user referring to the story at each task or to previous tasks, to refresh the information delivered there. "Long-time" questions ask a participant to write a small paragraph about "The story that I remember/liked the most ... (describe what it was about and which elements were misleading in as many details as you can)". These small paragraphs were evaluated and recalculated into numerical memorability. A 100% numerical memorability can be only reached when the paragraphs contain key phrases about the general description of the visualization, intention, and each assumption misleading element. For example, 100% numerical memorability for story 7 could be reached with approximately these key phrases: "COVID-19 statistics in Austrian provinces", "political gain", "invalid comparison of 2022 and 2023 data", "missing province", "stacked bar chart", "number of death is too small", and "not normalized data". Each of these key phrases is equally weighted toward 100% memorability.

Within the *reinforcement* questions, we assess whether the reinforcement strategies (virtual currency system with points and achievement of badges) work on a participant. The participants are asked not only about their experience with the reinforcement strategy but also about the reasons why they continued or exited the game after task 4. The *engagement* aspect includes three sub-aspects: aesthetics, cognitive involvement, and captivation. *Aesthetics* is important for the general population, due to its attractiveness [GMF+21]. To evaluate this aspect, we asked the participants whether the game is intuitive to use and if it is visually appealing. The most important aspect of an education game is *cognitive involvement*. This aspect addresses the desire to learn and think over the information presented [MGS+22]. In the questionnaire, we asked if a participant learned something and got interested in learning more about misleading visualizations while playing the game. Also, we asked about the type of medical data visualization that was harder to analyze and in which areas they expanded their knowledge. The *captivation* questions were about whether it was easy to focus during the game and what was distracting a participant. At the end of the questionnaire, we asked open questions about participants' *likes and dislikes* while playing the game.

The questionnaire answers and game activity file included valuable information, which was processed and visualized. The results are included in the following Section 5.2. The open questions about subjective likes elucidated the substantial aspect of the game, while the dislikes pointed out some limitations and proposed ideas for improvement, which are summarized in Section 5.3.

## 5.2 Analysis of User Study Outcomes

This section is divided into two parts to present the results from the questionnaire and the game activity file separately. Google Forms automatically creates an Excel file with all the data. We used Python to extract all the necessary values from the game activity file into another Excel file. Then, the data from the Excel files were visualized in Tableau.

### 5.2.1 Questionnaire Results

The questionnaire was used to collect data about the participant's experience after they played "DeteCATive". The total number of user study participants is 21 and more than 50% of the participants are in their 20s (Figure 5.1). According to the Pew Research Center, people between 18 and 29 are in the leading position of using social media from 2015 to 2021 (no data available for 2022 and 2023) [Pew]. Therefore, younger people have higher chances of consuming visualizations in the wild (e.g., social media). The following paragraphs contain the results of the questionnaire relative to the "DeteCATive" evaluation aspects mentioned in Section 5.1.

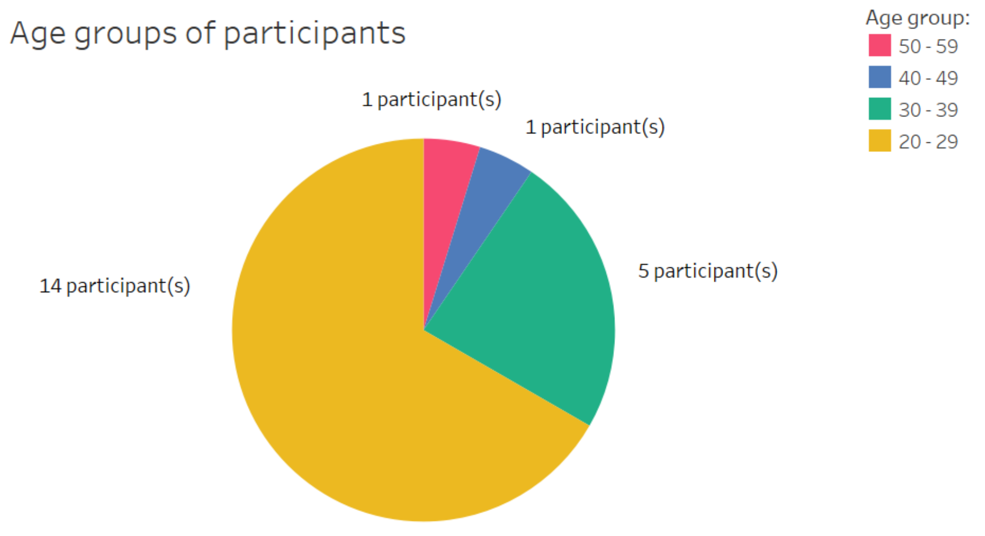



Figure 5.1: The distribution of age among the 21 user study participants.

**Memorability**. In Figure 5.2 the combined statistics of the memorability questions are shown. All four *imaging data stories* were remembered with a median (interquartile range) memorability of 28% (20–29%). This phenomenon could be due to the appeal of medical imaging visualizations in comparison to statistical representation. Only one non-imaging data story was mentioned two times and both times were described with a memorability rate reaching 70%. Most of the participants never returned to previous tasks, but sometimes they referred to the story when analyzing the assumptions or reading the explanations.

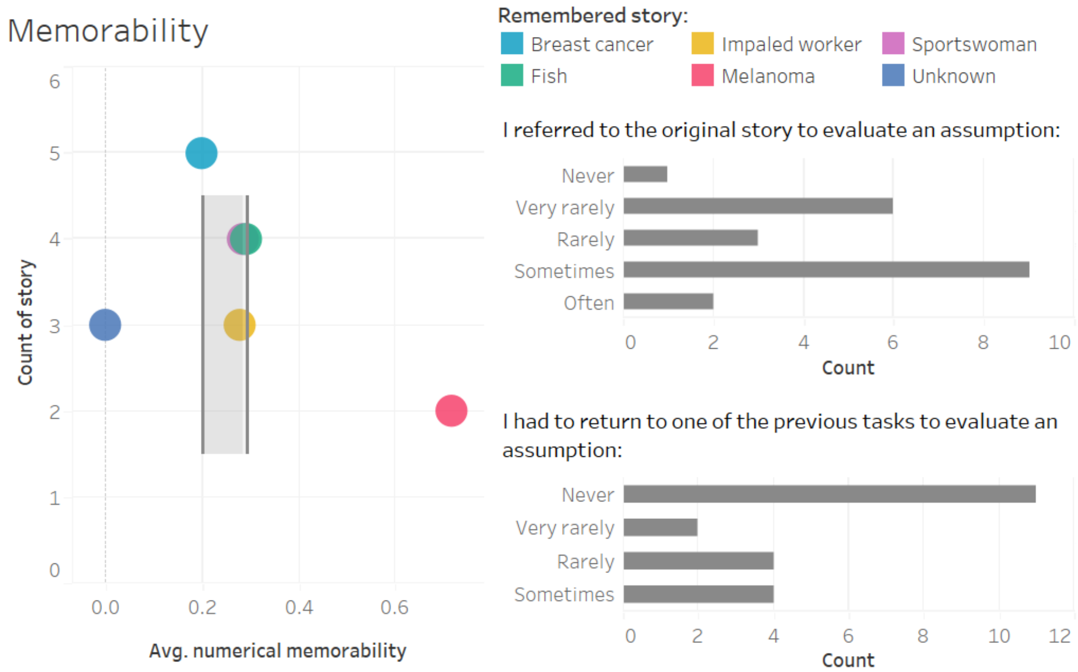



Figure 5.2: Answers to the memorability aspect questions, generated with Tableau from our user study raw data. On the left, the box and whisker plot contains a grey box representing the interquartile range split by median inside and the whiskers show the outliers borders.

**Reinforcement**. This aspect includes the answers to rating questions (Figure 5.3), the reasons why the participant continued the game after task 4 (Figure 5.4), and, if they did not, why they left the game. Most of the participants (16 participants who answered positively) wanted to play again to earn more cats and badges. During the game, 17 participants wanted to gain more badges. After gaining a badge, 10 participants were motivated to continue the game. Only two participants pressed “Spend points” after task 4 and terminated the game in the middle.

For 14 out of 21 participants, the motivation to continue was in most cases based on curiosity about the following stories. Additionally, gaining points, with consequential spending them for foster cats, was more popular than gaining badges. Only two participants exited the game earlier. One participant "didn't know it would end", which happened due to the inattention of reading a caption to a button "Spend points" that ends the test. For the second participant, the "session to reward loop was too long".

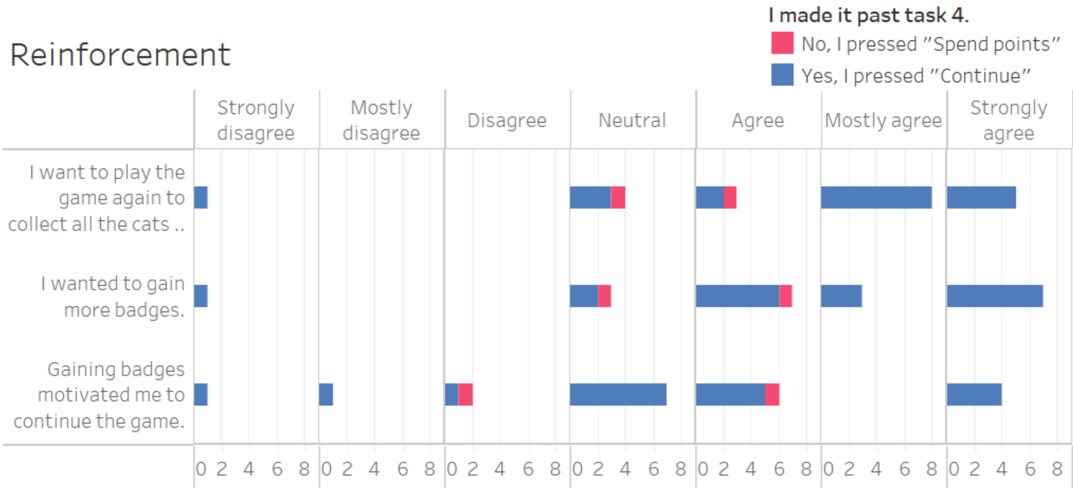



Figure 5.3: Answers to the reinforcement aspect questions, generated with Tableau from our user study raw data.

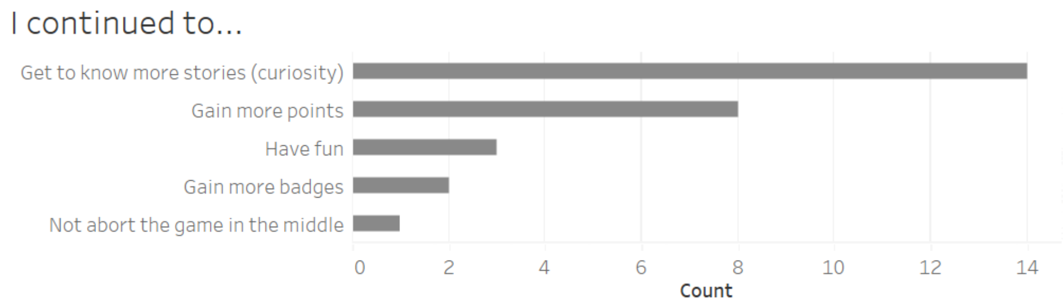



Figure 5.4: Reasons that led the participants to continue the game after the middle point was reached, generated with Tableau from our user study raw data.

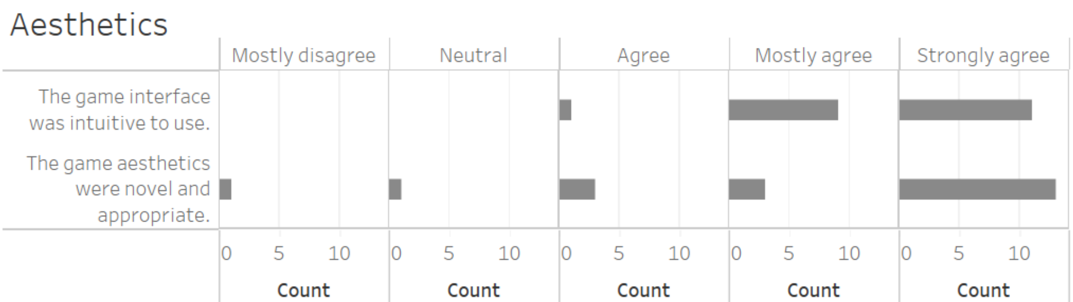



Figure 5.5: Answers to the engagement (aesthetic aspect) questions, generated with Tableau from our user study raw data.

**Engagement: Aesthetics**. The aesthetics got mostly positive feedback (Figure 5.5). All the participants agreed that the game interface was intuitive to use and that the aesthetics of the game were satisfactory.

**Engagement: Cognitive Involvement**. Answers to the questions of this aspect are shown in Figures 5.6 and 5.7. It is worth mentioning that most of the participants found the non-imaging data much harder to analyze. However, 5 experts in data visualization experienced difficulties with imaging data visualizations, although they claimed to have a school level of medicine/biology knowledge.

Most of the participants were strongly cognitively involved. For example, 19 participants agreed that they learned about misleading visualizations during the game, while 18 participants felt interested in learning more about misleading visualizations. Most of the participants indicated that they expanded their knowledge with regard to dirty data. By contrast, the color violation was well-known for the participants.

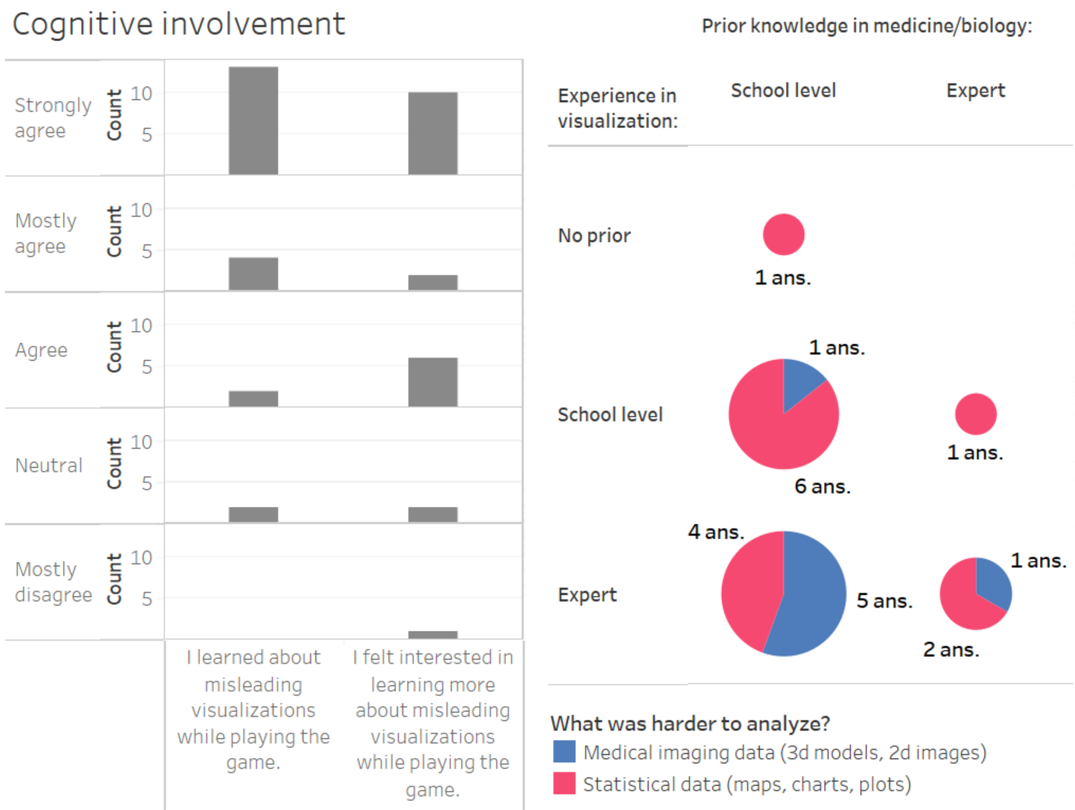



Figure 5.6: Answers to the engagement (cognitive involvement) aspect questions, generated with Tableau from our user study raw data.

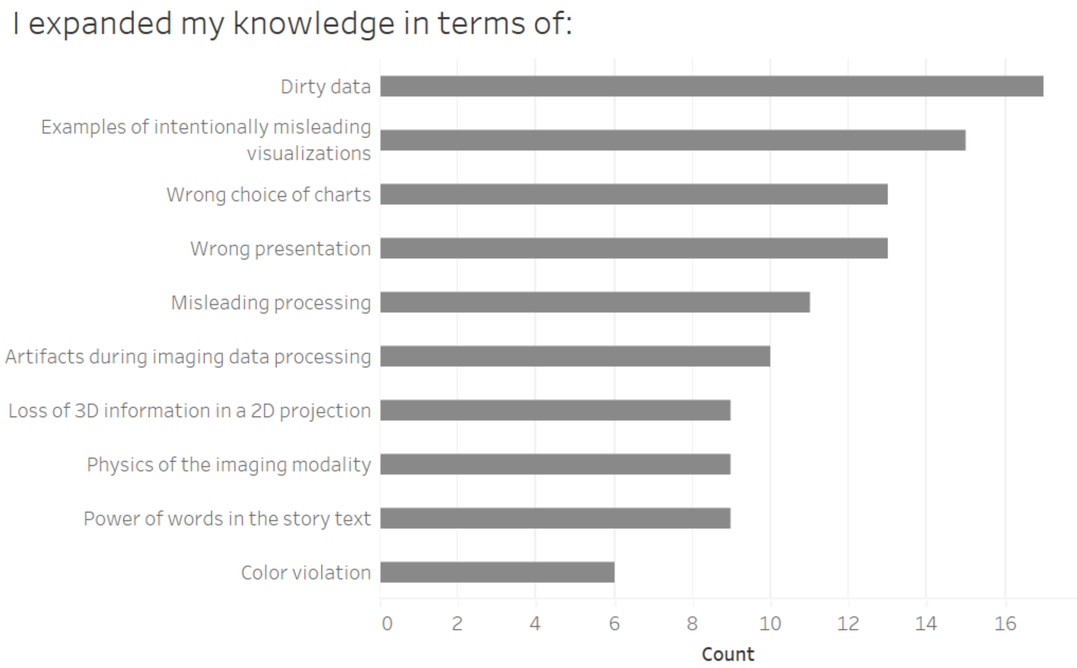



Figure 5.7: Frequencies of the topics that participants indicated that they learned a lot during the game, generated with Tableau from our user study raw data.

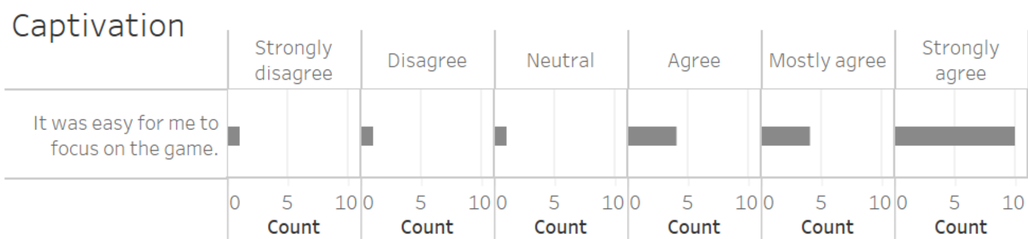



Figure 5.8: Answers to the engagement (captivation aspect) questions, generated with Tableau from our user study raw data.

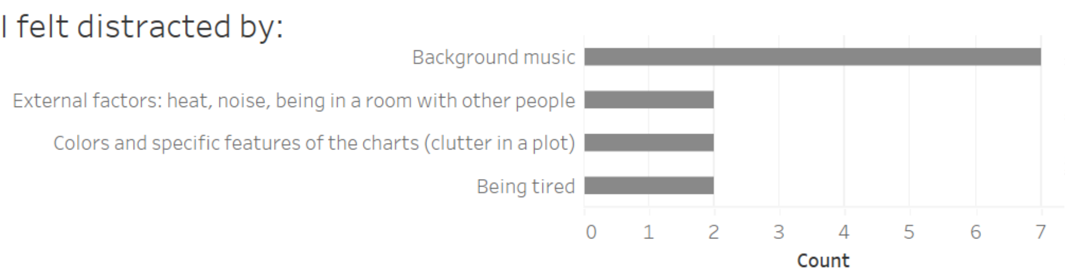



Figure 5.9: The reasons why participants felt distracted during the game, generated with Tableau from our user study raw data.

**Engagement: Captivation**. Most of the participants agreed that it was easy to focus on the game (Figure 5.8). However, 7 participants found the background music to be disturbing (Figure 5.9). During the user study, some participants changed the volume of the music and some just muted it in the game.

**Subjective Likes and Dislikes**. At the end of the questionnaire, there are two open questions about their likes and dislikes within the game respectively. The answers were sorted between several groups during the data processing step for convenience of representation. Figure 5.10 shows the general components that the participants liked. By contrast, Figure 5.11 contains information about dislikes.

The art style was the most often mentioned component by the participants as a positive aspect of the game. They also mentioned the "intuitive, not cluttered" user interface enriched with "appropriate music". Some also remarked on the idea of creating an educational game to inform about misleading visualizations in health care: "... it has a clear purpose; it is presented in an interesting way", "... I enjoyed playing the game and learning about how easy it is to fake data or its interpretation. The gamification clearly helped, but I believe it helped most in the way new knowledge was presented to the user, in a really fun way". Apart from learning new things, challenges while solving the tasks and the explanations for the assumptions were indicated to be appealing factors of the game.

A component that the participants did not like about the game, is the ambiguity of assumptions. Some participants were lacking a third option (i.e., "I don't know") when selecting whether an assumption is true or false. They also remarked game design problems included annoying art style, clutter, disturbing music, and not intuitive UI. Also, the mentioned weak reinforcement strategy included lack of interaction with the cats and inability to "... switch to the shop to get cats at all times. This would motivate for further play".

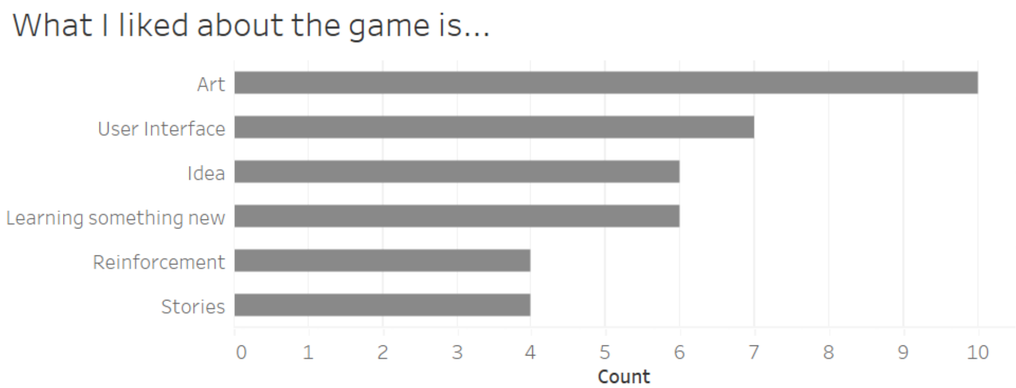



Figure 5.10: Components that participants liked about the game, generated with Tableau from our user study raw data.

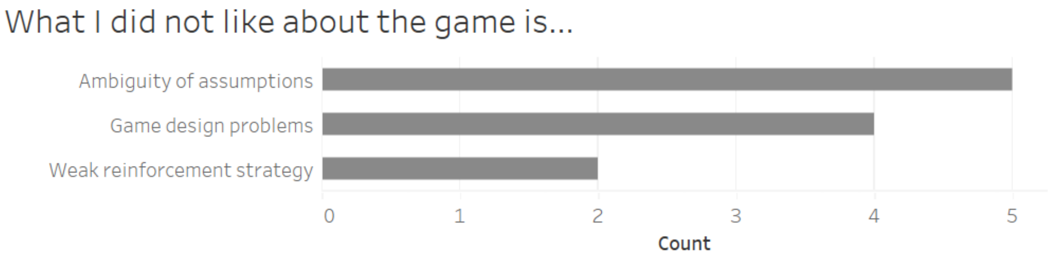



Figure 5.11: Components that the participants did not like about the game, generated with Tableau from our user study raw data.

### 5.2.2 Game Activity Results

This section contains the results of the game activity data including statistical testing. Most of the figures below include box and whisker plots. The box represents the interquartile range, the grey area with 50% of the distribution is with the median in the middle, dividing the box into the areas with different grey shades. The whiskers show the borders beyond which the outliers lay.

**Story Reading Time.** These results represent the data collected while the user was playing "DeteCATtive" (Figure 5.12). It is worth mentioning that among all tasks, tasks 5 and 7 contained the least text. Both tasks contained non-imaging data. However, the medians of time that participants needed to read the story of tasks 5 and 7 differ: 74 s (62–90 s) against 40 s (33–60 s), respectively. Task 5 contained the area chart visualizations, which were the hardest to read during the user study. By contrast, task 7 contained a stacked bar visualization that participants analyzed in task 1. The reason

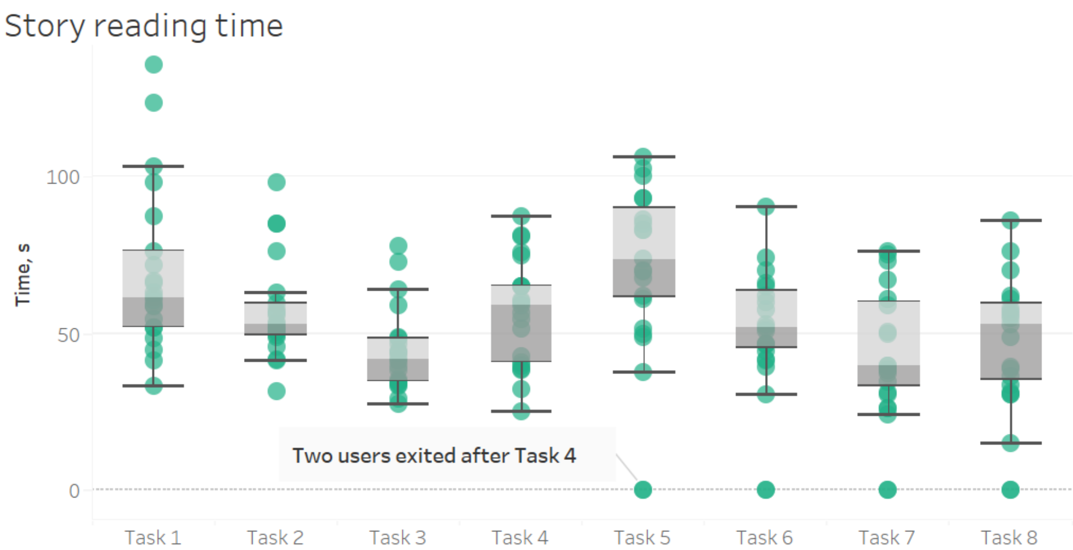



Figure 5.12: Distribution of story reading time for each task.

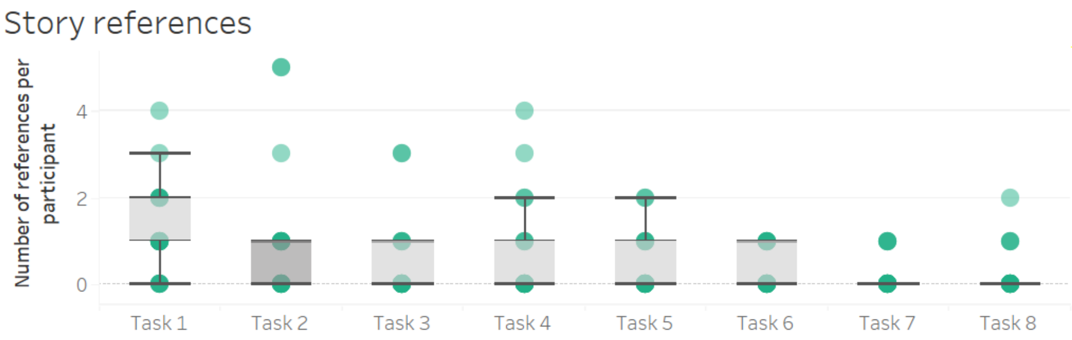



Figure 5.13: Frequency of story reference for each task, generated with Tableau from our user study raw data.

could be that the median story reading time in task 7 is shorter.

**Number of Referencing Back**. In Figure 5.13 the number of references to a story is shown. Each participant is represented with a circle with low opacity. Mostly, participants referred to the story in the first task, but later in the game, they referred less often. Maybe they became more confident during the game, thus not cross-referencing their answers. In tasks 7 and 8 the median number of referencing back to the story is zero maybe due to fatigue, boredom, or being discouraged due to previous mistakes.

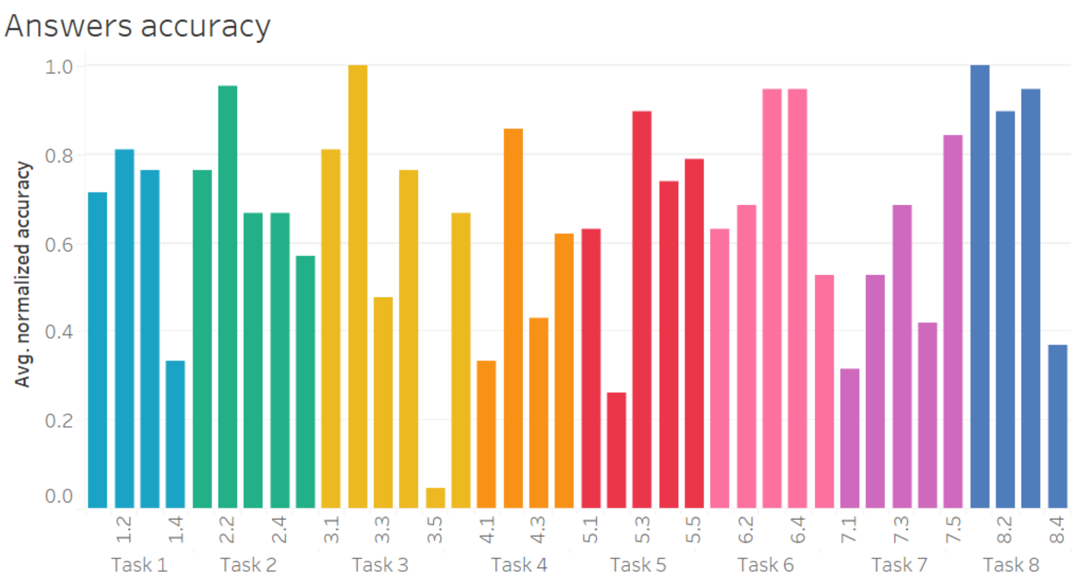



Figure 5.14: Accuracy of assessing whether an assumption is true or false for each assumption for all tasks, generated with Tableau from our user study raw data.

**Accuracy of Answers**. To calculate the accuracy of the answers, we normalized the total number of correct answers to an assumption by the number of participants: 21 participants for the first 4 tasks and 19 for the last 4 tasks. The accuracy of the answers is shown in Figure 5.14. For comparison, we divided the accuracy bar charts into two groups (Figure 5.15). The groups differ in the data used (imaging or non-imaging).

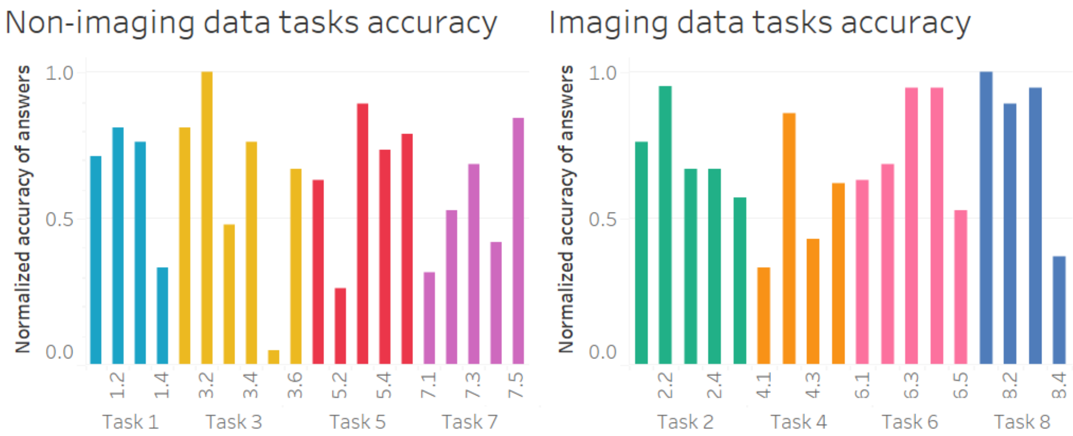



Figure 5.15: Answers accuracy separation based on the medical data used, generated for each task and assumption with Tableau from our user study raw data.

Some interesting details were found related to the assumptions assessed the most and the least correctly. As seen in Figure 5.15, assumptions 3.2 (golden, selective data, only man data are used, no woman data) and 8.1 (blueberry, fish skeleton visibility in an X-ray) were answered by all participants correctly. Regarding imaging tasks, assumptions 2.2 (jade, not labeled aorta), 4.2 (orange, X-rays in MRI), and 6.3/4 (pink, differentiation of bones and soft tissues in the CT scan) were correctly identified with a rate of 85%. The reason behind this could be the impact of prior knowledge.

The causes for low accuracy could be a lack of knowledge or attention when reading the visualization in assumptions 3.5 (golden, missing Burgenland), 5.2 (red, dirty data – Steiermark and Styria), 6.5 (pink, the 180° rotation of the CT scan), and 8.4 (blueberry, staples outside the fish), as shown in Figure 5.15. In the same figure, we see that low attention when reading the story text could cause missing some hints and answer incorrectly to assumptions 4.1 (orange, metal in MRI) and 7.1 (violet, comparison of the whole year statistics with several months one).

Assumption 7.4 (violet, the low magnitude of death) is imbalanced. There are no hints to the correct answer, so the chance to answer correctly is 50%. Also, assumption 1.4 (light blue, curvature of the cumulative bar chart) caused many ambiguities in perception, so the accuracy is lower than 50%. It is worth mentioning that the most complicated assumption was 3.5 (map missing Burgenland). Only one participant managed to notice it. This is a clear indication of a lack of attention.

**Assumption Reading Time**. The hypothesis $H_0$ is that the time of assumption reading is the same, regardless of whether an assumption is "easy" or "hard". We picked from each task the assumptions that have been assessed the most correctly and least correctly by the participants. We will hereby refer to them as "easy" and "hard", respectively. We compared the time of reading an assumption between these two groups for each task. For instance, we compare assumptions 1.2 and 1.4 for task 1 (Figure 5.16).

In task 1, one assumption reading time is not normally distributed due to the Shapiro-Wilk test for normality. Therefore, a t-test for two independent means is not applicable. Therefore, we implemented a two-tailed Mann-Whitney U test with a significance level set to 0.05. In the tests of the first four tasks, the sample size was 21, and in the last ones – 19 since two participants exited after the middle of the game.

In the Mann-Whitney U test, the normality approximation can be used if the sample size is larger than 20, so the Z-ratio can be used to calculate the p-value. In this case, when $p < .05$, the result is significant, and $H_0$ can be rejected. If the sample size is not larger than 20, then the significance of the result is determined by the critical U-value. For a sample size of 19 and a significance level of 0.05, the critical U-value equals 113. If the obtained U-value is lower than the critical U-value, then the result is significant and we reject $H_0$. We got the following results:

Mann-Whitney U test result for non-imaging tasks:

- Task 1 (1.2 vs 1.4): Z-Score = -0.25156, p-value = $.80258 > .05 \rightarrow$ not significant;
- Task 3 (3.2 vs 3.5): Z-Score = -1.86152, p-value = $.06288 > .05 \rightarrow$ not significant;
- Task 5 (5.2 vs 5.3): U-value = $84 <$ critical U-value $\rightarrow$ significant;
- Task 7 (7.1 vs 7.5): U-value = $129 >$ critical U-value $\rightarrow$ not significant;

Mann-Whitney U test result for imaging tasks:

- Task 2 (2.2 vs 2.5): Z-Score = 0.17609, p-value = $.85716 > .05 \rightarrow$ not significant;
- Task 4 (4.1 vs 4.2): Z-Score = -0.47796, p-value = $.63122 > .05 \rightarrow$ not significant;
- Task 6 (6.3 vs 6.5): U-value = $178 >$ critical U-value $\rightarrow$ not significant;
- Task 6 (6.4 vs 6.5): U-value = $71.5 <$ critical U-value $\rightarrow$ significant;
- Task 8 (8.1 vs 8.4): U-value = $138 >$ critical U-value $\rightarrow$ not significant.

Finally, in most cases, there was no significant difference. So we do not reject the null hypothesis, meaning that *the assumption reading time does not relate to how easy or hard an assumption is.*

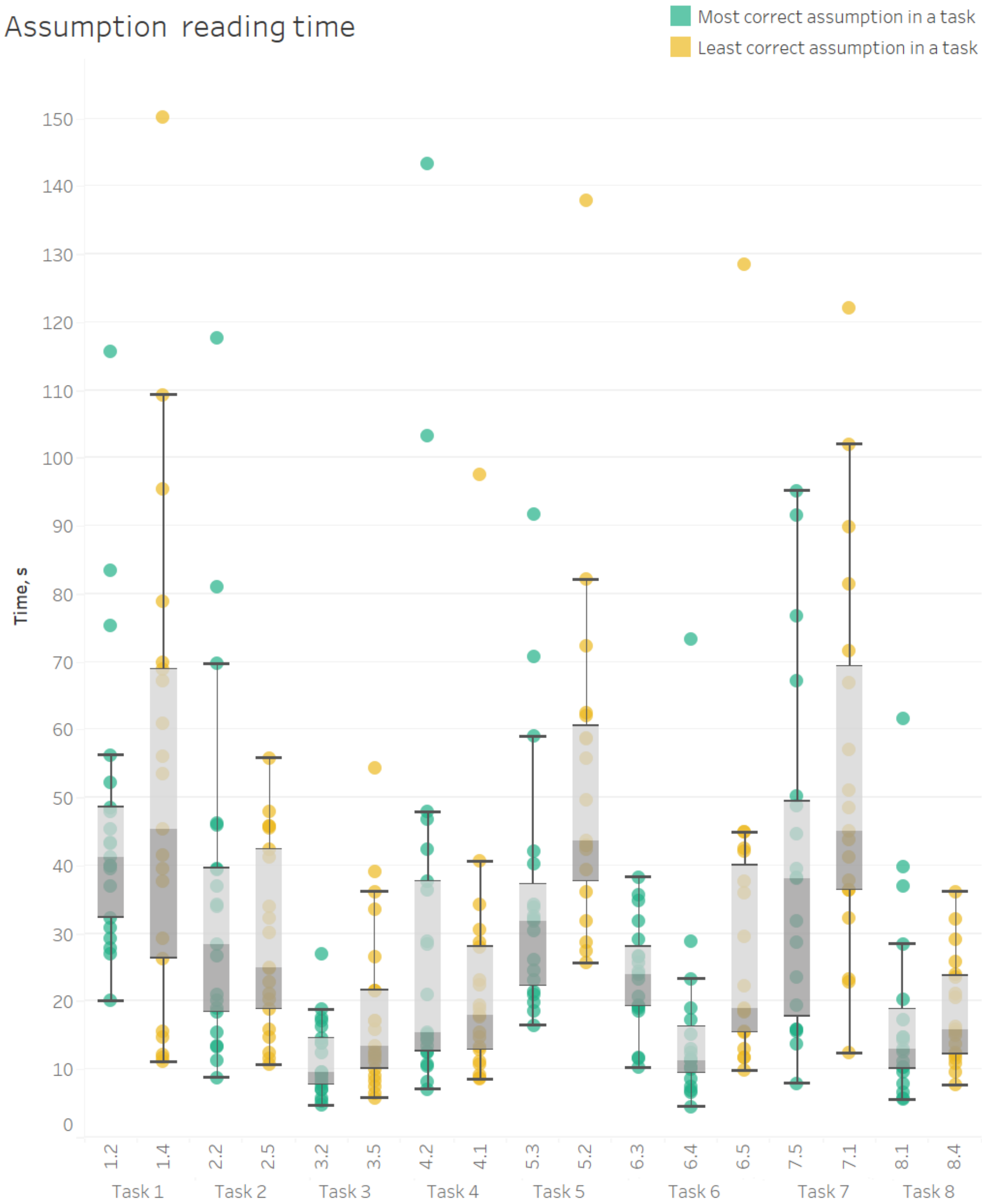



Figure 5.16: Distribution of assumption reading time for “easy” (i.e., most correct) and “hard” (i.e., least correct) assumptions in each task, generated with Tableau from our user study raw data.

**Explanation Reading Time**. In this case, the null hypothesis $H_0$ is that the time of explanation of an assumption reading is the same, regardless of whether an assumption is "easy" or "hard". For the explanation reading, we also used the Mann-Whitney U test with the same conditions as for the assumption reading time. The explanation reading time of selected "easy" or "hard" assumptions for each task are shown in Figure 5.17.

We checked the significant difference between the explanation reading times of "easy" and "hard" assumptions in each task. If the difference is not significant, then the null hypothesis $H_0$ is not rejected and the explanation time is not influenced by the accuracy of the answer. We obtained the following results:

Mann-Whitney U test result for non-imaging tasks:

- Task 1 (1.2 vs 1.4): Z-Score = -2.42753, p-value = .0151 $<$ .05 $\rightarrow$ significant;
- Task 3 (3.2 vs 3.5): Z-Score = -2.44011, p-value = .01468 $<$ .05 $\rightarrow$ significant;
- Task 5 (5.2 vs 5.3): U-value = 47.5 $<$ critical U-value $\rightarrow$ significant;
- Task 7 (7.1 vs 7.5): U-value = 100 $<$ critical U-value $\rightarrow$ significant;

Mann-Whitney U test result for imaging tasks:

- Task 2 (2.2 vs 2.5): Z-Score = -1.29552, p-value = .1936 $>$ .05 $\rightarrow$ not significant;
- Task 4 (4.1 vs 4.2): Z-Score = 2.28917, p-value = .02202 $<$ .05 $\rightarrow$ significant;
- Task 6 (6.3 vs 6.5): U-value = 41.5 $<$ critical U-value $\rightarrow$ significant;
- Task 6 (6.4 vs 6.5): U-value = 26.5 $<$ critical U-value $\rightarrow$ significant;
- Task 8 (8.1 vs 8.4): U-value = 54 $<$ critical U-value $\rightarrow$ significant.

As a result of the Mann-Whitney U test, the null hypothesis $H_0$ can be rejected for all tasks except for task 2. In other words, *the least correctly assessed assumptions explanations were thought over longer than the most correct ones*. Generally, this indicates that if a participant makes a mistake, the game engages the person to spend more time reading the explanation to the respective assumption. For task 2, we see no significant difference in the statistical test because the answers contain self-explanatory images with labels and short explanation texts.

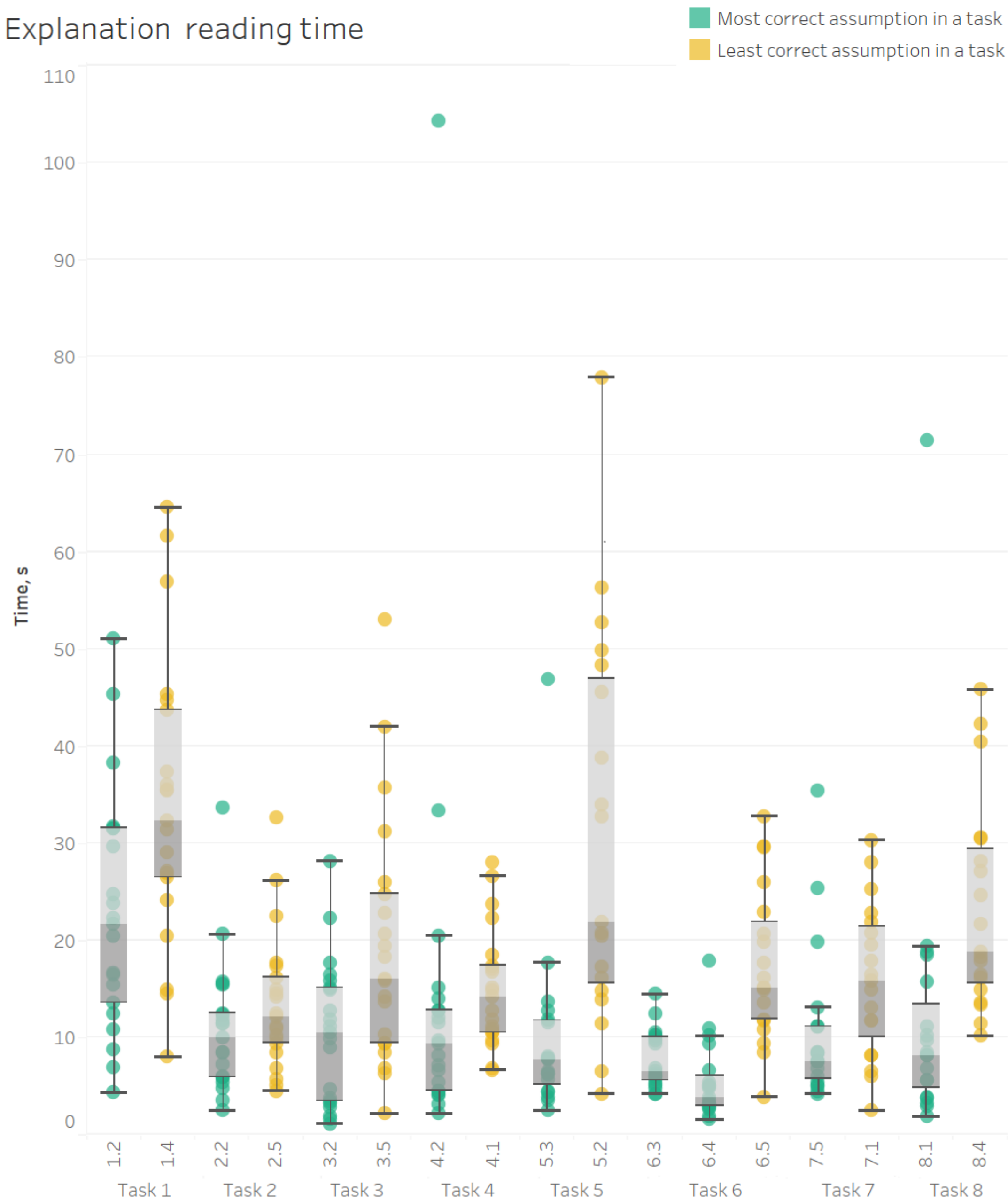



Figure 5.17: Distribution of explanation reading time for "easy" (i.e., most correct) and "hard" (i.e., least correct) assumptions in each task, generated with Tableau from our user study raw data.

## 5.3 Limitations

In this section, we discuss the limitations and potential improvement directions of our work. The limitations are distributed between three categories: user study, tasks, and game design, including UI and reinforcement system.

The *user study limitations* are primarily due to the small number of participants and low coverage of older age groups. Also, some results are based on the subjective feelings of the participants and can be biased. For example, it is very probable that many participants did not exit earlier not to be rude towards the study conductor.

Regarding the *tasks*, several participants commented on ambiguous assumptions. Some wanted to have a third option, instead of only true and false (for example, "I do not know"). A problem with a third option is how to assess correctness. Additionally, a third option could lower the motivation of players and can more probably lead to a situation when the goal of the game is not achieved. Moreover, some assumptions needed basic knowledge of Austrian geography. This was a limitation when participants were not familiar with the geographical aspects of Austria. We would also like to improve the tasks by adding more types of graphs in non-imaging tasks and making cinematic renderings of the medical imaging data.

A major limitation of the game design is no replayability. The current *game design* could benefit if the reinforcement was combined with interactivity. This could improve the retention of players (prevent them from exiting the game earlier). For example, at any time it should be possible to switch between the test and the final scene, which includes cats in boxes, results and achievements, and the garden views. This could allow participants to use already collected points for the cats. In other words, the players could see how the garden gets filled with cats, as they progress in the test. Also, the reinforcement strategy could include interaction with the cats at the end. Moreover, for the game end, one participant has also emphasized to "offer participants a list of all the tricks that can be used when visualizing data to spread awareness".

Based on the user feedback, another game limitation is that the reward loop was too long. This could be improved, for instance, by dividing the test into four subsets and giving rewards after the end of each subsection. Also, one participant offered "probably displaying a number of questions", so that it would be possible to track the progression within a task. Also, the badges should be able to be achieved at early game stages because in "DeteCATive", the first badge achievement possibility appears only after task 5. Additionally, the amount of points needed to rescue all cats should be reduced. In "'DeteCATive", a player should finish the test with 100% accuracy to get all cats, which is barely possible.

CHAPTER 6

# Conclusion

This chapter includes a summary of the work and a discussion of future research directions.

**Summary**. Misleading visualizations can appear while browsing the internet or reading an advertisement on the street. They can lead to inaccurate interpretations and insights, which can potentially cause harm to people. In a user study conducted by Zheng et al. [ZM22], the most commonly occurring reason for consumers being misinformed relates to a lack of attention to misleading elements. In this thesis, we explored (intentionally and unintentionally) misleading visualizations in the medical field and proposed a solution that improves the visual literacy of the general public.

Our work methodology included open-coding and educational approaches. Within the open-coding approach, we collected cases of misleading elements in medical visualizations of non-imaging and imaging data. Then we used five types of uncertainty, four of these proposed by Griethe et al. [GS05] and the fifth being incompleteness [ANI+20]. These uncertainty types, found in each step of the medical visualization pipeline, are the source of misleading elements. Additionally, we assess the potential intent behind their presence within the medical visualization pipeline.

Our educational approach aimed to combat these uncertainties including storytelling and gamification. The product is an educational game, "DeteCATive". In this game, players try to identify the misleading elements in 8 amusing and fictional stories created with specific intent. These stories imitate newspaper snippets containing a visualization of medical data filled with uncertainties. For each story, there is a list of accompanying assumptions. The players have to identify whether they are true or false and they gather points and badges, as rewards. Then, the players get the explanations for each assumption that focus the attention on the misleading element and its influence on the visualization perception.

to assess our game, a user study was conducted on 21 participants, who were mostly in their 20s. In a questionnaire, we checked for memorability, reinforcement, engagement, and preferences of the participants. The participants mostly commented on the medical imaging stories after playing the game, but not as detailed as the non-imaging story about melanoma. Reinforcement results demonstrated that the players wanted to follow the game objective, e.g., to foster all cats. All participants, except for two, finished all 8 tasks, mostly because they were curious to know more misleading stories. The second frequent reason to continue after the middle of the game was to gain more points, thus indicating the retaining performance of the reward system. The game engagement, divided into aesthetics, cognitive involvement, and captivation got mostly positive feedback in all three categories. The color violations in the visualization were well-known by most of the participants, while 17 participants commented that they expanded their knowledge about dirty data. Moreover, participants commented most often on the appealing art style of the game. Also, the ambiguity of assumptions in the game was most commonly indicated as what they did not like.

The game activity files indicated that the story with a confusing visualization including a cluttered stacked area chart took the longest to read. The accuracy statistics provided evidence that players get misled when assessing assumptions because visualization elements lead them to the story misinterpretation. The least correct assumption was about a missing province on the map of Austria, which is small and located close to the borders indicating errors due to lack of attention. While assumption reading showed no correlation with the assumption accuracy, explanation reading time increased for the least correct assumptions.

Our work has two contributions. The first one is a taxonomy of five uncertainty types (error, imprecision, non-specificity, incompleteness, and subjectivity) in the medical visualization pipeline arising from potential intents, which gives the answers to our first research question "Which types of uncertainty arise in the medical visualization pipeline and whether there is any intent behind those?". The second contribution is the design, development, and assessment of the educational game "DeteCATive", which communicates misleading visualizations enriched with uncertainty types arising from an intent to the general population. This game is the answer to our second research question "How can we inform the general population about the existence of visualization uncertainty?".

**Future Work**. As general directions, we propose investigating potential correlations between participants' information (age group, prior knowledge in medicine/biology, and experience in visualization) and the detection of a specific uncertainty type. In our user study, the participants sample is too small for such statistical tests. Moreover, it could be exciting to dive deeper into the intentions field and investigate which uncertainty is most commonly used for a specific intent, e.g., for political gains or lobbyism. Additionally, further investigations are needed on which uncertainty type is harder to detect in a misleading visualization.

To improve the existing solution, we would offer to switch the platform of the game from Windows to WebGL, so that the game is easier to distribute and play since only an internet connection would be required. Additionally, to expand the content, more stories can be invented and there are plenty of resources in the wild that can be used for further inspiration. Moreover, one could add interactivity to visualizations like zooming and highlighting variables similar to the Tableau interface. All limitations and improvement directions mentioned provide a place for researchers to improve our work since it is the first step towards an educational game about misleading visualizations in health care based on uncertainty types in the medical visualization pipeline.

# List of Figures


1.1 The piece of Dr. John Snow's map of London, where the hash marks represent the number of deaths at a specific address. The source of cholera, a contaminated pump, is highlighted in red. The screenshot originates from "John Snow and the 1854 Broad Street cholera outbreak" YouTube teaser video to the Harvard University online course: "PredictionX: John Snow and the Cholera Epidemic of 1854". . . . . . . . . . . . . . . . . . . . . . . . 2
1.2 Charts, which were used in a tweet about the inefficiency of COVID-19 vaccination. Original figures from [LPLK22], reconstructed in Our World in Data to achieve better image quality. . . . . . . . . . . . . . . . . . . . . . . 3

2.1 Each step of the VA pipeline creates an opportunity for errors caused by visualization creator choices to emerge, thus giving way to visualization mirages [MKC20]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
2.2 Example of inductive reasoning provided by Lisnic et al. [LPLK22]. In this case, two premises weakly support the general conclusion and provide no formal logical fallacies. To invalidate this conclusion, one needs to search for omitted facts. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
2.3 Five general taxonomies of dirty data combined into one by Gschwandtner et al. [GGAM12]. Multiple sources refer to a situation when multiple data sources have to be integrated/joined. . . . . . . . . . . . . . . . . . . . . . . 9
2.4 Medical visualization pipeline according to Ristovski et al. [RPHL14]. Each box represents a step (or substep), where uncertainty can appear. . . . . . 10
2.5 Steps within medical imaging, including image acquisition, transformation, or visualization. An uncertainty-aware visualization may be required to assess the uncertainties introduced at these steps and to communicate them to a clinician [GSWS21]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
2.6 Taxonomy of uncertainty sources in medical imaging [GSWS21]. The green colored lines coincide with the image acquisition step, while the orange and blue lines refer to the image transformation and visualization steps, respectively. 11
2.7 a) Spazio La Stampa exhibition in Turin, b) ExplorAR tablet with augmented reality application visualizing a dinosaur from its skeleton at National Museum Cardiff in Wales. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.8 Fragments from the graphic-based information brochure about coronary angiography procedure [BGH+19]. The fragments include pieces of information about the procedure details and possible complications during and after the intervention. The full version of the brochure is available on Annals of Internal Medicine. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

2.9 Fragments from top: pelvic fracture story, bottom: aneurysm storyboard. The stories include detailed anatomy with 3D models, condition-associated risks, and aspects of treatment [MGS+21]. . . . . . . . . . . . . . . . . . . . 13

2.10 Template proposed by Meuschke et al. to communicate disease data using a seven-step template in narrative medical visualization [MGS+21]. . . . . . 14

2.11 Complete liver cancer story [MGS+22]. . . . . . . . . . . . . . . . . . . . . 14

2.12 Games promoting visualization literacy. Top: do-it-yourself crafting game for learning through creation about visualizations [BVY+22], bottom left: role-playing game for learning how to read visualizations [HNGC20], bottom right: two card games for visualization design education [AGR21]. . . . . 15

2.13 Assembled nested papercraft of a human head (left) that reveals the different anatomical substructures, when color filters are used (right). Edited from [SKRW22]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

3.1 Medical data structures selected for the purposes of this work. . . . . . . 18

3.2 Fused PET and CT data used for arthritis inflammation assessment in a mouse. The CT scan delivers the anatomical information for the PET scan, while the SUV – standardized uptake value of [$^{18}$F]-FDG that shows metabolic activity. Edited from [Gro]. . . . . . . . . . . . . . . . . . . . . . . . . . . 19

3.3 Medical data visualization pipeline, inspired and adapted from [RPHL14, SPBR20, GSWS21, LGS+22]. Black blocks: non-imaging/imaging data visualization pipeline steps. Grey blocks: processes or attributes describing each step of the visualization pipeline. Colored blocks: uncertainty types categories used to sort misleading elements appearing at each pipeline step. . . . . . 21

3.4 ■ Receiver operating characteristic (ROC) curves for each individual marker showing different performance for lung cancer prediction [BPDCRG+17]. 23

3.5 Hints to reading ROC curves. . . . . . . . . . . . . . . . . . . . . . . . . . 23

3.6 ■ Crude bias (a measure of divergence of the data with imputed values from the true data) plotted against the percentage of missing data in the true data set for 5 different imputation methods. Edited from [FHG21]. . . . . . . . 25

3.7 ■ Different sample distributions can lose features when represented as a bar chart. The bar chart on the left obscures that each sample size is small. Observe, for example, that there is a large variance in group 1 and an outlier in group 3. These features are clearly visible in the scatter plot on the right. Edited from [WMWG14]. . . . . . . . . . . . . . . . . . . . . . . . . . . 26

3.8 ■ Comparison of ranking star rating of two movies (cases 5 and 6) by different statistical tests. The ordered probit model (top) describes the ordinal data more accurately. When applying the ordered probit test, the resulting difference in mean between the two movies is in the opposite direction to the outcome of the metric test (bottom) [LK18]. . . . . . . . . . . . . . . . . . 26
3.9 ■ Map of the United States of America with confirmed cases of an unknown disease. While data are not normalized on the population of the states, the hotspots are located in highly populated areas [LGS+b]. . . . . . . . . . . 27
3.10 ■ Comparison of COVID-19 cases incidence between age groups, not accounting for demographical distribution (there are more young people than elderly people) [LGS+c]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
3.11 ■ The COVID-19 death curve in Sweden, where the COVID-19 restrictions were relatively mild. The author of this visualization was arguing against COVID-19 restrictions by providing a single piece of evidence. A comparable neighbor with fewer deaths or with soft measures but more deaths is missing [LPLK22]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
3.12 ■ Visualization of COVID-19 statistics where a continuous variable (number of confirmed cases) has been discretized and ■ hardly differentiable colors were used [LGS+a]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
3.13 ■ Visualization with inconsistent tick intervals (random time ranges) and low time resolution, resulting in imprecision of extracted values for a specific year [LGS+f]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
3.14 ■ Visualization with data of different magnitudes, leading to a low resolution of data with a smaller magnitude (deaths and recoveries) [LGS+d]. . . . . 30
3.15 ■ Visualization with dual and truncated axes that make the orange graph look below the blue one, even though the orange line values are higher [LPLK22]. 31
3.16 ■ Visualization with missing title, labels, and annotations. It is impossible to know what the infographic is about [LGS+22]. . . . . . . . . . . . . . . . 31
3.17 ■ Visualization with cluttering due to line overplotting [LGS+e]. . . . . . 31
3.18 ■ Challenges when creating a color map with perceptual uniformity, when using a rainbow color map. Comparison of (a) "unscientific" color map (rainbow) and (b) "scientific" color map (batlow) shows over- or underestimation of incremental contrast of the rainbow color map. A sequential ordering of the color maps is shown in (c) and (d) [CSH20]. . . . . . . . . . . . . . 33
3.19 ■ Visualization of disability-adjusted life-years (DALYs) with ineffective representation that leads to imprecision when trying to obtain scalar values [C+22]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.20 ■ Bar chart visualization that hides the underlying data distributions causing a lack of data sample distinction. Edited from [WMWG14]. . . . . . . . . 34
3.21 ■ A humorous visualization pointing out the lack of visual hierarchy in infographics, where the enumerated consequential information units are ordered chaotically [MTV]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

3.22 Visualization with pictorial area encoding. The height of the silhouette with 30.1x (brown) looks only two times higher than the 5.2x (green), but the area should be 6 times larger [LGS+g]. . . . . . . . . . . . . . . . . . . 35

3.23 Variety of examples of CT artifacts. Top left: motion, top right: beam hardening, bottom left: ring (pointed by an arrow), bottom right: metal artifacts. Edited from [PSHK08]. . . . . . . . . . . . . . . . . . . . . . . . 38

3.24 Partial volume correction in PET scans. Partial volume correction reduces the imprecision of values in voxels. Top: original scan, middle: after partial volume correction, bottom: T1-weighted MR scan used in partial volume correction. The color legend represents regional standard uptake value ratios [YHG+17]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

3.25 The red-eared slider examined via CT (left) and MR (right) techniques. The skeleton is captured in a CT scan, a slice of which is depicted in (b), and finally rendered in (g). Due to the injection of contrast substances, the vascular system can be visualized by CT (d) and MR (e). The soft tissues are visible in the MR scan (c), while the CT scan contains the lungs due to air-tissue radiodensity difference [LHW+11]. . . . . . . . . . . . . . . . . . 39

3.26 MR scan of a whiteknee tarantula gastrointestinal tract. The structures are enhanced due to feeding the tarantula a contrast medium-filled cockroach [LHW+11]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

3.27 MR scans of the abdomen using different pulse sequences: (A) T2-weighted, (B) T1-weighted out-of-phase, (C) T1-weighted in-phase, (D) Diffusion-weighted imaging. The structure shown on the left is the liver because the CT image was captured in the direction from feet to head. Edited from [RCF+21]. . 40

3.28 Oversegmented putamen treated by segmentation editing. (a) 3D PET image, (b) oversegmented putamen, (c) voxels with the white matter as the second best choice in the segmentation editing widget, (d) classification after two editing iterations, (f) segmentation editing widget. Edited from [SMH10]. 42

3.29 Heat map of voxels registration outcomes. Red voxels (corresponding to 1 in the color legend) were correctly located in 50 perturbed PET-MRI registrations. Voxels with other colors varied their position or were absent across perturbations, indicating PET-MRI rigid registration imprecision [S+17]. 43

3.30 Results of lung segmentation that does not account for pathological variability. Above: CT scan slices, bottom: segmentation results [LZXP20]. . 43

3.31 Confidence scatter plot of prostate, bladder, vesicles, and rectum segmentation outcome. The scatter plot is based on mean error $\mu$ and the standard deviation $\sigma$. Segmentation outcomes with high standard deviation correspond to areas left undetected (cyan). Edited from [RMB+16]. . . . . . . . . . 44

3.32 Results of different segmentation techniques (a–f) for retinal vasculature indicating the influence of parametrization on the segmentation accuracy. Edited from [BAKS16]. . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

3.33 Comparison of interpolation methods. The left one (no interblock interpolation) contains block boundary artifacts, removed by the maximum distance method on the left. Sampling artifacts present in both images [LLY06]. . 46
3.34 The volume rendering pipeline contains four steps: ray casting, sampling, classification, and composing. Samples along the ray contribute to the composed value depending on their material properties. The interpolation method (type and order) of all sample values along the ray results in uncertainties. Edited from [SPBR20]. . . . . . . . . . . . . . . . . . . . . . . . . . . 47
3.35 Flattening technique used for aneurysm visualization. The aneurysm dome is cut at the aneurysm's neck (ostium) region (a). The surface of the dome is unfolded (b). The red color represents the areas with a high risk of rupture (wall thickness dependent) [MVB+16]. . . . . . . . . . . . . . . . . . . . 47
3.36 The smoothing of vessels (left). Due to the discontinuous nature of radiological data, the rendering contains a disrupted vessel representation without smoothing (right). Color coding reflects the diameter of a vessel [HPSP01]. 48
3.37 Volume renderings of intestines: (a) without halos, (b) with shadow-like and (c) with semi-transparent halos, (d) without shading, but with a smooth white and black contour [BG07]. Shape and depth perception might be affected by different representations. . . . . . . . . . . . . . . . . . . . . . . . . . . 48
3.38 Comparison of color legends and the way they are seen by color-blind people and people with color vision deficiency [CSH20]. . . . . . . . . . . . . . . 50
3.39 Pelvis and femora CT scan rendering of a patient with osteoporosis that caused a hip fracture. The color legend encodes the cortical thickness and poses perceptional uncertainty when trying to obtain an exact value for a specific color [PTM+12]. . . . . . . . . . . . . . . . . . . . . . . . . . . 51
3.40 CT scan rendering with specular reflection of blood vessels and isotropic phase function used for bones to minimize the uncertainty of material perception [JKRY12]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
3.41 Copy-move forgery in medical images. Top: original CT, MRI, PET, ultrasound, and digital X-ray scans. Bottom: forged images with different types of attacks. Red colored rectangles indicate tampered regions [DD22]. . . . . 54
3.42 The motivation behind attacking medical imaging data by manipulation. Targeted attacks aim for a specific patient, while untargeted attack has no specific target patient [MMSE19]. . . . . . . . . . . . . . . . . . . . . . . . 54
3.43 The cycle of interconnection between stories, uncertainty types, and assumptions used within our proposed approach. . . . . . . . . . . . . . . . . . . 56
3.44 Task 1 story with COVID-19 statistics used to support public health initiatives. The short answers to assumptions are in grey. . . . . . . . . . . . . . . . . . 58
3.45 Uncertainty types in the misleading visualization of task 1. . . . . . . . . 58
3.46 Task 2 story with undersegmented liver used to create a health warning. The short answers to assumptions are in grey. . . . . . . . . . . . . . . . . . . 59
3.47 Uncertainty types in the misleading visualization of task 2. . . . . . . . . 59

3.48 Task 3 story with melanoma incidence in Austrian provinces used to persuade people to buy real estate. The short answers to assumptions are in grey. . . 60
3.49 Uncertainty types in the misleading visualization of task 3. . . . . . . . . 60
3.50 Task 4 story with noise in an MR scan used to create sensation and draw attention to work safety measures. The short answers to assumptions are in grey. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.51 Uncertainty types in the misleading visualization of task 4. . . . . . . . . 61
3.52 Task 5 story with breast cancer statistics used to generate attention toward the rise of breast cancer cases. The short answers to assumptions are in grey. 62
3.53 Uncertainty types in the misleading visualization of task 5. . . . . . . . . 62
3.54 Task 6 story imitating dextrocardia in a human breast CT slice used to raise the selling rates. The short answers to assumptions are in grey. . . . . . . 63
3.55 Uncertainty types in the misleading visualization of task 6. . . . . . . . . 63
3.56 Task 7 story with COVID-19 statistics used to support a narrative for political gain. The short answers to assumptions are in grey. . . . . . . . . . . . . 64
3.57 Uncertainty types in the misleading visualization of task 7. . . . . . . . . 64
3.58 Task 7 story with a 2D topogram of a fish X-ray scan used to draw attention to lake pollution. The short answers to assumptions are in grey. . . . . . 65
3.59 Uncertainty types in the misleading visualization of task 8. . . . . . . . . 65
3.60 Stray cats that can become foster cats at the end of “DeteCATive” are used as the main objective to engage and reward the audience. . . . . . . . . . 66
3.61 Badges rewarding system. . . . . . . . . . . . . . . . . . . . . . . . . . . 67
3.62 Screenshot of a view when a player achieves a badge. . . . . . . . . . . . . 68

4.1 Tableau interface with the visualization for task 1. Cumulative data from Vienna is used for this bar chart. . . . . . . . . . . . . . . . . . . . . . . . . 71
4.2 Tableau interface with the visualization for task 7. COVID-19 data are summed up over three years: 2021, 2022, and 2023. Vorarlberg is excluded from the visualization. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
4.3 Tableau interface with the visualization for task 3. Four filters are applied to the data: neoplasm location, sex, province of residence, and reporting year. Burgenland is excluded from the provinces of residence list. . . . . . . . . 72
4.4 Tableau interface with the visualization for task 5. Filters are applied to the data – only malignant neoplasm of breast data for women is shown. The data of Styria is divided between two variables: Steiermark and Styria to create dirty data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
4.5 MeVisLab network for task 2. Organs are extracted from the original data via region growing and close gap approaches. Then indirect volume rendering is employed for the organs while the liver is surface rendered to hide the undersegmentation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

4.6 MeVisLab networks for task 4. The left-most network was used to create the noise inside the MR scan. The middle network was used to create a 3D volume rendering of a head impaled with a rod. For the explanatory visualizations, the right-most network was used. . . . . . . . . . . . . . . . . . . . . . . . 75
4.7 MeVisLab network for task 6. A SoClipBox allowed the creation of a slice out of the volume model, which then was rotated 180 degrees in the SoExaminerViewer to simulate a fake dextrocardia case. . . . . . . . . . . . . . . . . . 75
4.8 MeVisLab networks for task 8. The left-most network used CT data to create misleading visualization. For the explanatory visualizations, the right network used the 3D CT scan data. . . . . . . . . . . . . . . . . . . . . . . . . . 76
4.9 Screenshot of welcome view of the start scene of the educational game "DeteCATive". . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
4.10 Screenshots of a detective table view of the start scene of the educational game "DeteCATive". . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
4.11 Screenshots of the main scene of the education game "DeteCATive": overall instruction view (above) and task 2 (below). . . . . . . . . . . . . . . . . . . 79
4.12 Screenshots of the main scene of the education game "DeteCATive": task 4, assumption 1 (above) and the explanation to it (below). . . . . . . . . . . 80
4.13 Screenshots of final scenes of the education game "DeteCATive": choosing foster cats (above) and final results and awards views (below). In this case, a player collected 26 points and achieved two badges: "Eagle-eye" and "Scholar". Cats can be picked by spending collected points. The green squares under each task represent a correctly assessed assumption. . . . . . . . . . . . 82

5.1 The distribution of age among the 21 user study participants. . . . . . . . 86
5.2 Answers to the memorability aspect questions, generated with Tableau from our user study raw data. On the left, the box and whisker plot contains a grey box representing the interquartile range split by median inside and the whiskers show the outliers borders. . . . . . . . . . . . . . . . . . . . . . . . 87
5.3 Answers to the reinforcement aspect questions, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88
5.4 Reasons that led the participants to continue the game after the middle point was reached, generated with Tableau from our user study raw data. . . . 88
5.5 Answers to the engagement (aesthetic aspect) questions, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . 88
5.6 Answers to the engagement (cognitive involvement) aspect questions, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . 89
5.7 Frequencies of the topics that participants indicated that they learned a lot during the game, generated with Tableau from our user study raw data. . 90
5.8 Answers to the engagement (captivation aspect) questions, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . 90
5.9 The reasons why participants felt distracted during the game, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . 90

5.10 Components that participants liked about the game, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . 91
5.11 Components that the participants did not like about the game, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . 92
5.12 Distribution of story reading time for each task. . . . . . . . . . . . . . . 92
5.13 Frequency of story reference for each task, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
5.14 Accuracy of assessing whether an assumption is true or false for each assumption for all tasks, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
5.15 Answers accuracy separation based on the medical data used, generated for each task and assumption with Tableau from our user study raw data. . . . 94
5.16 Distribution of assumption reading time for "easy" (i.e., most correct) and "hard" (i.e., least correct) assumptions in each task, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . 96
5.17 Distribution of explanation reading time for "easy" (i.e., most correct) and "hard" (i.e., least correct) assumptions in each task, generated with Tableau from our user study raw data. . . . . . . . . . . . . . . . . . . . . . . . . . 98

# List of Tables


3.1 Summary of medical imaging modalities, indicating their working principle and examples of target conditions. . . . . . . . . . . . . . . . . . . . . . . . 20
3.2 Uncertainty type cases in the data acquisition step of the non-imaging medical data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . . . . 22
3.3 Uncertainty type cases in the transformation and processing step of the non-imaging medical data visualization pipeline. . . . . . . . . . . . . . . 24
3.4 Uncertainty type cases in representation step of the non-imaging medical data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
3.5 Uncertainty type cases in human processing and interaction step of the non-imaging medical data visualization pipeline. . . . . . . . . . . . . . . . . . 32
3.6 Uncertainty type cases in data acquisition step of the medical imaging data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
3.7 Uncertainty type cases in transformation and processing step of the medical imaging data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . 41
3.8 Uncertainty type cases in representation step of the medical imaging data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
3.9 Uncertainty type cases in human processing and interaction step of the medical imaging data visualization pipeline. . . . . . . . . . . . . . . . . . . . . . . . 49
3.10 Introduction of uncertainties to non-imaging medical data at different visualization pipeline steps. Legend: ✓: intentional, ◯: not known intention. . 52
3.11 Introduction of uncertainties to medical imaging data at different visualization pipeline steps. Legend: ✓- intentional, ◯ - not known intention, × - not intentional. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

# Bibliography


[AGR21] Lorenzo Amabili, Kuhu Gupta, and Renata Georgia Raidou. A taxonomy-driven model for designing educational games in visualization. *IEEE Computer Graphics and Applications*, 41(6):71–79, 2021.

[ANI+20] Shiva Alemzadeh, Uli Niemann, Till Ittermann, Henry Völzke, Daniel Schneider, Myra Spiliopoulou, Katja Bühler, and Bernhard Preim. Visual analysis of missing values in longitudinal cohort study data. In *Computer Graphics Forum*, volume 39, pages 63–75. Wiley Online Library, 2020.

[APT+22] Nikolas Aristokli, Irene Polycarpou, Sophia C. Themistocleous, Dimitris Sophocleous, and Ioannis Mamais. Comparison of the diagnostic performance of magnetic resonance imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: a review. *Radiography*, 28(3):848–856, 2022.

[ARB04] Ilka Antcheva, Fons Rademakers, and Rene Brun. Guidelines for developing a good GUI. *Computing in High Energy Physics and Nuclear Physics*, page 613, 2004.

[AZP+20] Nouf Alrasheed, Arun Zachariah, Shivika Prasanna, Deepthi Rao, and Praveen Rao. Deepfakes for histopathology images: myth or reality? In *Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–7. IEEE, 2020.

[BAKS16] Khan BahadarKhan, Amir A. Khaliq, and Muhammad Shahid. A morphological Hessian based approach for retinal blood vessels segmentation and denoising using region based Otsu thresholding. *PLOS One*, 11(7), 2016.

[BF12] F. Edward Boas and Dominik Fleischmann. CT artifacts: causes and reduction techniques. *Imaging in Medicine*, 4, 2012.

[BG07] Stefan Bruckner and Eduard Gröller. Enhancing depth-perception with flexible volumetric halos. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1344–1351, 2007.

[BGC22] Joshua Bierbrier, Houssem-Eddine Gueziri, and D. Louis Collins. Estimating medical image registration error and confidence: a taxonomy and scoping review. *Medical Image Analysis*, 2022.

[BGH+19] Anna Brand, Linde Gao, Alexandra Hamann, Claudia Crayen, Hannah Brand, Susan M Squier, Karl Stangl, Friederike Kendel, and Verena Stangl. Medical graphic narratives to improve patient comprehension and periprocedural anxiety before coronary angiography and percutaneous coronary intervention: a randomized trial. *Annals of Internal Medicine*, 170(8):579–581, 2019.

[BI07] David Borland and Russell M. Taylor Ii. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17, 2007.

[BKN+08] Gunnar Brix, Heinrich Kolem, Wolfgang R. Nitz, Michael Bock, Alexander Huppertz, Cristoph J. Zech, and Olaf Dietrich. *Basics of Magnetic Resonance Imaging and Magnetic Resonance Spectroscopy*, pages 3–167. Springer, 2008.

[Böt20] Michael Böttinger. Reaching broad audiences from a research institute setting. *Foundations of Data Visualization*, pages 307–318, 2020.

[BPDCRG+17] Sonia Blanco-Prieto, Loretta De Chiara, Mar Rodríguez-Girondo, Lorena Vázquez-Iglesias, Francisco Javier Rodríguez-Berrocal, Alberto Fernández-Villar, María Isabel Botana-Rial, and María Páez de La Cadena. Highly sensitive marker panel for guidance in lung cancer rapid diagnostic units. *Scientific Reports*, 7(1):41151, 2017.

[Bro90] Irwin D. Bross. How to eradicate fraudulent statistical methods: statisticians must do science. *Biometrics*, pages 1213–1225, 1990.

[BSB+18] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. Narrative design patterns for data-driven storytelling. In *Data-driven Storytelling*, pages 107–133. 2018.

[BVY+22] S. Sandra Bae, Rishi Vanukuru, Ruhan Yang, Peter Gyory, Ran Zhou, Ellen Yi-Luen Do, and Danielle Albers Szafir. Cultivating visualization literacy for children through curiosity and play. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):257–267, 2022.

[C+22] GBD 2019 Mental Disorders Collaborators et al. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet Psychiatry*, 9(2):137–150, 2022.

[CAA+20] Davide Ceneda, Natalia Andrienko, Gennady Andrienko, Theresia Gschwandtner, Silvia Miksch, Nikolaus Piccolotto, Tobias Schreck, Marc Streit, Josef Suschnigg, and Christian Tominski. Guide me in analysis: a framework for guidance designers. In *Computer Graphics Forum*, volume 39, pages 269–288. Wiley Online Library, 2020.

[CBM05] Mark D. Chatfield, Carol E. Brayne, and Fiona E. Matthews. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *Journal of Clinical Epidemiology*, 58(1):13–19, 2005.

[CCN17] Vanessa Cesário, António Coelho, and Valentina Nisi. Enhancing museums' experiences through games and stories for young audiences. In *Interactive Storytelling: 10th International Conference on Interactive Digital Storytelling*, pages 351–354. Springer, 2017.

[CSH20] Fabio Crameri, Grace E. Shephard, and Philip J. Heron. The misuse of colour in science communication. *Nature Communications*, 11(1):5444, 2020.

[DD22] Anuja Dixit and Rahul Dixit. Forgery detection in medical images with distinguished recognition of original and tampered regions using density-based clustering technique. *Applied Soft Computing*, 130, 2022.

[DOIG19] Nidal Drissi, Sofia Ouhbi, Mohammed Abdou Janati Idtissi, and Mounir Ghogho. Gamification-based apps for PTSD: An analysis of functionality and characteristics. In *ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–6. IEEE, 2019.

[Dri19] Natasha Morales Drissi. *Brain Networks and Dynamics in Narcolepsy*. Linköping University Electronic Press, 2019.

[DSSK19] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1):1–25, 2019.

[DWQW22] Dazhen Deng, Aoyu Wu, Huamin Qu, and Yingcai Wu. Dashbot: insight-driven dashboard generation based on deep reinforcement learning. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):690–700, 2022.

[EMW23] Johannes Eschner, Peter Mindek, and Manuela Waldner. Illustrative motion smoothing for attention guidance in dynamic visualizations. *arXiv preprint arXiv:2305.16030*, 2023.

[FHG21] Shuo Feng, Celestin Hategeka, and Karen Ann Grépin. Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic. *Population Health Metrics*, 19(1):1–14, 2021.

[GBR21] Aya Gamal, Sherif Barakat, and Amira Rezk. Standardized electronic health record data modeling and persistence: a comparative review. *Journal of Biomedical Informatics*, 114, 2021.

[GGAM12] Theresia Gschwandtner, Johannes Gärtner, Wolfgang Aigner, and Silvia Miksch. A taxonomy of dirty time-oriented data. In *Multidisciplinary Research and Practice for Information Systems*, pages 58–72. Springer, 2012.

[GMF+21] Laura Garrison, Monique Meuschke, Jennifer E. Fairman, Noeska N. Smit, Bernhard Preim, and Stefan Bruckner. An exploration of practice and preferences for the visual communication of biomedical processes. In *VCBM*, pages 1–12, 2021.

[Gro] Groep Biomedische Wetenschappen KU Leuven. Home: Equipment: PET/PET/CT: Example images. `https://gbiomed.kuleuven.be/english/corefacilities/mosaic/equipment/molecubes`. Accessed: 10-August-2023.

[GS05] Henning Griethe and Heidrun Schumann. Visualizing uncertainty for improved decision making. In *Proceedings of 4th International Conference on Perspectives in Business Informatics Research*, volume 20, 2005.

[GSWS21] Christina Gillmann, Dorothee Saur, Thomas Wischgoll, and Gerik Scheuermann. Uncertainty-aware visualization in medical imaging — a survey. In *Computer Graphics Forum*, volume 40, pages 665–689. Wiley Online Library, 2021.

[HAS09] Susan E. Hardy, Heather Allore, and Stephanie A. Studenski. Missing data: a special challenge in aging research. *Journal of the American Geriatrics Society*, 57(4):722–729, 2009.

[HNGC20] Elaine Huynh, Angela Nyhout, Patricia Ganea, and Fanny Chevalier. Designing narrative-focused role-playing games for visualization literacy in young children. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):924–934, 2020.

[HP17] Ya-Hsin Hung and Paul Parsons. Assessing user engagement in information visualization. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1708–1717, 2017.

[HPB13] Ammar Hindi, Cynthia Peterson, and Richard G. Barr. Artifacts in diagnostic ultrasound. *Reports in Medical Imaging*, pages 29–48, 2013.

[HPSP01] Horst K. Hahn, Bernhard Preim, Dirk Selle, and H.-O. Peitgen. Visualization and interaction techniques for the exploration of vascular structures. In *Proceedings Visualization*, pages 395–578. IEEE, 2001.

[HS20] Jinling Hua and Rajib Shaw. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: the case of China. *International Journal of Environmental Research and Public Health*, 17(7):2309, 2020.

[Hul19] Jessica Hullman. Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2019.

[HW21] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.

[Jel17] Bjørn Petter Jelle. Reviewing the learning process through creative puzzle solving. *Creative Education*, 8:2009–2035, 2017.

[JKRY12] Daniel Jönsson, Joel Kronander, Timo Ropinski, and Anders Ynnerman. Historygrams: enabling interactive global illumination in direct volume rendering using photon mapping. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2364–2371, 2012.

[JSYR14] Daniel Jönsson, Erik Sundén, Anders Ynnerman, and Timo Ropinski. A survey of volumetric illumination techniques for interactive volume rendering. In *Computer Graphics Forum*, volume 33, pages 27–51. Wiley Online Library, 2014.

[KAK22] Robert M. Kwee, Maan T. Almaghrabi, and Thomas C. Kwee. Scientific integrity and fraud in radiology research. *European Journal of Radiology*, 156, 2022.

[KLK18] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[KLS+21] Julia Keizer, Christian F. Luz, Bhanu Sinha, Lisette van Gemert-Pijnen, Casper Albers, Nienke Beerlage-de Jong, and Corinna Glasner. The visual dictionary of antimicrobial stewardship, infection control, and institutional surveillance data. *Frontiers in microbiology*, 12, 2021.

[KMM+18] Julian Kreiser, Monique Meuschke, Gabriel Mistelbauer, Bernhard Preim, and Timo Ropinski. A survey of flattening-based medical visualization techniques. In *Computer Graphics Forum*, volume 37, pages 597–624. Wiley Online Library, 2018.

[KSG+22] Kandamaran Krishnamurthy, Nikil Selvaraj, Palak Gupta, Benitta Cyriac, Puvin Dhurairaj, Adnan Abdullah, Ambigga Krishnapillai, Halyna Lugova, Mainul Haque, Sophie Xie, and Eng-Tat Ang. Benefits of gamification in medical education. *Clinical Anatomy*, 35(6):795–807, 2022.

[KSM+09] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P. W. Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2009.

[KSM+22] Anna Kleinau, Evgenia Stupak, Eric Mörth, Laura Ann Garrison, Sarah Mittenentzwei, Noeska Natasja Smit, Kai Lawonn, Stefan Bruckner, Matthias Gutberlet, Bernhard Preim, et al. Is there a tornado in Alex's blood flow? A case study for narrative medical visualization. Eurographics, 2022.

[KSO+23] Sohail Ahmed Khan, Ghazaal Sheikhi, Andreas L. Opdahl, Fazle Rabbi, Sergej Stoppel, Christoph Trattner, and Duc-Tien Dang-Nguyen. Visual user-generated content verification in journalism: an overview. *IEEE Access*, 2023.

[LACA18] Richard N. Landers, Elena M. Auer, Andrew B. Collmus, and Michael B. Armstrong. Gamification science, its history and future: definitions and a research agenda. *Simulation & Gaming*, 49(3):315–337, 2018.

[LBI+11] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Seven guiding scenarios for information visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 2011.

[LGS+a] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 1: Input/Lie with Stats/Discretized Continuous Variable. `https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%201:%20Input/Lie%20with%20Stats/Discretized%20Continuous%20Variable/002752_85c2add9b0e62b4e_missing%20normalization-discretized%20continuous%20variable-ineffective%20color%20scheme.png`. Accessed: 09-August-2023.

[LGS+b] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 1: Input/Lie with

Stats/Missing Normalization. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%201%3A%20Input/Lie%20with%20Stats/Missing%20Normalization/002735_ce81b12eea586e95_missing%20normalization.png. Accessed: 09-August-2023.

[LGS+c] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 1: Input/Lie with Stats/Missing Normalization. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%201%3A%20Input/Lie%20with%20Stats/Missing%20Normalization/001808_bec1b49c613ecaa1_missing%20normalization.jpg. Accessed: 09-August-2023.

[LGS+d] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 2: Visualization Design/Choice of Axis/Data of Different Magnitudes. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%202:%20Visualization%20Design/Choice%20of%20Axis/Data%20of%20Different%20Magnitudes/000419_a02d352d2d1bd6d6_data%20of%20different%20magnitudes.jpg. Accessed: 09-August-2023.

[LGS+e] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 3: Plotting/Chaotic Canvas/Cluttering. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%203%3A%20Plotting/Chaotic%20Canvas/Cluttering/000333_ca9525e9166f892e_cluttering-illegible%20text.jpg. Accessed: 09-August-2023.

[LGS+f] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 3: Plotting/Inconsistency/Inconsistent Tick Intervals. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%203:%20Plotting/Inconsistency/Inconsistent%20Tick%20Intervals/001125_aec26e7e303e650c_inconsistent%20binning%20size-inconsistent%20tick%20intervals.png. Accessed: 09-August-2023.

[LGS+g] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Github bad-vis-browser, Stage 4: Perception/Visual Illusion/Pictorial Area Encoding. https://github.com/leoyuholo/bad-vis-browser/blob/master/gallery/Stage%204:%20Perception/Visual%

20Illusion/Pictorial%20Area%20Encoding/005813_d126d8983ec11ff1_pictorial%20area%20encoding.png. Accessed: 09-August-2023.

[LGS+22] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. Misinformed by visualization: what do we learn from misinformative visualizations? In *Computer Graphics Forum*, volume 41, pages 515–525. Wiley Online Library, 2022.

[LH11] Joey J. Lee and Jessica Hammer. Gamification in education: what, how, why bother? *Academic Exchange Quarterly*, 15(2):146, 2011.

[LHW+11] Henrik Lauridsen, Kasper Hansen, Tobias Wang, Peter Agger, Jonas L. Andersen, Peter S Knudsen, Anne S. Rasmussen, Lars Uhrenholt, and Michael Pedersen. Inside out: modern imaging techniques to reveal animal anatomy. *PLOS One*, 6(3), 2011.

[Liv12] Lefteris Livieratos. Basic principles of SPECT and PET imaging. In *Radionuclide and Hybrid Bone Imaging*, pages 345–359. Springer, 2012.

[LK18] Torrin M. Liddell and John K. Kruschke. Analyzing ordinal data with metric models: what could possibly go wrong? *Journal of Experimental Social Psychology*, 79:328–348, 2018.

[LLY06] Patric Ljung, Claes Lundström, and Anders Ynnerman. Multiresolution interblock interpolation in direct volume rendering. In *EuroVis - Eurographics/IEEE VGTC Symposium on Visualization*. The Eurographics Association, 2006.

[LPLK22] Maxim Lisnic, Cole Polychronis, Alexander Lex, and Marina Kogan. *Misleading beyond visual tricks: how people actually lie with charts*. OSF Preprints, 2022.

[LR06] Ulf Lindgren and Olof Rehnström. CPU-based volume rendering of large medical data sets with level-set clipping. *Linköping University Electronic Press*, 2006.

[LYI+21] Crystal Lee, Tanya Yang, Gabrielle D. Inchoco, Graham M. Jones, and Arvind Satyanarayan. Viral visualizations: how coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2021.

[LZXP20] Caixia Liu, Ruibin Zhao, Wangli Xie, and Mingyong Pang. Pathological lung segmentation based on random forest combined with deep model and multi-scale superpixels. *Neural Processing Letters*, 52:1631–1649, 2020.

[MA14] Silvia Miksch and Wolfgang Aigner. A matter of time: applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014.

[Mau19] Oliver Mauroner. Gamification in management and other non-game contexts—understanding game elements, motivation, reward systems, and user types. *Open Journal of Business and Management*, 7(4):1815–1830, 2019.

[MEB+23] Lidwine B. Mokkink, Iris Eekhout, Maarten Boers, Cees P. M. van der Vleuten, and Henrica C. W. de Vet. Studies on reliability and measurement error of measurements in medicine–from design to statistics explained for medical researchers. *Patient Related Outcome Measures*, pages 193–212, 2023.

[MGS+21] Monique Meuschke, Laura Garrison, Noeska Smit, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. Towards narrative medical visualization. *arXiv preprint arXiv:2108.05462*, 2021.

[MGS+22] Monique Meuschke, Laura A. Garrison, Noeska N. Smit, Benjamin Bach, Sarah Mittenentzwei, Veronika Weiß, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. Narrative medical visualization to communicate disease data. *Computers & Graphics*, 107:144–157, 2022.

[MKC20] Andrew McNutt, Gordon Kindlmann, and Michael Correll. Surfacing visualization mirages. In *Proceedings of the CHI Conference on Human Factors in Computing System*, pages 1–16, 2020.

[MMAF14] Hamid Reza Marateb, Marjan Mansourian, Peyman Adibi, and Dario Farina. Manipulating measurement scales in medical statistical analysis and data mining: a review of methodologies. *Journal of Research in Medical Sciences*, 19(1):47, 2014.

[MMSE19] Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. CT-GAN: malicious tampering of 3D medical imagery using deep learning. In *USENIX Security Symposium*, 2019.

[MPB+09] Cynthia H. McCollough, Andrew N. Primak, Natalie Braun, James Kofler, Lifeng Yu, and Jodie Christner. Strategies for reducing radiation dose in CT. *Radiologic Clinics*, 47(1):27–40, 2009.

[MPB+20] Neal Mangaokar, Jiameng Pu, Parantapa Bhattacharya, Chandan K. Reddy, and Bimal Viswanath. Jekyll: attacking medical image diagnostics using deep generative models. In *European Symposium on Security and Privacy (EuroS&P)*, pages 139–157. IEEE, 2020.

[MTV] MTV show Guy Code. Most infographics are stupid. http://nicolecheriehess.overblog.com/good-infographic-bad-infographic. Accessed: 09-August-2023.

[Mul14] Michael Muller. Curiosity, creativity, and surprise as analytic tools: Grounded theory method. In *Ways of Knowing in HCI*, pages 25–48. Springer, 2014.

[MVB+16] Monique Meuschke, Samuel Voss, Oliver Beuing, Bernhard Preim, and Kai Lawonn. Combined visualization of vessel deformation and hemodynamics in cerebral aneurysms. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):761–770, 2016.

[NJD20] Vinh T. Nguyen, Kwanghee Jung, and Tommy Dang. Revisiting common pitfalls in graphical representations utilizing a case-based learning approach. In *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*, pages 1–5, 2020.

[NN21] K. A. Narayan and M. Siva Durga Prasad Nayak. Need for interactive data visualization in public health practice: examples from India. *International Journal of Preventive Medicine*, 12, 2021.

[OEI+22] Chinenye Okpara, Chidozie Edokwe, George Ioannidis, Alexandra Papaioannou, Jonathan D. Adachi, and Lehana Thabane. The reporting and handling of missing data in longitudinal studies of older adults is suboptimal: a methodological survey of geriatric journals. *BMC Medical Research Methodology*, 22(1):122, 2022.

[OYA+21] Omer Onder, Yasin Yarasir, Aynur Azizova, Gamze Durhan, Mehmet Ruhi Onur, and Orhan Macit Ariyurek. Errors, discrepancies and underlying bias in radiology with case examples: a pictorial review. *Insights into Imaging*, 12(1):1–21, 2021.

[PAA+22] Iris Cathrina Abacan Pilares, Sami Azam, Serkan Akbulut, Mirjam Jonkman, and Bharanidharan Shanmugam. Addressing the challenges of electronic health records using blockchain and IPFS. *Sensors*, 22(11), 2022.

[PB07] Bernhard Preim and Dirk Bartz. *Visualization in medicine: theory, algorithms, and applications*. Elsevier, 2007.

[Peo] People's Daily post on Weibo. Special reminder: do not rush to buy and take Shuanghuanglian oral liquid. https://www.weibo.com/2803301701/Is8xy5pp0?filter=hot&root_comment_id=0&type=comment#_rnd1580541069478. Accessed: 01-February-2023.

[Pew] Pew Research Center, a nonpartisan American think tank. Social media use by age, 2005–2021. https://www.pewresearch.org/internet/chart/social-media-use-by-age/. Accessed: 20-August-2023.

[PL20] Bernhard Preim and Kai Lawonn. A survey of visual analytics for public health. In *Computer Graphics Forum*, volume 39, pages 543–580. Wiley Online Library, 2020.

[PMN+14] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. The persuasive power of data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2211–2220, 2014.

[PRSL23] Bernhard Preim, Renata Raidou, Noeska Smit, and Kai Lawonn. *Visualization, Visual Analytics and Virtual Reality in Medicine: State-of-the-art Techniques and Applications*. Elsevier, 2023.

[PS09] John J. Palmieri and Theodore A. Stern. Lies in the doctor-patient relationship. *Primary Care Companion to the Journal of Clinical Psychiatry*, 11(4):163, 2009.

[PSHK08] Robert Popilock, Kumar Sandrasagaren, Lowell Harris, and Keith A. Kaser. CT artifact recognition for the nuclear technologist. *Journal of Nuclear Medicine Technology*, 36(2):79–81, 2008.

[PTM+12] Kenneth E. S. Poole, Graham M. Treece, Paul M. Mayhew, Jan Vaculík, Pavel Dungl, Martin Horák, Jan J. Štěpán, and Andrew H. Gee. Cortical thickness mapping to identify focal osteoporosis in patients with hip fracture. *PlOS One*, 7(6), 2012.

[Rai18] Renata G. Raidou. Uncertainty visualization: recent developments and future challenges in prostate cancer radiotherapy planning. In *EuroVis workshop on reproducibility, verification, and validation in visualization (EuroRV3)*. The Eurographics Association, 2018.

[RAS21] Neda Rostamzadeh, Sheikh S. Abdullah, and Kamran Sedig. Visual analytics for electronic health records: a review. In *Informatics*, volume 8, page 12. MDPI, 2021.

[RCF+21] Matteo Renzulli, Mario Casavola, Alberto Foà, Carmine Pizzi, and Rita Golfieri. Imaging of biliary involvement in sarcoidosis: computed tomography, magnetic resonance cholangiopancreatography, and gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid-enhanced magnetic resonance imaging findings. *Tomography*, 7(4):783–791, 2021.

[RMB+16] Renata G. Raidou, Freek J. J. Marcelis, Marcel Breeuwer, M. Eduard Gröller, Anna Vilanova, and Huub M. M. van de Wetering. Visual analytics for the exploration and assessment of segmentation errors. In *VCBM/MedViz*, pages 193–202, 2016.

[RPHL14] Gordan Ristovski, Tobias Preusser, Horst K. Hahn, and Lars Linsen. Uncertainty in medical visualization: towards a taxonomy. *Computers & Graphics*, 39:60–73, 2014.

[RRGT+13] Lilia L. Ramírez-Ramírez, Yulia R. Gel, Mary Thompson, Eileen de Villa, and Matt McPherson. A new surveillance and spatio-temporal visualization tool SIMID: SIMulation of infectious diseases using random networks and GIS. *Computer Methods and Programs in Biomedicine*, 110(3):455–470, 2013.

[RSDG20] Alessandro Romano, Chiara Sotis, Goran Dominioni, and Sebastián Guidi. The scale of covid-19 graphs affects understanding, attitudes, and policy preferences. *Health Economics*, 29(11):1482–1494, 2020.

[RZ04] Linda Rozmovits and Sue Ziebland. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Education and Counseling*, 53(1):57–64, 2004.

[RZR+17] Bhavya Rehani, Yi C. Zhang, Madan M. Rehani, András Palkó, Lawrence Lau, Miriam N. Mikhail Lette, and William P. Dillon. Radiology education in Europe: analysis of results from 22 European countries. *World Journal of Radiology*, 9(2):55, 2017.

[S+15] Mark E. Schmidt et al. The influence of biological and technical factors on quantitative analysis of amyloid PET: points to consider and recommendations for controlling variability in longitudinal data. *Alzheimer's & Dementia*, 11(9):1050–1068, 2015.

[S+17] Christopher G. Schwarz et al. Contributions of imprecision in PET-MRI rigid registration to imprecision in amyloid PET SUVR measurements. *Human Brain Mapping*, 38(7):3323–3336, 2017.

[SH10] Edward Segel and Jeffrey Heer. Narrative visualization: telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[SHMB17] Peter B. Sachs, Kelly Hunt, Fabien Mansoubi, and James Borgstede. CT and MR protocol standardization across a large health system: providing a consistent radiologist, patient, and referring provider experience. *Journal of Digital Imaging*, 30:11–16, 2017.

[SKRW22] Marwin Schindler, Thorsten Korpitsch, Renata G. Raidou, and H.-Y. Wu. Nested papercrafts for anatomical and biological ddutainment. In *Computer Graphics Forum*, volume 41, pages 541–553. Wiley Online Library, 2022.

[SMH10] Ahmed Saad, Torsten Möller, and Ghassan Hamarneh. Probexplorer: uncertainty-guided exploration and editing of probabilistic medical image segmentation. In *Computer Graphics Forum*, volume 29, pages 1113–1122. Wiley Online Library, 2010.

[SPB21] Gunjan Singh, Raj H. Patel, and Joshua Boster. Root cause analysis and medical error prevention. *StatPearls*, 2021.

[SPBR20] Matthias Schlachter, Bernhard Preim, Katja Bühler, and Renata G. Raidou. Principles of visualization in radiation oncology. *Oncology*, 98(6):412–422, 2020.

[SPS11] David Spiegelhalter, Mike Pearson, and Ian Short. Visualizing uncertainty about the future. *Science*, 333(6048):1393–1400, 2011.

[SSHW05] Paul Schoenhagen, Arthur E. Stillman, Sandra S. Halliburton, and Richard D. White. CT of the heart: principles, advances, clinical uses. *Cleveland Clinic Journal of Medicine*, 72(2):127–138, 2005.

[Sza18] Danielle Albers Szafir. The good, the bad, and the biased: five ways visualizations can mislead (and how to fix them). *ACM Interactions*, 25(4):26–33, 2018.

[Tel14] Alexandru C. Telea. *Data Visualization: Principles and Practice*. CRC Press, 2014.

[TOI+20] Yoshihiro Tanaka, Masataka Oita, Shinichiro Inomata, Toshiaki Fuse, Yuichi Akino, and Kohei Shimomura. Impact of patient positioning uncertainty in noncoplanar intracranial stereotactic radiotherapy. *Journal of Applied Clinical Medical Physics*, 21(2):89–97, 2020.

[TP03] Eric J. Thomas and Laura A. Petersen. Measuring errors and adverse events in health care. *Journal of General Internal Medicine*, 18:61–67, 2003.

[Tut] Kathleen Tuthill. John Snow and Broad street pump on the trail of an epidemic. `https://www.ph.ucla.edu/epi/snow/snowcricketarticle.html`. Accessed: 22-February-2023.

[TZ09] Guoxin Tan and Zheng Zhong. The application of storytelling in digital museums. In *10th International Conference on Computer-Aided Industrial Design & Conceptual Design*, pages 1638–1641. IEEE, 2009.

[VMPDV09] Vivianne H. M. Visschers, Ree M. Meertens, Wim W. F. Passchier, and Nanne N. K. De Vries. Probability information in risk communication: a review of the research literature. *Risk Analysis*, 29(2):267–287, 2009.

[WL22] Qiru Wang and Robert S. Laramee. EHR STAR: the state-of-the-art in interactive EHR visualization. In *Computer Graphics Forum*, volume 41, pages 69–105. Wiley Online Library, 2022.

[WMWG14] Tracey L. Weissgerber, Natasa Milic, Stacey J. Winham, and Vesna D. Garovic. Beyond bar graphs: time for a new data presentation paradigm. In *Journal of Women's Health*, volume 23, pages 872–872, 2014.

[WS11] Hao Wang and Chuen-Tsai Sun. Game reward systems: gaming experiences and social meanings. In *Digital Games Research Association Conference*, volume 114, 2011.

[YA21] Ar Anil Yasin and Asad Abbas. Role of gamification in engineering education: a systematic literature review. In *Global Engineering Education Conference (EDUCON)*, pages 210–213. IEEE, 2021.

[YHG+17] Jiarui Yang, Chenhui Hu, Ning Guo, Joyita Dutta, Lucia M. Vaina, Keith A. Johnson, Jorge Sepulcre, Georges El Fakhri, and Quanzheng Li. Partial volume correction for PET quantification and its impact on brain network in Alzheimer's disease. *Scientific Reports*, 7(1), 2017.

[ZEP11] Qi Zhang, Roy Eagleson, and Terry M. Peters. Volume visualization: a technical overview with a focus on medical applications. *Journal of Digital Imaging*, 24:640–664, 2011.

[ZM22] Chengbo Zheng and Xiaojuan Ma. Evaluating the effect of enhanced text-visualization integration on combating misinformation in data story. In *15th Pacific Visualization Symposium (PacificVis)*, pages 141–150. IEEE, 2022.

[ZSP+21] Yixuan Zhang, Yifan Sun, Lace Padilla, Sumit Barua, Enrico Bertini, and Andrea G. Parker. Mapping the landscape of COVID-19 crisis visualizations. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–23, 2021.