HAW HAMBURG

BACHELORTHESIS

Thi Ngan Chau Vu

Exploration und Implementierung von Visualisierungstechniken für Covid-19-Daten

FAKULTÄT TECHNIK UND INFORMATIK
Department Informatik

Faculty of Computer Science and Engineering
Department Computer Science

HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG
Hamburg University of Applied Sciences

Thi Ngan Chau Vu

Exploration und Implementierung von Visualisierungstechniken für Covid-19-Daten

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung
im Studiengang *Bachelor of Science Informatik Technischer Systeme*
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Marina Tropmann-Frick
Zweitgutachter: Prof. Dr. Stefan Sarstedt

Eingereicht am: 23 May 2023

Thi Ngan Chau Vu

Title of Thesis

Exploration and Implementation of Visualisation Techniques on Covid-19 Data

Keywords

Visualization, Data, Covid-19, Visualization Fundamentals, Visualization Principles, Design Process, Visualization Evaluation

Abstract

Visualization techniques have played a prominent role in conveying scientific information about COVID-19 to a broad audience, including policymakers, scientists, healthcare providers, and the general public. The main objective of this thesis is to explore the application of data visualization techniques in analyzing and effectively communicating the impact of the COVID-19 pandemic. The research starts by establishing the fundamentals of data visualization, which include essential terminologies, concepts, and general data visualization principles. The upcoming chapters will illustrate the process of designing data visualizations, emphasizing the various stages required to produce impactful visual representations. Further, the thesis explores different methodologies for data visualization and presents a methodological framework and a guide for information visualization techniques. Through a case study employing an exploratory data analysis approach, this thesis demonstrates implementing and evaluating data visualizations in COVID-19 analysis. The evaluation aims to determine how compelling the visualizations are in conveying information and facilitating data exploration. This thesis enhances our comprehension of the potential of data visualization in conveying and comprehending the complicated consequences of the COVID-19 pandemic.

Acknowledgements

I would like to thank all those who accompanied and supported me during my thesis journey. I would like to express my gratitude to my professors at Hamburg University of Applied Sciences (HAW) for their motivation and support throughout this project. Their lectures and projects were instrumental in inspiring me to complete this work. Especially I leave my special thank to Prof. Dr. Marina Tropmann-Frick, for her kindness, encouragement, counselling, time, consideration, and patience. And to Prof. Dr. Stefan Sarstedt for kindly accepting my request to work under his supervision at the HAW. This thesis would not have been possible without them.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Introduction

1.1 Motivation

We live in a data-driven culture, where data influences everything we see and do. Every day, trillion MB of data are produced, which will continue to grow mainly in the future. According to experts, internet users created 1.7 MB of data on average for every individual per second on the planet.[6] Furthermore, the increase in the information system has led people with heaps of knowledge and fewer ways to capture, investigate and utilize data properly. People are expecting to move from monitoring the data (ex. what is the current covid infection rate? how many people have been vaccinated in the country?) to mining this data to comprehend the situation better (ex. Which impact does these numbers have on household and individuals?). In other words, having the ability to obtain, interpret, and analyze data, as well as present insights and information from this analysis, is becoming increasingly vital in everyday life. In his well-known quote, legendary engineer W. Edwards Deming, who is widely acknowledged as the leading management thinker in the field of quality, stated that: "Without data, you're just another person with an opinion."[4]

Without the backup of data, people make poor judgments based on instinct, supposition, or conventional belief. As a result, they are at risk of behaving based on prejudices or incorrect assumptions. Data is the key to making intelligent decisions and monitoring performance, and nearly every innovative technology today relies heavily on it. However, this potentially valuable resource will be rendered useless unless we can create a practical approach to delve into data and extract its learning. This is where data visualization comes in to assist us in addressing the problem.

The important strength of data visualization is not that it makes data more beautiful but that it provides insight into complex data sets by communicating their key aspects more intuitively in meaningful ways. Data visualization has been identified as a critical 21st-century research skill. The following sections introduce shortly how data will be interpreted and displayed using data visualization techniques.

1.2 Structure of the thesis

The COVID-19 pandemic has significantly impacted societies worldwide, resulting in significant damage to health, economies, and daily life. To effectively respond to the global crisis, it is crucial to analyze and communicate the vast amount of data related to the pandemic. The application of data visualization, as a potent instrument, play a vital role in examining and disseminating the effects of the COVID-19 pandemic.

The second chapter of this thesis explores the fundamental principles of data visualization, providing a solid basis for comprehending its importance in effectively communicating data. This chapter will delve into the essential terms and concepts related to data visualization. The focus goes into the fundamental terminology and principles linked to data visualization, emphasizing its significance in rendering information easily comprehensible and accessible to diverse groups of individuals. This chapter covers general principles of data visualization and highlights the significance of clarity, accuracy, and compelling storytelling through visual representations.

Chapter 2 delves deeper into the process of designing data visualizations, building upon the foundation established in the previous chapter. This section provides an overview of the crucial steps required to create powerful visualizations. Researchers can follow the design process to communicate insights and facilitate data exploration effectively. I also present a methodological framework that aims to assist researchers in selecting suitable visualization techniques based on their unique data and research goals.

Chapter 3 is concentrated on the implementation and assessment of data visualizations within the framework of COVID-19 analysis. This chapter employs an exploratory data analysis methodology to showcase a case study demonstrating the practical application of data visualization methods in exploring the COVID-19 outbreak. By effectively applying these methods, researchers can learn more about the virus's effects on many facets of society.

In conclusion, my thesis attempts to examine the core principles of data visualization, establish its design process, introduce diverse methodologies, and provide a practical case study for implementation and assessment.

1.3 Research question

Visualization of data is a reliable way to communicate and interpret information. Analysts often use data visualization to explore and understand complex data sets, making it possible to identify patterns, trends, and relationships previously hidden in the raw data. They consider data visualization as a fundamental component of data analysis and find widespread use in numerous industries and disciplines, including business, economics, and the social sciences. The ability to clearly and concisely communicate information through data visualization makes it a crucial tool for comprehending and making decisions upon data-driven insights. Therefore, this thesis will primarily aim to answer the following question:

"How can data visualization techniques be effectively used to analyze and communicate the impact of the COVID-19 pandemic on different types of data and audiences, and how do these techniques vary in their ability to convey different information types?"

This thesis aims to answer the research question by thoroughly examining the fundamental principles of data visualization, establishing a comprehensive design process, introducing various methodologies, and presenting a practical case study with COVID-19 related information for implementation and assessment.

2

Data Visualization

2.1 Data Visualization Fundamentals

2.1.1 Term and Concept

Data Visualization represents viewers' messages and information through pictures, charts, graphs, or any visual presentation. It is standing at the crossroads of communication science, information science, and design. Research visualization visually provides blocks of abstract data to enhance human perception. It is not new to use graphics to represent data, and it has been utilized in maps, scientific drawings, and data visualizations for thousands of years. Computer graphics have been used to examine scientific topics since its debut but their benefits were restricted in the early days due to a lack of graphical capability. Data visualization is the process of mapping data to vision [11], and the crafting principle is to mine data and graphically portray its values.

In an article presented at the smashing magazine, Vitaly Friedman pointed out that the primary goal of data visualization is to convey information clearly and efficiently to the user through selected graphics such as tables or charts[10]. While tables are often used when evaluating or measuring the value of a component, graph types display the results or relate data of one or more components. They assemble complex data that is more manageable to

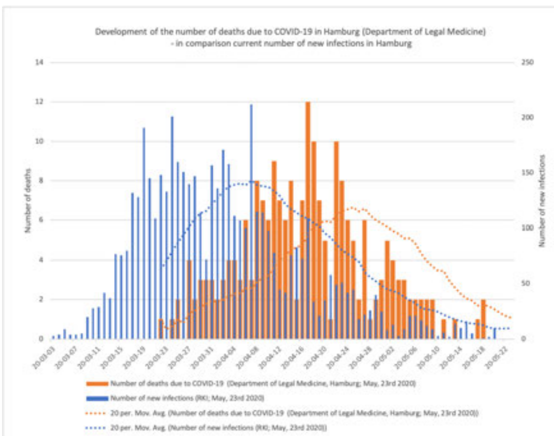


Abbildung 2.1: Statistics of infections and SARS-CoV-2 fatalities in Hamburg/Germany [60]

understand and use. Therefore, users can perform particular analysis actions to better comprehend data, such as creating data comparisons.

An example of infections and SARS-CoV-2 fatalities in Hamburg/Germany data visualization is shown in Figure 2.1. Each bar represents the number of infection cases or fatalities in area Hamburg.

According to Johns Hopkins Sheridan Libraries, data visualization includes the following classes [65]:

Information Visualization - Presentation of data, which can be separated into :

Exploratory: Exploratory data visualization is a process for producing descriptive and graphical data summaries. The advantage of exploratory data visualization is its ability to explore raw data without presumptions. It is a valuable tool in many fields to examine and research ill-defined matters and study and visualize data. It does not, however, provide final or conclusive answers to the problems at hand.

Explanatory: Visualizations that try to explain why things work the way they do are called explanatory visualizations. Explanatory visualization is used to answer the "why"question and explains why certain phenomena happen. This visualization type aims to connect diverse ideas in order to comprehend the nature of cause and effect associations. Explanatory visualization, in other words, analyzes specific issues and explains the patterns of relationships between variables.

Scientific visualization - Scientific visualization is the representation of data with a spatial component to visualize scalar, vector, and tensor fields. Some examples of typical scientific visualization applications include computational fluid dynamics, medical imaging and analysis, and meteorological data analysis.

Designer - Reader - Data trinity in data visualization

The concept of designer - reader - data trinity was introduced by Noah Iliinsky and Julie Steele in their book Designing Data Visualizations.[38] A three-legged chair-like configuration of designer, viewer, and data could be interpreted to facilitate data visualization. To build a steady, effective visualization, each components performs following its function[38]. Each chair leg is linked to the other two legs. Determining the appropriate visualization technique involves evaluating the significance of relationships, as depicted in Figure 2.2.:

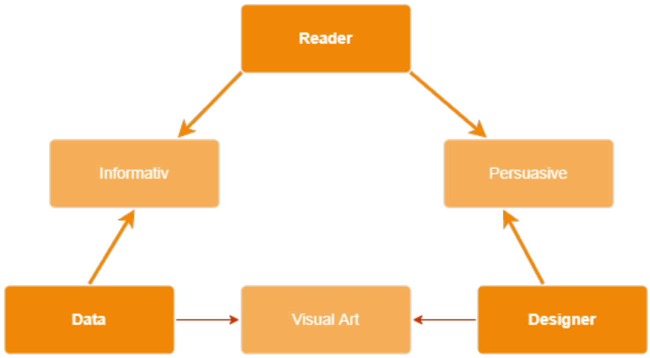


Abbildung 2.2: The relationship between the 3 components, adopted from from Iliinsky and Steele [p. 9]

Designer - Designers understand the goals of visualizations and develop them for those purposes. A designer who possesses a sense of drive, purpose, and priorities is more likely to create a successful visualization that goes beyond just presenting data.

Reader - The second component of the relationship is the reader. In this trilogy, the audience's receptive expectations play a specific role, which could be either the most significant assist or the worst impediment, or both, in reaching the aim of clearly expressing a message. In the data visualization process, it is necessary to understand from the reader's point of view.

Data - Data is the third-factor affecting visualization design. The ideal visualization is one that demonstrates exciting aspects of the data set researchers are working on. Different data requires different methods, decoding, display techniques to represent its characteristics. While default visualizations have a specific role and can happen when the proper design is selected initially, data can occasionally provide additional knowledge when another visualization method or form is chosen. The way we choose the visualization format that represents the best aspect of the data is to be data savvy and respectful of the data. Instead of trying to convert to a suitable format, consider the inherent values, relationships, and structure of the data. Having a clear understanding of the nature of the data can lead to better design choices. Each property and relation of the data could be visualized with a suitable interface and characterization of each dimension data leads to the right interface.

2.1.2 Role of data visualization

Foremost, Data visualization makes sharing and presenting information easy and fast, bypassing barriers and language limitations, thereby enabling readers and listeners to observe and absorb the data.

Data visualization helps users quickly and thoroughly capture the information presented in tables and graphs about any issue. Data visualization visualizes the relationship between research users in the data, uncovering uncertainties, anomalies, and trends of research users.

Today's organizations use data visualization and support tools to question business problems better and make better decisions. More programs and applications that support data visualization are born, making it easier for people to learn more about their data and make better data-driven business decisions. Furthermore, data visualization helps control and monitor performance indicators, KPIs, company's performance based on dashboards, showing the importance of taking advantage of and exploiting data assets for service.

Data visualization is also the foundation for the company to move towards data-driven, data-oriented. Data visualization will help employees read reports charts, understand the data, grasp information about the current company's operation status, and be comfortable participating in recommendations and suggestions. Furthermore, the utilization of data visualization facilitates the effective communication of information to both readers and listeners. When individuals follow the domestic and international media, broadcasters and reporters providing information often tell more about the story behind the numbers and visualization shown on TV. Data visualization serves as an intermediary between data and narratives, thereby facilitating the practice of data storytelling, a critical competency for analysts to effectively communicate data to their intended audience through the use of specialized tools and engaging storytelling techniques.

The advantages of data visualization are numerous, as are its applications in various businesses and industries.

2.1.3 General data visualization principles

Good data visualizations greatly enhance the ability to reason and think about data effectively. However, when dealing with complicated data sets, information visualization is not relatively as straightforward as it might appear initially. So what are the requirements for presenting a robust representation of the insights gained from the data?

While there are different sets of rules to build a robust data visualization like the Gestalt Principle of Perception by Todorovic [69], Grice's Maxim of Communication or known as the Cooperative principle by Grice [35], beginners are advised to follow the four principles listed by Edward Tufte from the Interaction Design Foundation[5]. Edward Tufte is renowned as a pioneer in the field of data visualization and for his writings on information design. His publications are regarded as the most significant works in information visualization, particularly Visual Display of Quantitative Information[70], Envisioning Information[73], Visual Explanations[72] and Beautiful Evidence[71]. His works contain key rules for the representation of visual information, including the following factors:

- Graphical Integrity
- Chart Junk and Data-Ink Ration
- Data density and Small Multiples

Graphical Integrity

When Tufte uses the term "Graphical Integrity", he is referring to a practical principle, in which the representation should not alter the underlying facts or give the viewer a false sense of it [70, p.56]. He accomplishes this by figuring out a graph's "Lie Factor", which is found by dividing the size of the effect shown in the graph by the size of the effect in the data.

The simple rule is that the graphic can accurately depict the true meaning of the data if the ratio is one. If the ratio is more than one or less than one, then the data is either overstated or understated.

where

Here is an example of a graph with poor graphical integrity provided by Tufte.

$$\textit{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Abbildung 2.3: Lie Factor Formula

$$\textit{size of effect} = \frac{|\text{second value - first value}|}{\text{first value}}$$

Abbildung 2.4: Size of Effect Formula

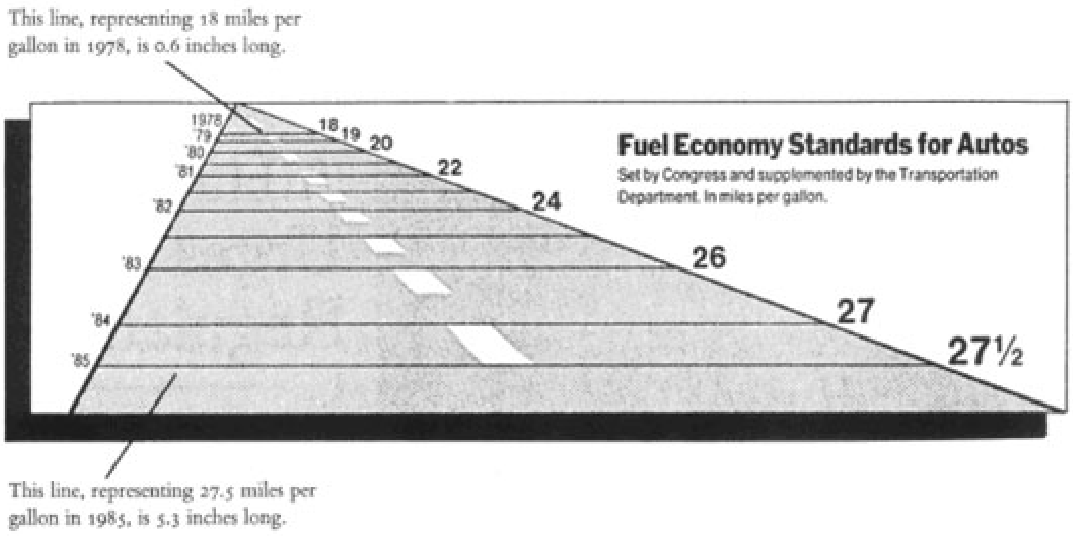


Abbildung 2.5: Fuel economy standards for autos. Reprinted from Visual Display of Quantitative Information by Edward Tufte. Copyright 2001 by the authors.

The above picture illustrates a classic example of fuel economy standards for autos, and the data is represented from 1978 to 1985. In the graph, the standard said that the mileage had to go up from 18 to 27.5, which is a 53% rise. However, the graph's rise in magnitude is 783%, giving it a lie factor of 14.8, which is very large.

$$LieFactor = \frac{\frac{5.3-0.6}{0.6}}{\frac{27.5-18}{18}} = 14.8$$

Given this lie factor, this particular visualization is not representing the data accurately and has very low integrity.

Chart Junk and Data-Ink Ration

Tufte suggested reducing redundant features from visualizations to make them more transparent and engage audience attention. The theory is that borders, backgrounds, the use of 3D and other unnecessary elements might only distract the user's attention away from the actual information. To Tufte, these irrelevant components are called chart junk. If lines and other visual components do not fit into the minimal set required to communicate the information effectively, they can be referred to as chart junk. The most typical type of chart junk is a Moiré effect when analysts like to work with various visual patterns to give the chart the appearance of being vibrated. Nevertheless, this effect does not result in a more precise data visualization. Instead, it causes the reader to frequently switch their focus between the legend and the graph, producing a physiological tremor to the eye.

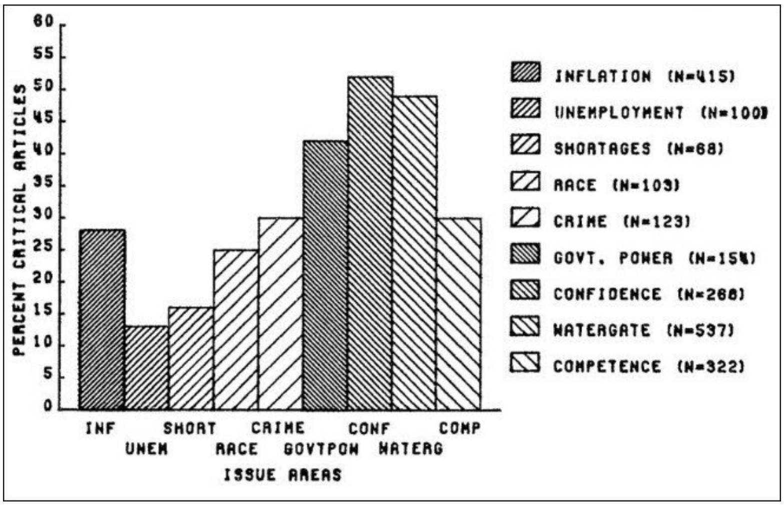


Abbildung 2.6: The Moiré effect. Illustration taken from the 2nd edition of from Visual Display of Quantitative Information by Edward Tufte (page 120). Copyright 2001 by the authors.

He also proposes a data-ink ratio mathematical formula to reduce chart junk:

$$\frac{data\text{–}ink}{total\text{–}ink} = \frac{Elements\ conveying\ data\ information}{All\ elements\ in\ the\ chart}$$

Abbildung 2.7: Data-ink Ration Formula

The typically high data-ink ratio means that a large portion of the components used in visualization is used to describe the data points themselves. A low data-ink ratio describes the visualization in which most elements are not used to describe the data but are describing something else. Generally, the higher the data-ink ratio, the better the data visualization.

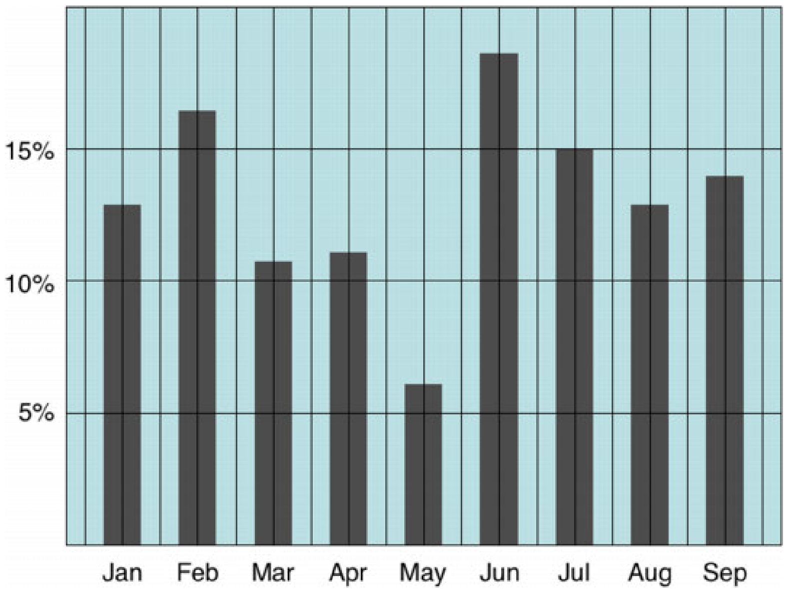


Abbildung 2.8: Low Data-ink Ration, Reprinted from Visual Display of Quantitative Information by Edward Tufte. Copyright 2001 by the authors.

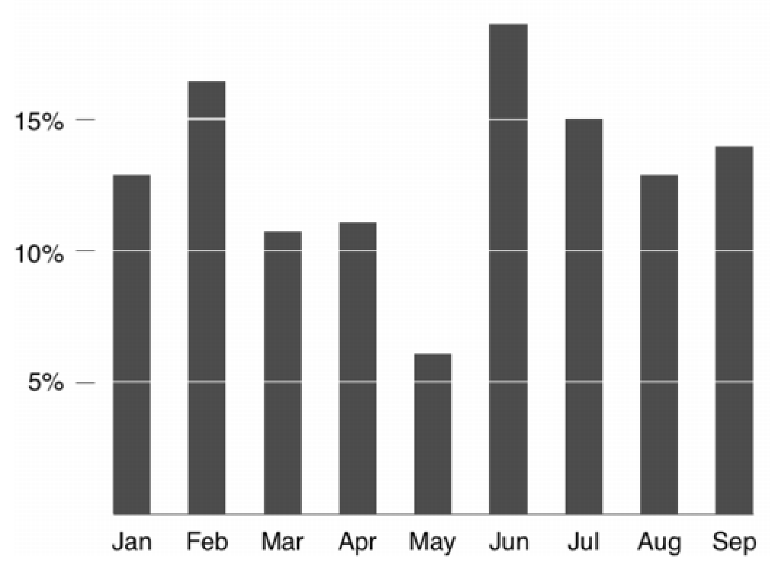


Abbildung 2.9: High Data-ink Ration, Reprinted from Visual Display of Quantitative Information by Edward Tufte. Copyright 2001 by the authors.

The same data is shown in the two graphs above. The graph's poor data-ink ratio on the left-hand side is caused by unnecessary components like the blue backdrop, the horizontal and vertical grid lines, and the outer black border. Many of the components used in this visualization do not directly describe the data at hand. On the right-hand side, the same data is visualized with a high data-ink ratio and uses only an element that genuinely describes the data directly. When comparing both visualizations side-by-side, the visualization on the right is far more aesthetically beautiful and simpler to comprehend.

Data density and Small Multiples

How much of a graph is used to display information is referred to as the graph's "data density". Accordingly, Tufte draws two conclusions from this:

- Maximize data density and the size of the data matrix within reason [70, p.168]
- Graphics can be shrunk way down [70, p.169]

In order to achieve a high-density data visualization, Tufte believes that most graphs may be significantly downsized without losing information or legibility.

The term “small multiple” is also mentioned by Tufte, which refers to a visual pattern in which a similar small graph is repeated numerous times. According to Tufte, small multiples are an excellent technique for displaying data sets with various dimensions and a significant amount of information. By using small multiple, readers can easily compare the whole data in parallel.

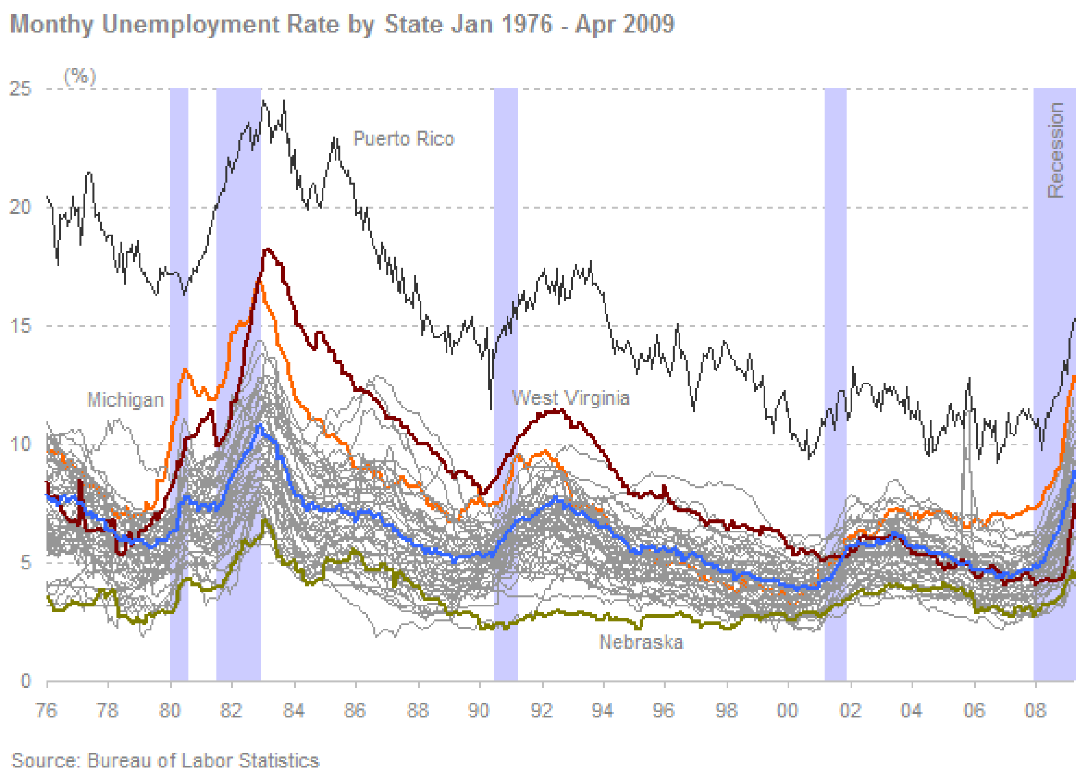


Abbildung 2.10: Monthly Unemployment Rates by State 1976-2009.

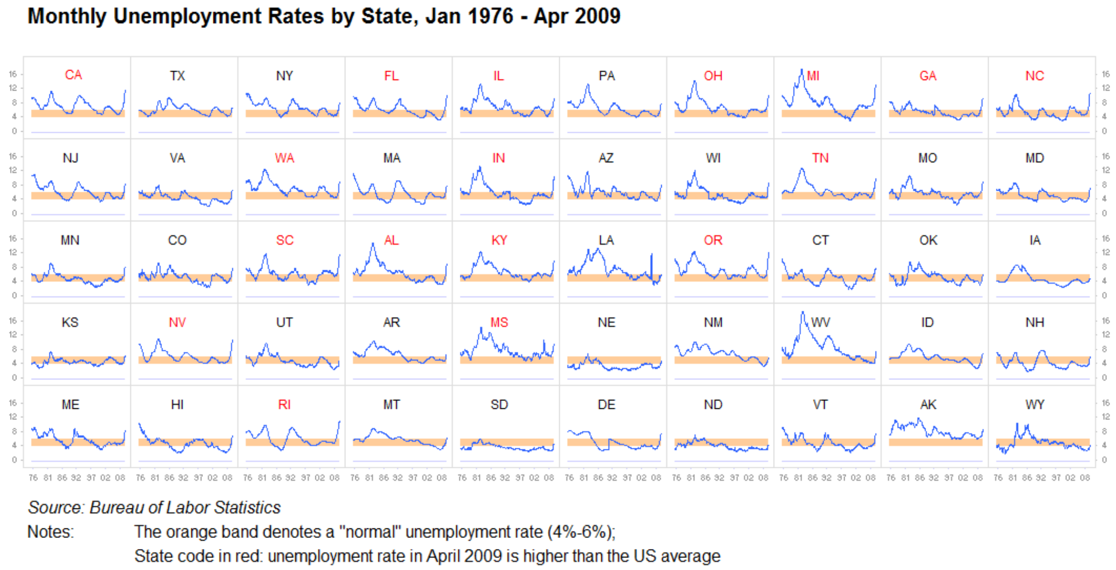


Abbildung 2.11: Monthly Unemployment Rates by State 1976-2009.

The monthly unemployment rates by states since 1976 to 2009 are shown in two different ways. The first graph in this Jorge Camoes example [1] is ideal for identifying the overall trends, the range from lowest to highest, the outliers, and the slopes. In addition, a small-multiple version enables the user to concentrate on particular states and contrast them with the standard bar. For instance, while seven out of 10 in the first row are higher than the April US average, only one is above the US average in the final row.

In brief, Tufte's principles revolve around keeping data visualization as simple and honest as possible. Therefore, his rules are not restrictive but aim to help information visualization professionals develop useful information visualizations.

2.2 Data Visualization Design Process

Under the guidance of Kirk from Data Visualization: a successful design process[44], he analyzed how the process of data visualizations in the following steps:

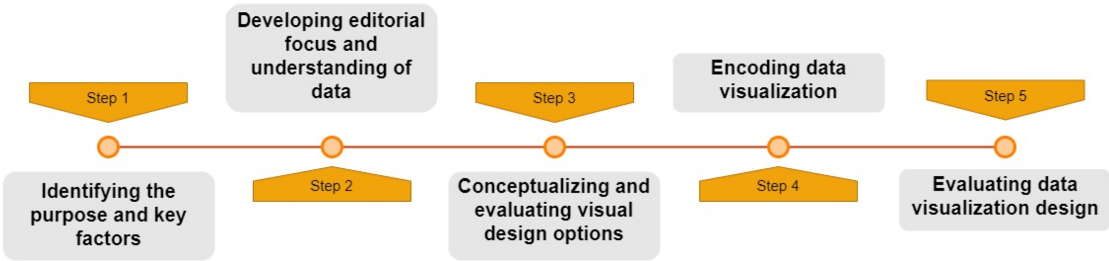


Abbildung 2.12: The process of designing data visualizations suggested by Andy Kirk [44]

2.2.1 Identifying the purpose and key factors

It is essential to identify the objectives and requirements in the first step. This step can be broad and deep in expertise, such as understanding why error reports are not trending down or estimating the reasonable time in the production process. For example, showing the amount of attention to "data visualization"content could be represented by the number of articles published in journals over time.Before embarking on data visualization, analysts should have a clear understanding of the project's underlying purpose, which includes identifying its intended audience and the specific needs it aims to address. This consideration is directly linked to the extent of their work.

Besides, to carry out the analysis and deliver accurate results, analysts need to define the project's key factors to draw out specific directions for the visualization methods in the projects. This will assist them in identifying all of the constraints, qualities, and requirements associated with their project that will impact their approach towards it. Various methods can be employed to examine crucial factors; however, it is essential that they are closely linked to the subject matter being investigated.

Kirk emphasizes that analysts should always consider the following question throughout the project: What is the intention and the reasoning behind undertaking the data visualization project?

Analysts can compile a comprehensive list of terms to signify various objectives associated with the creation of a visual representation of data. This approach is beneficial in identifying the relevant purposes and essential factors involved in the process. Figure 2.13 shows, for instance, a list of important phrases for a data visualization project.

Abbildung 2.13: Collection of key factors. Reprinted from Data Visualization: a successful design process (p. 45) by Andy Kirk, 2012, Packt publishing LTD. Copyright 2012 by the authors.

The visualization's function should also be taken into consideration, whether its function to explanatory, exploration.

Explanatory data visualization is the process of providing information to a reader in a way that is focused on a specific and targeted storyline. For example of explaining data, visualization could be illustrated by the graphic in articles, explaining the examining numbers of infected, fatalities, and recoveries of COVID-19 throughout the world.

Exploration data visualization is the process of comprehending the different facets of data. This visualization allows individuals to explore and interact with the information. Exploratory visualizations perform effectively for users with a specific purpose in mind, such as learning about the COVID-19 situation in their nation. The Figure 2.14 illustrates the geo-dashboard shown on the online endpoint owned by the World Health Organization to inform visitors of the COVID-19 pandemic spread as an example for the exploratory visualization.

Filtering, sorting, brushing (selecting or isolating specific data), variable modification, and view customization are essential features and methods to assist a user with data exploration.

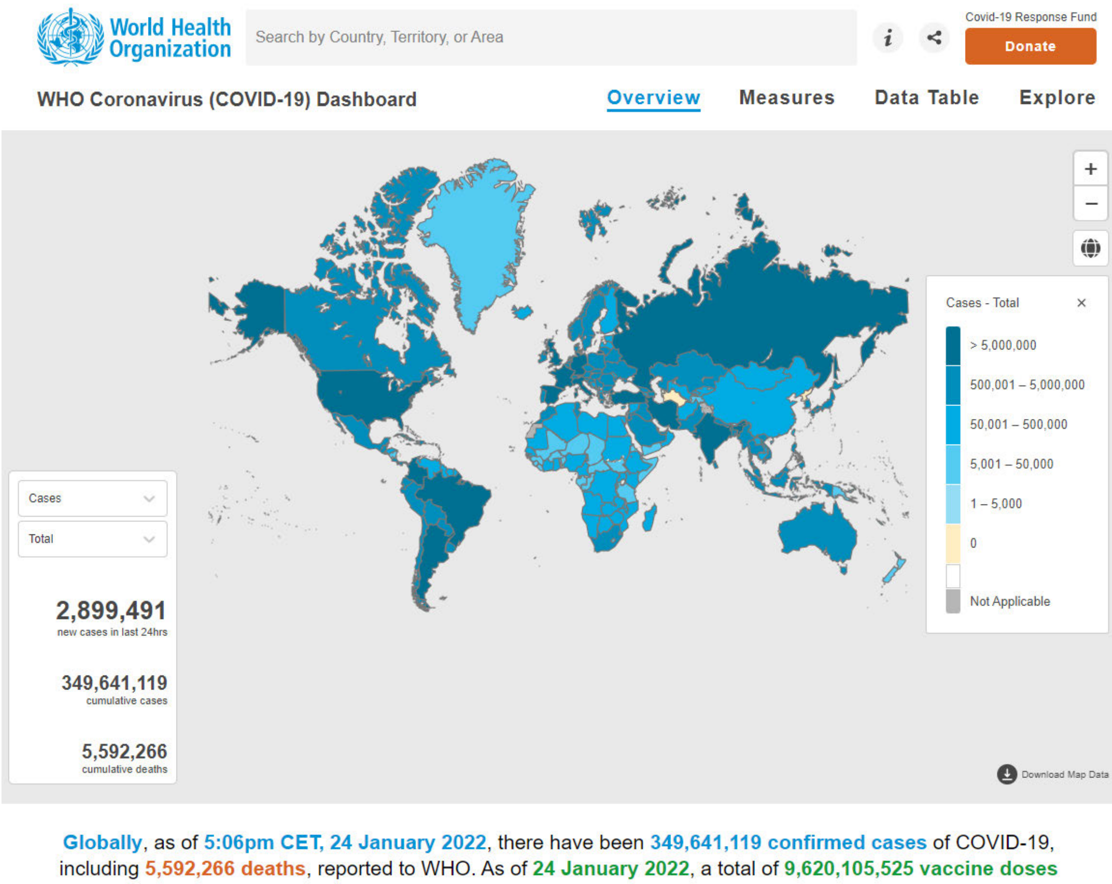


Abbildung 2.14: Representation of the geo-dashboard showed on the online endpoint owned by the World Health Organization to inform visitors on the COVID-19 pandemic spread. Screenshot from World Health Organization on 12 December 2021 [15]

The more specific the data visualization's function, the more uncomplicated it is for analysts to design their data visualization.

In any project and organization, planning, preparation, and scoping is the crucial state of any process. Without this preliminary effort, we risk undermining the efficacy and efficiency of our future design process, something no designer can afford.

2.2.2 Developing editorial focus and understanding of data

After defining the project's objectives and identifying the project's contributing factors, analysts will come to the next phase of the process - developing editorial focus and understanding of data. This is the state where they will seek to strengthen and enhance their editorial focus on the main communication components of their visualization problem. It is essential to ensure the data visualization is always tailored to the intended audience. Editorial focus sets particular standards and expectations for content based on the audience's desires. By applying editorial focus, analysts can provide appropriate content to the appropriate audience at the appropriate moment.

Another critical action is preparing appropriate data, ensuring it will be suitable for the project's objectives. In a world where data is overwhelmed, analysts should prioritize the relevant data set and consistently iterate the process of selecting and examining it. Data can be obtained from a variety of sources, including coworkers, clients, or other third-party entities, structured systems and websites, or even personally cultivated and researched. After collecting data, a comprehensive assessment will establish the degree of trust in the suitability of the obtained data set. In this data evaluation, analysts should consider these two factors: the completeness and the quality of the data. When it comes to the completeness, they assess the size and the shape of the data collection, while the quality factor is to make sure there are no noticeable errors, no incomplete or missing items, no duplicates, as well as no unusual outliners. Following the examination phase is the transforming task, where analysts work to correct any detected problems to improve the condition of the data they will be working with for their design. For instance, filling the missing data, deleting duplicates, cleaning error data, and other necessary actions to improve the data quality. Lastly, analysts went over how to use visual analysis to make sense of their data, uncover stories, and present those insights to others.

2.2.3 Conceptualizing and evaluating visual design options

This step is helpful as a framework to adhesive our thinking and guides us on the right path. From Andy Kirk's perspective on picking reasonable visual design options, we should choose the proper visualization methods, consider the properties of our data, evaluate the level of accuracy in the interpretation and generate a suitable design metaphor.

People use different diagrams to design data. However, the utilization of appropriate visualization techniques can aid in simplifying the complex nature of the data while effectively communicating information and captivating the audience.

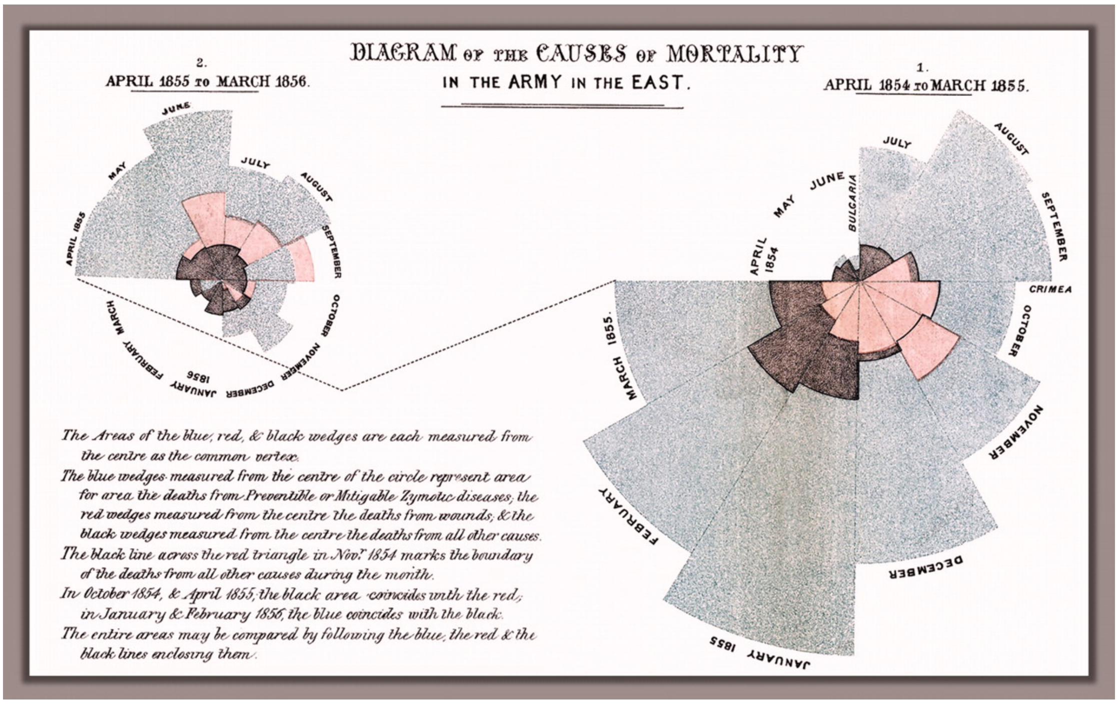


Abbildung 2.15: Florence Nightingale's 'coxcomb' diagram on mortality in the army.

One of the most well-known visualizations was made by English nurse Florence Nightingale, who was the first to use statistical graphics as effective tools for persuasion and policy change. Her most famous chart was the 1858 Diagram of the Causes of Mortality in the Army in the East, which detailed causes of mortality among soldiers fighting in Turkey during the Crimean War. This particular chart, a variation on Playfair's pie chart, is called a Polar Area Diagram or, in her honour, a Nightingale Rose diagram. As a result of her presentation, Queen Victoria formed a sanitary commission to visit Turkey and remove dead animals from the water, restore rotting floors, and enhance ventilation. As a result, the mortality rate in the area decreased from 52% to 20%, making this graph arguably the one that saved the most lives.

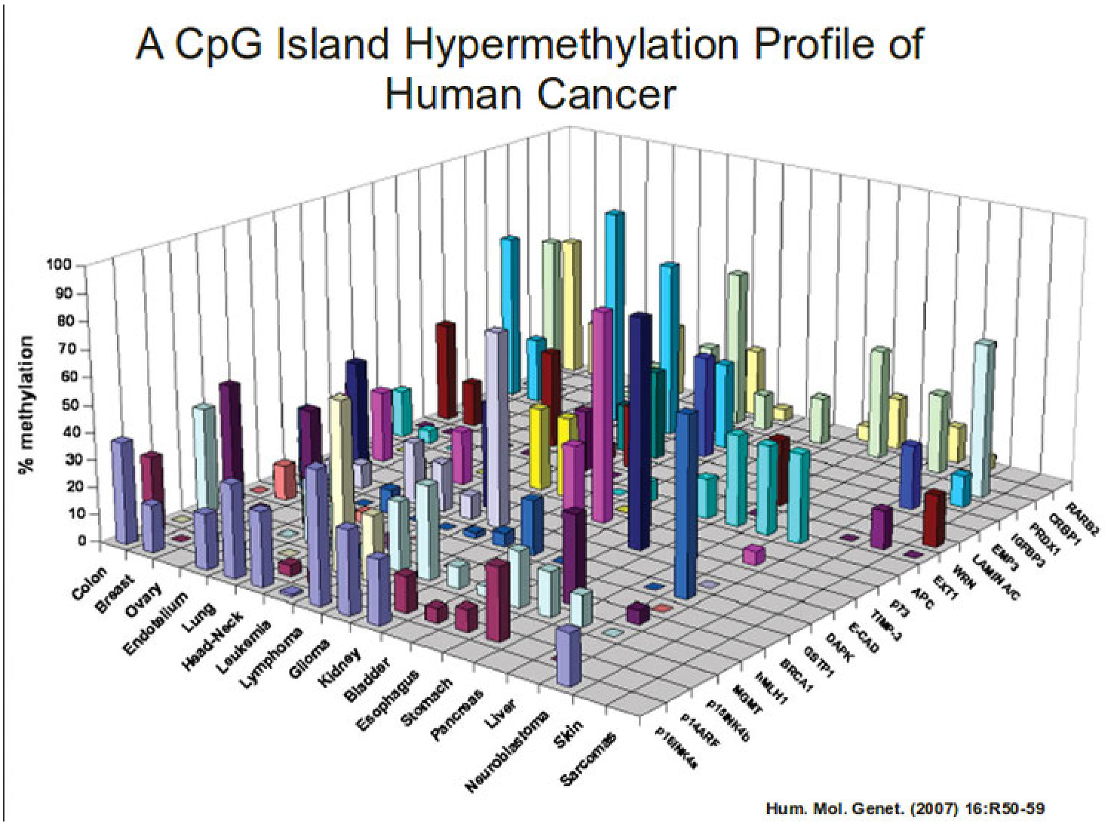


Abbildung 2.16: A CpG island hypermethylation profile of human cancer. Y-axis, frequency of hypermethylation for each gene in each primary[32]

Selecting an appropriate chart or graph is a pivotal aspect in the development of effective data visualizations. In the illustration above, displaying data visualization in 3D has been considered a bad data visualization. Too many variables in the graph make it difficult for the user to understand it at first glance. The 3D design has also resulted in excessive complexity since certain bars appear to be concealed behind the ones visible in the front. Due to the difficulty in understanding the data, it has failed in its intended role of displaying the data.

There are numerous ways to categorize the various types of data visualization. However, the following common taxonomy of information visualization techniques from Keim's research[41] are well-known to many information visualization professionals:

- Data to be visualized
- Visualization Technique

- Interaction and Distortion Technique

The above techniques will be discussed in greater depth in chapter 2, Data Visualization Methodologies.

2.2.4 Encoding data visualization

This chapter's objective is to briefly introduce visual encoding. The goal of encoding is to make the data easy to understand and interpret, so that people can quickly and accurately draw insights from it.

Transforming the data into a visual element is known as encoding data visualization. An accurate data visualization encoding means that anyone viewing your visualizations will comprehend what you're attempting to say or portray. Before successfully turning data into a visual representation, we need to know what visualization parts are involved and how these parts affect information visualization.

Much of the foundation for data categorization and the visual representation of data was systematized in 1973 when Jacques Bertin released The Semiology of Graphics[22]. He introduced marks as the fundamental components for representing information in visualization. Marks can take the form of points, lines, areas, surfaces, or volumes, and their appearance can be altered based on visual variables. According to Jacques Bertin, two distinct categories that make up visual encoding variables are planar and retinal. Planar variable is the position that can be accurately perceived by placing any graphic object on the plane (x,y or z). It is one of the most prominent visual encoding attributes that viewers can see in data visualization. There are six factors that the retina takes into account: shape, size, color hue, color value, orientation, and texture.

Furthermore, each visual characteristic offers its perception and interpretation technique. These qualities make it easier for each variable to show specific information. They are also closely related to their function in the visual hierarchy and their ability to display data. There are empirical levels of measurement that can be applied to data, depending upon the type of the variable being measured: nominal, ordinal, and quantitative.

- Nominal data contains information that can be labeled or classed into discrete categories. A meaningful order cannot be established among these categories. For instance, nominal data in breeds of dogs can be Dalmatian, Doberman, Labrador, and German Shepherd. Bertin divides the nominal scale level into associative and selective subcategories.

  - Associative variables are the lowest level of organization. It enables the grouping of all variables' components with distinct values.
  - Selective variables are the next higher level and the opposite of association. It allows the viewer to choose one component category, focus their attention on the placements of items inside that category, and ignore other component categories.

- The term ordinal data describes sets of information that can be organized in a hierarchy and then ranked. For example, ordinal data for the size of a dog can be small, medium, and large.

- Quantitative variables represents the highest level. Quantitative data is information that can be measured and counted in specific units and quantities. This data format is extensively used in mathematical calculations, algorithms, and statistical analysis. Here are some instances of quantitative data: the number of dogs in a household and the weight of a dog.

The visualization designer should determine a match between the data type and the used visual variables. The table below summarizes which attributes can be associated with certain data characteristics.

Type	Variable	Data Characteristics				
		Associative	Selective	Nominal	Ordinal	Quantitative
Planar	Postion	Y	Y	G	G	G
Retinal	Size	N	Y	P	G	G
	Shape	Y	N	G	P	P
	Color Value	N	Y	P	G	M
	Color Hue	Y	Y	G	M	M
	Orientation	Y	Y	G	M	M
	Texture	Y	Y	G	M	M

G = Good; M = Marginally Effective; P = Poor; Y = Yes; N = No

Tabelle 2.1: Visual variables' effectiveness for each data type. Based on MacEachren (1994)[49] and Slocum et al. (2009)[67]

2.2.5 Evaluating data visualization design

The visualization community employs a range of metrics to evaluate data visualization. These metrics will be discussed later in Chapter 3.

2.3 Data Visualization Methodologies

Choosing the correct visualization method is an essential part of presenting data, which will define and influence the purpose of visualization communication. Figure 1.2 shows the classification of information visualization methodologies, illustrating the data to be visualized, the visualization techniques and interaction and distortion technique.

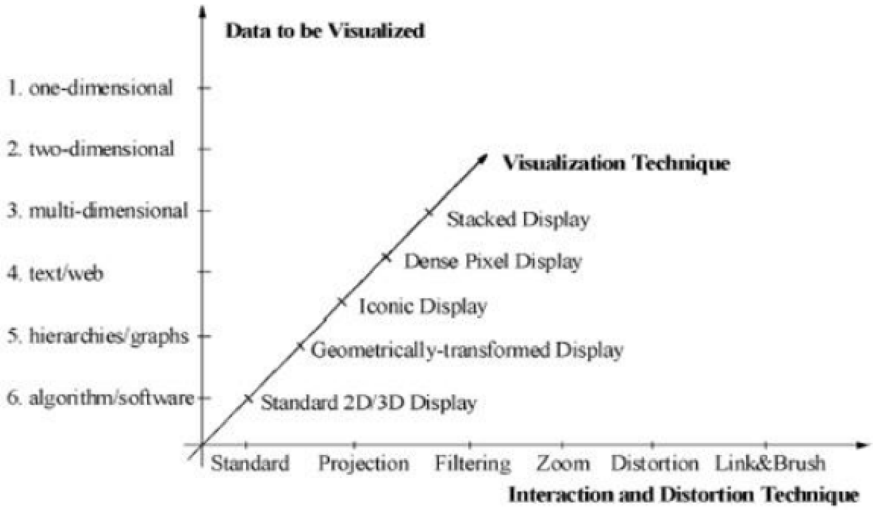


Abbildung 2.17: Visualization methodologies classification. (Keim, 2002)[41]

It is important to note that the three categorization dimensions might be taken to be orthogonal, as shown by Keim (2002) [41]. For any form of data, orthogonality means that any visualization strategy can be used in conjunction with any interaction technique, as well as any distortion technique. A particular system might be built to accommodate several data types and make use of a combination of various visualization and interaction strategies.

Based on a research paper from Akanmu and Jamaludin (2013)[27], this chapter focuses on information visualization techniques usage, which are designed to assist visual analytic professionals and users in specific domains.

2.3.1 The methodological framework

Keim's five classes of visualization techniques are a way to categorize and understand the different types of visual representations that can be used to convey information. These five classes of visualization techniques include standard 1D-3D graphics, iconographic, geometric, pixel-oriented, and graph-based or hierarchical techniques. Each category serves a unique purpose and is used to achieve different goals regarding understanding and communicating data.

Standard 1D - 3D graphics These are the most common types of visualizations and include visualization like bar charts, line graphs, scatter plot, etc and 3D models. Visualizations of this nature are used to present data comprehensibly and interpretably and can effectively demonstrate trends, patterns, and interrelationships among diverse data sets. For example, a bar chart can be used to compare the sales of different products, while a line graph can be used to show the trend of a stock's price, and a 3D model can visualize a building or a molecule. Standard 1D-3D graphics are commonly utilized in diverse fields including business, science, engineering, and education.

Standard 1D visualizations include bar charts, line graphs, and pie charts. They provide a standardized method of numerically displaying data and allowing comparisons across data sets. Bar charts are a suitable tool for comparing the sales of various products, whereas line graphs are a good choice for displaying the trend of a stock's value over a period of time. Pie charts are a useful tool for visually representing the proportions of various data sets.

Standard 2D visualizations include scatter plots and heat maps. These illustrate the connection between two variables and help spot anomalies and trends. Scatter plots, for example, can show how a person's age and wealth are related, while heat maps can show how a variable is distributed.

Standard 3D visualizations include 3D models and 3D diagrams. Data could be presented in a more immersive and interactive manner using these. 3D models have the potential to provide visual representations of complex structures such as buildings or molecules. For instance, 3D diagrams can be used to explain the complex structure of computer networks.

The primary advantage of using standard 1D-3D graphics is their ability to efficiently and unambiguously convey information. They give context to the information and make it simpler to see relationships and tendencies. In addition, they have the potential to convey intricate data in a comprehensible way.

However, that even the highest quality of 1D-3D graphics has its limits. They may not be able to express information in great depth and can only display a small quantity of data at a time. In addition, it is possible that certain types of data may not be well-suited for these representations, especially those that are highly complex..

Iconographic techniques involve using standardized symbols and icons to represent different types of data, facilitating the quick identification and comprehension of the information presented. For example, a standardized icon could be used to represent a bar in a bar chart, while a standardized icon could be used to represent a line in a line chart. This facilitates a quick comprehension of the data's significance for individuals unfamiliar with the particular symbols employed.

Iconographic techniques are useful for conveying information clearly and understandably. Additionally, they can be employed to demonstrate the connections between various data sets. Maps can be utilized to display the locations of various cities, flowcharts can be employed to illustrate the manufacturing process of a system, and diagrams can be utilized to clarify the structure of a computer network. Iconographic techniques are commonly utilized in various fields, including cartography, information design, and technical communication.

Geometric techniques These types of visualizations use geometric shapes and patterns to represent data. Some examples of visualizations are scatter plots, heat maps, and contour plots. Geometric techniques can be advantageous in displaying patterns and relationships among various data sets and also help in detecting trends and outliers. While scatter plots help display the correlation between two variables, heat maps effectively illustrate a variable's distribution, and contour plots are ideal for depicting the level of a variable. Geometric techniques find extensive application in various fields, including statistics, data mining, and machine learning.

Keim's visualization techniques are based on simple geometric figures such as points, lines, and polygons. Shapes are utilized to represent diverse types of data and can be employed to generate a broad range of visualizations. A scatter plot typically employs points to represent individual data points, whereas a line chart generally uses lines to depict the relationship between different data points.

Another important aspect of Keim's geometric visualization techniques is the use of spatial relationships.These techniques utilize geometric shapes and patterns to depict the spatial connections among various data points. A heat map can utilize color to represent the concentration of data points in a particular region, whereas a choropleth map can use color to indicate the comparative values of data points across various regions. This feature simplifies the process of identifying patterns and trends in the data, allowing individuals to comprehend the connections between various data points

Keim's geometric visualization techniques also often make use of interactive elements, such as buttons and sliders, which allow users to explore the data in more depth. This can be particularly useful for large data sets, where it can be difficult to understand everything at once. For example, a user might be able to zoom in on a specific area of a graph or filter the data to only show certain data points. This allows users to explore the data in more depth, making it easier to understand and interpret.

Pixel-oriented techniques use pixels to represent data. These can include things like digital images, videos, and virtual reality experiences. The aim of these techniques is to analyze data efficiently and effectively by focusing on individual pixels, rather than the entire dataset. Digital images can effectively display the surface of a planet, while videos can aptly capture the movement of a crowd. Additionally, virtual reality experiences can provide an immersive view of the interior of a building. Pixel-based techniques are widely utilized in various fields such as computer graphics, multimedia, and virtual reality.

Pixel-oriented techniques have advantage in their capacity to manage substantial amounts of data. These methods efficiently visualize and analyze datasets that would otherwise be too large to handle by focusing on individual pixels. Furthermore, pixel-oriented techniques have the capability to offer a significant amount of detail, enabling the identification of patterns and trends in the data that may be challenging to discern otherwise.

Another advantage of pixel-oriented techniques is their flexibility. These methods can be applied to a wide variety of data types, including images, text, and numerical data. Additionally, pixel-oriented techniques can be applied in various settings such as scientific research, business intelligence, and data journalism.

There are several specific techniques that fall under the umbrella of pixel-oriented techniques. Some examples include scatter plots, heat maps, and treemaps. Scatter plots are used to visualize the relationships between two or more variables, while heat maps are used to visualize the distribution of values across a dataset. On the other hand, treemap are used to visualize hierarchical data structures.

One of the most widely used pixel-oriented techniques is the scatter plot matrix, also known as a pairs plot or scatterplot matrix. This technique is used to visualize the relationships between multiple variables in a dataset.A scatter plot matrix refers to a visual representation consisting of a grid of scatter plots. Each scatter plot displays the correlation between two variables. Through analyzing the patterns depicted in scatter plots, scholars are able to identify correlations between variables that may not be immediately apparent.

Another popular pixel-oriented technique is the heat map. A heat map is a graphical representation of data where individual values are represented as colors. This technique is particularly useful for visualizing data where there are many variables and it's difficult to see patterns. Heat maps facilitate the identification of patterns and trends in data by utilizing color to represent the data.

Graph-based or hierarchical techniques represent data by using structures such as graphs or hierarchies. These methods are specifically designed to visualize relationships among datasets, facilitating the identification of trends and patterns. In particular, these methods specialize in displaying interrelationships between data sets, making it easier to spot trends and patterns. Tree diagrams help illustrate the classification of species, network diagrams are practical in depicting the relationships within a social network, and hierarchical maps are helpful in depicting the structure of a business or other organization. Because of their effectiveness in evaluating complex data, graph-based or hierarchical techniques are extensively employed in various fields, including biology, sociology, and management.

A Graph-based or hierarchical techniques offer a noteworthy benefit in terms of their versatility, as they can be effectively employed across diverse data formats such as images, text, and numerical data. Graph-based or hierarchical methodologies have diverse applications in various contexts, including scientific research, business intelligence, and data journalism.

SThere exist a number of distinct techniques that can be categorized as either graph-based or hierarchical techniques. For instance, consider the dendrogram, tree map, and node-link diagram. Both dendrograms and tree maps can be used to display hierarchical data structures, although they are most commonly applied to illustrate the relationship between nodes in a database. However, node-link diagrams depict the interrelationships among nodes within a graph.

Choosing the proper visualization technique depends on the specific data and its questions. A scatter plot, for example, would be an excellent choice if the purpose is to show the relationship between two variables. At the same time, a pie chart would be better suited if the purpose is to depict the proportion of distinct groups within a whole. Since different visualization approaches may be more appropriate for various audiences and settings, it is also essential to consider the audience viewing the visualization and in which context.

In conclusion, Keim's classification of visualization techniques is a helpful framework for understanding the various visualization approaches that are accessible, and it can assist researchers and practitioners in selecting the most suitable technique for a given dataset and purpose. Since most visualization problems are multi-dimensional and require a combination of techniques, it is essential to be aware of the advantages and disadvantages of each type before deciding on a solution.

2.3.2 Information visualization technique guide

Based on the classification of information visualization techniques by Keim D.(2002)[41], the following academic works have been studied by Akanmu Semiu A. and Zulikha Jamaludin(2013)[20]: Robinson et al. (2005)[61], Roth et al. (2010)[62], Mansmann et al. (2009)[52], Meyer (2012)[55], Pinto et al. (2012)[59], Wang et al. (2012), Lirong et al. (2011), Simon et al. (2011)[66], Kohlhammer et al. (2010), Schaefer et al. (2011), Stoffel et al. (2010)[68], Oelke et al. (2009)[57], Wanner et al. (2009), Ziegler et al. (2008)[42] and Krstajic et al. (2012)[46] to create a mapping of data types with the suitable interaction and visualization techniques that can be used in specific domains and for their intended purposes.

Tabelle 2.2: Data Visualization Techniques

Domain Application	Data types	Function of IV tool	Interaction Techniques	Visualization Techniques
Geographical Information Science	Multidimensional	The goal is to analyze geographic health data to gain a better understanding of the complicated and interdependent factors that contribute to the outbreak of epidemics.	Brushing, Selection,classification, and colour scheme	Scatter plot, bivariate map, time series plot,and parallel coordinate plot.
		For spatiotemporal crime analysis. Used to explore geographical data for the detection and prevention of criminal activity.	Linear and Composite animation functions, Temporal legend, Zooming	Scatter plots, interactive geographical map
	Geospatial data	To effectively communicate research findings to a diverse audience through interactive means.	Explore, Selection,View	Interactive geographical map,
Network Security	Hierarchical /Graph	To examine network traffic and find intrusion events	Filtering, classification, selection, Colour Scheme, Drag and Drop, Threshold	Tree Map layout, Line chart, grouped line-wise pixel plots
Crime Prevention	Spatial and Temporal Data	To comprehend crimes statistics in cities	Zooming, Selection, Filtering	Scatter plot, Map
Risk Management	Multidimensional	A multi-functional platform for data discovery and visual analytics		Tree Map
Biology Bacteriology	Graphs, Table, and Tree	To explore molecular biological data	Selection, Filtering	Spatial encoding, Line chart, Curve map
		To identify overlapping genes in bacterial genomes	Selection, Filtering, Zooming	Dense Pixel Display

Domain Application	Data types	Function of IV tool	Interaction Techniques	Visualization Techniques
Higher Education Institutions	Multidimensional	It is used as a classification tool of Higher Education Institutions.	Filtering, Selection, Explore, Boolean, Selection, Drag and Drop	Sunburst layout,Multiset bar chart, Ringbar chart
		A global multi-ranking instrument used to select the performance indicators of Higher Education Institutions.	Filtering, Selection, Explore, Boolean, Selection, Drag and Drop	Sunburst layout,Stacked Bar charts, Pie Chart
Higher Education Institutions	Multidimensional	To examine the Higher Education systems regarding their academic, organisation,financial and staffing autonomy.	Explore	Sunburst layout,Multi set bar chart, Ringbar chart
Aircraft Engineering	Table, Tree,Text	To manage information regarding Aircraft Product Development and evolution	Filtering, Selection, Translation ,Zooming	Stacked Tree Map
Commerce: Credit Information	Text	To analyse credit related web pages, and present the information according to the users‘selection.	Browse, Zooming, Drag and Drop, Colour Scheming	Tree Map, Dense Pixel display of Radial Graph
Online Shop	Text	To support the analysis of products‘ review	Selection	Dense Pixel Display.Colour Scheming
Banking	Multidimensional	To identify fraud in mortgage accounts	Selection, Filtering	Line Chart, Colour-Scheme
Governance and Policy Modelling	Text	To analyse and examine opinions and interactively simulate policy decisions	Filtering, Selection, Analysis	TreeMap, Hierarchical
News Reporting	Text	To analyse sentiment in Newsfeed	Selection, Filtering	Bar chart, Colour-Scheming
		To analyse sentiment in Newsfeed and emotional content of RSS news feed	Zooming, Details on-demand, Similarity-search, Filtering	Colour Scheming,Geometrical shape,Lines

Domain Application	Data types Text	Function of IV tool	Interaction Techniques	Visualization Techniques
News Reporting		To show the temporal characteristics of news story with its detail, allows incremental updates of the display, and sorts out newsbased on the time of its release.	Selection, Filtering, Zooming	Geometrical object-display
Library and Archives	Text, Image	To retrieve and visualizebooks, images, file cards andother documents kept in the archive	Selection	Colour Scheming
Financial Market	Multidimensional Time-Series Data	To visualize financial data and support analysts in making long term investment decision	Zooming, Selection, Filtering	Dense Pixel Display,Colour mapping, Line
Software Development	Software code in Hierarchical compound graph	To visualize relational information representing large software views.	Perspective zooming, Explore, Multi-level views	Graph model, Glyphs

3

Implementation and Evaluation of Data visualizations

3.1 The Case: COVID-19 Analysis Using an Exploratory Data Analysis Approach

3.1.1 Vision

As a picture is worth a thousand words, visualization helps transform numerous information into a broad picture that can increase the understanding of the information. COVID-19 is a disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and was first identified in humans in December 2019. Studying COVID-19's epidemiological characteristics is critical for continuing global efforts to control the virus. With the help of data visualization, political leaders and government employees could convey the situation and justify their choices during the COVID-19 epidemic. This chapter aims to discuss how data visualization techniques can be effectively used to explore and communicate the impact of the COVID-19 pandemic on different types of data and audiences and how these techniques vary in their ability to convey different information.

3.1.2 Overview of COVID-19 pandemic

On December 1, 2019, a patient in Wuhan, China, began displaying signs of a novel coronavirus, according to doctors.[3] That patient is believed to be the virus's first documented case. Several pneumonia cases with unknown causes have been reported throughout the month in Wuhan. Some of the symptoms include a dry cough, laboured breathing, and fever. At the end of the month, the Chinese government reports to the World Health

Organization (WHO) that Wuhan has been experiencing an outbreak of cases of pneumonia. Between December 31 and January 3, an estimated 44 cases were reported to the World Health Organization.[14]Furthermore, the Chinese government also discovered that many patients had regularly visited an open-air seafood and animal market.[53] As a result, they shut down the market to prevent the spread of this new virus.

On January 5 2020, the WHO reported that the infection was possibly caused by contact with live animals.[14] According to reports in the official media of China, researchers in China have found evidence that some patients are infected with a novel strain of coronavirus. Corona viruses are a large virus family that includes the common cold. At this point, there have not been any human-to-human transmissions that have been documented. On January 11, China reported the first death from the disease in Wuhan.[2] With millions of Chinese celebrating the upcoming lunar new year, finding a way to control the virus is becoming increasingly urgent. Wuhan is linked to other major Chinese cities Beijing ,Guangzhou, Shenzhen and Hong Kong and made up nearly 22,000 miles of high-speed rail. Furthermore, it is linked to the rest of the world by an international airport that transports 24.5 million passengers annually and connects to 30 cities worldwide.

After arriving in Bangkok, Thailand, a Chinese tourist from the Wuhan area is discovered to have a fever. On January 13, it was confirmed that this person was the first person outside of mainland China to have the new coronavirus.[14] Over the next week, cases appear across Asia, the United States and Europe. According to the WHO, there are 581 confirmed cases worldwide as of January 23. The first case in Germany was confirmed on January 27, 2020, near Munich, Bavaria.[3] By mid-February, the newly-emerging cluster of cases was completely confined. Baden-Württemberg reported many cases associated with the Italian epidemic on February 25 and 26. On February 25, a man who had been at a carnival in Heinsberg, North Rhine-Westphalia, on February 15 and was later found to be positive was able to be connected to the outbreak.[16] However, the German government could no longer track most of the likely chains of infection. On January 30 2020, WHO proclaimed COVID19 the sixth international public health emergency, following H1N1 (2009), polio (2014), Ebola in West Africa (2014), Zika (2016), and Ebola in the DRC (2019).[76]

As the rising concern about the outbreak of novel SARS, epidemiological data can be evaluated with the help of the exploratory data analysis (EDA) methodologies and the visualization model since these will raise the general public's situational awareness about the novel SARS-CoV2 outbreak. Overall, it is crucial to provide information to begin evaluating the risks and beginning containment with the intention that health workers, governments, and the general public can act together on a global scale to prevent its spread.

3.1.3 Materials and methods

The methodologies provided by the data visualization design process will be used for the project outlined previously in this chapter. This chapter is organized into five major components, each corresponding to a phase in the process as outlined in the data visualization design process. The five sectors are as follows:

- Identifying the purpose and key factors
- Developing editorial focus and understanding of data
- Conceptualizing and evaluating visual design options
- Data visualization with methods
- Building and evaluating of design solution

Each part begins with details on the relevant process step and then moves on to its techniques, the execution and meaning of which are first discussed, then implemented, and lastly summarized.

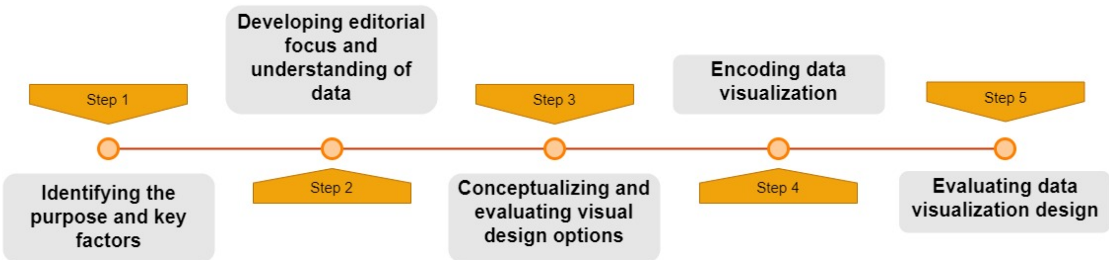


Abbildung 3.1: Data Visualization Design Process

Identifying the purpose and key factors

The chapter proposed an exploratory data visualization to analyze and portray the impact of the COVID-19 pandemic on various categories of data, such as epidemiological and economic data. First of all, epidemiological data includes information about the spread of the virus, such as the number of confirmed cases, deaths, and recoveries. Second, economic data, such as unemployment rates, GDP, and the stock market's performance, show how the pandemic has affected the economy. This exploratory data visualization will help audiences understand the data set and find the story inside it. For example, researchers require data visualization techniques that help them understand the underlying factors contributing to the virus's spread. On the other hand, the general public requires data visualization techniques that can present the impact of the pandemic on their daily lives in an easy-to-understand way.

Developing editorial focus and understanding of data

Data sources

For analysis of epidemiological data, I used the open extracted dataset of SARS-CoV-2[7] provided by acknowledged organizations:

- The Robert Koch Institute provided the data set SARS-CoV-2[45] infections in Germany, which includes daily updates of infections, fatalities, and recoveries recorded by German health authorities. The aggregated COVID-19 cases used in this study are gathered at the county level until July 8th, 2022.
- The National Platform for Geographic Data provided the COVID-19 dashboard and data access.[8]
- The Federal Statistical Office (Statistisches Bundesamt) and their open data platform, GENESIS, are responsible for facilitating access to precise demographic data[13] related to Germany.
- The Federal Agency for Cartography and Geodesy provided geospatial shape files.[9]

Data description

Dataset	Description	Columns
covid_de.csv	This file provides daily updated COVID-19 cases and deaths. The original dataset is being gathered by Germany's Robert Koch Institute	state, county, age_group, gender, date, cases, death, recovered
demographics_de.csv	This file provides general demographic data of Germany at the national and state level. The original dataset has been collected from Germany's Federal Office for Statistics	state, county, age_group, population
covid_de_vaccines.csv	This file provides the details of daily doses, broken down cumulatively by manufacturer, as well as the cumulative number of people receiving their first and complete vaccination	date, doses, doses_first, doses_second, pfizer_cumul, moderna_cumul, astrazeneca_cumul
de_state.∗	This file provides geospatial shape files for Germany's 16 federal states from Germany's Federal Agency for Cartography and Geodesy	persons_first_cumul, persons_full_cumul

Tabelle 3.1: Table represents different sources of SARS-CoV-2 datasets.

Data Observation

Analysts should first have an overview of the dataset they study. Following are the part of the exploited datasets for this analysis:

	state	county	age_group	gender	date	cases	deaths	recovered
2211084	Thueringen	SK Weimar	80-99	M	2022-06-08	1	0	1
2211085	Thueringen	SK Weimar	80-99	M	2022-06-10	1	0	1
2211086	Thueringen	SK Weimar	80-99	M	2022-06-14	2	0	2
2211087	Thueringen	SK Weimar	80-99	M	2022-06-16	1	0	1
2211088	Thueringen	SK Weimar	80-99	M	2022-06-22	1	0	0
2211089	Thueringen	SK Weimar	80-99	M	2022-06-23	1	0	0
2211090	Thueringen	SK Weimar	80-99	M	2022-06-24	1	0	0
2211091	Thueringen	SK Weimar	80-99	M	2022-06-27	1	0	0
2211092	Thueringen	SK Weimar	80-99	M	2022-06-28	1	0	0
2211093	Thueringen	SK Weimar	80-99	M	2022-06-30	2	0	0
2211094	Thueringen	SK Weimar	80-99	M	2022-07-01	1	0	1

Tabelle 3.3: Part of dataset covid_de

Columns	Description
state	German federal states of the observation
county	German Landkreis (LK) or Stadtkreis (SK), or counties of the observation
age_group	Age groups of the observation. The COVID-19 data is being reported for 6 age groups: 0-4, 5-14, 15-34, 35-59, 60-79, and above 80 years old
gender	Gender is divided as male (M) or female (F)
date	Date and time of the observation in YYYY-MM-DD
cases	Number of confirmed cases
deaths	Number of deaths
recovered	Number of recovered cases
population	Population counts for the respective categories
date(*covid_de_vaccines.csv*)	Date of vaccinations
doses	Daily count of total administered doses
doses_first	Daily count of first administered doses
doses_second	Daily count of second administered doses
pfizer_cumul	Daily cumulative number of administered vaccinations by manufacturer BioNTech/Pfizer
moderna_cumul	Daily cumulative number of administered vaccinations by manufacturer Moderna
astrazeneca_cumul	Daily cumulative number of administered vaccinations by manufacturer AstraZeneca
persons_first_cumul	Daily cumulative number of people having received their 1st shot
persons_full_cumul	Daily cumulative number of people having received fully vaccination

Tabelle 3.2: Description of SARS-CoV-2 datasets

Overview of the demographics datasets

	state	gender	age_group	population
191	Thueringen	male	80-99	57340
140	Saarland	male	15-34	113059
23	Bayern	male	80-99	298226
70	Hamburg	male	60-79	152316
85	Mecklenburg-Vorpommern	female	05-14	66830
5	Baden-Wuerttemberg	female	80-99	419471
63	Hamburg	female	35-59	325303
124	Rheinland-Pfalz	female	60-79	472082
147	Sachsen	female	35-59	658461
135	Saarland	female	35-59	171839
163	Sachsen-Anhalt	male	05-14	92088
112	Nordrhein-Westfalen	female	60-79	2009976
65	Hamburg	female	80-99	64937
171	Schleswig-Holstein	female	35-59	517809

Tabelle 3.4: Part of dataset demographics_de

Overview of the vaccine datasets

	date	doses	doses_first	doses_second	pfizer_cumul	moderna_cumul	astrazeneca_cumul	persons_first_cumul	persons_full_cumul
546	2022-06-26	3188	251	362	134794254	31205821	12794561	64698131	63342454
547	2022-06-27	24210	806	1240	134814943	31209133	12794562	64698937	63343694
548	2022-06-28	55276	2389	2968	134864081	31214611	12794589	64701326	63346662
549	2022-06-29	54108	1394	2183	134911453	31220996	12794602	64702720	63348845
550	2022-06-30	60681	1568	2831	134965667	31227140	12794604	64704288	63351676
551	2022-07-01	35233	1281	1907	134995843	31231858	12794604	64705569	63353583
552	2022-07-02	10285	506	873	135004072	31233778	12794622	64706075	63354456
553	2022-07-03	2639	100	202	135006273	31234175	12794622	64706175	63354658
554	2022-07-04	20674	594	957	135024044	31236964	12794626	64706769	63355615
555	2022-07-05	46473	1086	1702	135065444	31241857	12794627	64707855	63357317
556	2022-07-06	51329	1226	1963	135110711	31247661	12794628	64709081	63359280
557	2022-07-07	58919	1482	2802	135162859	31254093	12794641	64710563	63362082
558	2022-07-08	38743	1138	1695	135196487	31258915	12794641	64711701	63363777
559	2022-07-09	9935	500	652	135204356	31260867	12794641	64712201	63364429
560	2022-07-10	2443	79	149	135206328	31261282	12794644	64712280	63364578

Tabelle 3.5: Part of dataset covid_de_vaccines

Upon the first examination of the data sets and the related information, the following observation has been made:

Dataset	**Observations for the datasets**
covid_de.csv	• There are 2.890.909 entries, and eight columns • 3 Numerical columns : 'cases', 'deaths', and 'recovered' • 4 Categorical columns : 'state', 'county', 'gender' and 'age_group' • 1 Timestamp column: 'date'
covid_de_vaccines.csv	• There are 771 entries, and nine columns • 3 Numerical columns : 'cases', 'deaths', and 'recovered' • 8 Categorical columns : 'dose', 'doses_first', 'doses_second', 'pfizer_cumul', 'moderna_cumul', 'astrazeneca_cumul', 'persons_-first_cumul' and 'persons_full_cumul' • 1 Timestamp column: 'date'
demgraphics_de.csv	• There are 192 entries, and four columns • 3 Numerical columns : 'population' • 4 Categorical columns : 'state', 'gender' and 'age_group'

Tabelle 3.6: Description of SARS-CoV-2 datasets

Data Pre-processing

Before analyzing, improving the data's quality through preprocessing is a vital step in allowing the uncovering of meaningful information. Data preprocessing is used before any analysis is performed to clean and organize the raw data so that it may be analyzed effectively.

Check Missing Values In the next phase, the data is checked if there are any missing values.

	state	county	age_group	gender	date	cases	deaths	recovered
933405	Hamburg	SK Hamburg	05-14	F	2022-01-30	230	0	230
934135	Hamburg	SK Hamburg	05-14	M	2022-01-30	235	0	235
934413	Hamburg	SK Hamburg	05-14	NaN	2022-01-30	1	0	1
935170	Hamburg	SK Hamburg	15-34	F	2022-01-30	352	0	352
935984	Hamburg	SK Hamburg	15-34	M	2022-01-30	314	0	314
936387	Hamburg	SK Hamburg	15-34	NaN	2022-01-30	7	0	7

Tabelle 3.7: Example of rows with missing data

The number of missing values in each dataset was calculated by counting all the null elements.

Dataset	Colum	Percentage of missing data rows
covid_de.csv	gender	4,35%
	age_group	0,48%

Tabelle 3.8: Observation for missing values in SARS-CoV-2 datasets.

The table above reflects the percentage of missing data from exploited datasets. Vaccine and demographic datasets have no missing information. There are 13.876 values that are missing from the age_group row, which makes up only 0,48 percent of all the entries in the row. Along these, there are a total of 125.625 variables in gender row that are missing, which is 4,35 percent of the total values.

Tidying up Fields in the Data The missing data from row age_group in the covid data set accounts for just 0,48 percent of the data set, which is insignificant. We can remove those missing data as their existence in the dataset could be caused by user mistakes or data corruption. Nevertheless, missing data from the row gender counts for up to 5 percent of the data set due to human error, such as doctors failing to question and record the patient's gender or patients choosing not to tell their gender. Therefore, we could drop the uninformative data for the age_group and replace the missing data with UNKNOWN in the column gender, implying that the gender is unknown.

The datasets must be deducted from the unnecessary data (uninformative, repetitive, irrelevant). In order to remove duplicate data, we check if the copies of the same observation (rows) exist. In all datasets, there are no duplicated rows.

It has come to our attention that some of the columns contain data of the incorrect type. Therefore, the datasets followed specific standards to fit the model, which is the usage of capitalization and data format should be consistent. Moreover, the variable date was converted into the DateTime format.

To Summarize A total of 13.876 rows of data were eliminated, representing around 0,48 percent of the overall covid dataset. In addition, no duplicates were discovered across any of the three datasets. The variables contained in the datasets were transformed into the appropriate format. Once everything is taken into account, we can move on to the next step, which is to visualize and analyze this data collection.

Encoding and evaluating data visualization

Data visualization for Epidemiological data In order to raise sufficient awareness about the COVID-19 pandemic in Germany and the world, several EDA methodologies were used to analyze and visualize the datasets. The analysis was generated using the data set SARS-CoV-2 from the Robert Koch Institute and demographic data of Germany from the Federal Statistical Office of Germany. This study attempts to visualize and analyze data from 02 January 2020 to 02 February 2023.

From the dataset, it is reported that the first reported COVID-19 case in Germany was on 02 January 2020. The first dose of vaccine in Germany was distributed on 27 December 2020, nearly a year later. In the period from 27 December 2020 to 05 February 2023, the most commonly used vaccine in Germany is the Pfizer vaccine with a total of nearly 87 million doses (79.4% of total doses), followed by AstraZeneca at 12.7 million doses (11.6%) and Moderna at 9.8 million doses(9%). There have been 37.7 million confirmed cases of COVID-19 in Germany from 02 January 2020 to 02 February 2023. 98% of these cases have recovered from the virus (37.4 million recoveries). The Infection Fatality Rate (IFR), which calculates the percentage of cases with a death outcome to the number of cases, is 0.004% in Germany. This means that 4 out of every 1000 confirmed COVID-19 cases has a fatal outcome. The Mortality Rate (MR) for COVID-19 in Germany is the total deaths

(166.016 deaths) divided by the total population of Germany (83.019.213 population), which results in 0.00199%. This translates to 199 deaths per 100.000 population, or 1 death every 502 people.

This mortality rate is smaller compared to other countries such as Bulgaria with a MR of 357 deaths per 100.000 population, The United States of America with 230 deaths per 100.000 population, but it's also higher than certain countries such as Vietnam and Egypt with respectively 23 and 19 deaths per 100.000 population.

The number of active cases is equal to the number of confirmed cases minus the number of cases recovered cases and deaths. It refers to how many instances are still regarded as contagious. From the reported dataset, from 02 January 2020 to 02 February 2023, there are 220.112 active COVID-19 cases.

Visualization with Bar Chart The COVID-19 data is frequently reported as a daily count of newly confirmed infections over the preceding 24-hour period. A practical approach to contextualizing this data is using a bar chart, which visually displays how many people are infected per day.

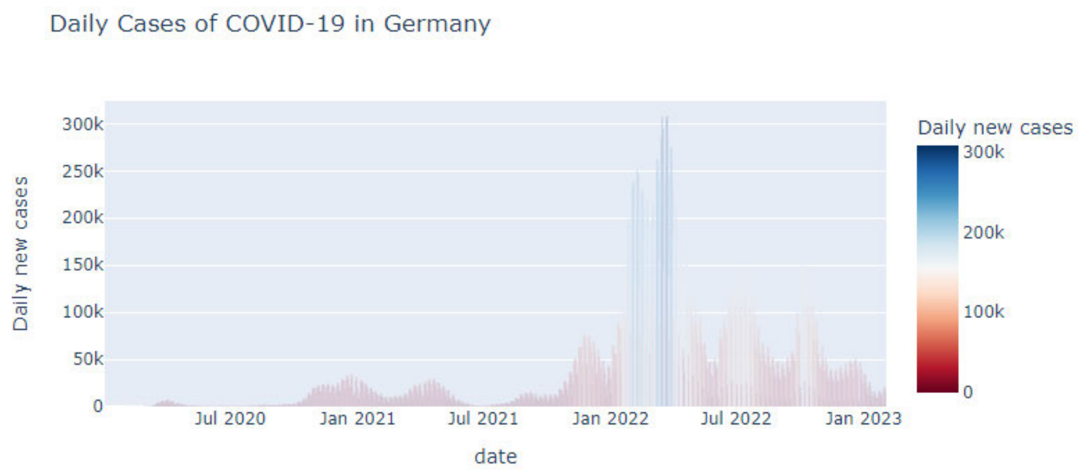


Abbildung 3.2: Daily New Cases in Germany

Figure 3.2 depicts the daily incidence of new cases in Germany from January 02, 2020, to February 05, 2023. The graphical representation illustrates that Germany has experienced several COVID-19 virus outbreaks since the onset of the pandemic. Each of these outbreaks is represented by the highest point on the graph. The initial surge happened in March and April in 2020, whereas the subsequent wave occurred in December 2020 and January 2021. The epidemic experienced a third peak during the spring of 2021, attributed to the emergence of the Delta variant, which was more contagious. In August 2021, the country has entered its fourth wave of the pandemic, with the majority of cases once again originating from younger age groups. The pandemic experienced a significant fifth wave, primarily caused by the Omicron variant. Daily case numbers surged to over 200,000 by mid-February 2022 and have remained at a high level throughout March.

It shows that the outbreaks are separated by brief infection periods. The time intervals between the peaks were approximately 9 months between the first and second, 4 months between the second and third, and 7 months between the third peak and the start of the fourth, and 5 months between the fourth and the fifth outbreak. The waves that have taken place so far follow a predictable pattern, with increases happening during the Easter season and again between Christmas and New Year's.

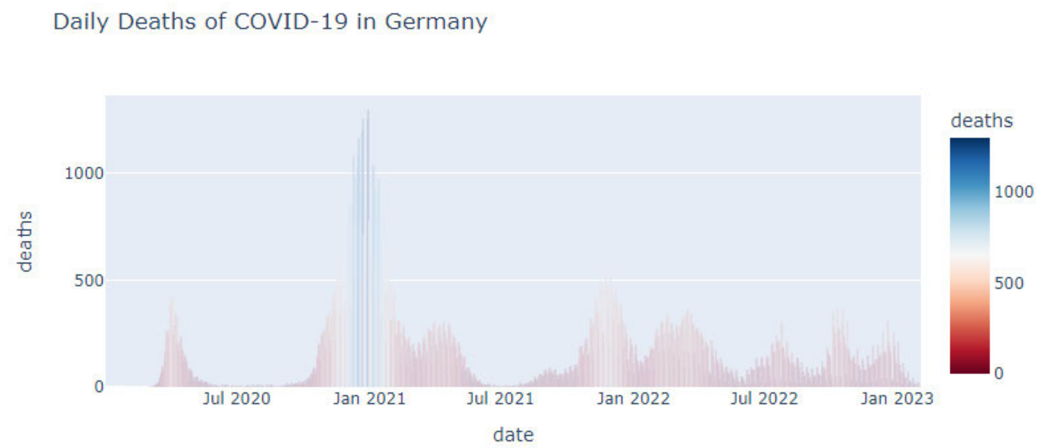


Abbildung 3.3: Daily New Deaths in Germany

The daily mortality graph of COVID-19 in Germany indicates a gradual increase in the number of deaths from March to April 2020. The highest recorded number of fatalities was observed in early April, exceeding 300 deaths per day. Afterward, there was a continuous decline in the number of fatalities that occurred each day until it reached its lowest point in late June 2020, when less than 10 deaths occurred each day.

During the autumn of 2020, the graph depicting the daily mortality rate of COVID-19 in Germany exhibited a sudden increase in fatalities, reaching its highest point of over 500 deaths per day towards the end of December. This surge can be attributed to the second wave of the pandemic, which was more severe than the initial wave and occurred later.

There is a third peak in the number of fatalities seen on the daily death graph of COVID-19 in Germany during the spring of 2021, reaching its highest point of over 1000 deaths per day . This rise is driven by the spread of the more virulent Delta variant. Following a peak in late April 2021, wherein the daily death count exceeded 300, a gradual yet consistent decline in fatalities persisted until the summer of 2021.

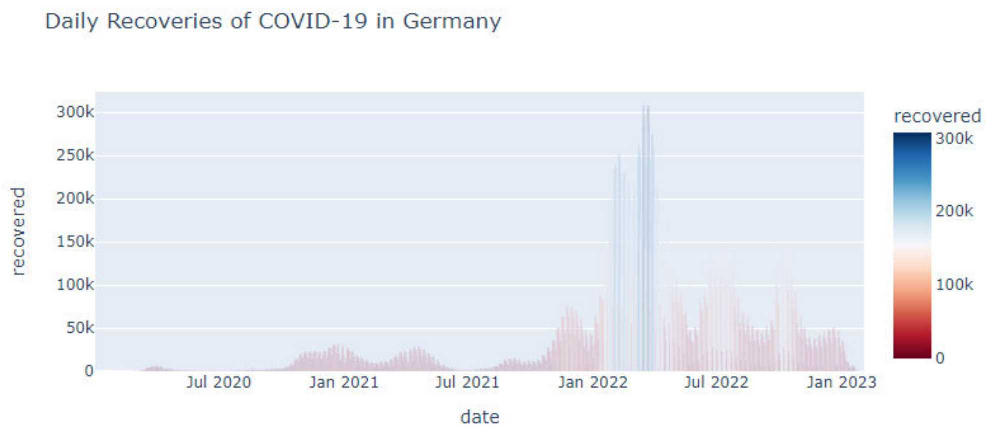


Abbildung 3.4: Daily New Recoveries in Germany

COVID-19 daily recoveries in Germany show an upward tendency from March 2020 to May 2020, with occasional waves due to large-scale recoveries or improved testing. During the summer of 2020, the implementation of public health policies and containment measures yielded favorable outcomes, as evidenced by the rapid pace of daily recoveries.

Beginning in the fall of 2020, the rate of daily COVID-19 recoveries in Germany climbed progressively, with a few falls that could be attributable to changes in testing or reporting protocols. Despite the fact that the second wave of the pandemic was more severe than the first, an upward trend in the daily recovery graph demonstrated that Germany was making progress in stopping the virus's spread.

The implementation of vaccines and other public health measures in Germany has accelerated the daily recovery rate of COVID-19 during the spring of 2021. Whilst short-term fluctuations may be observed in the graph due to various factors, the general trend indicates a progressive improvement.

Evaluating These figures are accurate for illustrating the scale of the pandemic, but they are less helpful for showing how the situation changes over time.

Although this visualization type has the ability to reveal significant patterns quickly, it is essential to take caution when analyzing smaller trends as they may be attributed to chance fluctuations in daily counts.

Visualization with Line chart The daily new case counts provide insight into the rate of expansion of the epidemic; however, they do not provide information regarding the general scale of its growth. It is essential to acquire a graphical illustration that portrays the cumulative number of cases up to the present moment. Thus, using a cumulative graph in conjunction with a line chart is considered the most suitable approach for representing the data in this specific context.

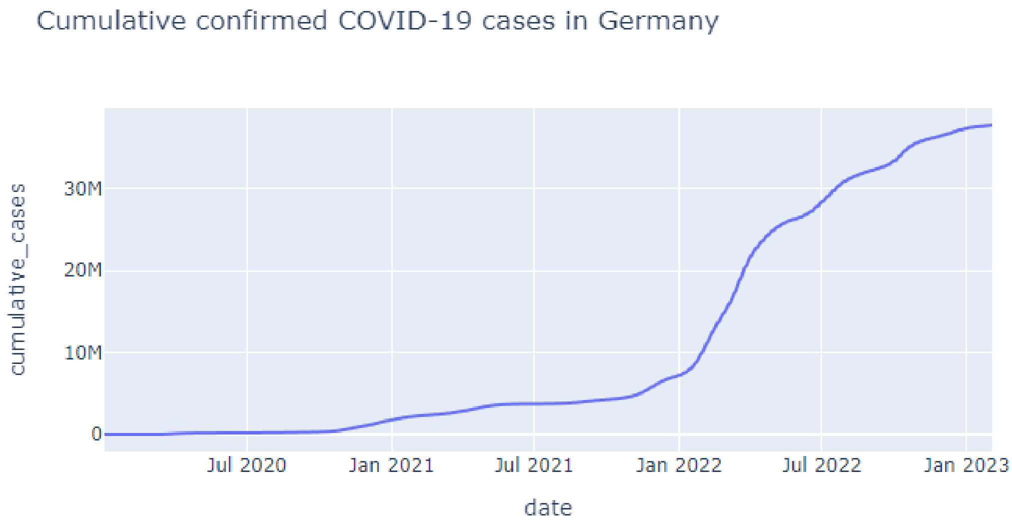


Abbildung 3.5: Daily cumulative case number of patients with COVID-19 in Germany

Cumulative graphs possess a significant attribute wherein they can be used to identify the trend of new cases, whether it is increasing, decreasing, or remaining constant. Figure 3.5 presents a comparative analysis of the spread of COVID-19 across various states in Germany. It is observed that the states of Nordhein Westfalen, Saarland, and Bayern have reported the most wide-spread of COVID-19 cases in Germany.

However, it is important to accurately analyze the information when using line charts to display cumulative data in order to avoid misinterpretation. A sudden rise in the number

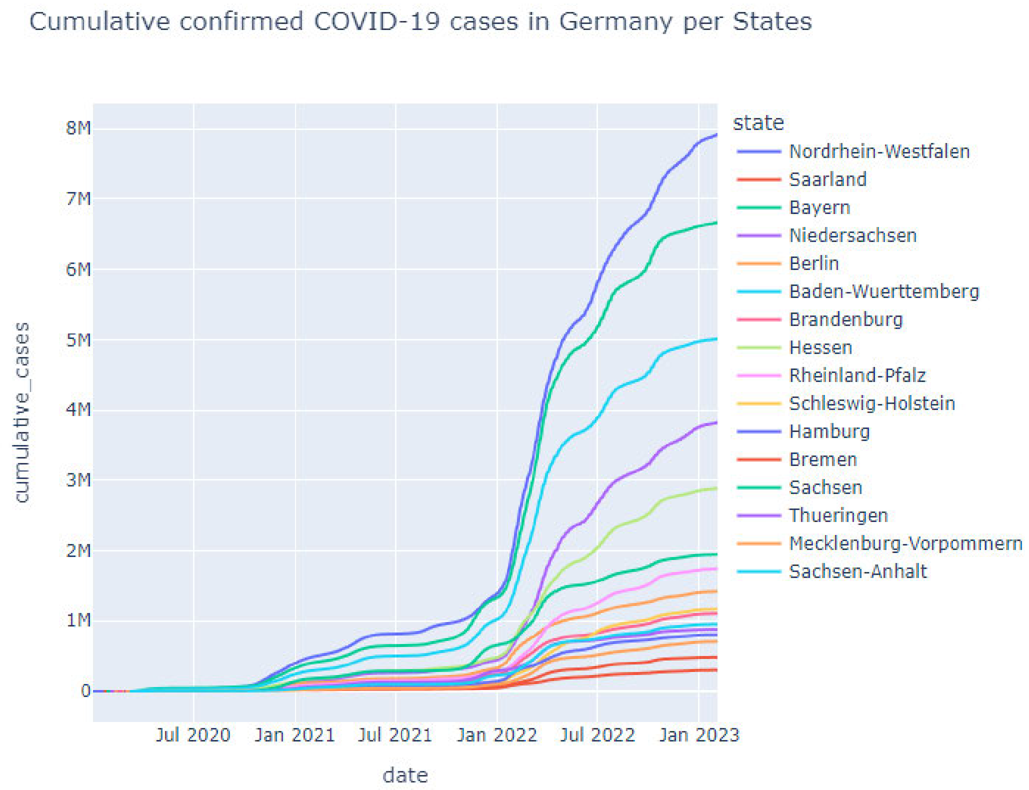


Abbildung 3.6: Cumulative confirmed COVID-19 cases by states

of illustrations may not always indicate a substantial increase in the number of new cases. Alternatively, the increase in numbers could be attributed to an increase in testing or improved reporting methods.

Visualization with Pie Chart Age is a significant factor in determining the severity and outcomes of COVID-19. It is reported that individuals over 80 have a 70 times higher risk of mortality from COVID-19, while those aged 70-79 have a 50 times higher risk compared to individuals under 40[12]. Based on Figures 3.7 and 3.8, the largest proportion of cases is attributed to the 35-59 age group. The mortality rate among individuals in this particular age group is comparatively low, not only within the confines of Germany but also across various global regions. Notably, regardless of their relatively larger population size, this demographic continues to make up most reported cases and exhibits the highest positivity rate. The accuracy of the statistic is confirmed that more than 90 percent of the documented fatalities in Germany occurred among individuals aged 60 years and above.

It is possible that Germany's testing efforts may not be sufficiently targeting the most vulnerable demographics, potentially resulting in a significant number of unreported cases and fatalities. Based on the available dataset, it seems unlikely that a substantial variation exists in the positivity rate among different age groups. However, it is essential to carry out further examination before entirely disregarding the likelihood. A precise enumeration can be attained at the conclusion of the period by comparing the average number of deaths with those during these months.

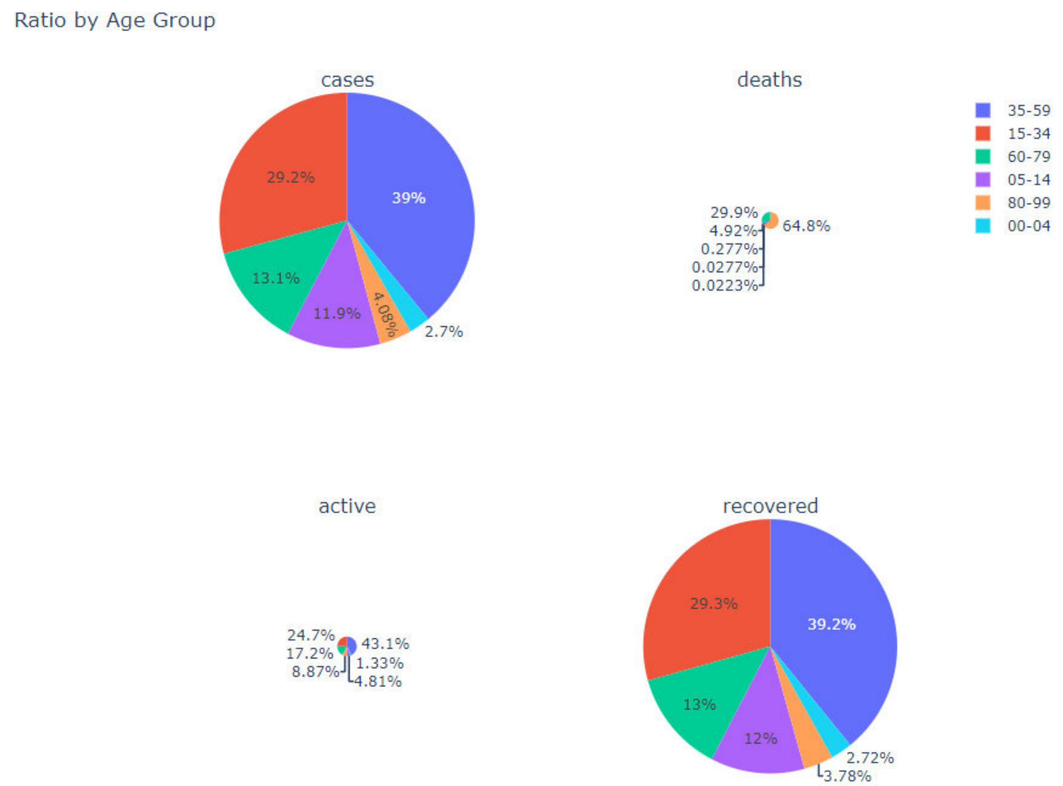
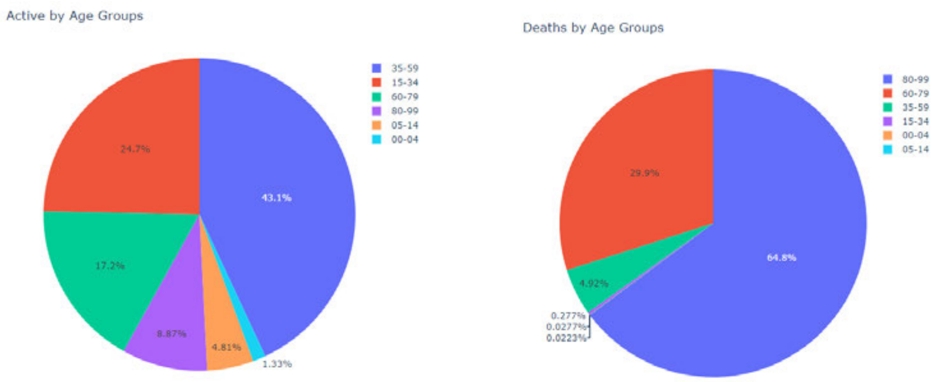


Abbildung 3.7: Distribution of confirmed, active, mortality and recovered cases by age groups in Germany



(a) COVID-19 active cases distribution by age groups

(b) COVID-19 mortality distribution by age groups

Abbildung 3.8: Distribution of active and mortality cases by age groups

Visualization with Dot Chart In the figures 3.9, the first dot chart illustrates the correlation between age group and the ratio of fatalities to confirmed cases of COVID-19. The second graph depicts the correlation between age groups and the proportion of confirmed COVID-19 cases in males compared to females.

It is widely acknowledged that older individuals are at a significantly higher risk of mortality compared to younger individuals. The ratio of males to females is fluctuating significantly, possibly due to high statistical uncertainty, particularly among the very young and very old age groups.

Visualization with the combination of various charts Whilst the virus had notable effects on individuals between the ages of 15-34 and 25-59, those belonging to the age group of 80-99 were found to be at a heightened risk of mortality. The mortality rate among individuals aged 80 and above who caught the virus was approximately 20 percent, whereas those under 60 exhibited considerably lower mortality rates.(Figure 3.10)

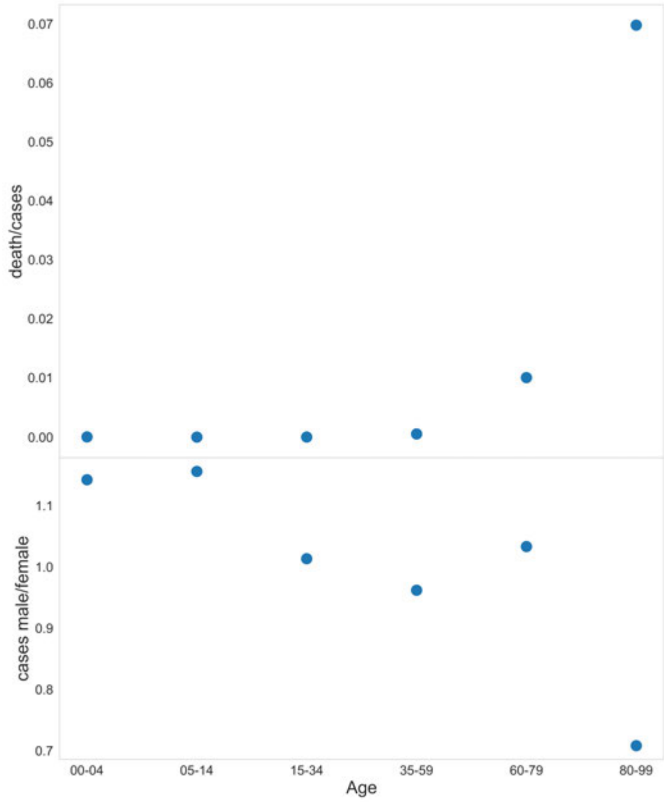


Abbildung 3.9: Relation of age and gender on COVID-19 severity

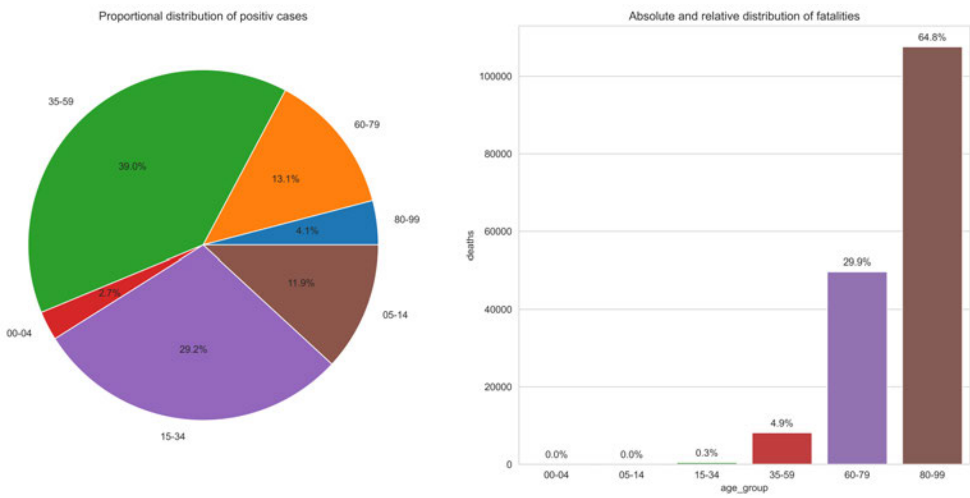
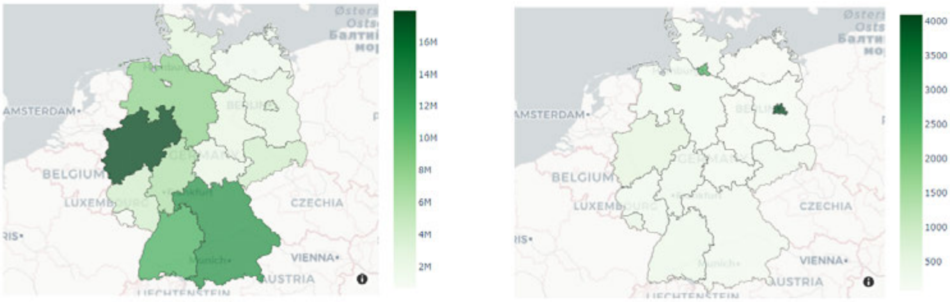


Abbildung 3.10: Distribution of confirmed and fatality cases by age groups



(a) Population per states

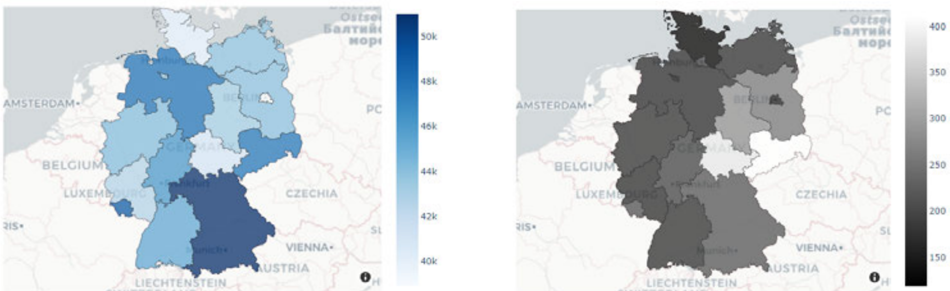
(b) Population density per states (ppl/sqkm)

Abbildung 3.11: Population and Population density per states

Visualization with Heatmap Heatmaps and tree maps are practical visualization tools for displaying the distribution of COVID-19 cases within various regions such as states, provinces, or cities. This methodology can aid in identifying hotspots and areas that require increased attention regarding healthcare and containment measures.

The figure 3.11 illustrates that while the population size in regions such as Nordrhein-Westfalen is large, the population density in urban centers like Hamburg or Berlin is significantly greater.

The incidence of cases and fatalities is primarily concentrated in states with high population densities. In the following, one can assess the normalization with respect to population and population density. It is noteworthy that the impact of the previously mentioned event is comparatively lower in the eastern region of Germany as opposed to its western counterpart.

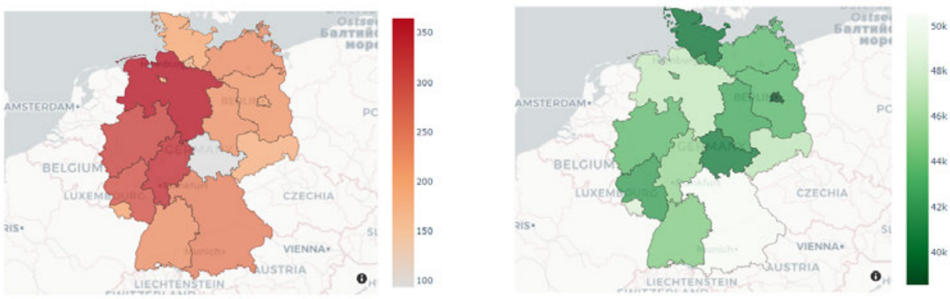


(a) Confirmed cases per states

(b) Deaths per states

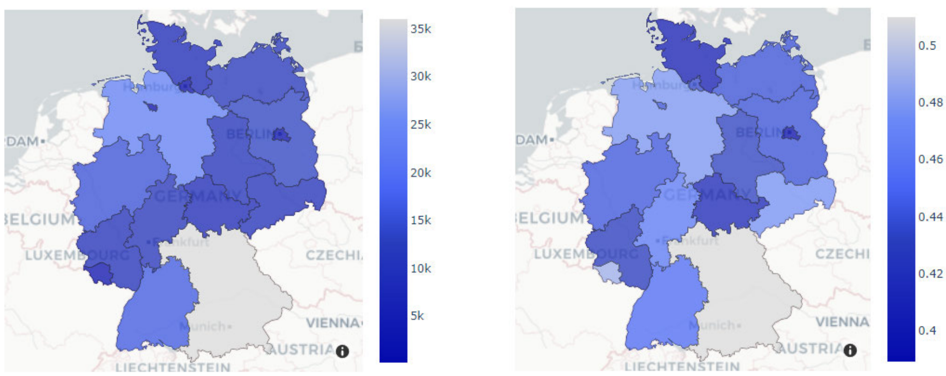
Abbildung 3.12: Confirmed cases and deaths per states

By dividing the number of cases with the population (Figure 3.13.a), it becomes apparent that the southern region has a higher number of cases. Bavaria is the leading state when population density is taken into account for normalization. This could be attributed to the fact that the region is relatively larger in comparison to other states.



(a) Confirmed cases per states (b) Deaths per states

Abbildung 3.13: Confirmed cases and deaths per states



(a) Confirmed cases per state's population (b) Confirmed cases per state's population density

Abbildung 3.14: Average COVID-19 Reported Cases per Population and per Population Density in Germany

Data visualization for Economical Data This section highlights various data visualizations depicting the impact of COVID-19 on different areas of the economy, as portrayed in works of literature related to the pandemic ([28],[50], [21], [39], [34],[19]) It also explores how these data visualizations represent the economic situation amidst the COVID-19 pandemic.

The COVID-19 pandemic is currently posing a significant threat to the global economy, as acknowledged by the OECD. This impact is comparable to the great recession experienced in 2008–2009. [21] Figure 3.15 shows the growth in the GDP and how it has responded to the different crises from 1995 to 2020. Using a line chart as a data visualization tool allows the audience to easily understand the overall trend of the economy and detect any changes or variations in the growth rate.

Padhan and Prabheesh claimed that the impact of COVID-19 on the economy can be classified into two main categories: supply and demand effects. The loss of working hours leads to supply effects, while the decline in aggregate demand results from reduced income due to unemployment caused by lockdowns. This claim is based on the research conducted by Maliszewska et al.. In their work, the pandemic has impacted the economy through various channels, which include (1) A direct reduction in employment, (2) The increasing

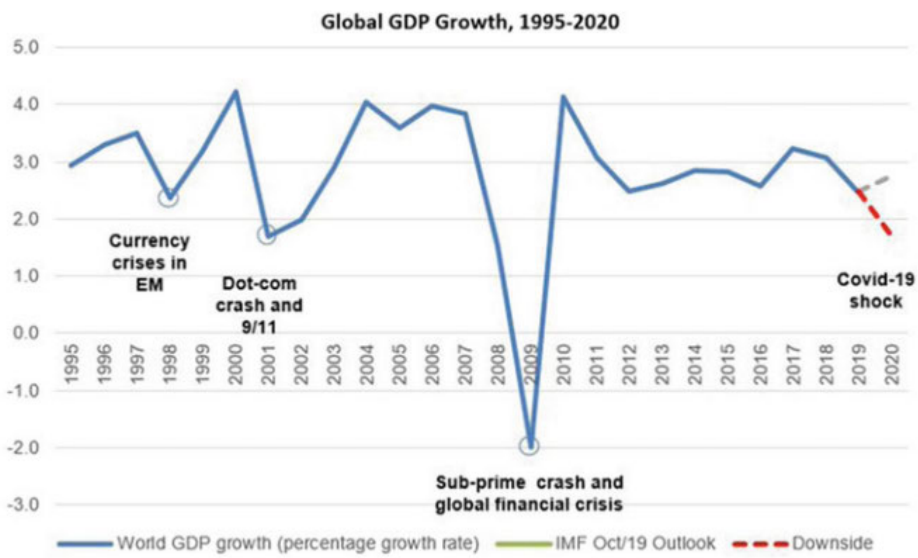


Exhibit 3.1 Global GDP growth, 1995–2020. *Sources* UNCTAD, March, 2020

Abbildung 3.15: Global GDP growth from 1995 to 2020 expenses associated with importing and exporting goods and services lead to declining trade and productivity, (3) The decline in travel and international tourism, (4) A decrease in demand for services that require proximity between individuals.

The research mainly presents the data through bar charts and stacked bar charts (Figure 3.16, Figure 3.17). For instance, the loss percentage of GDP in various regions is indicated by the height or length of the bar chart in the below fig. Bar graphs are used here since this particular data visualization enables data comparison across different regions.

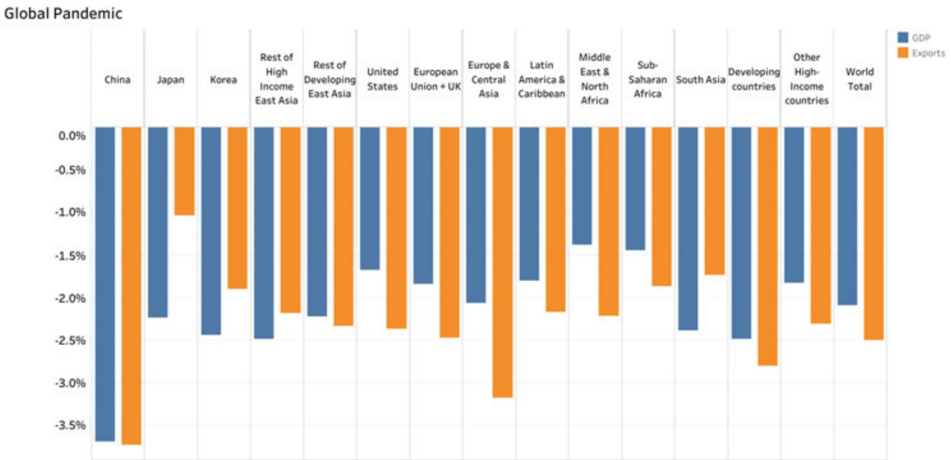


Abbildung 3.16: GDP and export implications of the global pandemic scenario (percent deviation from the benchmark)

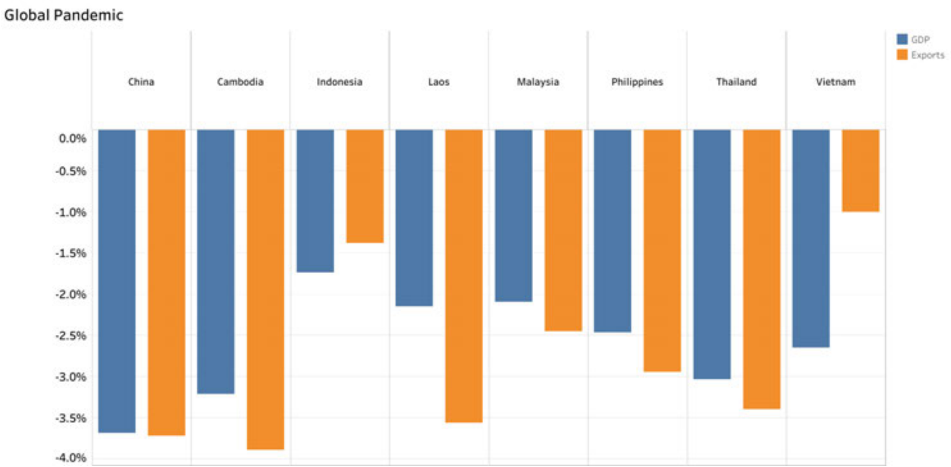


Abbildung 3.17: GDP and export implications of the global pandemic scenario for EAP countries (percent deviation from the benchmark)

In the midst of the ongoing global pandemic, China has been significantly affected, and unfortunately, this has also had a ripple effect on other countries. Experts predict that the ongoing pandemic will adversely affect China's GDP, leading to a decrease of 3.7 percent. The reason behind this is a supply shock that will result in a decline in production and exports, higher trade expenses, decreased tourism, and a move toward manufacturing. A supply shock can lead to a decrease in GDP, resulting in reduced employment and capital. Additionally, increased trade costs can cause both imports and exports to rise. The tourism industry is experiencing a decline in both inbound and outbound travel. Furthermore, there has been a shift in the types of expenditures being made due to decreased demand for sectors that have been impacted by social distancing measures and an increase in demand for goods. A decline in competitiveness and income may result in a 3.5 percent decrease in overall exports. In the same direction, there can also be a decrease of 3.2 percent in imports.

The worldwide Gross Domestic Product (GDP) is anticipated to decrease by 2.1 percent. Developing nations are expected to experience a decline of 2.5 percent, while high-income countries are projected to see a decrease of 1.9 percent. East Asia and Pacific (EAP) countries suffer significant reductions in their Gross Domestic Product (GDP) due to their deep integration through trade and the direct impact on tourism. There will be a 2.5 percent reduction in global exports, with China experiencing the most significant decline at 5.2 percent. The Lao People's Democratic Republic, Cambodia, is also expected to face losses at 3.6 percent and 4.4 percent, respectively.

Some countries are witnessing an increase in demand for their tourism exports due to the shift of tourism away from the EAP region. Nevertheless, tourism is experiencing a decline in all areas, and the EAP region has seen a decrease of about 30 percent in exports. The slight increase in tourism exports between the two countries has been negated due to the widespread effects of the shock that originated from China and East Asia, which has now affected other regions worldwide.

The article also utilizes a stacked bar chart to illustrate the effects of COVID-19 on exports between the USA, China, and other regions (Figure 3.18, Figure 3.19) These graphs provide a clear visualization of the composition of various industry sectors, highlighting the number of goods and services exported by the USA and China to other regions. Using a stacked bar chart allows the comparison of multiple industry sectors side-by-side, thereby enabling the effortless discovery of dissimilarities and similarities. The graph illustrates that the most significant decreases in exports are anticipated in

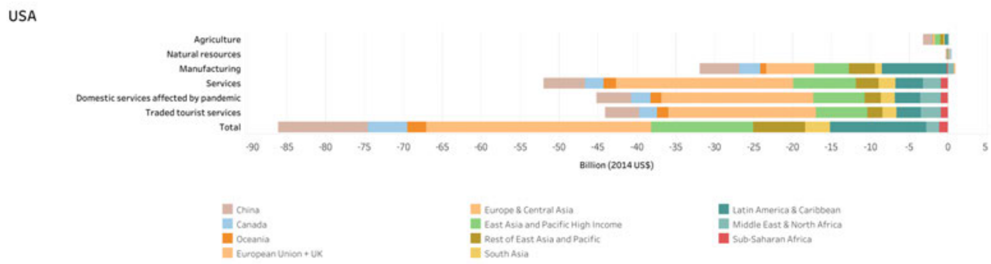


Abbildung 3.18: Impacts on US exports in the amplified global pandemic scenario

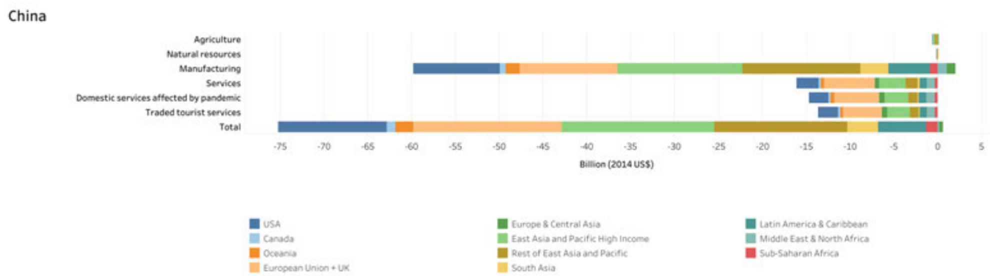


Abbildung 3.19: Impacts on Chinese exports in the amplified global pandemic scenario

Europe and the EAP due to the economic downturn and reduced demand in those areas. These regions are the primary markets for US service exports. The same issue happened in China, where manufacturing items and exports to the US, EU, and EAP countries have experienced the sharpest declines in volume.

Another indicator of the pandemic's impact on the economy is the performance of global stock markets.

Figure 3.20 displays that the Dow Jones has experienced a decrease of 19 percent, the Nikkei has experienced a decline of 27 percent, and the FTSE has experienced a 33 percent reduction. The reason for the decline in Nikkei was the apprehension and unease felt by Japanese investors regarding the impact of the COVID-19 pandemic on corporate profits.[17] On March 9, 2020 (Monday), during the final hour of trading, the FTSE-100 index experienced a significant drop of approximately 7.7 percent. This drop was attributed to widespread fear and panic caused by the coronavirus outbreak and the potential for the economy to enter into a recession. [36]

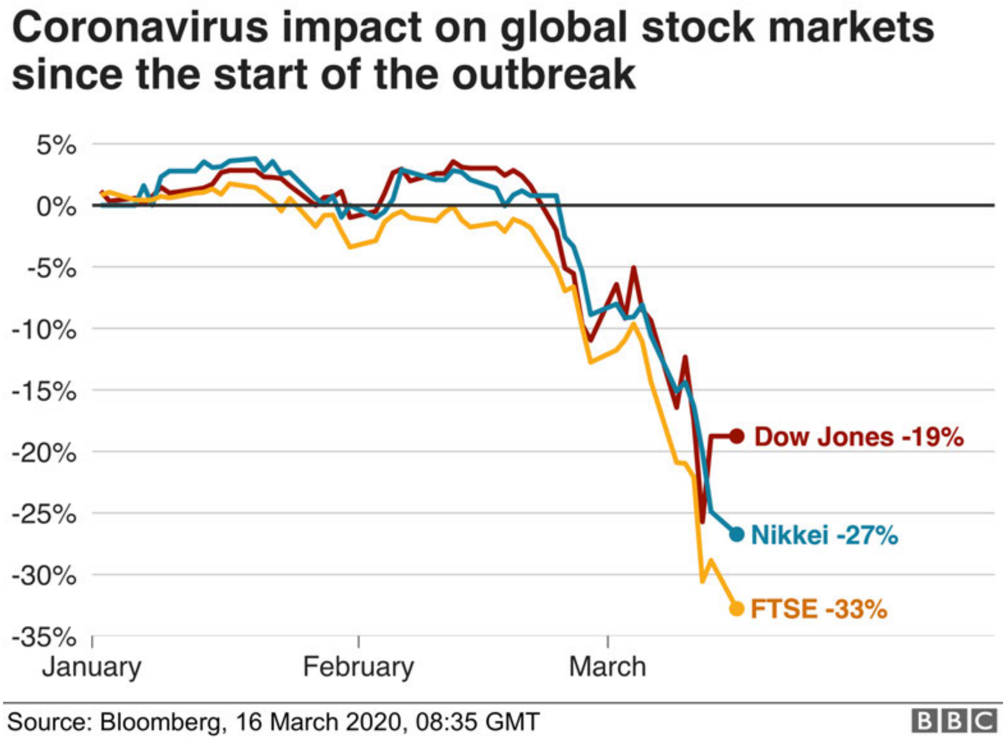


Abbildung 3.20: COVID-19 Impacts on global stock markets

Line charts are a popular choice for depicting the impact of COVID-19 on stock markets due to their ability to display trends and changes over time effectively. The stock market is a dynamic system that experiences changes daily. Hence, a line chart is a suitable option for visualizing the overall trend in the market by plotting the daily closing price of a stock or an index over a period of time. Another advantage is that it demonstrates how the market has reacted to specific events, such as the pandemic, the introduction of a vaccine, or changes in governmental regulations.

Meyer et al. documented and evaluated the reaction of businesses to the COVID-19 crisis in the period of August 2020. They suggest that the COVID-19 pandemic has not only caused a decrease in demand but has also significantly impacted inflation expectations. The pandemic has caused a shift in concerns about inflation among businesses and expert forecasters, who are now less worried about it in the coming year. Nevertheless, households have become increasingly concerned about inflation. They use a histogram (Figure 3.21) to illustrate their assertion, displaying the distribution of anticipated values among the respondents. Starting in April 2020, there was a noticeable decrease in expectations.

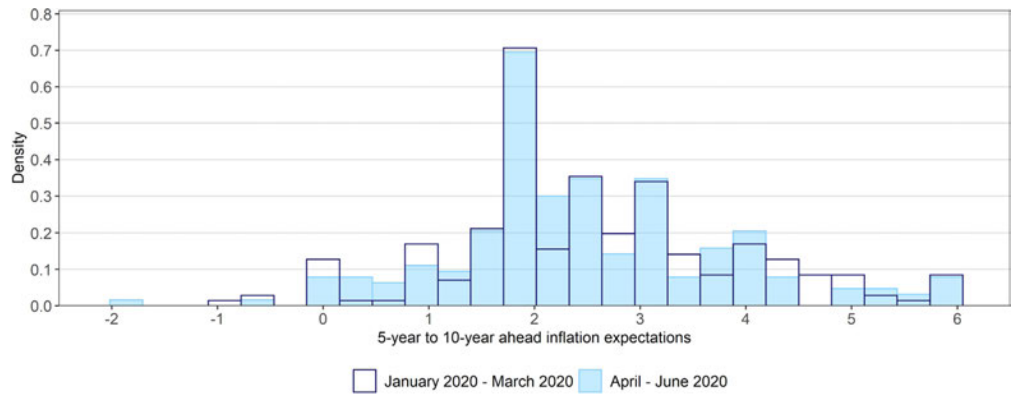


Abbildung 3.21: Distribution of firms' short-run inflation expectations from January to August 2020

It is reported that there is no significant correlation between companies' short-term and long-term expectations during the pandemic. To support their assertion, the researchers chose a scatter plot (Figure 3.22) as a suitable data visualization. Dots are used in the graph to symbolize the values of the variables, which are the short- and long-term expectations of the firms.

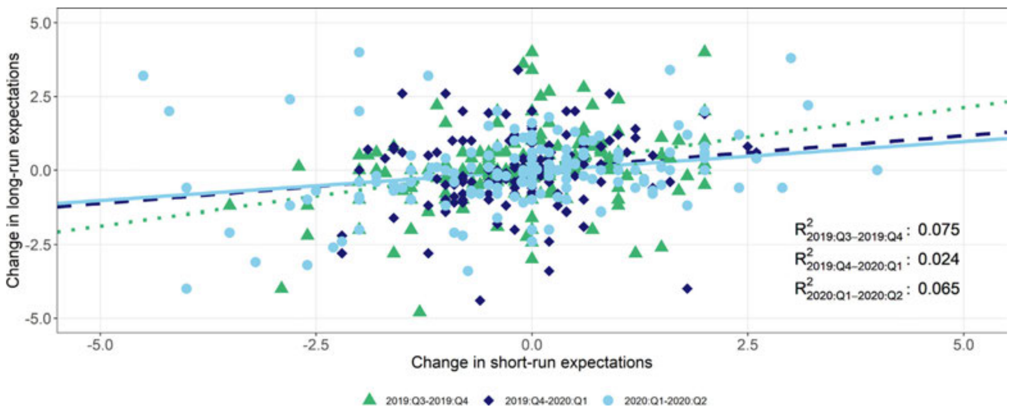


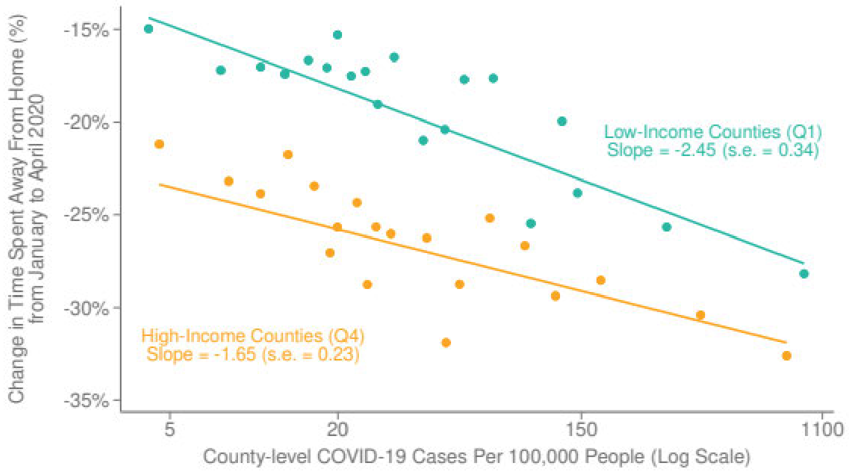
Abbildung 3.22: Changes in long-run and short-run inflation expectations. Notes: Both the x-axis and y-axis report the difference in percentage points. The solid fitted line belongs to the period December 2019 to March 2020 (pre-COVID), while the dashed fitted line belongs to the period from March to June 2020

Several studies have examined the impact of COVID-19 on consumer spending and employment rates ([28],[?],[31]) One example is the decline in business revenue and ongoing decrease in low-wage employment resulting from a significant reduction in spending by high-income individuals [28].

Notes: This figure disaggregates spending changes by income and sector using debit and credit card data from Affinity Solutions and national accounts (NIPA) data. Panel A plots daily spending levels for consumers in the highest and lowest quartiles of household income by combining total card spending in January 2020 (from NIPA Table 2.3.5) with our Affinity Solutions spending series. See the notes to Appendix Table 5 for details on this method. Panel B disaggregates the sectoral shares of seasonally-adjusted spending changes (left bar) and pre-COVID spending levels (right bar). See Appendix B.3 for the definitions of the sectors plotted in Panel B. Panel C decomposes the change in personal consumption expenditures (PCE) in the Great Recession and the COVID-19 Recession using NIPA Table 2.3.6. PCE is defined here as the sum of durable goods, non-durable goods and services in seasonally adjusted, chained (2012) dollars. The peak to trough declines are calculated from December 2007 to June 2009 for the Great Recession and from January 2020 to April 2020 for the COVID-19 Recession. Data sources: Affinity Solutions, NIPA.

Abbildung 3.23: Spending Behavior Changes Due to the COVID Epidemic

Data visualization aid in addressing the decline in corporate income and jobs. For instance, Chetty et al. provides a detailed breakdown of expenditure fluctuations based on income and industry categories, utilizing information obtained from debit and credit card transactions. The data is presented through different visualizations (Figure 3.23), including a line chart that displays spending changes by income quartile, a stacked bar chart that shows spending changes by sector, and a bar chart that compares spending changes by sectors during the COVID-19 pandemic and the Great Recession. Each chart provides a unique perspective to enhance the reader's grasp of the subject. The line chart displays the credit and card spending trend from January to November 2020. On the other hand, the stacked bar chart presents an overview of the spending distribution across different sectors. It shows that 57 percent of the decrease in spending was due to reduced expenditure on goods or services that require in-person contact. The bar chart in Figure 3.23.c shows the shift in spending across different categories throughout the COVID crisis and the Great Recession of 2009-2010. During the Great Recession, most of the decrease in consumer spending was due to reduced spending on goods. However, spending on services remained relatively stable. During the COVID recession, a significant portion of the decrease in total expenditures, precisely 71 percent, was due to decreased spending on services.



Notes: This figure presents a county-level binned scatter plot, constructed as described in Figure 2. The y-axis presents the change in time spent away from home from the base period (Jan 3-Feb 6 2020) to the three-week period of March 25-April 14 2020 (see Appendix H for details on the time-away-from-home series from Google Community Mobility Reports). The x-axis variable is the logarithm of the county's cumulative COVID case rate per capita as of April 14, 2020; with axis labels showing the levels on a logarithmic scale. We plot values separately for counties in the top and bottom quartiles of median household income (measured using population-weighted 2014-2018 ACS data). Data sources: Google Community Mobility Reports, New York Times.

Abbildung 3.24: The relationship between COVID-19 incidence and changes in consumer spending

Chetty et al. analyzed the impact of revenue loss on labor demand through using job posting data from Lightcast. The scatter plot in Figure 3.24 investigates the correlation between low-income job opportunities and consumer expenditure among individuals residing.

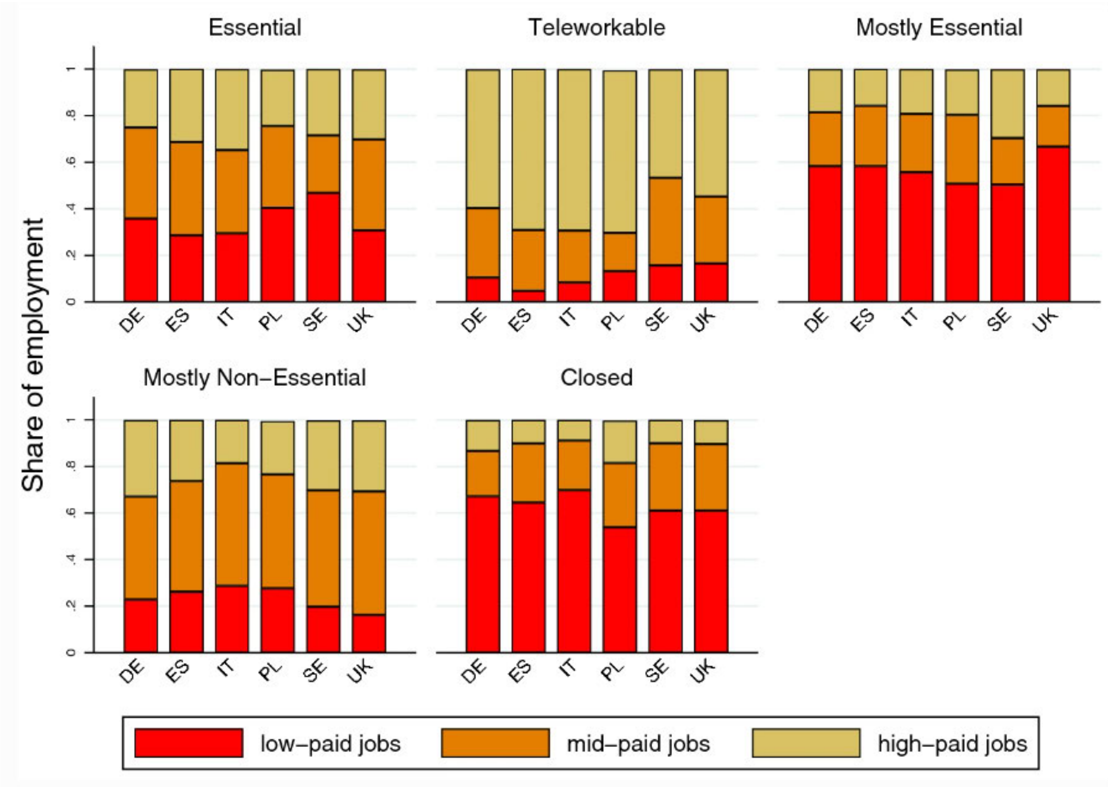


Abbildung 3.25: Share of self-employment by sector category and country

Fana et al. observed that the impact of employment is asymmetric both within and between countries. Due to their labor market structures and productive specialization, the most significant pandemic-hit nations like Spain, Italy, and the UK will likely experience the worst employment consequences.

There are numerous examples of utilizing data visualization to analyze COVID-19 data, and new ones are emerging regularly. This summary aims to provide noteworthy illustrations, direct readers to additional sources, and encourage further exploration of alternative implementations.

Literature summary

Tabelle 3.9: Summary of Data Visualization Techniques

Title & Authors	**Objective**	**Data visualization Purpose**	**Interaction Techniques**	**Visualization Techniques**
The Potential Impact of COVID-19 on GDP and Trade (Maryla Maliszewska, Aaditya Mattoo, Dominique van der Mensbrugghe)	Global economy research : Demonstrate the various transmission channels and diverse effects of COVID-19 on output and trade under different circumstances.	GDP and export implications of the global pandemic scenario	-	Bar Chart
		Impacts on US, China, Thai exports in the amplified global pandemic scenario	-	Stacked bar Chart
The Great Lockdown: Worst Economic Downturn Since the Great Depression (Gita Gopinath)	Global economy research : Compare the impact of great lockdown 2022 and global finacial crisis 2009 on the global economy	Comparation GDP growth in the great lockdown 2020 and global crisis 2009	-	Bar Chart
		The cumulative loss to global GDP	-	Line Chart
		Comparation GDP growth in the great lockdown 2020 and global crisis 2009 in different regions	-	Bar Chart
The impact of the COVID-19 pandemic on business expectations (Meyer et al.)	Global economy research : Assess how businesses have responded to the COVID-19 crisis up until August 2020.	Firms' mean quantitative sales gap by firm size	-	Bar Chart
		Level of disruption by activity type	-	Bar Chart
		Firms' experienced and expected wage changes.	-	Bar Chart
		Distribution of firms' short-run inflation expectations from January to August 2020	-	Histogram
		Inflation expectations and uncertainty of consumers, firms, and professionals	-	Line Chart
		Consumer price index component price change distribution	-	Histogram

		Pandemic impact on price indexes	-	Bar Chart
		Changes in long-run and short-run inflation expectations.	-	Scatter Chart
		Firms' experienced and expected wage changes by expected duration of the pandemic	-	Faceted Bar Chart
		Firms' experienced and expected wage changes by quantitative sales gap and level of sales disruption	-	Line Chart, Bar Chart,Histogram, Scatter plot
The Impact of COVID-19 on the Economy in Different Countries (Li et al.)	Economy in USA, China and UK analyse: compare macroeconomic indicators of different countries and evaluate the performance of financial markets.	The unemployment rate	-	Line Chart
		The rate of change ofconsumer price index	-	Line Chart
		The rate of change of realGDP	-	Line Chart
How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data (Chetty, Raj and Friedman, John N and Hendren, Nathaniel and Stepner)	Weekly statistics report on consumer expenditures, business revenues, open positions, and employment rates by county, sector and income group	Changes in Consumer Spending During the COVID Pandemic	-	Line Chart, Stacked Bar Chart, Bar Chart
		Association Between COVID-19 Incidence and Changes in Consumer Spending	-	Scatter Chart
		Changes in Small Business Revenues vs. Median Two Bedroom Rent	-	Scatter Chart
		Changes in Job Postings and Employment Rates vs. Rent	-	Faceted Scatter Chart
		Changes in Employment by Wage Quartile	-	Faceted Line Chart

		Evolution of the Association between Low-Education Job Postings and Low-Wage Employment with Rent	-	Faceted Line Chart, Faceted Scatter Chart
		Effects of Stimulus Payments on Spending: Event Studies	-	Faceted Line Chart
		Impacts of Stimulus Payments on Spending, by Income Quartile	-	Bar Chart
		Changes in Employment and Consumer Spending for Low-Income House holds vs. Workplace Rent	-	Faceted Line Chart
		Effects of COVID-19 on Educational Progress by Income Group	-	Line Chart
		Correction of Structural Breaks in Spending, Small Business Revenue, and Employment	-	Faceted Line Chart
		Industry Shares of Consumer Spending and Business Revenues Across Datasets	-	Bar Chart
		Consumer Spending in National Accounts vs. Credit and Debit Card Data	-	Bar Chart, Line Chart
		Relationship between Employment and Rent	-	Box plot
		Changes in Small Business Revenue	-	Heat Map
		Changes in Low-Wage Employment	-	Heat Map, Scatter plot

Employment impact of COVID-19 crisis: from short term effects to long terms prospects(Fana et al.)	Evaluating the impact of the COVID-19 crisis on employment by categorizing economic sectors based on the lockdown measures implemented in three European countries: Germany, Spain, and Italy.	Employment distribution across sector categories and country, (%)	-	Stacked Bar Chart
		Share of temporary employment by sector category and country	-	Faceted Bar Chart
		Share of self-employment by sector category and country	-	Faceted Bar Chart
		Jobs-wage terciles by country and sector category	-	Faceted Bar Chart
Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data (Dunn et al.)	Analysis of daily card transaction data to assess the impact of the COVID-19 pandemic on the economy	Growth Rates for Retail Based on the Fiserv Series and the Monthly Retail Trade Survey	-	Line Chart
		Study for Aggregate Retail and Food Services Sales Exluding Nonstore Retail	-	Line Chart
		Study for Accommodations	-	Line Chart
		Study for Food and Beverage	-	Line Chart
		Study for Gas Stations	-	Line Chart
		Study for Ambulatory Services	-	Line Chart
		Study for Hospital Services	-	Line Chart
		Study for Restaurants	-	Line Chart

3.2 Evaluation

3.2.1 Visualizations and tools

Data visualization is a crucial aspect of the information field, as it effectively conveys complicated messages by presenting simplified yet impactful data representations.[25]. Assessing the effectiveness of data visualization techniques can yield valuable insights and enhance their development and implementation. This chapter aims first to (1) review the existing definitions of the efficacy of visual information and later provide (2) an overview of prevalent metrics and methodologies for evaluating data visualization techniques' efficacy per definition. In addition, we discuss the (3) benefits and boundaries of each metric and provide examples of scientific articles involving these metrics.

The research by Zhu suggests that effective visualization depends on either a data-centric or task-centric approach. Several researchers embrace a data-centric perspective and propose that visualizations' effectiveness primarily depends upon the degree of alignment between the displayed data and the visualization itself. At the same time, other researchers adopt a task-centric perspective and hold the opinion that the efficacy of visualization relies on the specific task.

In 2002, Dastani published a research paper stating that visualization can be effective if it takes full advantage of the human visual system's capabilities[30]. The effectiveness of visualization depends on how well the data's intended structure aligns with the visualization's perceptual structure. In other words, visualization is effective if it presents the input data clearly and understandably. In addition, a procedure model for efficient data visualization is provided in the article, which will be further discussed in this chapter. It states that perception uses visual attribute values to establish perceivable interactions between visual components. Turk [70] supports the same opinion that the visualization structure should be consistent with the data structure.

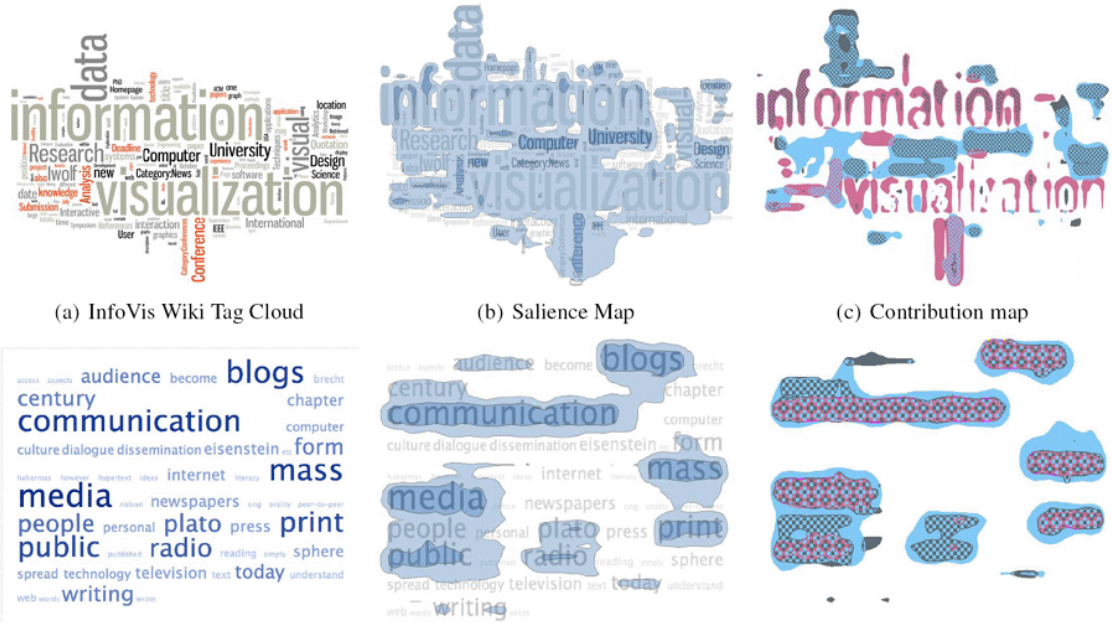
In his famous book The Semiology of Graphics, Bertin proposes that visualization designers should have a clearly defined objective in mind during the process of designing the visualization.[22]. Casner claimed that a visualization be tailored to a particular task and that the most effective visualizations enhance task efficiency. Zhu also pointed out that the empirical studies conducted in psychology and computer-human interaction appear to endorse the task-centric perspective. Numerous psychological studies have demonstrated that the efficacy of visualizations is influenced by the task ([56], [64], [48], [29]).

The findings of the subsequent studies indicate that the efficacy of visual encoding is primarily determined by the perceptual task undertaken by the users rather than the specific data presented.[56]

Researchers can decide on the appropriate approach based on the data type and research objectives, whether data-centric or task-centric. Depending on the established definitions, different metrics, and methodologies exist for assessing the effectiveness of data visualization techniques.

Data-centric evaluation approach Kennedy et al.'s study, which is based on both qualitative and empirical research, seeks to define success in visualization by examining the aspects that drive engagement in visualization's consumption and creation processes. They employed multiple methods, such as diaries, focus groups, and interviews, to gain insight into what makes people engage with visualizations. Six distinct factors that impact engagement with data visualizations were identified by implementing these methodologies. These factors include the subject matter, the location of the source or media, personal beliefs and opinions, time constraints, emotional responses, and the individual's confidence level and skills. [43] they argue that these factors have implications for the definition of 'effectiveness' concerning data visualizations. However, technical measures such as memorability, speed, recall accuracy, or comprehension consistency are considered insufficient in capturing what users perceive as an ëffective"visualization. This is because they need to consider other factors beyond the visualization that contribute to its effectiveness.

Another metric to measure the quality of data visualization is introduced in research by Jänicke and Chen. A well-crafted visualization effectively directs the viewer's focus toward the relevant components of the illustration. Therefore, the salience distribution across a visualization image is considered a metric for evaluating the visualization's quality. Jänicke and Chen offered a methodology to calculate a metric for a visualization image within the framework of a specific dataset. The author suggests that by utilizing various analysis tools such as salience overlays, contribution maps, and quality charts, it is possible to demonstrate the effectiveness of their quality metric in selecting the most suitable visualization from a given set of options. Additionally, this metric can be used to establish an optimal default setting for the visualization process. They recommend several procedures depending on whether users assess techniques or visualization. Given that the user is a visualizer and may be presented with various techniques, the following approach is recommended: (1) Use a method to visualize a sample data set and obtain a relevance map with the second above or third-level support. (2) Use the salience map, contribution map, and difference field to analyze the results of your experiments using varying approaches and parameters. Find the optimal approach and configuration of settings. (3) Making the current method and parameters the default for future exploration of similar data. If the user needs to assess the effectiveness of visualization in communicating information, the visualizer can use salience and contribution maps to objectively and visually evaluate each visualization created.



(a) InfoVis Wiki Tag Cloud (b) Salience Map (c) Contribution map

(d) Tag Cloud on communication (e) Salience Map (f) Contribution map

Abbildung 3.26: Implementation of A Salience-based Quality Metric for TagCloud visualizations.[40]. The TagClouds represent keywords of (a) the InfoVis Wiki and (b) communication. (b,e) Overlay of the original image with the salience map. (c,f) Contribution maps (color opponency (red), intensity change (gray) and orientation (blue). The contribution map (f) displays a more homogenous pattern and is better data visualization.

The study "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design,"by Jeffrey Heer and Michael Bostock [37], investigates the application of crowdsourcing, focusing on the Amazon Mechanical Turk platform as a tool for evaluating visualization design. The concept of crowdsourcing is presented as a scalable and cost-effective approach for gathering data and insights from a significant number of participants. The authors explain that formal user study approaches, such as in-person experiments, may result in substantial costs, consume significant time, and display constraints to sample size. Hence, crowdsourcing platforms like Mechanical Turk offer a valuable opportunity to collect various opinions and insights from a more extensive and diverse group of participants. Crowdsourcing offers the benefit of efficiently gathering data from a vast and heterogeneous group of individuals at a reduced expense compared to conventional approaches. However, they acknowledge potential limitations such as participant quality control, task complexity, and the demand to design tasks and instructions carefully.

Martin Wattenberg[75] took another approach to measure the efficacy of data visualization by introducing the Gestalt principles of perceptual organization, which outline how humans instinctively perceive and organize visual components according to particular relationships and patterns. The Gestalt theory comprises six distinct principles: similarity, continuation, closure, proximity, figure/ground, and symmetry order. These principles significantly impact how the audience perceives and comprehends the visual components of an information graphic.

Task-centric evaluation approach Burns et al. presented a structured framework based on Bloom's taxonomy and illustrated its application in generating a series of questions that methodically assess visualizations' efficacy. The effectiveness of data visualizations is evaluated using a framework based on the six stages of knowledge acquisition (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5)Synthesis, and (6) Evaluation)[23]

The Figure 3.27 summarize the application of Bloom's taxonomy as a framework for creating a series of questions that methodically evaluate visualizations' affordances. Figure 3.28 depicts a case study on COVID-19 that demonstrates how the framework enhances the current methods to comprehensively evaluate the effectiveness of a visualization design in enhancing a viewer's comprehension of visualizations.

In their study [74], Wall et al. proposed a framework that integrates usability heuristics with value-driven criteria to evaluate the efficacy of visualization. The study relied on an iterative design process involving input from domain experts and other stakeholders to determine value-driven criteria relevant to the selected application domain. The criteria consist of various factors, such as (1) Insight: demonstrates how visualization can facilitate both intentional and unintentional insights (2) Time: defines how visualizations can enhance the speed and efficiency of data comprehension, both in terms of searching for specific information and browsing through more significant sets of data. (3) Essence:

Level	Description	Example Tasks
Knowledge	Recall basic facts and definitions.	• Retrieve points • Locate value • Identify axis labels
Comprehension	Understand the information in context.	• Summarize main message/take away • Describe content of visualization • Explain the topic of the visualization
Application	Apply knowledge to a new problem or represent it differently.	• Use a percentage and total population to calculate a number • Calculate the difference between two points • Translate the data in a chart to a table
Analysis	Break down a concept into parts and understand their relationship.	• Describe a trend • Describe the relationship between two variables • Identify what data was used to come to a conclusion
Synthesis	Use knowledge to create something new.	• Predict a future value • Generate a new visual representation
Evaluation	Judge the value of information, backed by evidence.	• Justify a conclusion based on data • Judge which design is more appropriate

Abbildung 3.27: The table lists the six original Bloom's taxonomy levels with brief descriptions and visualization-specific tasks.[24]

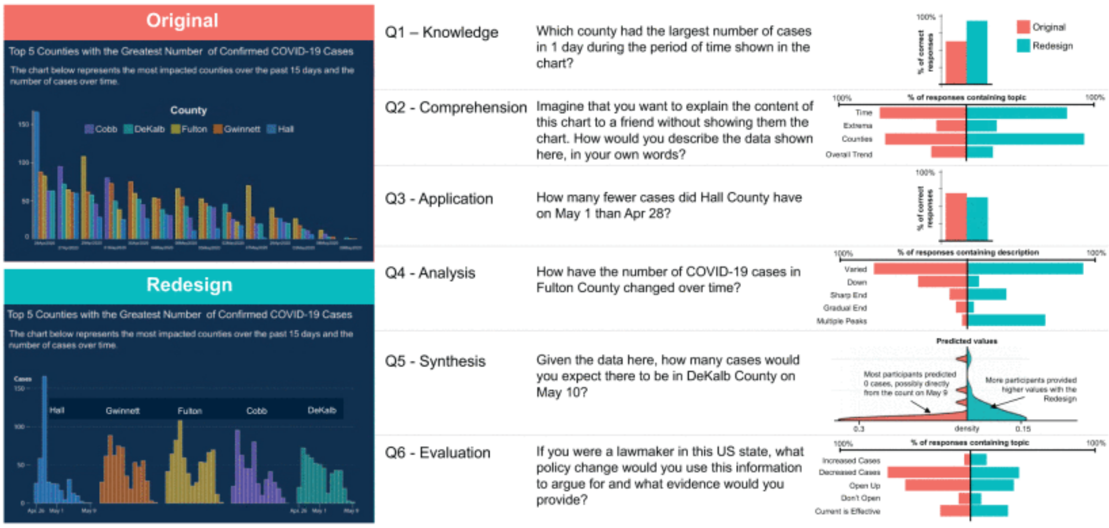


Abbildung 3.28: A summary of the experimental stimuli and questions used for the COVID charts. The graphs on the right display the outcomes for each examined level.

describes how a visualization effectively conveys the key information of a dataset in terms of providing an overview and context and (4) Confidence: describes how a visualization can enhance a user's confidence in their understanding of a dataset, particularly in terms of the quality of the data and quality of the visualization.(Figure 3.29)

This heuristic approach has several advantages. Firstly, it is capable of capturing the multifaceted nature of visualization effectiveness. Secondly, it is flexible and can adapt to diverse domains and tasks. The approach promotes collaboration between visualiza-

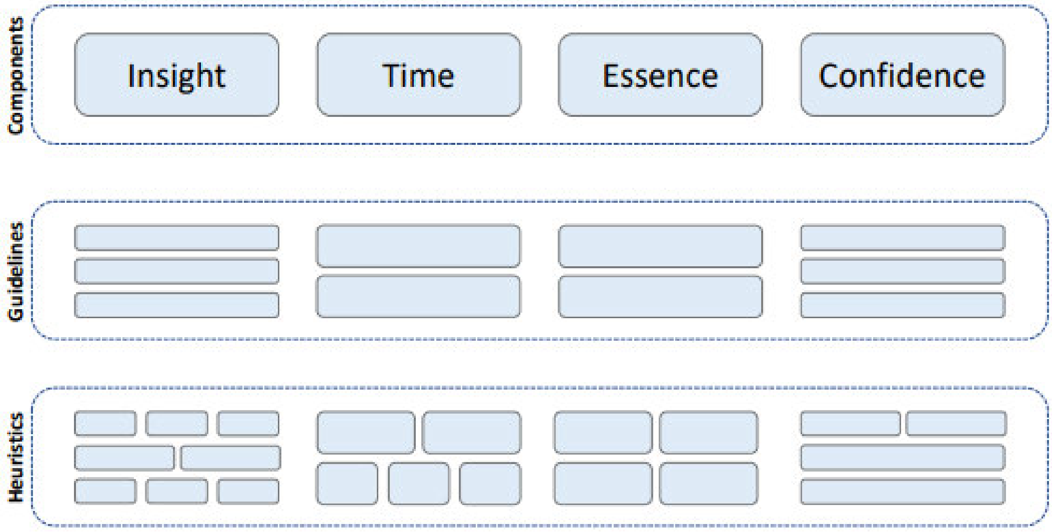


Abbildung 3.29: Hierarchical value framework terminology and structure

tion designers and domain experts, resulting in a more profound comprehension of user requirements and objectives. Furthermore, the authors address the limitations of their approach, including the possibility of subjective interpretation in the heuristic application and the difficulty of balancing multiple criteria driven by values.

Nowell et al. conducted an empirical study to assess the impact of different graphical encodings on user performance in interpreting visualizations. The research conducted a comparative analysis of various graphical encodings, such as position, size, angle, area, color hue, color intensity, and texture. These encodings were assessed considering various criteria, including precision, response time, and user preferences. The study's results indicate that various graphical encodings exhibit differing efficacy levels for conveying information. The encodings that were found to be highly effective were position and size, whereas angle and area were found to be relatively insufficient performers. However, the effectiveness of depicting nominal or quantitative data differs depending on the shape and size of the visual representation. When testing accuracy for nominal data, the shape is more effective than size. However, when it comes to time for task completion, the order reverses, and size becomes more important than shape. Therefore, the shape is placed behind the size. Regarding quantitative data, using shape for encoding is more efficient than using size in terms of task completion time and accuracy. In specific contexts, color hue and color intensity are effective, while texture has generally been found to be less effective. [56] The authors reflected on the study's limitations, which include the requirement for additional research and the controlled setting in which the experiment was conducted.

Although tables, line charts, bar charts, scatterplots, and pie charts are commonly used for data visualization, there is an absence of empirical research on their efficacy in assisting with diverse data analysis tasks. Hence, the study conducted by Saket et al. aimed to assess the efficacy of mentioned fundamental visualizations. The research employed experimental methodology, wherein a study was conducted through crowdsourcing to assess the efficacy of five two-dimensional visualization types with a small-scale set ranging from 5-34 data points for ten distinct visual analysis tasks. Saket et al. claimed that variations in the effectiveness of different visualizations depend on the specific task. They followed the low-level taxonomy by Amar et al. to define the set of low-level analysis tasks, which includes (1) Anomalies, (2) Clusters, (3) Correlation, (4) Compute Derived Value, (5) Characterize Distribution, (6) Extremum, (7) Filter, (8) Order, (9) Determine Range, (10) Retrieve Value.

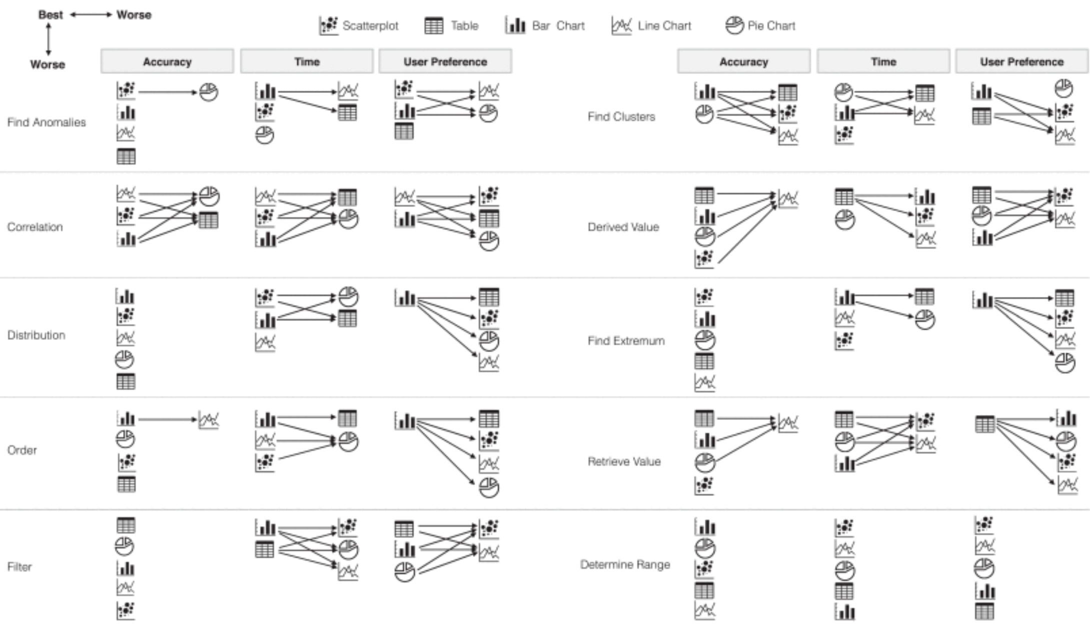


Abbildung 3.30: Relationship between visualizations types in terms of tasks and efficiency measures. The arrows indicate that the source is better than the target.[63]

For example, bar charts were a highly effective tool involving speed and accuracy, whereas scatter plots demonstrated more effective performance in identifying trends and outliers.

On the other hand, line graphs have shown outstanding efficiency in tasks involving temporal data analysis, especially in Correlation and Distribution tasks. The pie chart is a valuable visualization for representing the relationship between different parts of a whole and comparing them, mainly when the presented data is limited. (Figure 3.30)

The results of this study provide valuable insights for enhancing the selection and development of visual representations, enabling users to make informed decisions that align with the demands of their tasks.

3.2.2 Discussion

Evaluating data visualizations is a complex task that involves considering various metrics and methodologies. Each metric has advantages and limitations that affect the evaluation of the effectiveness of visualization. Table 3.10 presents the details of mentioned metric on assessing data visualization.

Tabelle 3.10: Summary of the visualization evaluation metric, including its advantages and limitations.

Category	**Research paper**	**Metrics**	**Methodology**	**Advantages**	**Limitations**
Data-centric	Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness.	Subject matter, source/media location; beliefs and opinions; time; emotions; and confidence and skills	Diary-keeping, focus groups and interviews	Providing a comprehensive assessment of visualization quality	Challenge of generalizing the metric across various visualization types and data domains

Data-centric	A Salience-based Quality Metric for Visualization [40]	Salience-based Quality Metric	Salience overlays, contribution maps and quality charts	- Adaptable to most popular visualization programs -Objectivity and ability to account for human perception factors - Providing a more comprehensive assessment of visualization quality compared to traditional metrics that solely focus on data accuracy or aesthetics	-Requirement for Eye-tracking equipment -Challenge of generalizing the metric across various visualization types and data domains
Data-centric	Analyzing Perceptual Organization in Information Graphics	Gestalt principles of perceptual organization: similarity, continuation, closure, proximity, figure/ground, and symmetry & order	Experimental and observational methods	Provide a quantitative and standardized approach to evaluating information graphics	Choosing appropriate metrics for evaluation requires subjective judgment, and the selection may vary depending on the goals and objectives of the evaluation.
Data-centric	Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design	Amazon Mechanical Turk platform	Experimental and observational methods	Low cost and Scalability	Requiring validation

Task-centric	A Heuristic Approach to Value-Driven Evaluation of Visualizations	Insight, Confidence, Essence,and Time	Literature Review, Brainstorm, Workshop, Affinity Map, Userstudy	-Capturing the multifaceted nature of visualization effectiveness -Flexibility in adapting to diverse domains and tasks	Difficulty of balancing multiple criteria driven by values
Task-centric	How to evaluate data visualizations across different levels of understanding [24]	Bloom's taxonomy: (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5)Synthesis, and (6) Evaluation	Experimental and observational methods	Capturing the nuanced differences in viewer interpretations and gaining insights into the impact of the visualization on different levels of understanding	-Challenge of recruiting and obtaining feedback from expert viewers -Dependence of Knowledge, Application, and the Analysis levels on participant's numeracy skills
Task-centric	Graphical encoding for information visualization: an empirical study	Visual elements such as position, length, angle, area, color hue, color intensity, and texture	Empirical study		Requirement for additional research and the controlled setting in which the experiment was conducted
Task-centric	Task-Based Effectiveness of Basic Visualizations	Categorized the effective of visualization based on tasks: Retrieve, Range, Order, Filter, Extremum, Derived and Cluster tasks	Experimental methodology	Providing a simple guideline for implementation of fundamental visualization	- Challenge of generalization of a specific visualization's performance on a particular task to all tasks -Limitations of conducting the study using static visualizations with a large number of visual marks

In summary, various metrics used to evaluate data visualizations have unique benefits and drawbacks. To ensure a thorough evaluation, utilizing a comprehensive approach that incorporates multiple metrics may be beneficial. Researchers and practitioners should carefully consider each metric's goals, context, and limitations when evaluating data visualizations to ensure a more nuanced understanding of their effectiveness.

4

Conclusions

In conclusion, we have explored the fundamentals of data visualization and its crucial role in effectively communicating information. My thesis attempts to provide a comprehensive understanding of the key concepts and principles of data visualization, highlighting the significance of data visualization in making complex data accessible and comprehensible.

This thesis outlines the data visualization design process, which involves the following steps: (1) Identifying the purpose and key factors, (2) Developing an editorial focus, (3) Conceptualizing design options and encoding the data effectively, and lastly, (4) Evaluating the data visualization. By adhering to these recommended steps, data visualizations can be designed to communicate information and insights to various audiences effectively.

Furthermore, my thesis explores various data visualization methodologies and presents a methodological framework and a guide for information visualization techniques. These resources provide guidance on selecting suitable visualization techniques based on the data and research objectives, which helps researchers to explore and present data related to the COVID-19 pandemic effectively.

With the purpose of demonstrating the implementation and evaluation of data visualizations, my thesis presents a case study on COVID-19 analysis employing an exploratory data analysis strategy. This case study demonstrates the application of data visualization techniques in analyzing the COVID-19 pandemic, enabling a more profound comprehension of the pandemic and its impact.

The assessment of the visualizations and tools used in the case study has revealed their effectiveness in communicating information and aiding in data exploration. Researchers can identify the strengths and weaknesses of visualizations, make necessary improvements, and ensure that they effectively serve their intended purpose through evaluation.

Overall, my thesis has contributed to the comprehension of data visualization as a potent instrument for investigating and conveying the effects of the COVID-19 pandemic. Researchers can effectively communicate complex information and insights by using data visualization techniques and following the design process and methodologies discussed. This can facilitate better understanding and decision-making during global challenges.

Literaturverzeichnis

[1] analyticjournalism.com» blog archive » some nifty unemployment charts from jorge camoes. URL http://www.analyticjournalism.com/2009/06/19/some-nifty-unemployment-charts-from-jorge-camoes/.

[2] China reports first death from new virus - the new york times, . URL https://www.nytimes.com/2020/01/10/world/asia/china-virus-wuhan-death.html.

[3] Coronavirus: Bayerische behörden bestätigen ersten fall in deutschland. . ISSN 2195-1349. URL https://www.spiegel.de/wissenschaft/medizin/corona-virus-erster-fall-in-deutschland-bestaetigt-a-19843b8d-8694-451f-baf7-0189d3356f99.

[4] Startup Bobbie Proves There's Profit In Baby Formula Despite The Dominance Of Industry Giants. URL https://www.forbes.com/sites/maggiemcgrath/2023/05/19/startup-bobbie-proves-theres-profit-in-baby-formula-despite-the-dominance-of-industry-giants/?

[5] Guidelines for good visual information representations | interaction design foundation (IxDF). URL https://www.interaction-design.org/literature/article/guidelines-for-good-visual-information-representations.

[6] How to Fix and Avoid TLS/SSL Errors. URL https://venafi.com/blog/how-fix-and-avoid-tlsssl-errors/.

[7] COVID-19 tracking germany | kaggle, . URL https://www.kaggle.com/datasets/headsortails/covid19-tracking-germany.

[8] COVID-19 datenhub, . URL https://npgeo-corona-npgeo-de.hub.arcgis.com/.

[9] BKG - homepage, . URL https://www.bkg.bund.de/EN/Home/home.html.

[10] Data visualization and infographics, . URL https://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/. Section: General.

[11] ECharts official tutorial (eight) [visual mapping of data] - programmer sought, . URL https://www.programmersought.com/article/11856463460/.

[12] Health inequalities: Age and gender and COVID-19 | Local Government Association, . URL https://www.local.gov.uk/age-and-gender-impact-covid-19.

[13] Homepage - german federal statistical office, . URL https://www.destatis.de/EN/Home/_node.html.

[14] Pneumonia of unknown cause – china, . URL https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229.

[15] WHO coronavirus (COVID-19) dashboard, . URL https://covid19.who.int.

[16] Carnival-going german couple contract coronavirus, authorities fear it has spread | reuters. URL https://www.reuters.com/article/us-china-health-germany-couple-idUSKCN20K2E3.

[17] Nikkei drops as investors anxious about pandemic's impact on earnings. *Reuters*, Apr. 2020. URL https://www.reuters.com/article/japan-stocks-idUSL3N2C10NN.

[18] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005.

[19] P. Asare and R. Barfi. The impact of covid-19 pandemic on the global economy: emphasis on poverty alleviation and economic growth. *Economics*, 8(1):32–43, 2021.

[20] A. S. Ayobami and Z. Jamaludin. Information visualization techniques usage model. In *Proceedings of the 4th International Conference on Computing and Informatics, ICOCI*, pages 28–30, 2013.

[21] B. Bagchi, S. Chatterjee, R. Ghosh, D. Dandapat, B. Bagchi, S. Chatterjee, R. Ghosh, and D. Dandapat. Impact of covid-19 on global economy. *Coronavirus Outbreak and the Great Lockdown: Impact on Oil Prices and Major Stock Markets Across the Globe*, pages 15–26, 2020.

[22] J. Bertin and M. Barbut. *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Gauthier Villars, 1973. URL https://books.google.fr/books?id=F4weAAAAMAAJ.

[23] B. S. Bloom and D. R. Krathwohl. *Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain.* longman, 2020.

[24] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pages 19–28. IEEE, 2020.

[25] S. Carpendale. Evaluating information visualizations. *Information visualization: Human-centered issues and perspectives*, pages 19–45, 2008.

[26] S. M. Casner. Task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics (ToG)*, 10(2):111–151, 1991.

[27] M. Causse and C. Hurter. The physiological user's response as a clue to assess visual variables effectiveness. volume 5619, pages 167–176, 07 2009. doi: 10.1007/978-3-642-02806-9_20.

[28] R. Chetty, J. N. Friedman, N. Hendren, M. Stepner, and T. O. I. Team. *How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data*, volume 91. National Bureau of Economic Research Cambridge, MA, 2020.

[29] R. Cox. Representation construction, externalised cognition and individual differences. *Learning and instruction*, 9(4):343–363, 1999.

[30] M. Dastani. The role of visual perception in data visualization. *Journal of Visual Languages & Computing*, 13(6):601–622, 2002.

[31] A. Dunn, K. Hood, and A. Driessen. *Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data.* US Department of Commerce, Bureau of Economic Analysis, 2020.

[32] M. Esteller. Epigenetic gene silencing in cancer: the dna hypermethylome. *Human molecular genetics*, 16(R1):R50–R59, 2007.

[33] M. Fana, S. Torrejón Pérez, and E. Fernández-Macías. Employment impact of covid-19 crisis: from short term effects to long terms prospects. *Journal of Industrial and Business Economics*, 47:391–410, 2020.

[34] N. Fernandes. Economic effects of coronavirus outbreak (covid-19) on the world economy. 2020.

[35] H. P. Grice. Logic and conversation. In *Speech acts*, pages 41–58. Brill, 1975.

[36] G. Healthcare. Financial markets look to be affected by Covid-19 as FTSE-100 index falls., Mar. 2020. URL https://www.pharmaceutical-technology.com/comment/covid-19-financial-markets/.

[37] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 203–212, 2010.

[38] N. Iliinsky and J. Steele. *Designing data visualizations: Representing informational Relationships*. Ö'Reilly Media, Inc.", 2011.

[39] J. K. Jackson. Global economic effects of covid-19. Technical report, Congressional Research Service, 2021.

[40] H. Jänicke and M. Chen. A salience-based quality metric for visualization. In *Computer graphics forum*, volume 29, pages 1183–1192. Wiley Online Library, 2010.

[41] D. A. Keim. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[42] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. *Lecture notes in computer science*, 4404:76–90, 2008.

[43] H. Kennedy, R. L. Hill, W. Allen, and A. Kirk. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, 21(11), 2016.

[44] A. Kirk. *Data Visualization: a successful design process*. Packt publishing LTD, 2012.

[45] R. Koch-Institut. SARS-CoV-2 infektionen in deutschland. URL https://zenodo.org/record/6800491. Version Number: 2022-07-06 Type: dataset.

[46] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler. Getting there first: Real-time detection of real-world incidents on twitter. 2012.

[47] R. Li, Z. Ma, and T. Sui. The impact of covid-19 on the economy in different countries. In *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, pages 1257–1264. Atlantis Press, 2022.

[48] G. L. Lohse. The role of working memory on graphical information processing. *Behaviour & Information Technology*, 16(6):297–308, 1997.

[49] A. M. MacEachren et al. Visualization in modern cartography: setting the agenda. *Visualization in modern cartography*, 28(1):1–12, 1994.

[50] S. Maital and E. Barzani. The global economic impact of covid-19: A summary of research. *Samuel Neaman Institute for National Policy Research*, 2020:1–12, 2020.

[51] M. Maliszewska, A. Mattoo, and D. Van Der Mensbrugghe. The potential impact of covid-19 on gdp and trade: A preliminary assessment. world bank policy research working paper,(no. 9211). *SSRN Electronic Journal. https://ssrn. com/abstract*, 3573211, 2020.

[52] F. Mansmann, F. Fischer, D. A. Keim, and S. C. North. Visual support for analyzing network traffic and intrusion detection events using treemap and graph representations. In *Proceedings of the Symposium on Computer Human Interaction for the Management of Information Technology*, pages 19–28, 2009.

[53] A. Maxmen et al. Wuhan market was epicentre of pandemic's start, studies suggest. *Nature*, 603(7899):15–16, 2022.

[54] B. H. Meyer, B. Prescott, and X. S. Sheng. The impact of the covid-19 pandemic on business expectations. *International Journal of Forecasting*, 38(2):529–544, 2022.

[55] M. Meyer. Designing visualizations for biological data. *Leonardo*, 46(3):270–271, 2013.

[56] L. Nowell, R. Schulman, and D. Hix. Graphical encoding for information visualization: an empirical study. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, pages 43–50. IEEE, 2002.

[57] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.-E. Haug, and H. Janetzko. Visual opinion analysis of customer feedback data. In *2009 IEEE symposium on visual analytics science and technology*, pages 187–194. IEEE, 2009.

[58] R. Padhan and K. Prabheesh. The economics of covid-19 pandemic: A survey. *Economic analysis and policy*, 70:220–237, 2021.

[59] M. Pinto, R. Raposo, and F. Ramos. Comparison of emerging information visualization tools for higher education. In *2012 16th international conference on information visualisation*, pages 100–105. IEEE, 2012.

[60] K. Püschel and J. P. Sperhake. Corona deaths in hamburg, germany. 134(4):1267–1269. ISSN 1437-1596. doi: 10.1007/s00414-020-02332-x. URL https://doi.org/10.1007/s00414-020-02332-x.

[61] A. C. Robinson, J. Chen, E. J. Lengerich, H. G. Meyer, and A. M. MacEachren. Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science*, 32(4):243–255, 2005.

[62] R. E. Roth, K. S. Ross, B. G. Finch, W. Luo, and A. M. MacEachren. A user-centered approach for designing and developing spatiotemporal crime analysis tools. In *Proceedings of GIScience*, volume 15. sn, 2010.

[63] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512, 2018.

[64] M. Scaife and Y. Rogers. External cognition: how do graphical representations work? *International journal of human-computer studies*, 45(2):185–213, 1996.

[65] D. Services. Guides: Data visualization: Introduction to data visualization. URL https://guides.library.jhu.edu/datavisualization/home.

[66] S. Simon, D. Oelke, R. Landstorfer, K. Neuhaus, and D. A. Keim. Visual analysis of next-generation sequencing data to detect overlapping genes in bacterial genomes. In *2011 IEEE Symposium on Biological Data Visualization (BioVis).*, pages 47–54. IEEE, 2011.

[67] J. W. Slocum and D. Hellriegel. *Principles of organizational behavior*. South-Western Cengage Learning Mason, OH, 2009.

[68] A. Stoffel, D. Spretke, H. Kinnemann, and D. A. Keim. Enhancing document structure analysis using visual analytics. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 8–12, 2010.

[69] D. Todorovic. Gestalt principles. *Scholarpedia*, 3(12):5345, 2008.

[70] E. R. Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985.

[71] E. R. Tufte. *Beautiful evidence*. Graphis Pr, 2006.

[72] E. R. Tufte and D. Robins. *Visual explanations*. Graphics Cheshire, CT, 1997.

[73] E. R. Tufte, N. H. Goeler, and R. Benson. *Envisioning information*, volume 126. Graphics press Cheshire, CT, 1990.

[74] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):491–500, 2018.

[75] M. Wattenberg and D. Fisher. Analyzing perceptual organization in information graphics. *Information Visualization*, 3(2):123–133, 2004.

[76] J.-H. Yoo. The fight against the 2019-ncov outbreak: an arduous march has just begun. *Journal of Korean medical science*, 35(4), 2020.

[77] Y. Zhu. Measuring effective data visualization. In *Advances in Visual Computing: Third International Symposium, ISVC 2007, Lake Tahoe, NV, USA, November 26-28, 2007, Proceedings, Part II 3*, pages 652–661. Springer, 2007.

A Anhang

The complete source code of the work can be found on the enclosed CD. In the appendix now follow some code examples

```
base_stats_fig = go.Figure()

for column in base_stats.columns.to_list()[2:6]:
    color_dict = {
      "cases": "#118ab2",
      "active": "#ef476f",
      "recovered": "#06d6a0",
      "deaths": "#073b4c"
      }
    base_stats_fig.add_trace(
        go.Scatter(
            x = base_stats['date'],
            y = base_stats[column],
            name = column,
            line = dict(color=color_dict[column]),
            hovertemplate ='<br><b>Date</b>: %{x}'+'<br><i>Count</i>:'+'%{y}',
        )
    )

for column in base_stats.columns.to_list()[2:6]:
    color_dict = {
      "cases": "#149ECC",
      "active": "#F47C98",
      "recovered": "#24F9C1",
      "deaths": "#0C6583"
      }
    base_stats_fig.add_trace(
        go.Scatter(
            x = base_stats['date'],
            y = base_stats['index'].apply(lambda x: (base_stats[column][x-7:x].sum())/7 if x>7 else (base_stats[column][0:x].sum())/7),
            name = column+" 7-day Moving Avg.",
            line = dict(dash="dash", color=color_dict[column]), showlegend=False,
            hovertemplate = '<br><b>Date</b>: %{x}'+'<br><i>7-day moving avg.</i>: %{y}'
        )
    )
```

Abbildung A.1: Code for Interactive COVID-19 daily new confirmed, deaths, recovered, active cases in Germany

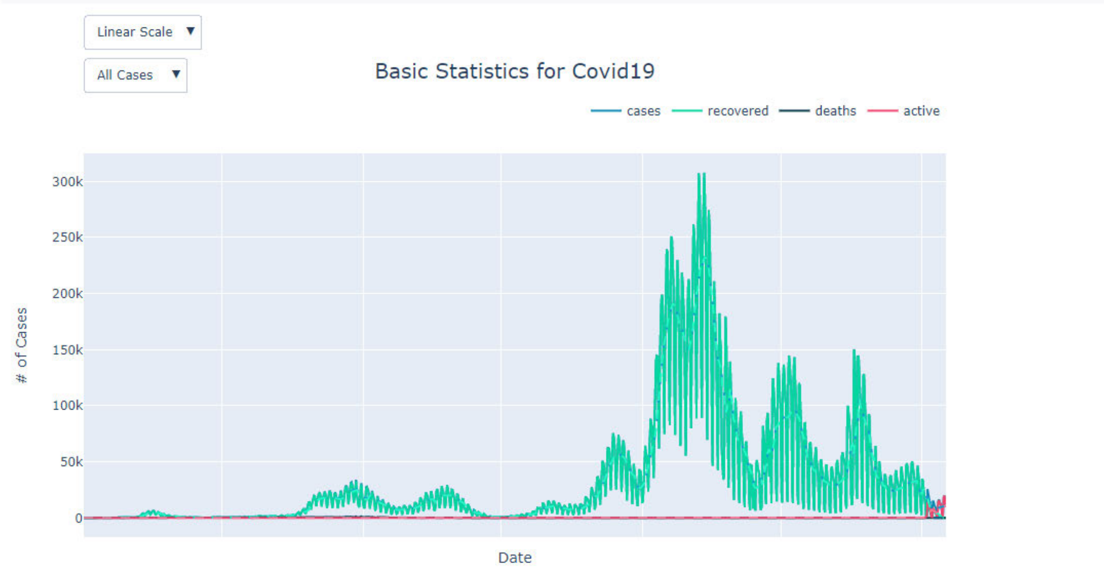


Abbildung A.2: Interactive COVID-19 daily new confirmed, deaths, recovered, active cases in Germany

```
# Calculating "active"
covid_data['active'] = covid_data['cases'] - covid_data['recovered'] - covid_data['deaths']
# Display
display(covid_data.sample(5))
#
covid_cases_sum = covid_data.cases.sum()
covid_deaths_sum = covid_data.deaths.sum()
covid_recovered_sum = covid_data.recovered.sum()
covid_active_sum = covid_data.active.sum()

# visualization with graph
latest_stats_fig = go.Figure()
latest_stats_fig.add_trace(go.Treemap(labels = ['Confirmed','Active','Recovered','Deaths'],
                                     parents = ['','Confirmed','Confirmed','Confirmed'],
                                     values = [covid_cases_sum, covid_active_sum, covid_recovered_sum, covid_deaths_sum],
                                      branchvalues="total", marker_colors = ['#29c8d6','#ef476f','#21cc54','#0b0d4a'],
                                      textinfo = "label+text+value",
                                      outsidetextfont = {"size": 30, "color": "darkblue"},
                                      marker = {"line": {"width": 2}},
                                        pathbar = {"visible": False}
                                     ))
latest_stats_fig.update_layout(width=1000,
                               height=300)
latest_stats_fig.show()
```

Abbildung A.3: Code for creating Tree map to illustrating the total confirmed, deaths, recovered, active cases in Germany

```
# Grouping the data by date
cases['cumulative_cases'] = cases.cases.cumsum()
display(cases)
# Plot bar chart
fig = px.line(cases.reset_index(), x = 'date', y = 'cumulative_cases',height = 400,
              title = 'Cumulative confirmed COVID-19 cases in Germany',
             labels={'cases': 'Daily new cases'})
fig.show()
```
Executed in 755ms, 22 May at 10:51:45

11 rows 1094 rows × 3 columns pd.Dataframe

	date	cases	cumulative_cases
0	2020-01-02	1	1
1	2020-01-23	1	2
2	2020-01-28	2	4
3	2020-01-29	2	6
4	2020-01-31	4	10
...	...	...	...
1089	2023-01-29	1566	37747019
1090	2023-01-30	18028	37765047
1091	2023-01-31	19968	37785015
1092	2023-02-01	15201	37800216
1093	2023-02-02	9408	37809624

This output uses HTML that may be stripped because the notebook is not trusted

Abbildung A.4: Code for creating Line chart to illustrating the Daily cumulative case number of patients with COVID-19 in Germany

```

# Create subplots, using 'domain' type for pie charts
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
labels = covid_by_age['age_group']

pie_fig = make_subplots(rows=2, cols=2,
                        specs=specs,
                        subplot_titles=['cases', 'deaths','active', 'recovered']
                        )
pie_fig.add_trace(go.Pie(labels=labels, values=covid_by_age.cases, scalegroup='one'),1,1)
pie_fig.add_trace(go.Pie(labels=labels, values=covid_by_age.deaths, scalegroup='one'),1,2)
pie_fig.add_trace(go.Pie(labels=labels, values=covid_by_age.active, scalegroup='one'),2,1)
pie_fig.add_trace(go.Pie(labels=labels, values=covid_by_age.recovered, scalegroup='one'),2,2)

pie_fig.update_layout(title_text='Ratio by Age Group',
                      height=800,  # adjust height of entire subplot
                        width=1000,   # adjust width of entire subplot
                        margin=dict(l=20, r=20, t=20, b=20),  # set margins
                      )
pie_fig.show()
Executed in 482ms, 22 May at 10:51:47
```

Abbildung A.5: Code for creating Pie chart to illustrating the Distribution of confirmed, active, mortality and recovered cases by age groups in Germany

Add Code Cell | Add Markdown Cell

```
# Grouping the data by date
cases = covid_data.groupby('date', as_index = False).agg({'cases' : 'sum'})
# Plot bar chart
fig = px.bar(cases, x = 'date', y = 'cases',height = 400, width=800,
             color = 'cases',
             color_continuous_scale = px.colors.diverging.RdBu,
             title = 'Daily Cases of COVID-19 in Germany',
             labels={'cases': 'Daily new cases'})
fig.show()
```

Executed in 478ms, 22 May at 10:51:44

This output uses HTML that may be stripped because the notebook is not trusted

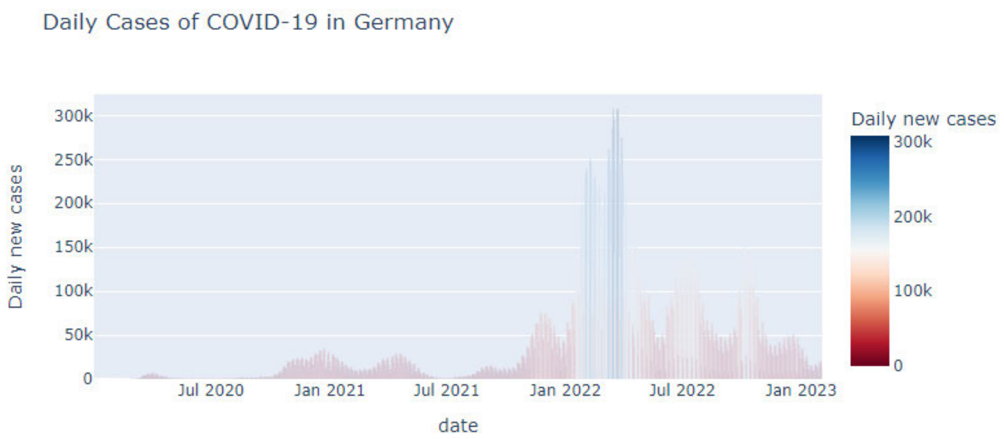


Abbildung A.6: Code for creating Bar chart to illustrating the Distribution of confirmed, active, mortality and recovered cases by age groups in Germany

Add Code Cell | Add Markdown Cell

```
# choropleth mapbox with button

mycustiondata2 = np.stack((merge_df['state'], merge_df['cases_per_100k'], merge_df['deaths_per_100k'], merge_df['recovered_per_100k'], merge_df['active_per_100k']), axis=-1)

title = 'Average Covid-19 Reported Cases, Deaths, Recovered and Active per 100.000 People per States in Germany '

situation_heatmap_fig = go.Figure(go.Choroplethmapbox(
                                  geojson=german_geo_data,
                                  locations=merge_df['state'],
                                  z=merge_df['cases_per_100k'],
                                  featureidkey='properties.name',
                                  coloraxis="coloraxis",
                                  customdata=mycustiondata2,
                                  marker_opacity=0.75, marker_line_width=0.5))
situation_heatmap_fig.update_layout(coloraxis_colorscale='blues',
                  mapbox=dict(style='carto-positron',
                              zoom=4.5,
                              center = {"lat": 51.095123 , "lon":10.271483 },
                              ))
situation_heatmap_fig.update_layout(title_text=title,
                  title_x=0.5,
                  margin={"r":10,"t":60,"l":0,"b":0});
```

Executed in 592ms, 22 May at 10:51:54

Abbildung A.7: Code for creating Geo Heat Map to illustrating the Reported Cases per Population and per Population Density in Germany (1)

```
# Define the buttons for an updatemenu

button1 = dict(method= 'update',
               label='confirmed cases',
               args=[
                    {"z": [merge_df['cases_per_100k']]},
                    {"coloraxis.colorscale":"Blues" }  #dict for  layout attribute update
                 ])

button2 = dict(method= 'update',
               label='deaths',
               args=[
                   {"z": [merge_df['deaths_per_100k']]},
                   {"coloraxis.colorscale": "Greys"}
           ])

button3 = dict(method= 'update',
               label='recovered',
               args=[
                   {"z": [merge_df['recovered_per_100k']]},
                   {"coloraxis.colorscale":"Greens" }
           ])

button4 = dict(method= 'update',
               label='active',
               args=[
                   {"z": [merge_df['active_per_100k']],
                   },
                   {"coloraxis.colorscale": "Reds"} #update layout attribute
           ])
Executed in 591ms, 22 May at 10:51:54
```

```
situation_heatmap_fig.update_layout(updatemenus=[dict(active=0,
                                    buttons= [button1, button2, button3, button4])]
                                     );
situation_heatmap_fig.show()
Executed in 599ms, 22 May at 10:51:54
```

Abbildung A.8: Code for creating Geo Heat Map to illustrating the Reported Cases per Population and per Population Density in Germany (2)

```
fig = px.scatter_geo(
         merge_df,
        geojson=german_geo_data,
        featureidkey= 'properties.name',
        locations='state',
        color='state',
        locationmode='geojson-id', # We pass the parameter of determining the country on the map (by name)
        hover_name='state',  # Passing values for the signature on hover
        size='cases', # Passing a column with values
        scope='europe',
        center = {"lat": 51.095123 , "lon":10.271483 }
)

fig.update_layout(
    # Set the name of the map
    title_text='Confirmed cases by country <br>',
    legend_orientation='h', # Place the legend caption under the chart
    legend_title_text='', # Remove the name of the legend group
    # Determine the map display settings (remove the frame, etc.)
    geo=dict(
        showframe=False,
        showcoastlines=False,
        projection_type='equirectangular'
    ),
    # Setting parameters for the text
    font=dict(
        family='TimesNewRoman',
        size=18,
        color='black'
    )
)
fig.update_geos(fitbounds="locations", visible=True)
fig.show()
Executed in 1s, 22 May at 10:51:55
```

Abbildung A.9: Code for creating Geo Scatter plot to illustrating the Reported Cases in Germany

Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

[redacted]

Ort	Datum	Unterschrift im Original