Principles of Data Visualization in the Social Sciences
Literature Review

Allison Towey, The University of Chicago

Abstract

Data visualization refers to the use of graphical and visual representations to communicate complex data effectively. The aim is to present data in a way that makes it easy to comprehend, analyze and draw insights. Data visualization has become increasingly important in recent years due to the growth in data volumes, and the need to extract valuable insights from large datasets. Drawing on research from scholars such as Edward Tufte, Alberto Cairo, and Stephen Few, this literature review explores important principles of visualization including simplicity, clarity, and accuracy.

I. Introduction

Data visualization has become increasingly important across a wide range of fields. Practitioners in academia, journalism, business, healthcare, entertainment, and many more are turning to data visualization as a powerful tool for communicating complex data sets in a clear, concise, and visually appealing manner. Visual techniques such as graphs, charts, and maps can help to facilitate communication of complex data to a variety of audiences. (Chen 2016)

By presenting information in a visually appealing, accessible, and understandable format, data visualization helps to transform large amounts of raw data into clear, interpretable, and meaningful insights. Data visualization can help elucidate patterns and relationships in the data that might be hidden in raw data, making it easier to identify key trends and gain insights. (Ware 2013, Few 2012) As such, data visualization can help researchers and decision-makers alike gain a better understanding of the data, leading to more informed and effective decisions. (Kirk 2016)

Data visualization has become increasingly important in the social sciences particularly as researchers seek new ways to 1) find patterns or inferences in large datasets and 2) communicate complex findings to a wide range of audiences, from published articles in academic journals to general-public-facing materials. Data visualization can be utilized throughout the research process, from finding trends in the data during the initial stages of research to presenting findings to the intended audience at the culmination of the project.

With the rise of the availability of large datasets, social scientists can leverage data visualization to identify key trends that might not be immediately apparent from raw data. By leveraging the various visualization techniques available, social scientists can gain insights from their data that may not be immediately apparent through traditional statistical analysis. For instance, line charts and scatter plots are useful for identifying trends over time or across different variables. Visualization can also help social scientists compare groups across different variables using charts such as bar charts, pie charts, and box plots to compare group sizes and proportions. Social scientists can use data visualization to examine the distribution of their data; histograms and density plots can help identify the shape of a distribution, whether it is skewed, bimodal, or normal. Additionally, it can assist in identifying outliers, or data points that are significantly different from the rest of the data using methods such as box plots and scatter plots and noting data points out of the norm. Visualization can also help social scientists explore relationships between different variables in their data, plotting scatter plots and heat maps to show how two or more variables are related to each other. Using these methods, social scientists can seek to identify patterns and explore relationships in their datasets.

Beyond the identification and inference stages of research, social scientists can also use visualization to communicate findings both to the academic community and the general public writ-large depending on their intended audience. For academic audiences, social scientists may utilize complex visualizations and statistical graphics, such as regression tables and heat maps, to present their findings in a more sophisticated way. These types of visualizations can help communicate complex statistical relationships and patterns in data to other experts in the field. Alternatively, when communicating with a broader, perhaps less technical, audience, social scientists can use simpler, more intuitive visualizations such as infographics, charts, and graphs, to make their findings more accessible to the general public. Infographics, in particular, can be highly effective at presenting data in an engaging and visually appealing way, while still conveying important information to non-expert audiences.

As social science is a broad field of disciplines, a range of visualization techniques are used to specifically address data and audience-specific needs. For example, in sociology, network graphs are being used to illustrate the structure and dynamics of social relationships and their impact on individual behavior (Knoke & Yang, 2008). In political science, choropleth maps can be used to show variations in voting patterns across different regions (Carty & Eagles, 2003), while line graphs and bar charts can be used to track changes in public opinion over time (Holbrook & Hill, 2005). In psychology, scatter plots and box plots can be used to identify patterns and relationships in experimental and survey data (Healy, 2018), and in economics, line graphs and heat maps can be used to track changes in economic indicators like GDP, inflation, and unemployment (Mankiw et al., 2019). By using techniques tailored to the specific data and audience, social scientists can effectively communicate their research findings and enhance the impact of their work.

Overall, data visualization is a powerful tool for social scientists to not only analyze and interpret data but also to communicate their findings to different audiences in an effective and impactful way. As technology continues to advance, data visualization is likely to become even more important in the social sciences and beyond. Data visualization aims to leverage our visual perception and cognitive abilities to enhance the understanding of complex data. In practice, there are various trade-offs that practitioners must consider when deciding the form, function, and design of the visualization. Effective data visualization requires a thoughtful and intentional approach that takes into account both the data and the human factors involved in its interpretation. Simplicity, clarity, and accuracy are fundamental to the design and effectiveness of visualizations.

II. Simplicity

The principle of simplicity suggests that simpler explanations or interpretations of observations are generally better than complex ones. In the field of data visualization, simplifying a visualization includes removing extraneous elements or unnecessary detail, organizing data logically, and presenting data that requires minimal cognitive load. (Kosslyn 2006) When done correctly, simplification can increase comprehension, and ultimately lead to better storytelling by the practitioner and decision-making by the audience.

In the field of cognitive science, simplicity has been shown to greatly enhance the viewer's understanding of complex information. The prevailing idea is that the human mind naturally gravitates towards the most straightforward explanation of observations. (Feldman 2016) This insight is bolstered by psychological and neuroscientific studies: in Stephen Few's "Show Me the Numbers: Designing Tables and Graphs to Enlighten" (2012), he highlights research that found that when presented with complex data, the brain's response was to tune out, leading to a failure to understand and engage with the information.

The concept of simplicity in data visualizations has been often discussed in academic literature. Edward Tufte is an American statistician and professor emeritus of political science, statistics, and computer science at Yale University. He is best known for his work in the field of data visualization and information design, and is considered a leading expert on the visual display of quantitative information. Tufte is known for his emphasis on the importance of clarity and simplicity in visualization, and for his principle of "data-ink ratio", or the proportion of ink on a page or screen that is used to display actual data, as opposed to ink used for non-data elements such as labels, gridlines, and other decorations. (Tufte 1997)

The goal of maximizing the data-ink ratio is to increase the clarity and effectiveness of the visualization by reducing visual clutter and directing the viewer's attention to the data itself. In turn, as the data-ink ratio in a visualization is increased, the visualization becomes more minimalist in style. In practice, maximizing the data-ink ratio involves eliminating or minimizing any elements that do not directly contribute to the representation of the relevant data. This can include removing gridlines, reducing the size and complexity of labels, using minimalist design principles, and avoiding the use of unnecessary decoration such as background images or patterns. By doing so, Tufte claims, the resulting visualization is less distracting and more effective at communicating the intended message.

One way to ensure a high data-ink ratio is to minimize what Tufte calls "chartjunk", the unnecessary elements that do not add value to the data being presented, but rather serve only to distract or confuse the viewer. (Tufte 2001) Examples of chart junk include heavy, patterned, or dark grid lines, unnecessary or lengthy text, overly complex or illegible fonts, decorative chart axes, frames, pictures, backgrounds, or icons within data graphs, ornamental shading, and unnecessary dimensions. Tufte argued that chartjunk can lead to confusion, misinterpretation, and misrepresentation of data. He advocated for a minimalist approach to chart design, where the focus is on presenting the data clearly and effectively, without unnecessary distractions, regardless of their potential artistic or visually appealing merit. (Tufte 2001)

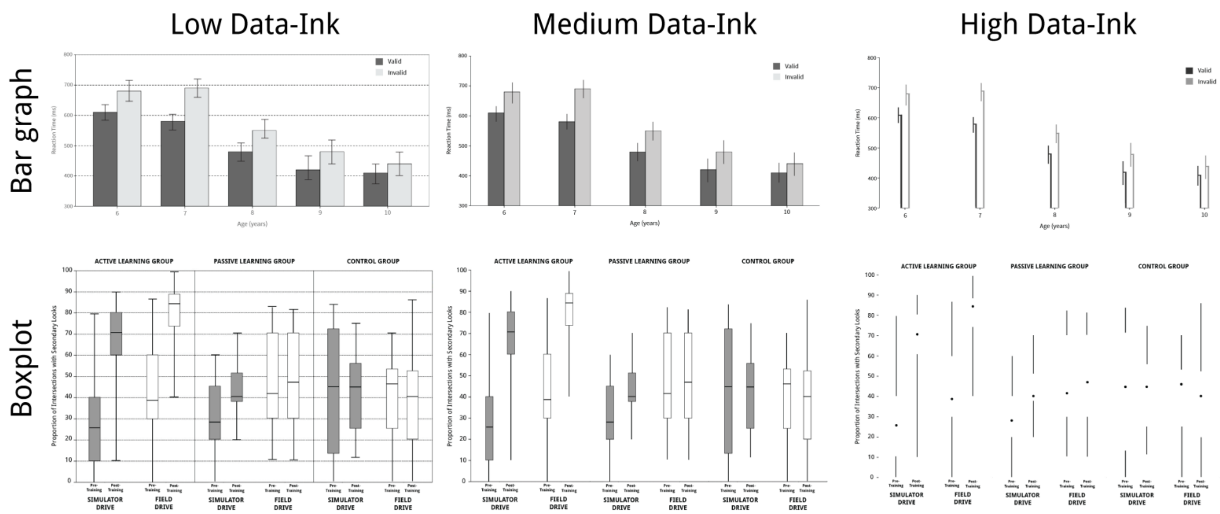

Figure 1: Examples of low, medium, and high data-ink bar graph and boxplots developed by Kevin McGurgan in "Data-ink Ratio and Task Complexity in Graph Comprehension" (2015)

Empirical studies have supported Tufte's idea that simplicity is fundamental to the effectiveness of visualizations. In Inbar et al. (2007), researchers examined Tufte's philosophy of minimalism via the data-ink ratio by crafting a survey presenting four bar charts with data-ink ratios ranging from low to high according to the researchers' subjective rankings. The researchers then asked participants to evaluate each chart on the basis of the dimensions of beauty, clarity, and simplicity and found that respondents preferred charts with lower data-ink ratios. Likewise, Blasio and Bisantz (2002) proposed that a higher data-ink ratio can effectively decrease the time required to identify an outlier event in a visualization and substantiated by empirical evidence from an experiment. Sorensen (1993) discovered that including a background image, which lowers the data-ink ratio, adversely affects the perceived quality of a chart. Furthermore, Hullman, Adar, and Shah (2011) suggest that it is preferable to use minimalist design principles when creating graphics that may be viewed by individuals with accessibility or disability concerns.

While minimizing chartjunk so as to maximize the data-ink ratio can be a useful principle in data visualization, it is not without its limitations. In some cases, non-data elements such as labels or annotations can be necessary to provide context or aid in interpretation. Additionally, some visualizations may benefit from the use of color or other decoration to aid in differentiation or highlighting of certain data points or trends. Essentially, it is important to balance the data-ink ratio with other considerations such as context, interpretation, and the specific goals of the visualization. (Wilkinson 2005)

Other empirical studies have shown the limits and potential challenges of minimalism at the expense of artistic merit and explanatory power. Despite previous research suggesting that a high data-ink ratio leads to better recall of data, Bateman et al. (2013) found that the use of "chartjunk" (lower data-ink ratio) in the form of embellishment actually resulted in better recall without adversely affecting interpretation of the data, according to subjective surveys with open-ended responses. (*See Figure 2*) Similar results were also found by Li and Moacdieh (2014) and McGurgan (2015). Borkin et al. (2013) further supported this by finding that embellished graphs improved memorability compared to plain or standard graphs, particularly with graphs targeted at the general public as opposed to practiced professionals.

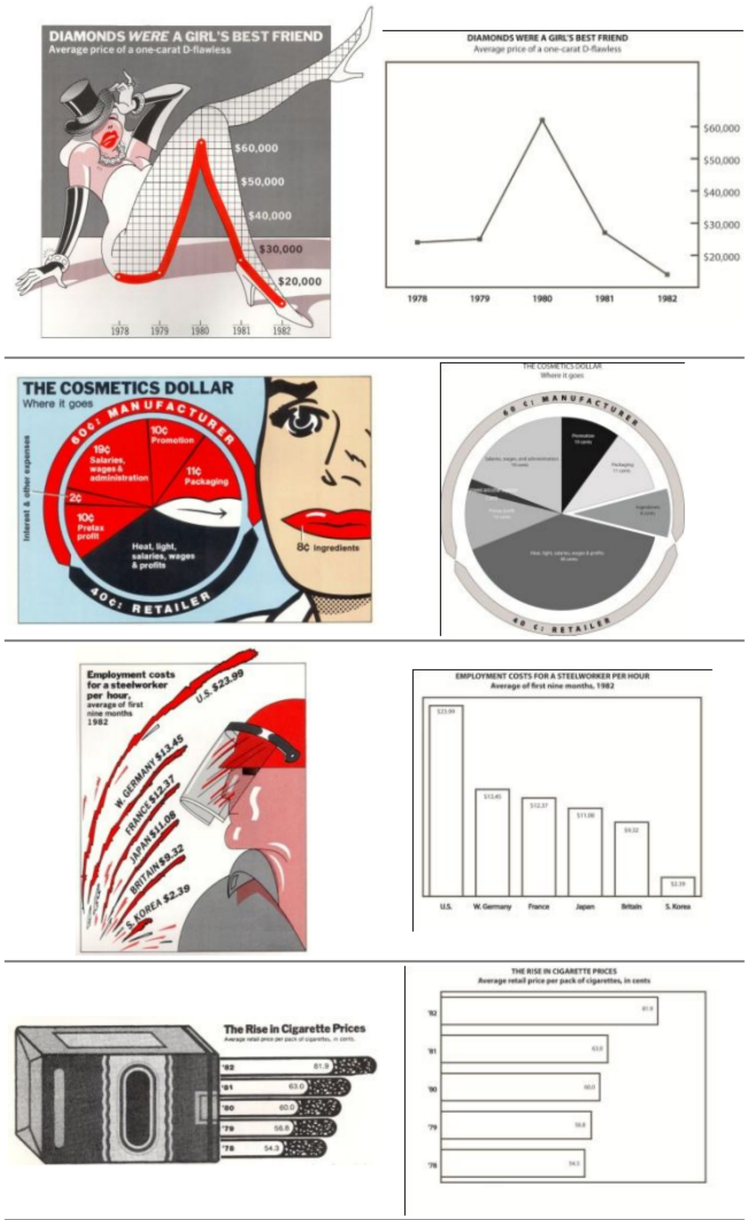

Figure 2: Examples of "chart junk" or embellished graphs compared to more simplistic depictions from Bateman et al 2013's "Useful junk?: the effects of visual embellishment on comprehension and memorability of charts"

Such vast research in the field of data visualization suggests that while simplicity and minimalism have their benefits, there is a delicate balance to achieve in order to optimize the effectiveness of a visualization. It is important to note that simplicity should not come at the expense of accuracy or completeness. A visualization that is too simplistic may oversimplify the data, leading to misinterpretation or misunderstanding. Thus, simplicity should be balanced with other considerations such as context, interpretation, and the specific goals of the visualization. Overall, simplicity is an important principle in data visualization as it can help to improve the effectiveness of the visualization and make it more accessible to a wider audience.

III. Clarity

A related but distinct principle of data visualization is clarity, which refers to the ability to effectively communicate the intended message to the audience. A clear visualization should convey the main insights and patterns in the data in a way that is easy to understand, without causing confusion or ambiguity. (Wilkinson 2005) There are a variety of methods to ensure visualizations are clear and concise, such as simplification, appropriate labeling and annotations, consistency in design and representation, the use of contrasting design elements and emphasis, and the inclusion of context.

Simplicity

Simplicity, as previously discussed, can allow the audience to understand the main messages of the visualization without unnecessary confusion. Minimizing clutter and presenting data in a clear and straightforward way can reduce cognitive load and improve comprehension. When a visualization is overloaded with too much information or cluttered with unnecessary elements, it can be difficult for the viewer to differentiate between important and unimportant information. (Kosslyn 2006) A simplified visualization can enable the viewer to focus on the key data points, trends, and patterns that the visualization designer intends to highlight, thus improving clarity.

Labels and Annotations

Labels and annotations too are important for clarity in data visualization, as they provide context and help the viewer to understand the data being presented. Labels are used to identify what is being represented, whether it is a particular data point or an axis on a graph. Annotations provide additional information, such as explanations of unusual data points or trends, and help the viewer to interpret the data correctly.

Without clear and concise labels, it can be difficult for the viewer to understand what the visualization is trying to convey, which can lead to confusion and misunderstandings, and ultimately a lack of clarity. Labels should be placed close to the relevant data points, with appropriate formatting and font size, to make them easy to read and distinguish from one another. Annotations can be used to provide more detailed information about the data being presented. For example, if there is a sudden spike in the data that is not immediately apparent, an annotation can be added to explain what caused the spike. This helps the viewer to understand the context and the underlying factors behind the data. Annotations can also be used to draw attention to important trends or patterns in the data, making it easier for the viewer to see and understand these key insights.

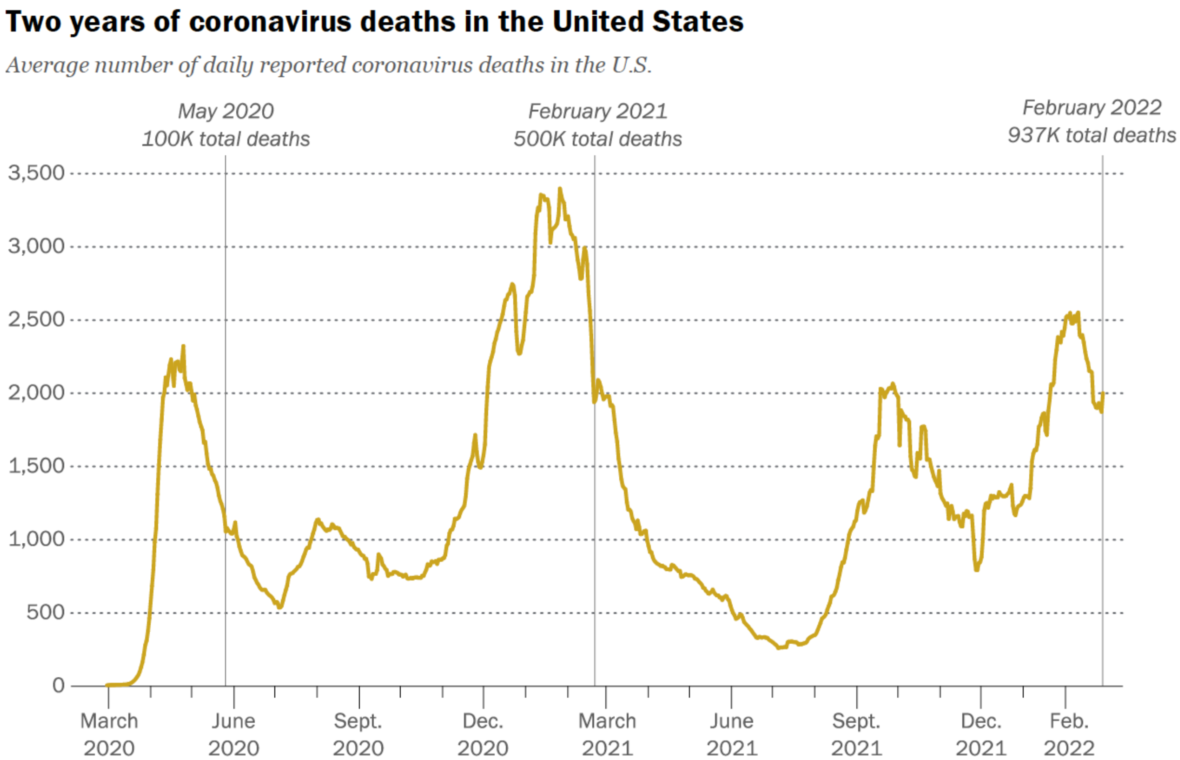

Figure 3: A line graph depicting the average daily deaths due to the coronavirus from Parker et al's "The Changing Political Geography of COVID-19 Over the Last Two Years"(2022). The graph uses annotations to highlight specific data points for the audience.

In *Figure 3*, we can see the use of annotations to tell a story to the audience. Pew Research Center is depicting the average daily reported coronavirus deaths in the United States and highlights milestones in the time period when cumulative deaths reached specific points. By using annotations, the designers draw attention to the times that they want the audience to remember. By providing this additional context, labels and annotations can help to make the visualization more intuitive and easier to interpret, ultimately improving clarity.

Consistency

Consistency is another valuable aspect of ensuring clarity in data visualization. Using a consistent visual style and layout throughout a visualization can help avoid confusion and make it easier to compare different elements. When a visualization is consistent in its use of colors, fonts, labels, and other visual elements, it creates a predictable and familiar experience for the viewer. This familiarity allows the viewer to focus on the data being presented, rather than trying to decipher inconsistencies or variations in the visualization. (Kelleher and Tierney 2018) Maintaining consistency is important within a single data visualization, across several visualizations, and across fields with established standards. Within one visualization or a group of visualizations in one article or portfolio, consistency of color scheme, labeling, layout, and fonts are important to ensure ease of interpretation and visual appeal.

Consistent use of color in data visualization is important because it helps the viewer to easily identify and differentiate between different data points or categories. For example, if purple represents men and green represents women in one graph, changing the color palette to something else or switching what colors represent which group (i.e. using green for men and purple for women) can be confusing. Consistent use of color can help to reduce confusion and errors in interpreting the data. If different colors are used to represent the same category or data point across different graphs or charts, the viewer may mistakenly assume that the colors represent different categories or data points, leading to misinterpretation of the data. (Cairo 2016)

Another aspect of color that is important to data visualization is cross-field interpretations of color choice. For instance, it is standard to use red to denote Republicans in the United States and blue to denote Democrats. It is then expected that when making a chart, practitioners use these colors. Likewise, using colors that don't accurately represent the data can be misleading. For example, using red to represent a positive trend can confuse viewers who associate red with negative trends. Understanding industry standards in color choices is a useful step in the visualization design process.

Consistency in labeling is also important. For example, if "U.S"" is used to represent the United States in one graph, it is recommended to continue using that labeling strategy instead of switching to "USA" in another graph in the same portfolio. Additionally, when displaying similar or the same data for different groups, say years, the labels should be consistently formatted and located in the same place. This ensures that viewers can easily understand the meaning of labels and terminology and helps to avoid confusion or misunderstandings that may arise from using different terms or labels for the same data points in different visualizations. Consistent labeling also helps to maintain the integrity and accuracy of the data, as using the same labels and terminology ensures that data is being compared and analyzed in a consistent manner. (Wickham 2010)

Consistent layout is another important factor in effective data visualization; the visual elements of the visualization should be placed in a predictable and familiar manner. For example, if a legend is used to explain the meaning of colors or symbols in a chart or graph, it should be placed in the same location in each chart or graph. This allows the viewer to quickly locate the legend and understand the meaning of the visual elements. The use of the same scale or axes is an important principle in Tufte's work: for instance, if one chart has a y-axis that ranges from 0 to 100 and another chart has a y-axis that ranges from 0 to 10, it can be difficult for the viewer to compare the two charts accurately. By using a consistent axis scale, the viewer can easily compare and interpret the data. (Tufte 2001) Consistent layout allows the viewer to focus on the data being presented, rather than trying to decipher variations in the layout or design. It creates a predictable and familiar experience for the viewer, which makes the visualization easier to understand and interpret.

Using consistent fonts means using the same font type, size, style, and color across all visualizations in a project. Consistent fonts help maintain the overall aesthetic appeal of the visualization and make it easier for viewers to read and understand the data. (Cairo 2016) When multiple fonts are used, it can make the visualization look cluttered and chaotic, which can confuse the audience and make it difficult to interpret the information being presented. (Kosara 2007) Maintaining the same font size and color for each type of text ensures that viewers can easily distinguish between different types of information: for instance, the title of the graph should remain at the top of the graph and be the most prominent text, as this is a consistent and understandable use of text.

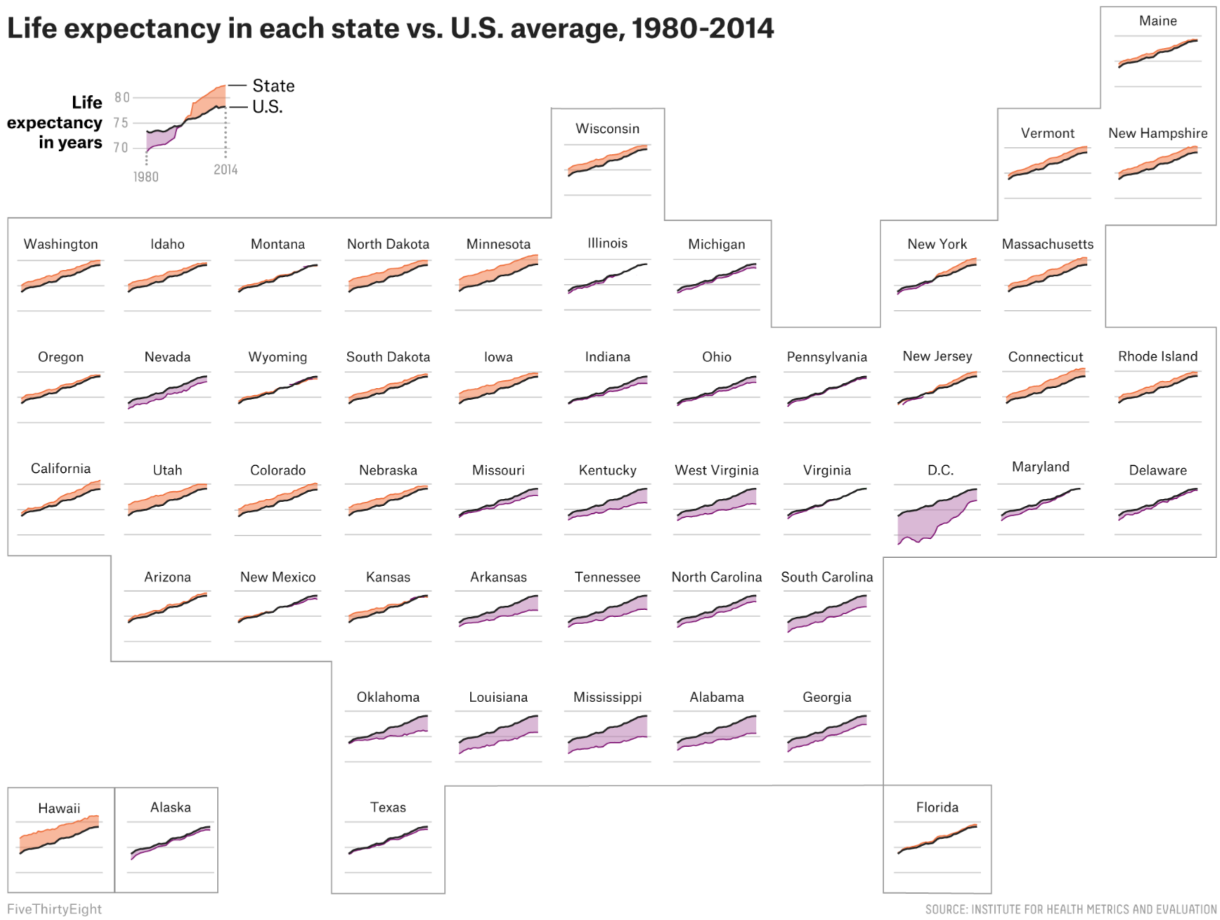

Figure 4: An example of a small multiples graph with consistent labeling, coloring, scale, font, and layout from FiveThirtyEight (Barry-Jester 2017).

By implementing these measures of consistency, practitioners can craft interpretable, familiar data visualizations that effectively communicate messages and ensure audiences better understand the data. Overall, consistency in data visualizations can improve clarity, reduce cognitive load, and enhance the viewer's ability to understand and interpret the data, all important in crafting effective, meaningful visualizations.

IV. Accuracy

Accuracy in data visualization refers to the degree to which the information presented in the visualization is true, correct, and representative of the underlying data. Particularly in social science, it is important that the data displayed in the visualization is accurate and not misleading, as the insights and decisions based on the visualization will only be as reliable as the underlying data. This involves ensuring that the data is collected, processed, and analyzed correctly, and that any visual representations of the data accurately reflect its meaning. Accuracy can be improved by verifying the data, checking for errors or inconsistencies, and using appropriate statistical methods to summarize and analyze the data.

Inaccuracies can be intentional or unintentional. Unintentional inaccuracy can occur when factors such as human error, flawed data sources, or technical issues are undetected by the practitioner. To ensure accuracy in data visualization, it is important to take a careful and rigorous approach to data collection, analysis, and presentation. This includes double-checking calculations, verifying data sources, and testing visualizations on multiple devices and platforms to ensure consistency and accuracy. It is also important to be transparent about the limitations and potential sources of error in the data, and to provide clear and detailed explanations of the methods used to generate the visualizations. Including source material in the visualization is a useful inclusion that allows the audience to view the underlying data and potentially recreate the visualization.

It is crucial to identify and address common causes of misleading information in graphs, whether they are intentional or unintentional inaccuracies. Misleading inaccuracies can arise due to various reasons, such as an inadequate or inappropriate selection of data, the use of misleading scales, and the presentation of data in a biased or incomplete manner. Purposeful misleading inaccuracies are often used to manipulate the audience or to push a particular agenda. For instance, a person or organization may use selective data to misrepresent a situation or issue to their advantage. Such tactics are often used in advertising or political campaigns. Incidental inaccuracies, on the other hand, are unintentional and can result from a lack of knowledge or awareness of best practices in data visualization.Examples of methods that may cause misleading visualizations include choices about chart type, axes and scale, data selection, units and labels, and colors.

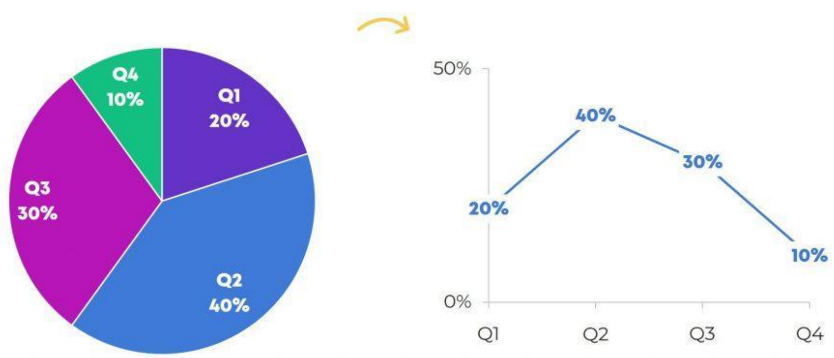

Figure 5: An example of a pie chart being incorrectly used to depict temporal data from Emery 2015's "When Pie Charts Are Okay (Seriously): Guidelines for Using Pie and Donut Charts"

Selecting an inappropriate type of graph for the data can result in distortion and mislead the audience about the observed pattern in the data. For instance, if the data involves comparing multiple categories, a bar chart or a stacked bar chart should be used instead of a line chart or scatter plot. Utilizing a line chart or scatter plot may suggest a relationship or trend between the categories that does not exist, thereby compromising accuracy. Additionally, using a pie chart to represent changes over time can be misleading as it fails to indicate the temporal relationship between the data points, as we see in *Figure 5*. It is essential to match the graph type with the data being presented to ensure accuracy. Failing to do so may result in confusion or even misinterpretation of the data.

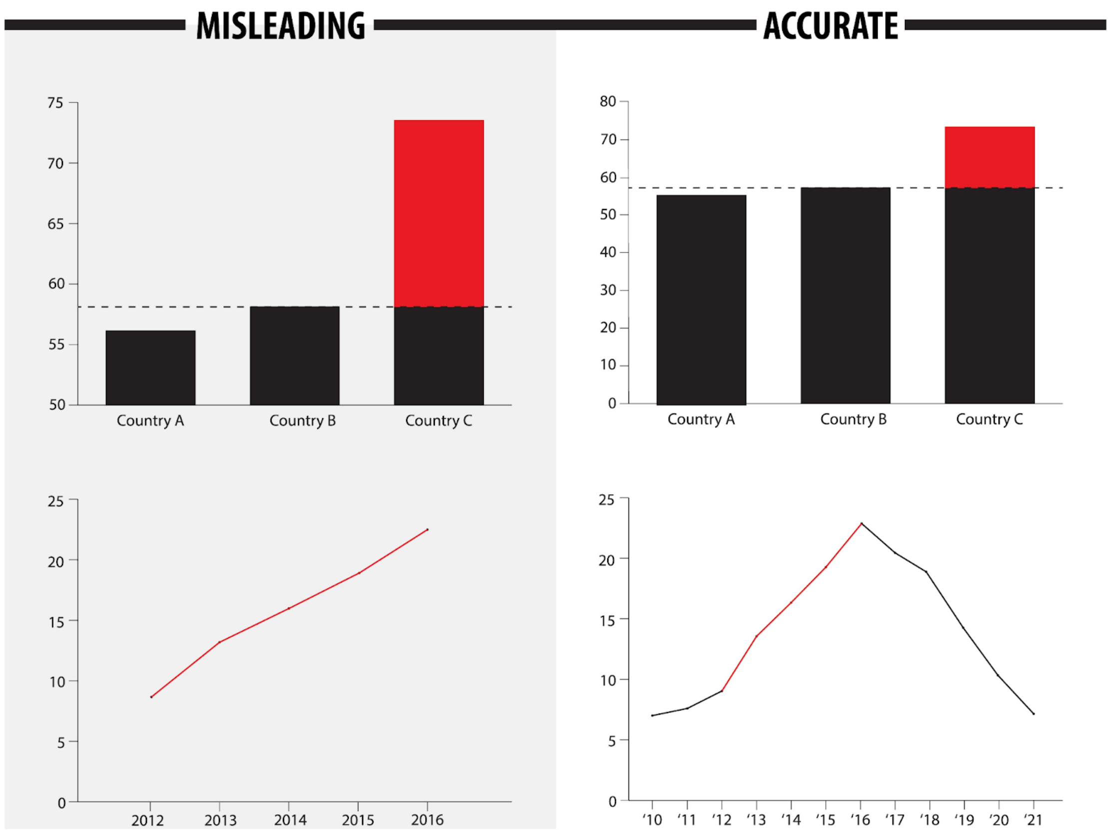

Figure 6: Examples of hypothetical misleading and accurate graphs from the "Using factchecks to combat misleading graphs" research project at the Universitet Leiden. (Smeets et al 2023) From the authors: "The top left graph is poorly designed by omitting the baseline. The bottom left graph displays insufficient data by cherry picking."

The choice of scale on axes may too lead to inaccuracy or misleading messaging. Changing the scale of an axis can alter the appearance of the data. One common concern in data visualization is the scaling of the y-axis, as the starting point of the y-axis can greatly impact how the data is perceived. Starting the y-axis at a value that is not zero can exaggerate the differences between data points and make the differences appear larger than they actually are as we see with the top charts in *Figure 6*. Tufte emphasizes the importance of using appropriate scales and axes in data visualizations. He suggests that scales should be chosen to reflect the range of data being represented, rather than being manipulated to exaggerate differences or create a desired effect (Tufte 2001). Similarly, he argues that axes should be labeled clearly and accurately to avoid confusion or misinterpretation. Visualization designers should ensure that the intervals between the tick marks on the axis are consistent. Inconsistent intervals can create confusion and make it difficult to compare data points accurately, including jumps or breaks in the axis for brevity.

Another design choice that may lead to inaccuracies or misleading messages is the choice of which data or subset of data to use. Selectively including data that supports a particular conclusion while excluding data that contradicts it can create a misleading graph, as we see in the bottom charts of *Figure 6*. While using all data available may not always be feasible, ensuring that the data selected does not support untrue conclusions is an important facet of ensuring accuracy in data visualization.

Accurate data visualizations are essential for effective communication with a range of audiences. Visualizations are often used to convey complex information to a non-technical audience, and if the information is inaccurate, it can be difficult to understand and make informed decisions. Accurate data visualizations can help to make complex information more accessible and understandable to the audience, which can lead to better decision-making and more effective communication. Inaccurate visualizations can have significant consequences, especially in fields such as medicine, finance, and politics, where incorrect decisions can have a profound impact.

V. Additional Considerations in Visualization

Audience Selection

In his book "The Functional Art," Alberto Cairo introduces the concept of the Visualization Wheel, a framework for understanding and comparing visualization tradeoffs.

The wheel consists of two halves that represent a fundamental complexity spectrum on which data visualizations can be placed. The top half of the wheel represents visuals that contain deeper, more complex data, while the bottom half represents visuals that provide more accessible but shallower data. This spectrum is at the core of Cairo's philosophy about data visualization. In presenting to a knowledgeable, technical audience, more complexity in the visualizations is to be expected and, to some degree encouraged. For more general audiences, more simple, easily digestible visualizations that may sacrifice some technical complexity may be warranted.

Beyond the basic complexity spectrum, the wheel also highlights other tradeoffs in visualization. Each section represents a fundamental spectrum on which data visualizations can be placed, and understanding where a particular visualization should be on each spectrum can help in designing effective and engaging visualizations. As crafted by Alberto Cairo in The Functional Art and described by Ryan Wingate in his article "Alberto Cairo's Visualization Wheel" (2019), these tradeoffs are:

- *Abstraction and Figuration*: Figurative visuals use physical representations such as photographs or drawings, while abstract visuals use more conceptual representations of phenomena.
- *Functionality and Decoration*: Functional graphics have little to no embellishments and are closer to a direct representation of the data, while visuals with significant decoration contain more artistic embellishments and visual appeal.
- *Density and Lightness*: Dense visuals convey a lot of information, and are intended to be studied in some depth, like those in scientific journals. By contrast, light visuals convey less nuance and less information, but express the core message quickly.

- *Multidimensional and Unidimensional*: Multidimensional visuals illustrate many different aspects of a phenomenon, and likely illustrate the phenomenon as a whole better than unidimensional visuals. Unidimensional visuals, contrastingly, illustrate only single items associated with a phenomenon.
- *Originality and Familiarity*: Original graphics do not readily conform to the most common, well-known visualization patterns, whereas familiar, commonplace, broadly-understood visuals include bar charts, line charts, and scatter plots.
- *Novelty and Redundancy*: Redundant graphics employ several methods to convey the same information. For instance, a bar may depict the highest value and be highlighted in a distinctive color. Novel graphics, in contrast, convey each phenomenon in the visualization using only one approach.

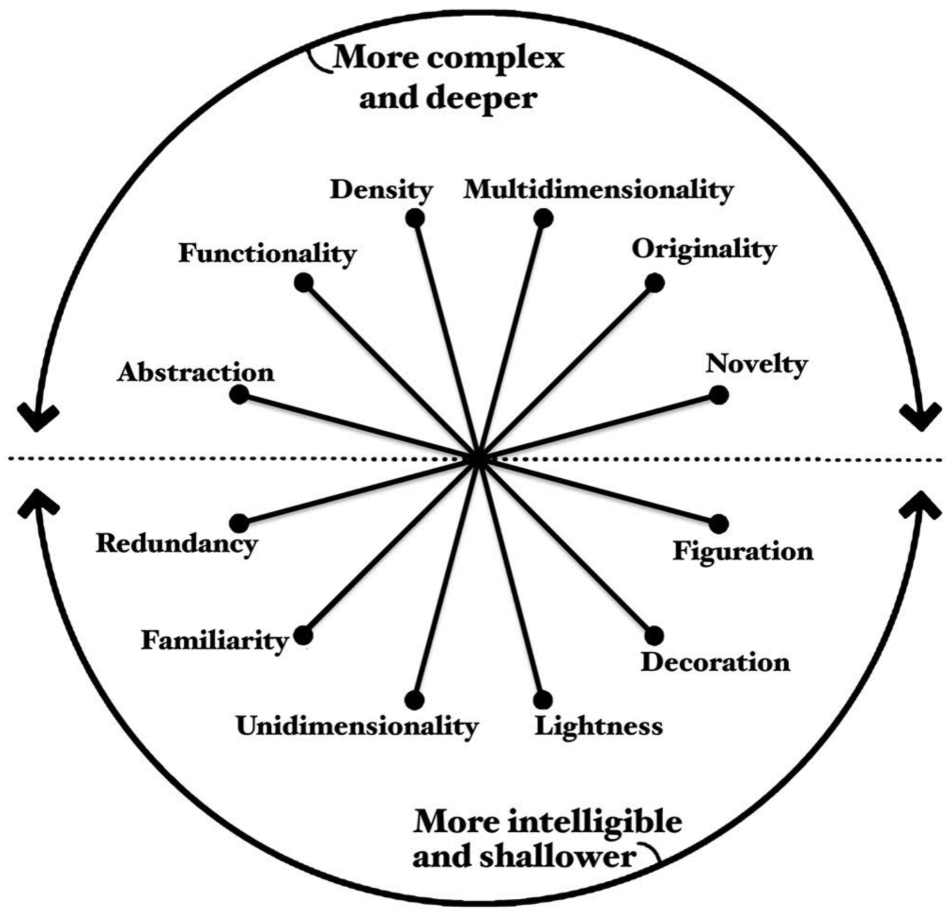

Figure 7: Alberto Cairo's Visualization Wheel.

According to Cairo, visualizations that are highly abstract, highly decorative, highly dense, highly multidimensional, highly original, or highly novel may be more difficult for viewers to understand, whereas visualizations that are highly figurative, highly functional, highly light, highly unidimensional, highly familiar, or highly redundant may be more accessible. There is no one right balance of all of these tradeoffs; instead, different data and different audiences may adjust the balance according to need. Different audiences likely prefer different types of visualizations. For instance, more technical or scientific researchers may prefer visuals that are dense, multidimensional, and have high functionality, as they are already familiar with the basic information and can understand even complex representations. Contrastingly, artists, graphic designers, and journalists are likely to prefer visuals that include decoration, lightness, and figuration, as this style is easier to understand and more visually appealing. (Wingate 2019)

Even within the social sciences, understanding which specific audience is the target of the visualization is fundamental to crafting an effective visualization. Different potential target audiences for social scientists such as policymakers, researchers, or members of the general public have different levels of expertise and familiarity with the subject matter, and may require different levels of detail and complexity in the visualization. When publishing in a technical academic journal where the intended audience is expected to have a strong foundation in the field of research and desires specificity, more complex, deeper visualizations may be apt. By contrast, when targeting the general public, perhaps through journalism or an infographic, more familiar forms with light, easily digestible visualizations may be more effective.

Overall, Cairo's framework can help to understand the tradeoffs of visualization complexity and style. Understanding the intended audience prior to designing a visualization is useful, as it helps to ensure that the visualization is designed with appropriate tradeoffs in mind. The social sciences are vast and cover many different fields, and different audiences within those fields may have varying levels of familiarity with certain visualization techniques or statistical concepts. By understanding the intended audience, a designer can make informed decisions about the level of complexity and the choice of visual representation that may be most effective in helping the audience understand the data.

Accessibility

Accessibility in data visualization refers to the design of visualizations that are usable and understandable by people with disabilities or limitations, such as visual impairments, color blindness, or cognitive difficulties. Accessible data visualizations are designed in a way that allows all users to interact with and interpret the information presented, regardless of their abilities or the technology they use to access the visualization. This can include the use of alternative text descriptions for images, high-contrast color schemes, and providing interactive elements that can be navigated through keyboard commands. (Lee et al 2020)

Crafting data visualizations that are engaging and accessible for all audiences involves taking into consideration factors such as color contrast, font size, the use of alternative text descriptions for images, and the provision of interactive elements that can be navigated through keyboard commands. Users with visual, auditory, intellectual, or motor disabilities may interact with the visualizations using accessible technology, and thus keeping these considerations in mind could allow more users to interact with the visualization. (UW 2023) For example, using high-contrast color schemes can help users with color blindness or low vision to distinguish between different elements in a visualization. Providing alternative text descriptions for images can make it possible for users with visual impairments to understand the content of the visualization through assistive technologies such as screen readers. Similarly, labels should be used to describe the data being presented where appropriate, as this helps users understand the information being conveyed and is especially important for people who are visually impaired or intellectually disabled. (Elavsky et al., 2022)

Throughout the design process, it can be important to test data visualizations with users with disabilities to ensure that they are accessible and easy to use. Designing with accessibility in mind does not always mean changing a graphic completely. Instead, adding certain labels, patterns, and descriptions are small ways that may make a large difference, as we see in *Figure 8*. Though this may not square with a strict interpretation of Tufte's principle of simplicity, it actually may improve clarity for users and be a more effective data visualization.

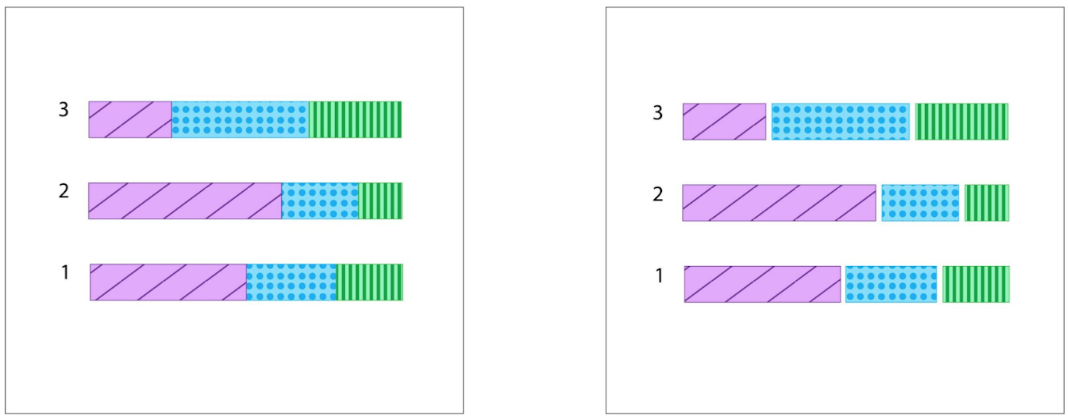

Figure 8: An example of accessible data visualization with multiple modalities. The designer uses different colors, patterns, and a white dividing separation line for the visualization on the left to make the bar graphs more accessible.

In addition to improving the accessibility of data visualizations for users with disabilities, designing for accessibility can also benefit all users by creating visualizations that are clearer, more legible, and easier to navigate. Stephen Few is a well-known data visualization expert, author, and consultant who has been instrumental in shaping the field of data visualization. Few's approach to data visualization emphasizes the importance of using simple and clear visual elements that are easy to interpret, rather than relying on flashy or complex designs. He believes that data visualizations should be designed to support the specific tasks and goals of the viewer, and that designers should carefully consider the audience's background knowledge, cognitive abilities, and goals when designing visualizations. Simple, clear, and easy to understand graphics help all people understand the visualizations better, not only those with disabilities.

VI. Concluding Remarks

Data visualization is a particularly important tool for social scientists to 1) interpret social data and 2) present their research findings in an accessible and compelling way in academia, policy, and public spheres. Social scientists often work with complex data sets that can be difficult to understand without visualization. Through data visualization, researchers can create graphical representations of their data that make it easier to identify patterns and relationships. One of the primary benefits of data visualization in the social sciences is that it can help researchers identify patterns and relationships in their data that might not be apparent through other means. For example, a scatterplot can reveal a correlation between two variables that might not be obvious by just looking at a table of numbers. Likewise, they can communicate these findings with a variety of technical and non-technical audiences as desired. When designed carefully and intentionally, visualization is a powerful tool for communicating complex information in an accessible and understandable way. Social scientists often interface with both academic and general audiences, and may need to present information accordingly. Data visualization is a tool that can be fine-tuned to the needs of the specific audience, researcher, and information.

In creating effective and meaningful visualizations, it is important to follow key principles such as using accurate and relevant data, choosing the right type of graph, maintaining consistency in labeling, layout, and design, and avoiding misleading information. Simplicity, clarity, and accuracy are three defining principles of data visualization that help to ensure the data is presented in a clear, understandable, and truthful manner, regardless of the specific tool or technique being used to present the data. By adhering to these principles, practitioners can create visualizations that not only convey information clearly, but also provide insights and promote informed decision-making tailored to whichever audience they so choose.

As data becomes increasingly central to modern life and work, the ability to create and interpret data visualizations will only become more important. Therefore, it is essential for social scientists to continue to refine and improve practices in this area, in order to harness the full potential of data visualization and its role in shaping understanding of the world around us.

References:

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2013). Useful junk?: The effects of visual embellishment on comprehension and memorability of charts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1143-1152.

Blasio, A. J., & Bisantz, A. M. (2002). A comparison of the effects of data-ink ratio on performance with dynamic displays in a monitoring task. International Journal of Industrial Ergonomics, 29(6), 363-372.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable?. IEEE transactions on visualization and computer graphics, 19(12), 2306-2315.

Carty, R. K., & Eagles, M. (2003). Canadian parties in transition (3rd ed.). Toronto, ON: Nelson Thomson Learning.

Chen, H. (2016). Data Visualization: Principles and Practice. CRC Press.

Elavsky, F., Bennett, C., & Moritz, D. (2022, June). How accessible is my visualization? Evaluating visualization accessibility with Chartability. In Computer Graphics Forum (41-3:57-70).

Emery, A. K. (2015). When Pie Charts Are Okay (Seriously): Guidelines for Using PIE and Donut Charts. Depict Data Studio. https://depictdatastudio.com/when-pie-charts-are-okay-seriously-guidelines-for-using-pie-and-donut-charts/.

Few, S. (2012). Show me the numbers: Designing tables and graphs to enlighten. Analytics Press.

Franconeri, S. (2021). The Science of Visual Data Communication: What Works. Annual Review of Psychology, 72(1), 671-696.

Healy, K. (2018). Data visualization: A practical introduction. Princeton, NJ: Princeton University Press.

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. Communications of the ACM, 55(4), 45-54.

Holbrook, T. M., & Hill, S. J. (2005). Exploring survey data using visualization. In J. A. Tanur (Ed.), Questions about questions: Inquiries into the cognitive bases of surveys, 71-95. New York, NY: Russell Sage Foundation.

Hullman, J., Adar, E., & Shah, P. (2011). The impact of social information on visual judgments. In Proceedings of the SIGCHI conference on human factors in computing systems, 1461-1470.

Inbar, O., & Nachmias, R. (2007). Minimalism in information visualization: Attitudes towards maximizing the data-ink ratio. Proceedings of the 12th international conference on Information visualization, 6-8.

Kelleher, C., & Tierney, B. (2018). Data visualization: A guide to visual storytelling for libraries. Rowman & Littlefield.

Kirk, A. (2016). Data visualization: A successful design process. Packt Publishing.

Knoke, D., & Yang, S. (2008). Social network analysis (2nd ed.). Thousand Oaks, CA: Sage Publications.

Kosara, R. (2007). Visualization criticism-the missing link between information visualization and art. In 2007 11th International Conference Information Visualization (IV'07: 631-636). IEEE.

Kosslyn, S. M. (2006). Graph design for the eye and mind. Oxford University Press.

Lee, Bongshin, et al. (2020) "Reaching broader audiences with data visualization." IEEE Computer Graphics and Applications 40.2: 82-90.

Li, H., & Moacdieh, N. (2014). Is "chart junk" useful? An extended examination of visual embellishment. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 58, No. 1, pp. 1516-1520). Sage CA: Los Angeles, CA: Sage Publications.

Mankiw, N. G., Weinzierl, M., & Yagan, D. (2019). Optimal taxation in theory and practice. Journal of Economic Perspectives, 33(4), 103-130.

McGurgan, K. (2015). Data-ink ratio and task complexity in graph comprehension. Rochester Institute of Technology.

Parker, K., Horowitz, J. M., & Minkin, R. (2022, March 3). The Changing Political Geography of COVID-19 Over the Last Two Years. Pew Research Center. https://www.pewresearch.org/politics/2022/03/03/the-changing-political-geography-of-covid-19-over-the-last-two-years/.

Smeets, I. et al. (2023). Using Factchecks to Combat Misleading Graphs [Research project]. Universiteit Leiden Department of Biology.

https://www.universiteitleiden.nl/en/research/research-projects/science/using-factchecks-to-combat-misleading-graphs.

Sorensen, D. L. (1993). The Psychological Processes of Embellished Graph Reading. University of Idaho, Moscow, ID.

Tufte, E. R. (1997). Visual explanations: Images and quantities, evidence and narrative. Graphics Press.

Tufte, E. R. (2001). The visual display of quantitative information. Graphics Press.

UW–Madison Information Technology. (2023). Accessible Data Visualizations. https://it.wisc.edu/learn/make-it-accessible/accessible-data-visualizations/.

Wickham, H. (2010). A layered grammar of graphics. Journal of Computational and Graphical Statistics, 19(1), 3-28. doi: 10.1198/jcgs.2009.07098

Wilkinson, L. (2005). The Grammar of Graphics. Springer Science & Business Media.