THE UNIVERSITY OF CHICAGO

Data Visualization with Applications in the Social Sciences: A Portfolio

By

Allison Marie Towey

June 2023

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in

Computational Social Science

Faculty Advisor: Jean Clipperton
Preceptor: Pedro Arroyo

Table of Contents

Portfolio Overview

Link to Digital Portfolio (Github): https://atowey-uchi.github.io/data_visualization/

In this portfolio, you will find a literature review and three data visualization projects that showcase the primary theoretical insights from the literature across a range of social science domains. The portfolio aims to demonstrate how the theoretical principles can be applied to real-world scenarios and how effective data visualization can enhance understanding and communication of complex information for different audiences.

Creating effective visualizations requires an understanding of the subject matter, the audience, and the goals of the communication. In the social sciences, this means understanding the theories and concepts that underlie the data, as well as the specific audience that the visualization is intended to reach. By carefully considering these factors, data visualizations can be designed to convey key insights and facilitate understanding among different audiences, whether they are academics, policymakers, or the general public.

The first visualization, "Men's Alcohol Consumption and Women's Experiences of Intimate Partner Violence in Ukraine" is intended for an advanced technical social scientific audience. It is written as an intended published journal article like that of the Data Visualization Collection from the American Sociological Association's *Socius* Journal. The project includes one data visualization with 500 words of accompanying text describing the findings. This visualization uses a novel style, a modified waffle chart, to depict frequencies. This chart style, though perhaps not familiar to most viewers, clearly displays the relationships between men's alcohol consumption and their wives' experiences of intimate partner violence in Ukraine. Additionally, as the audience is a published social science research community, the accompanying text is more technical and sparsely written as this is the primary style in the community and emphasizes clarity and simplicity over embellishment.

The second visualization, "Global Access to Safe Drinking Water", has a different intended message, audience, and dataset. Instead of one novel visualization that elucidates relationships in the data, this infographic-style data visualization contains a series of familiar data visualization styles, including choropleth maps, line and bar graphs, and scatter plots from a variety of related sources. The use of multiple visual elements and data sources allows for a more complete understanding of the issue and highlights the interconnected nature of the challenges and potential solutions. The visualization is designed to communicate the urgency of the issue and the potential benefits of action, while also providing actionable information for policymakers, advocates, and community members. The use of familiar, more visually appealing graphics intends to reach the general public and highlight the issue overall as opposed to delving deeply into one aspect.

The third visualization, "Covid-19 and Transportation: How the Pandemic Affected Where We Went and How We Got There", has a focus on the general public as the audience and is written as a journalism piece. This set of visualizations explores the impact of the COVID-19 pandemic on transportation patterns using a combination of maps, line charts, and bar charts. It is designed to communicate the drastic changes in transportation patterns caused by the pandemic, from the sharp decrease in transportation of all kinds, including most significantly, public transit ridership and aviation. By presenting the data in a journalistic style, the visualization makes the data more accessible to a broad audience and emphasizes the significance of the findings. Like the second visualization, this article intends to inform broadly, and thus, a variety of data sources are used. The emphasis in this piece is the connection between text and graphics that augment the text, highlighting a specific trend in recent history in an accessible and visually appealing manner.

Each of these visualizations has an emphasis on understanding the audience and attempting to craft visualizations that are simple, clear, accurate, and effectively targeted. As shown throughout the portfolio, the choice of visualization type, color scheme, and level of detail is tailored to the audience's level of expertise and familiarity with the data. Overall, this portfolio seeks to demonstrate that effective data visualization requires an understanding of the audience and a deliberate, digestible design approach that takes into account the specific message, data, and intended impact.

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 can 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)

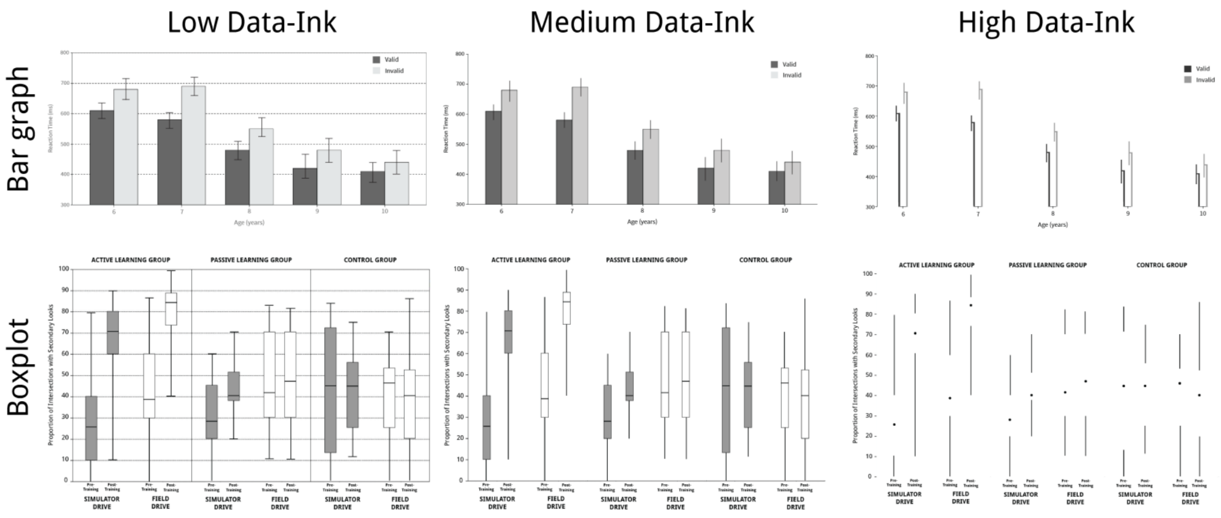


Figure 1: Examples of low, medium, and high data-ink bar graph and boxplots developed by Kevin McGurgan in "Data-ink Ratio and Task Complexity in Graph Comprehension" (2015)

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.

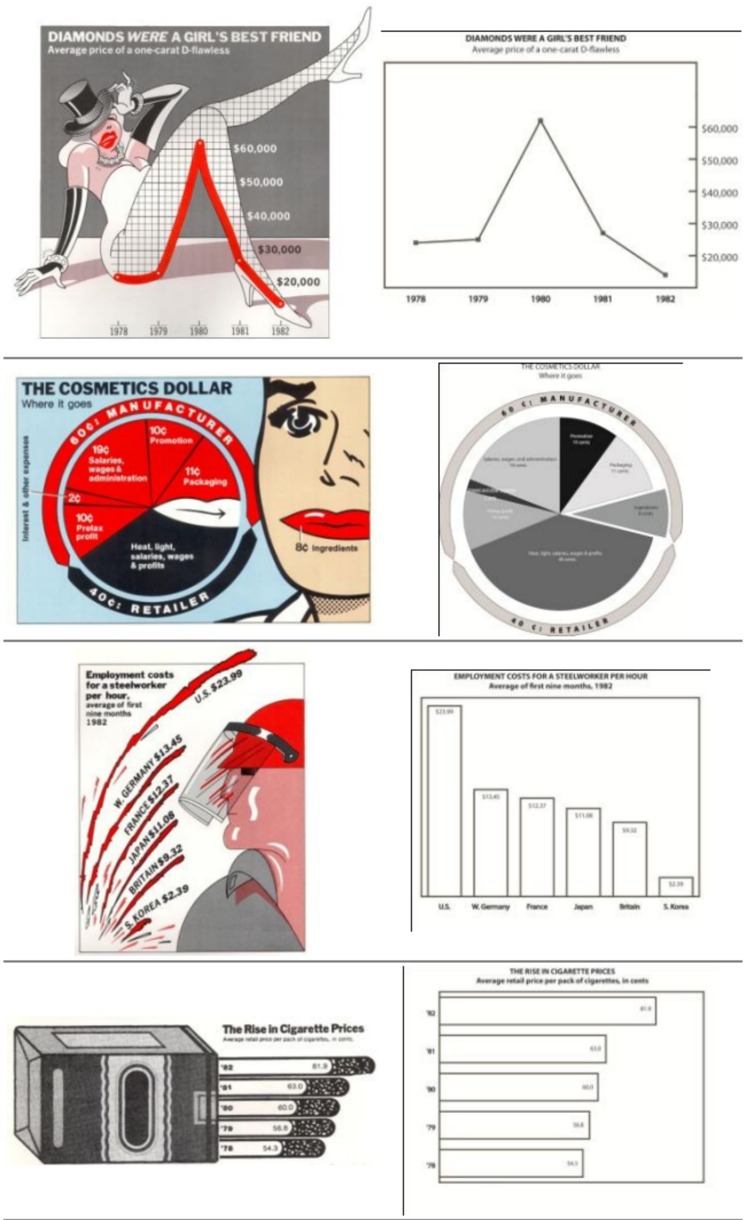


Figure 2: Examples of "chart junk" or embellished graphs compared to more simplistic depictions from Bateman et al 2013's "Useful junk?: the effects of visual embellishment on comprehension and memorability of charts"

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 and appealing 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 are also 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 assist the viewer in interpreting 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, allowing the viewer to see and understand these key insights without excessive effort.

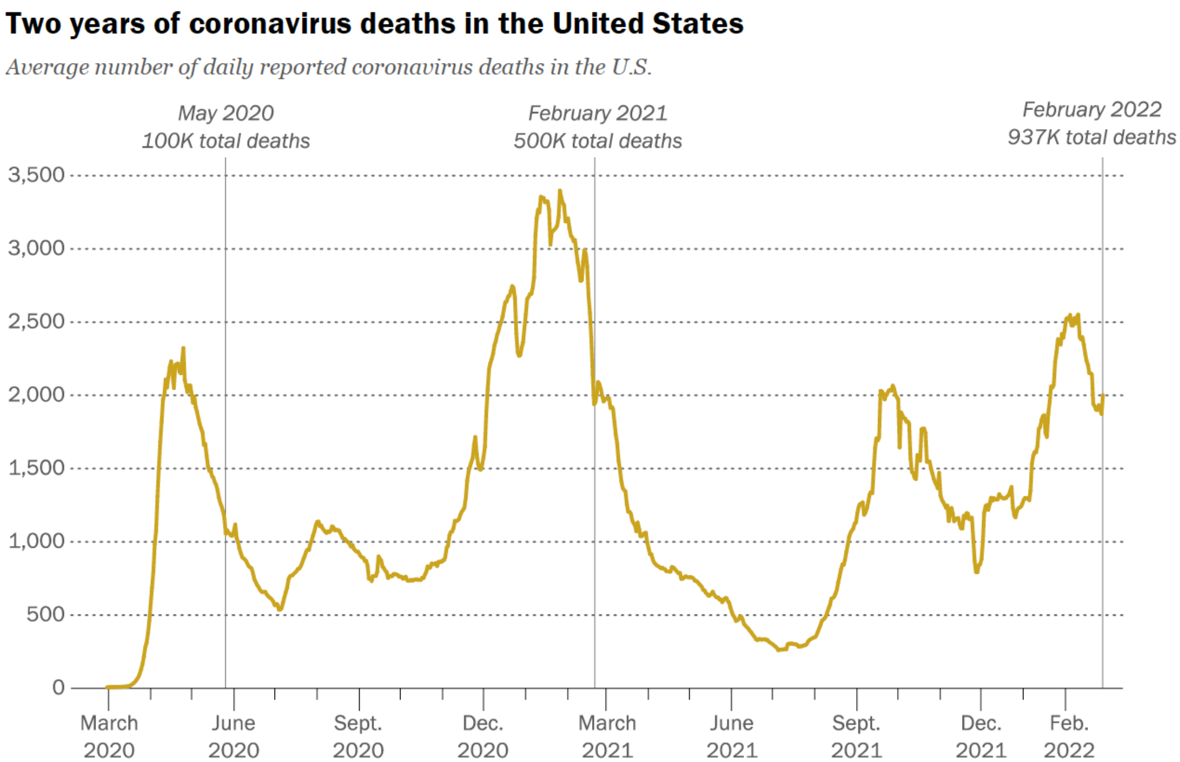


Figure 3: A line graph depicting the average daily deaths due to the coronavirus from Parker et al's "The Changing Political Geography of COVID-19 Over the Last Two Years"(2022). The graph uses annotations to highlight specific data points for the audience.

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, for instance 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.

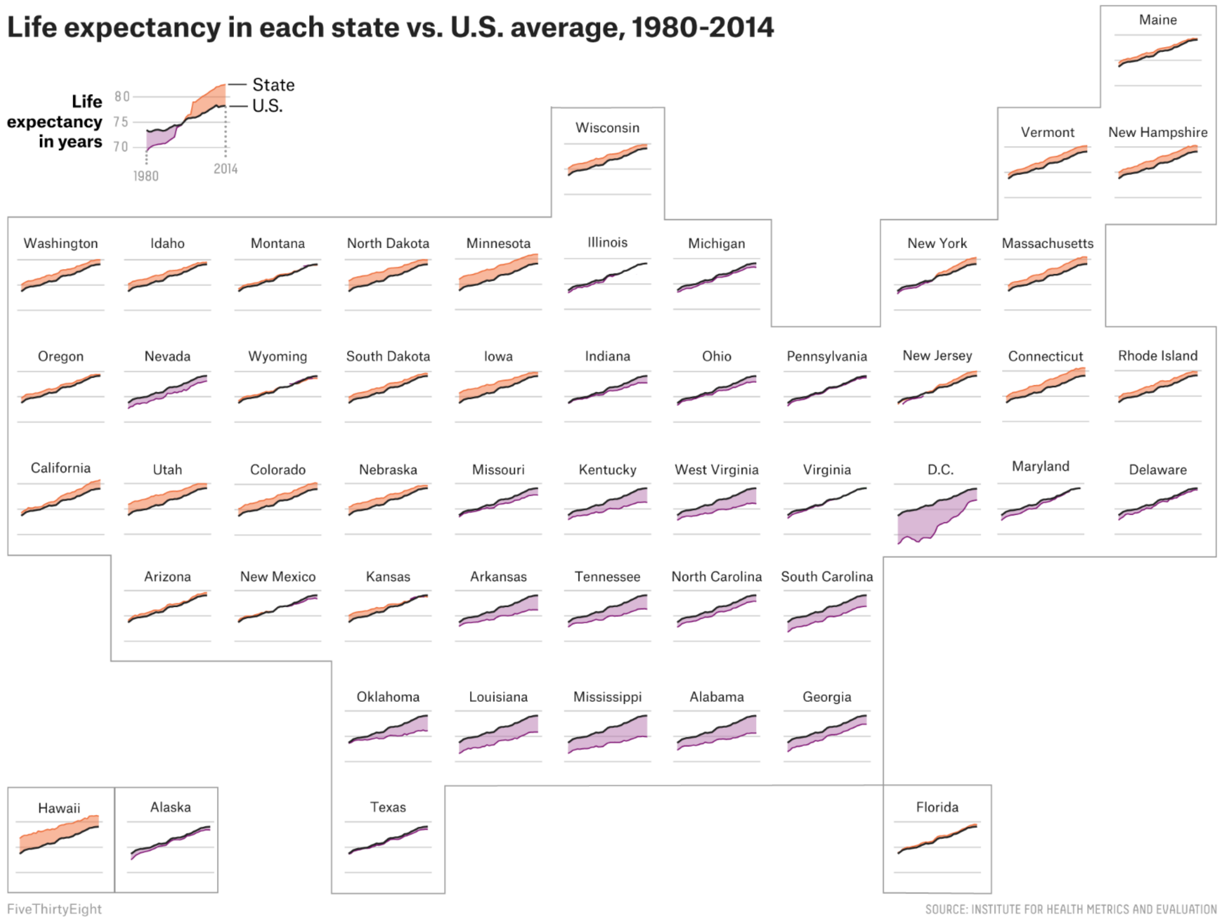


Figure 4: An example of a small multiples graph with consistent labeling, coloring, scale, font, and layout from FiveThirtyEight (Barry-Jester 2017).

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,which are 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.

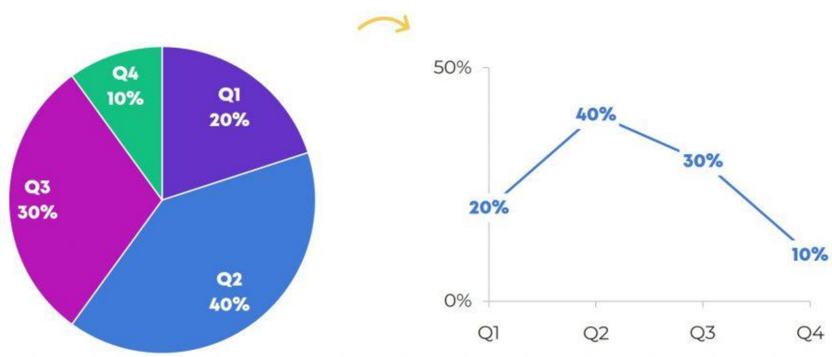


Figure 5: An example of a pie chart being incorrectly used to depict temporal data from Emery 2015's "When Pie Charts Are Okay (Seriously): Guidelines for Using Pie and Donut Charts"

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.

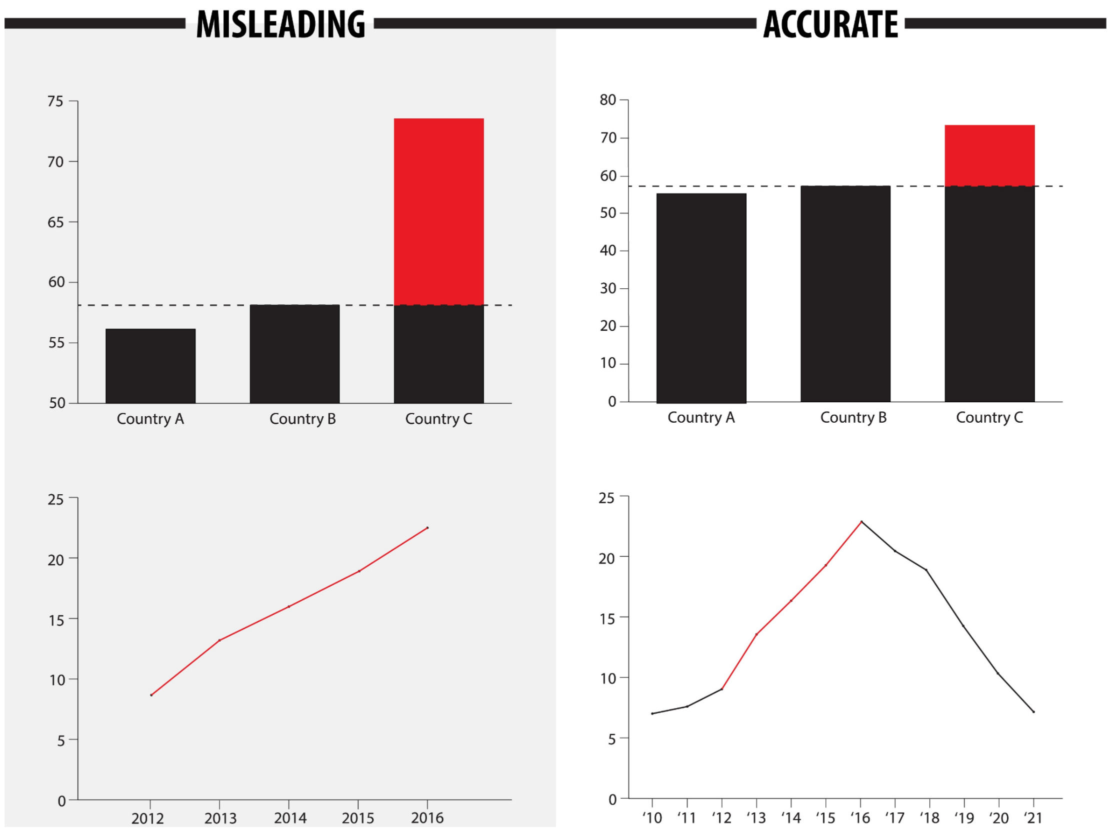


Figure 6: Examples of hypothetical misleading and accurate graphs from the "Using factchecks to combat misleading graphs" research project at the Universitet Leiden. (Smeets et al 2023) From the authors: "The top left graph is poorly designed by omitting the baseline. The bottom left graph displays insufficient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

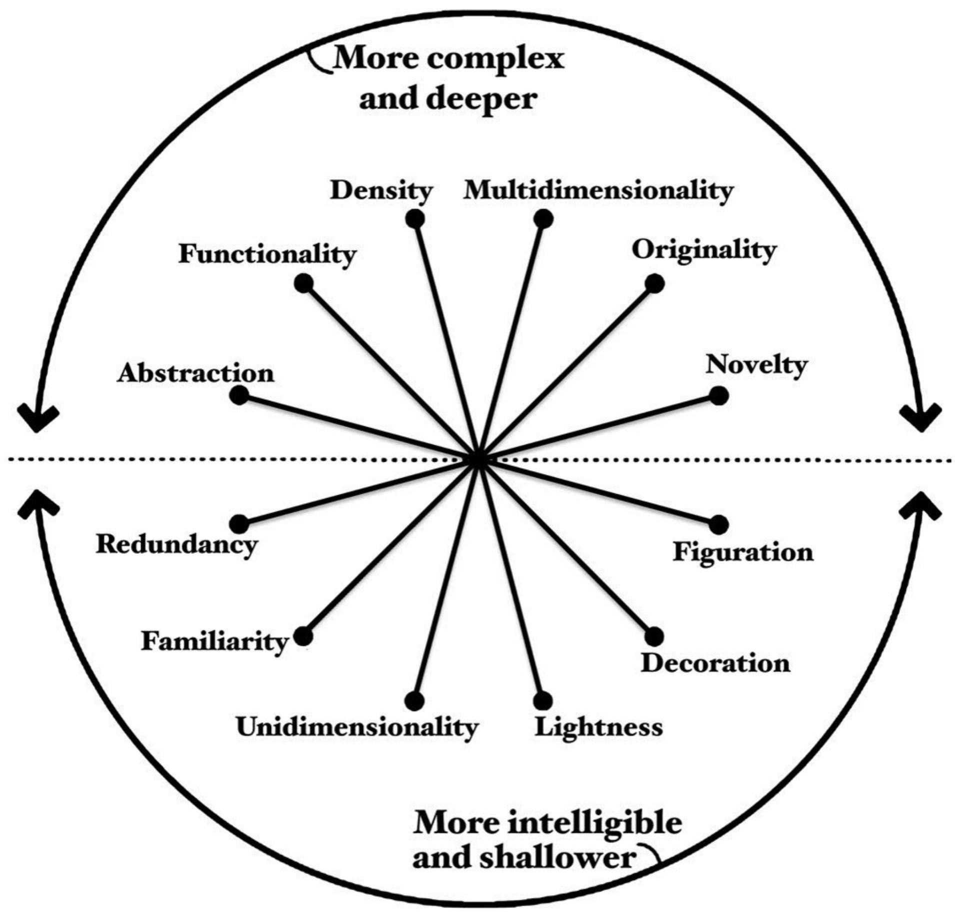


Figure 7: Alberto Cairo's Visualization Wheel.

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.

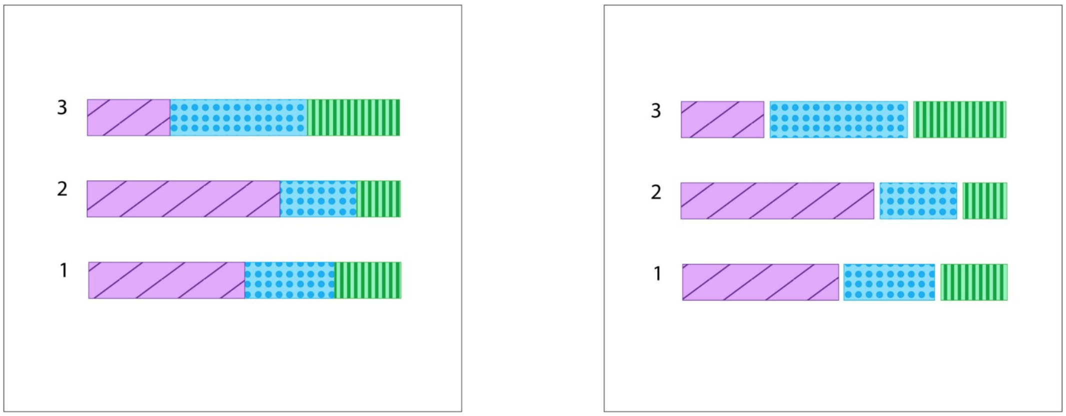


Figure 8: An example of accessible data visualization with multiple modalities. The designer uses different colors, patterns, and a white dividing separation line for the visualization on the left to make the bar graphs more accessible.

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 will provide insights and promote informed decision-making tailored to their intended audience.

As data becomes increasingly central to modern life and work, the ability to create and interpret data visualizations will only become increasingly 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.

Visualization 1: Men's Alcohol Consumption and Women's Experiences of Intimate Partner Violence in Ukraine

Overview and Methodology

Link to digital document (HTML):
https://atowey-uchi.github.io/data_visualization/ua-ipv/code%20%2B%20data/ua-ipv.html

Abstract:
Intimate Partner Violence (IPV) is a matter of concern across the globe, especially in Ukraine, where a fourth of ever-married women report having experienced some type of emotional, physical, or sexual violence perpetrated by a partner or husband. Leveraging data from the 2007 Demographic and Health Survey, this visualization summarizes women's experiences of physical, sexual, and emotional IPV according to their husbands' alcohol consumption. The proportionally drawn square matrix chart depicts the prevalence of IPV in Ukraine, by type, and confirms the strong association between alcohol consumption and IPV found in previous research. This study adds to an existing literature on the impact of alcohol on IPV perpetration and suggests a need for targeted interventions that address alcohol use in intimate partnerships. Read today in the current context of armed conflict in Ukraine, we are reminded of the urgent need for reliable and recent data on health and family life.

Methodology:

This data visualization aims to provide valuable insights into the prevalence of intimate partner violence (IPV) and its connection to alcohol consumption in Ukraine. It is developed in the style of a publishable journal article like that of the Data Visualization Collection from the American Sociological Association's Socius Journal, containing one graphic with approximately 500 words of accompanying text with a maximum of five references. Specifically tailored for readers of a social science academic journal with prior knowledge of the literature, this single data visualization offers a multidimensional examination of the prevalence of IPV and the role of men's alcohol consumption as a risk factor.

The square matrix chart illustrates the occurrence of intimate partner violence (IPV) among ever-married Ukrainian women. It is divided into three panels based on the frequency of their husbands' drinking: 'never,' 'sometimes,' or 'often.' Each square within the matrix represents 0.1% of the total weighted sample from the 2007 Demographic and Health Survey in Ukraine, totaling 1000 squares. The colors of the squares indicate the proportion of women who have experienced physical, sexual, emotional violence, or a combination thereof.

This visualization also seeks to exemplify simplicity, clarity, and accuracy as highlighted in data visualization literature. To ensure the graphic is simple, there are no extraneous decorative 'chartjunk' elements in order to maximize the data-ink ratio. The focus of the graphic is clearly on the colored boxes and their location on the page. As such, the graphic maintains a minimal style without excessive text, instead allowing the number and color of the boxes to explain the relationship. Simultaneously, however, to maintain clarity, there is a clear legend with the colors of the boxes listed and accompanying caption text stating: "Each square represents 0.1% of the total weighted sample of Ukrainian women from the 2007 Demographic and Health Survey." By including succinct explanatory text, the visualization is not only clearer, but allows for a more accurate reading of the visualization.

Accuracy is prioritized through the use of weighted sampling and proportional representation. By employing the weighted sample, the visualization offers a more accurate reflection of the entire Ukrainian population, reducing potential biases that may arise from a limited survey sample. It encompasses all of the available survey data of married women in Ukraine from the 2007 DHS, mitigating data-choice bias and providing a comprehensive perspective. To further ensure accuracy, the replication code is made publicly available, promoting transparency and validating the data analysis process. By incorporating simplicity, clarity, and accuracy, this visualization adheres to best practices in data visualization, fostering a better understanding of the subject matter while maintaining integrity and reliability.

Given the presumed academic background knowledge of the intended audience, this visualization delves into the "more complex and deeper" aspects of Albert Cairo's Visualization Wheel, namely abstraction, originality, and multidimensionality. The visualization employs abstraction by transforming the survey data into proportional squares that are weighted based on population representativeness. Instead of representing each individual woman's response, the squares represent 0.1% of the weighted sample of Ukrainian women. This abstraction enables a more condensed, accurate representation that emphasizes overall population-level proportions rather than individual responses in the survey.

The visualization demonstrates originality in its representation by utilizing the square matrix chart, a chart type that is less commonly employed and typically associated with illustrating proportions of a whole. It offers a distinctive, novel visual approach to convey multidimensional information; as such, the graphic brings a fresh perspective to the representation of multi-categorical survey data. The originality of this approach has the potential to enhance comprehension and uncover patterns or insights that might not be as readily apparent in more traditional chart forms, as it represents many quantities in one graphic.

Additionally, it encompasses multidimensionality by representing various aspects of the data simultaneously. It not only separates the entire survey sample into three categories based on the frequency of husbands' drinking but also further categorizes the weighted survey samples based on different forms of intimate partner violence (IPV) experienced by respondents. It therefore can represent multiple quantities in one chart: the proportion of married men in Ukraine who drink 'never,' 'sometimes,' or 'often'; the proportion of married women experiencing IPV in total; the proportion of women experiencing different (physical, sexual, emotional) forms of IPV; and the proportion of women in each category of their husband's drinking who experience IPV and which forms. As a result, visualization offers a comprehensive and efficient representation of the complex relationship between alcohol consumption and intimate partner violence.

Given its abstraction, originality, and multidimensionality, "Men's Alcohol Consumption and Women's Experiences of Intimate Partner Violence in Ukraine" serves as a valuable tool for academic audiences to understand the prevalence of intimate partner violence (IPV) and its connection to alcohol consumption in Ukraine. In understanding and catering to the audienceship, the visualization provides a comprehensive examination of IPV and the role of men's alcohol consumption as a risk factor.

Men’s Alcohol Consumption and Women’s Experiences of Intimate Partner Violence in Ukraine

April 25, 2023

Allison Towey
The University of Chicago

Abstract (150)
Intimate Partner Violence (IPV) is a matter of concern across the globe, especially in Ukraine, where a fourth of ever-married women report having experienced some type of emotional, physical, or sexual violence perpetrated by a partner or husband. Leveraging data from the 2007 Demographic and Health Survey, this visualization summarizes women's experiences of physical, sexual, and emotional IPV according to their husbands' alcohol consumption. The proportionally drawn square matrix chart depicts the prevalence of IPV in Ukraine, by type, and confirms the strong association between alcohol consumption and IPV found in previous research. This study adds to an existing literature on the impact of alcohol on IPV perpetration and suggests a need for targeted interventions that address alcohol use in intimate partnerships. Read today in the current context of armed conflict in Ukraine, we are reminded of the urgent need for reliable and recent data on health and family life.

Keywords (3)
intimate partner violence, alcohol consumption, gender

Text (517)

Intimate Partner Violence (IPV) is a matter of great concern across the globe, especially in Ukraine, where a fourth of ever-married women report having experienced some type of emotional, physical, or sexual violence perpetrated by a partner or husband. Additionally, Ukraine is characterized by elevated levels of alcohol use and alcoholism in comparison to other European countries, particularly among men (Bromet et al. 2005). This is concerning given that men’s alcohol use has been identified as the strongest risk factor associated with IPV perpetration in Ukraine and across former Soviet states (Barrett et al. 2012, Ismayilova 2015).

This figure leverages data from the most recent Demographic and Health Survey fielded in Ukraine; effectively, I summarize the experiences of currently-married Ukrainian women in 2007 according to their husbands’ levels of alcohol consumption. The measure of violence distinguishes women’s cumulative experience of physical, sexual, and emotional intimate partner violence.

The matrix chart is partitioned proportionally, according to women's reports of their husbands' drinking habits: never, sometimes, or often. Each square of the matrix represents 0.1% of the total weighted sample of currently-married Ukrainian women, for a total of 1000 squares. Within each group, the proportion of women who report intimate partner violence are denoted by the color of the square: those having experienced physical violence are denoted in yellow, emotional violence in blue, and sexual violence in red. Because women can and do experience multiple types of IPV, combinations of violence are represented by multi-colored squares.

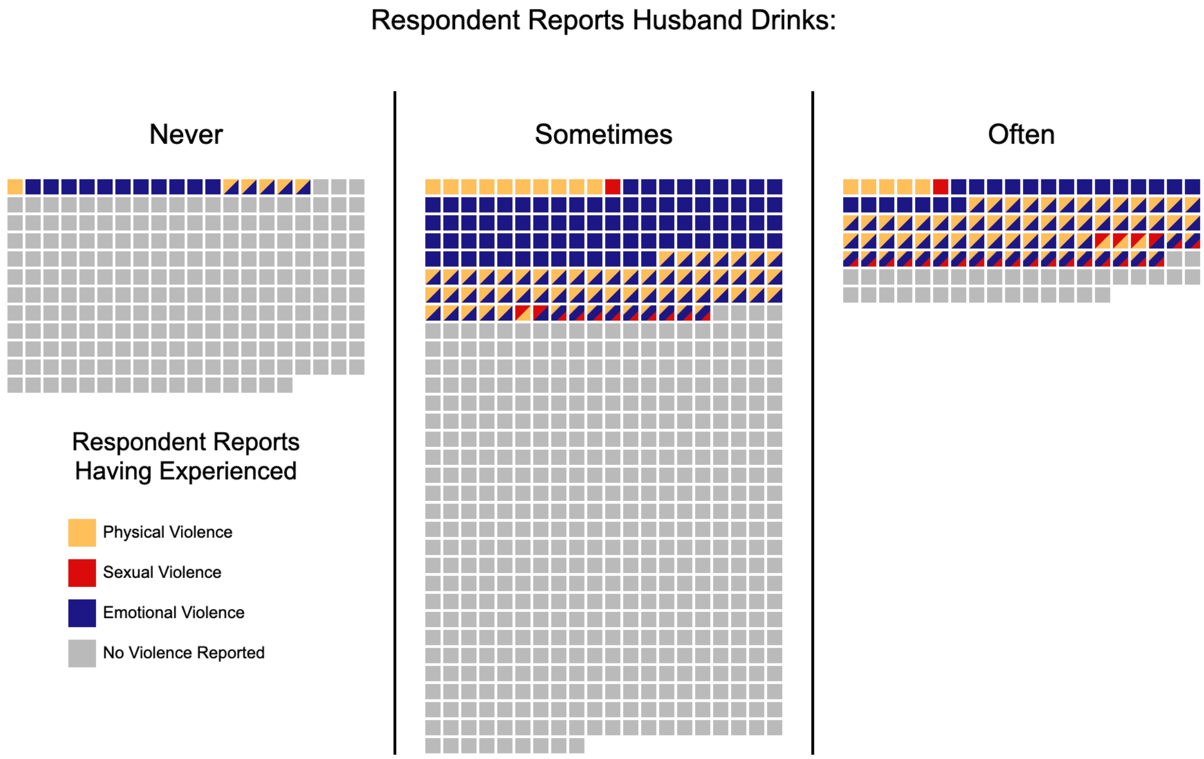
Overall, this visualization allows readers to 1) compare frequencies of how often women report their husbands drink, 2) compare frequencies of the forms of IPV women experience, and, importantly, 3) comprehend the relative proportion of IPV experiences in each alcohol-consumption group. For women with abstemious husbands, 7% report having experienced at least one form of IPV, with none of these women reporting sexual violence. Among women whose husbands sometimes drink, nearly a quarter (24.5%) experienced some form of violence. Nearly three-fourths (72.6%) of women whose husbands often drink report having experienced some form of IPV. Moreover, women with husbands who drink "often" experience extremely high levels of violence; with over one third (36.2%) reporting IPV experiences, even though they make up just 13% of the sample. The most prevalent form of IPV is emotional violence (24.9%), followed by physical (15.3%), and sexual (3.7%) violence.

Confirming the strong relationship between alcohol consumption and intimate partner violence shown in previous epidemiological studies (e.g. Foran and O'Leary 2008), this visualization underscores the prevalence of IPV in Ukraine and the role of alcohol as a risk factor in intimate partnerships. The current conditions of conflict in Ukraine have disrupted daily life, including the collection of data on health and familial relationships. Research from other conflict settings suggests that the stress of armed conflict exacerbates IPV, among other social problems (Annan and Brier, 2010). As the consequences of war are cataloged and examined by scholars and humanitarian workers, the prevalence and patterning of IPV among Ukrainian families should be counted among the many urgent questions that need to be answered.

References:

1. Annan, J., & Brier, M. (2010). The risk of return: intimate partner violence in Northern Uganda's armed conflict. *Social science & medicine*, *70*(1), 152-159.
2. Barrett BJ, Habibov N, Chernyak E. Factors affecting prevalence and extent of intimate partner violence in Ukraine: evidence from a nationally representative survey. Violence Against Women. 2012;18(10):1147–1176
3. Bromet, EJ., et al. Epidemiology of psychiatric and alcohol disorders in Ukraine. Social psychiatry and psychiatric epidemiology 40.9 (2005): 681-690.
4. Foran HM, O'Leary KD. Alcohol and intimate partner violence: a meta-analytic review. Clin Psychol Rev. 2008 Oct;28(7):1222-34.
5. Ismayilova, L. Spousal violence in 5 transitional countries: a population-based multilevel analysis of individual and contextual factors. American Journal of Public Health 105.11 (2015): e12-e22.

Figure 1: Ever-married Ukrainian Women's Experience of Intimate Partner Violence, Given How Often Her Husband Drinks, 2007



Source: Demographic and Health Surveys Program, Ukraine 2007
Note: This square matrix chart depicts the proportion of ever-married Ukrainian women who report having ever experienced intimate partner violence (IPV); the matrix is partitioned into three panels according to women's reports of how often their husbands drink: 'never', 'sometimes', or 'often.' Squares are colored to denote the proportion of women who have experienced physical, sexual, or emotional violence, or any combination thereof. Each square represents 0.1% of the total weighted sample of Ukrainian women from the 2007 Demographic and Health Survey. The matrix is composed of 1000 squares.

Visualization 2: Global Access to Safe Drinking Water

Overview and Methodology

Link: https://atowey-uchi.github.io/data_visualization/drinking_water/index.html

Abstract:
This data visualization infographic explores the issue of access to safe drinking water, a fundamental need and human right that remains unfulfilled for billions of people around the world. The visualization highlights the current state of global access to safe drinking water and sanitation services, as well as the progress that has been made in recent years. By presenting data on preventable illness and death rates, particularly among children, the visualization underscores the urgent need to address this ongoing crisis. Through a combination of maps, charts, and graphs, this visualization provides a comprehensive overview of the challenges facing communities without access to clean, easily accessible water, and the potential benefits of addressing this critical issue.

Methodology:

This data visualization is crafted with the intention of engaging both technical and non-technical audiences including policymakers and the general public, shedding light on the urgent matter of access to safe drinking water—a fundamental human right still unmet by billions worldwide. By presenting compelling data on preventable illnesses and fatalities, particularly among children, the visualization underscores the urgent need to address this ongoing crisis, calling the audience to action. Through the thoughtful integration of maps, line graphs, and other charts, this visualization offers a comprehensive and informative overview of the challenges faced by communities lacking access to clean and easily obtainable water, emphasizing the potential benefits of addressing this critical issue.

This visualization prioritizes clarity and visual appeal to attract viewers and concisely community why they should engage with the material. As opposed to a more academic or technical visualization, the goal is to persuade and engage the audience, thus inspiring the incorporation of more decoration and lightness from Alberto Cairo's Visualization Wheel. The use of large, bolded text in conjunction with colorful charts and well-designed graphics intends to draw attention and maintain focus of the audience.

The infographic employs a range of graphs that are both unidimensional and multidimensional, creating visual interest and appealing to a range of viewers. Some of the graphs, particularly the first graph that uses a series of proportional matrix charts to appear like water glasses, prioritize stylistic visual appeal, while others, such as the multidimensional scatterplot, prioritize depth and complexity of information. By incorporating a variety of chart types, the visualization caters to different ways of understanding and interpreting data, accommodating a broader audience.

Clarity is prioritized through the use of minimalistic, yet carefully designed chart styles. By adopting a consistent theme and color palette throughout the visualization, a cohesive visual identity is established, promoting a sense of unity. Consistency in data formatting and color usage helps convey the information clearly and avoids confusion. Clear and succinct titles are used for each chart, providing a quick and informative summary of what the chart represents. Bold formatting helps draw attention to the titles, making them easily identifiable. This approach allows viewers, including those who may be scanning the content, to quickly grasp the focus of each chart. By prioritizing clarity through simple and well-designed chart styles, consistent themes and color palettes, and clear titles, this infographic aims to facilitate understanding and engagement with the information presented.

Accuracy is exemplified by the careful choice of axes and dutiful representation of the underlying data. The choice of axes in the charts is important for accurately presenting the data. All of the graphs have an absolute zero point, which aids in contextualizing the data. The scales and intervals on the axes are selected to appropriately represent the range and magnitude of the data points. This ensures that the data is not distorted or misrepresented, allowing viewers to interpret the information accurately. Dutiful representation of the data includes accurately plotting data points, using precise measurements, and avoiding misleading visual techniques. Data points are plotted with precision, and the chart types are chosen to provide an accurate representation of the underlying data. For instance, line graphs are used to characterize temporal data. This type of chart is particularly useful for visualizing trends, identifying patterns, and understanding the relationship between variables over time. Accuracy is also promoted by the citations of data on the infographic itself, allowing for the audience to view the data for themselves and perform their own data analysis or replication.

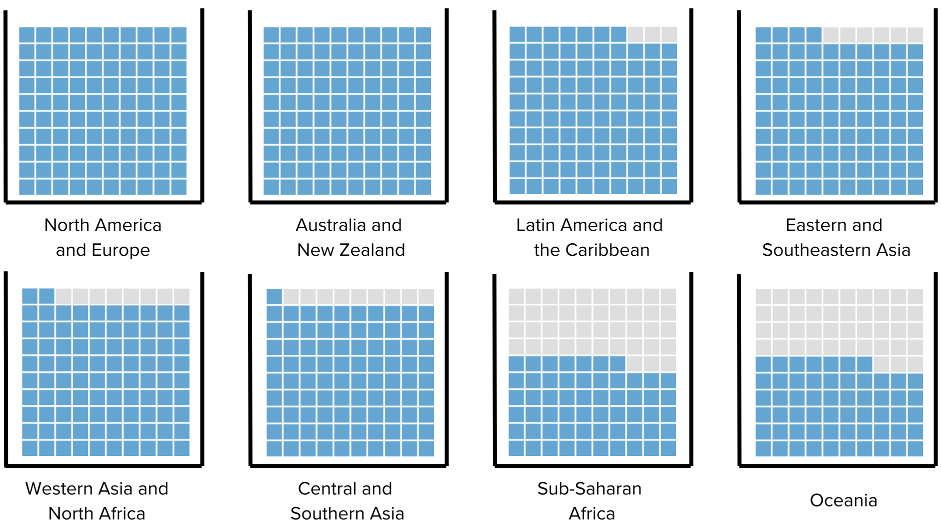
This infographic is distinct from more academic or technical visualizations because of its emphasis on design appeal. While Edward Tufte advocates for simplicity and minimalism, emphasizing the importance of conveying information effectively, other scholars like Bateman et al. (2013) argue that well-designed decorative elements can enhance engagement and retention of information. Consistency in the theme, color palette, and overall style helps create a cohesive and visually pleasing experience, contributing to a positive perception of the content. The thoughtful layout of text and charts plays a crucial role in maintaining visual appeal. Clear and concise titles, labels, and visibly sized text guide the viewers' attention and ensure information is easily understandable. The use of well-designed alignment helps create a balanced and visually pleasing composition. The inclusion of attractive decorative elements, such as the bolded, colorful title, serves to captivate the audience's attention and make the information more memorable. These elements can evoke emotions and create a connection with the viewers, enhancing their engagement with the material and increasing the likelihood of taking action. By considering the balance between simplicity and decoration, this infographic aims to engage the audience, communicate the message effectively, and encourage action.

GLOBAL ACCESS TO SAFE DRINKING WATER

Universal access to safe drinking water is a fundamental need and human right. Securing access yo clean, easily accessible water for all could have a major role in reducing preventable illness and death, particularly children. Though progress has been made to provide safe drinking water and sanitation to people throughout the world, billions of people still lack access to these services every day.

One-in-four people do not have access to clean drinking water.

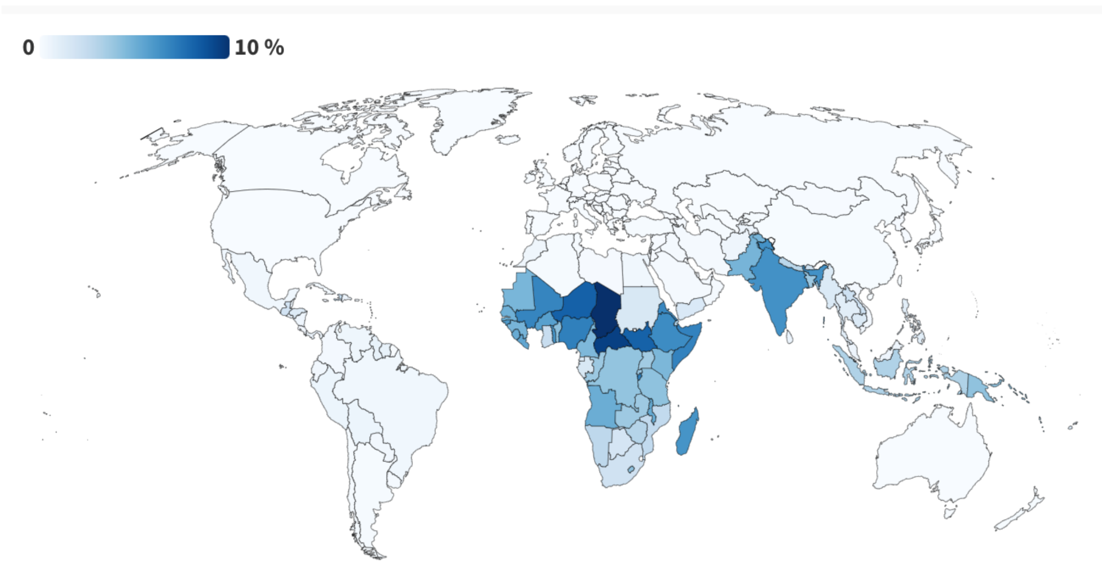
As of 2020, 2 billion people globally lack access to safely managed drinking water at home, with almost 800 million of those lacking even basic drinking water service.[1]



Share of households with at least basic level of drinking water service, 2020

Basic drinking water service means access to drinking water from an improved source such as a piped household water connection or public well, in which provided collection time is not more than 30 minutes for a round trip, including lining up and waiting.[2] There are considerable disparities in such access; while many developed countries have now achieved universal access, coverage with safely managed drinking water sources varies widely in developing regions.

Unsafe water sources are responsible for 1.2 million deaths each year.



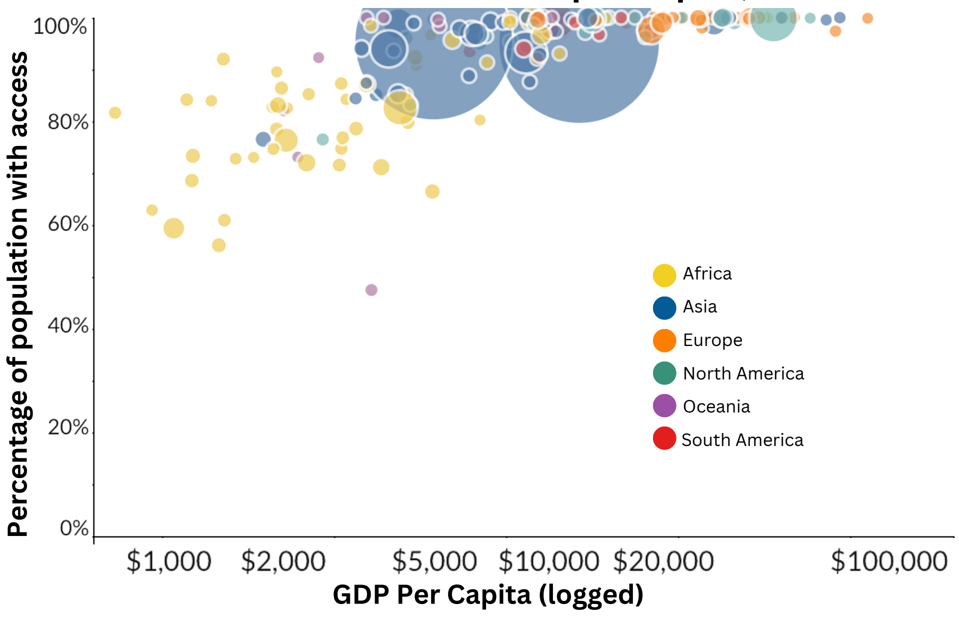
Percentage of deaths attributed to unsafe drinking water by country, 2020

Diseases from dirty water kill more people every year than all forms of violence, with 43% of those deaths are children under five years old.[1] Lack of access to safe water sources is a leading risk factor for infectious diseases, including cholera, diarrhea, dysentery, hepatitis A, typhoid and polio.[3] It also exacerbates malnutrition and childhood growth stunting.[4]

In 2017, the estimated 1.2 million people who died as a result of unsafe water sources represented 2.2% of global deaths. In low-income countries, it accounted for 6% of deaths.[1] According to the World Bank, it is estimated that universal access to safe drinking water, adequate sanitation, and hygiene has the potential to reduce the global disease burden by 10%.[5]

Only 29% of people in low income countries have safely managed water supply, compared to 98% in high income countries.

Share of population with access to improved water sources vs. GDP per capita, 2020

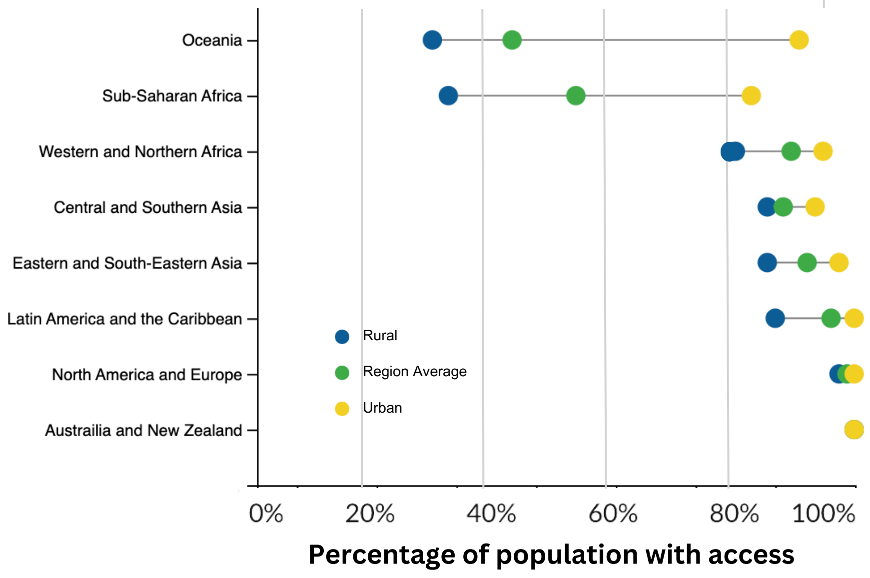


There are large disparities in both access to safely-managed clean water and deaths due to unsafe water between high-income and low-income countries. Safely managed drinking water is water from an improved water source which is located on premises, available when needed and free from contamination.

Most countries with greater than 90% of households with improved water have an average GDP per capita of more than $10,000. We see, however, that there is a range of levels of access within groups of countries with similar GDPs per capita, suggesting that there are other factors, such as population structure, natural location, governance and infrastructure, which may be contributing to the disparities.

8 out 10 people who continue to lack basic drinking water services live in rural areas.

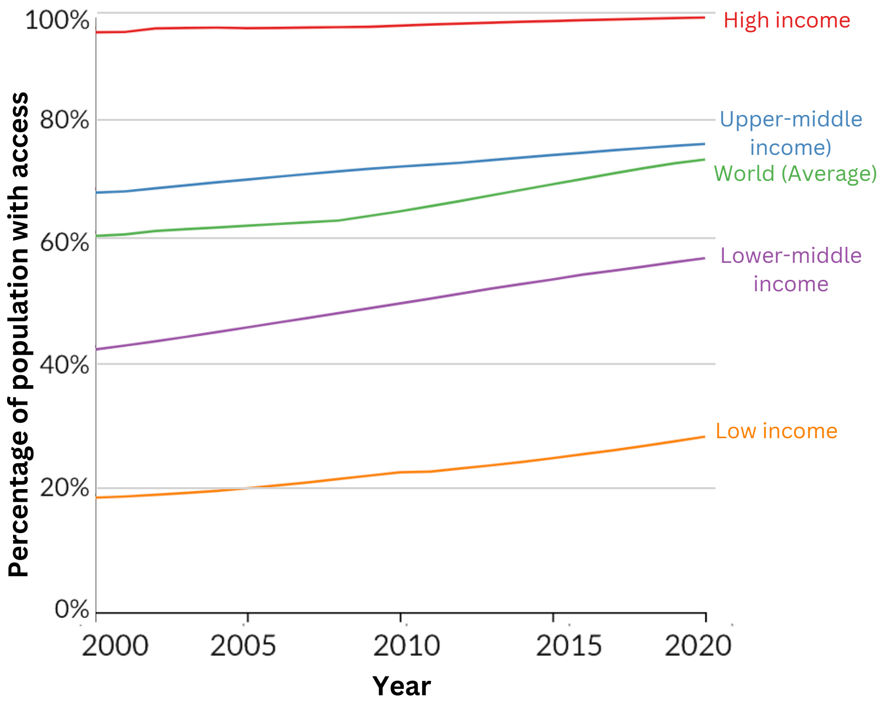
Share of population with access to at least basic water services per region, 2020



In addition to disparities across counties, there are also differences in access between rural and urban populations within countries. In 2020, 97 percent of the world's urban population had access to at least basic drinking water services, compared to just 82 percent of the rural population. This difference is even more stark for access to safely-managed drinking services, where 86% of urban populations had access, compared to 60% of the rural populations. While some of these differences may be due to urbanization and disproportionate economic growth in urban versus rural settings, infrastructure and natural boundaries also may play a role.

From 2015 to 2020, we saw an increase of 4% (70 to 74%) in global access to safely managed drinking water access.

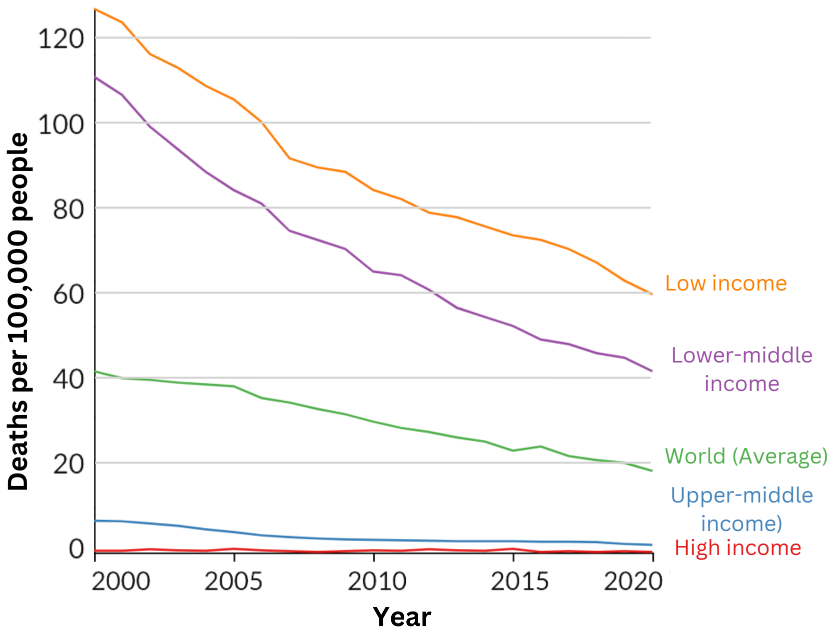
Share of population with access to safely-managed drinking water by World Bank Economic Classification, 2000-2020



We've seen many improvements over the past few decades in both an increase in access to safe drinking water and a reduction in the deaths attributed to unsafe water. These improvements have affected the lives of millions. Between 2015 and 2020, 107 million people gained access to safely managed drinking water at home, and 115 million people gained access to safe toilets at home.

The United Nations Sustainable Development Goal Target 6.1 is to “achieve universal and equitable access to safe and affordable drinking water for all” by 2030. When the SDGs began in 2015, 70% of the global population had safe drinking of the global population had safe drinking water. In 2020, five years later, we've seen an increase of four percentage to 74%. At the current rate, however, the SDG target *will not* be reached. Further work and monitoring is necessary.

Deaths per 100,000 people attributed to unsafe drinking water (age-standardized), 2000-2020



Achieving global universal access to safe water will be especially challenging for the 41 countries where over one-fifth of the population used unimproved drinking water sources in 2015, which are largely concentrated in sub-Saharan Africa. Special emphasis in these countries and the global community writ-large should be placed on rural and impoverished populations, as significant disparities fall upon these lines. While progress has undoubtedly been made, additional work to reach the most vulnerable is of great importance.

Allison Towey,
The University of Chicago

March 2023

Sources

[1] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
[2] WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP)
[3] World Health Organization. Guidelines for drinking-water quality: second addendum. Vol. 1, Recommendations. World Health Organization, 2008.
[4] World Health Organization. WHO (2019) – Fact sheet – Sanitation. Updated June 2019.
[5] World Health Organization. Global costs and benefits of drinking water supply and sanitation interventions to reach the MDG target and universal coverage. Geneva: World Health Organization (WHO), 2012.

Data from: The World Bank, World Health Organization/UNICEF's JMP, and the Institute for Health Metrics. For more information on the data, please visit https://github.com/atowey-uchi/data_visualization/tree/main/drinking_water

Visualization 3: Covid-19 and Transportation: How the Pandemic Affected Where We Went and How We Got There

Overview and Methodology

Link:
https://atowey-uchi.github.io/data_visualization/covid-transpo/code%20%2B%20data/index.html

Abstract:
This data visualization article explores the enduring effects of the COVID-19 pandemic on transportation patterns in the United States. Through the use of maps, charts, and graphs, it delves into the changes observed in how people travel as a result of the pandemic. The article highlights the significant disruptions caused by the pandemic, such as the widespread adoption of remote work and virtual learning. These changes led to short and long-term shifts in transportation behaviors overall, with some regional disparities such as the Southeast exhibiting higher mobility rates throughout the pandemic compared to other regions. The article highlights April 2020 as a period when mobility rates hit their lowest point; this decline can be attributed to various factors, including school closures, the implementation of telework arrangements, and the temporary shutdowns of restaurants and bars. Although travel has shown signs of recovery since the early stages of the pandemic, it has not yet reached pre-pandemic levels, with a particularly large effect on public transportation and aviation. Throughout the article, the data is presented in a way that is accessible to a general audience, enabling readers to understand the lasting changes in transportation patterns brought about by the COVID-19 pandemic in the United States.

Methodology:

This data visualization article adopts the format of a data journalism piece tailored to a general audience with limited familiarity with the data being presented. The article's emphasis on simplicity, clarity, interactivity, and accuracy works to ensure effective communication of its message. The article strives to offer valuable insights about the influence of the COVID-19 pandemic on transportation patterns in the United States for the purpose of general knowledge.

To achieve simplicity, the article utilizes well-known, unidimensional chart types: choropleth maps, multi-line line graphs, and an area chart. These familiar chart types allow for the general audience to comprehend the information being presented without undue cognitive load. This approach aligns with the principles outlined in Alberto Cairo's Visualization Wheel, which highlights the importance of using familiar, unidimensional visual representations to enhance intelligibility, particularly when targeting a broad audience. (Cairo 2011) In addition to using familiar chart types, the article ensures simplicity by organizing the text and graphics in a logical and understandable manner. The flow of information is carefully structured to guide the audience through the key findings and implications of the data. By weaving the text and graphical representations in a clear and coherent way, the article helps the general audience navigate the complexities of the data and grasp the main insights being communicated. Likewise, the minimalist style of the charts contributes to achieving the goal of Tufte's high data-ink ratio and simplicity. (Tufte 1997) By limiting the use of decorative elements and focusing on essential information, the charts maintain a clean and uncluttered appearance, which aids in emphasizing the data itself. The limited use of text in the graphics, primarily reserved for labels and annotations, enhances clarity by providing concise and relevant information.

In addition to simplicity, clarity is a crucial aspect of data visualization emphasized in this article. By utilizing multi-modal data representations and implementing tooltips for specificity, the charts enhance clarity by providing additional context and details when needed. The consistent style, color palette, and data format throughout the article contribute to clarity by creating a cohesive visual language that facilitates easy comprehension and comparison of data across different charts.

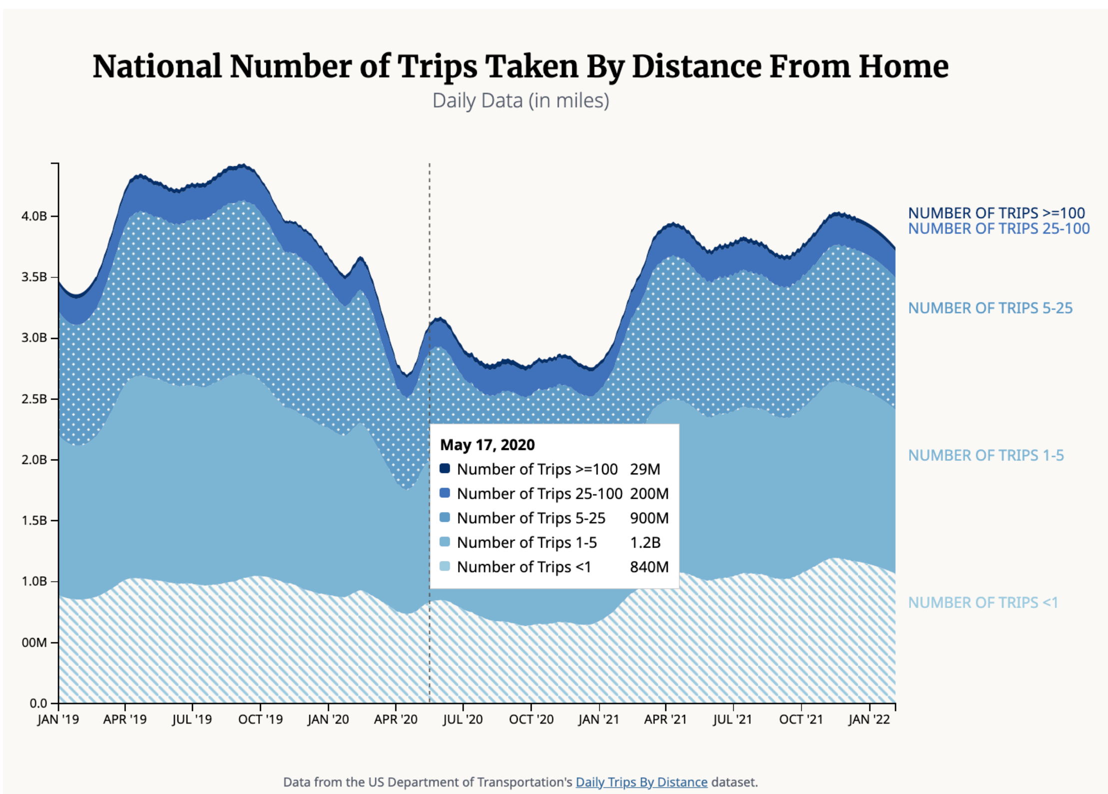


Figure 1: Shades of the same color (blue) are used to denote discrete values of numerical data. Additionally, a tooltip allows for more specificity and user interaction.

The choice of colors and patterns in data visualization plays a crucial role in enhancing clarity and accessibility. By using a consistent color palette and applying shades of the same hue to represent discrete numerical data, the visualization achieves clarity in communicating the values of the Number of Trips Taken By Distance From Home chart. The gradual transition from light to dark blue effectively conveys the values of the discrete numerical data. In the case of the Percent Change by Transportation Type chart, different colors are utilized to differentiate between categorical variables. This color differentiation helps viewers distinguish between the lines representing different transportation types. The inclusion of line patterns further aids in enhancing clarity and accessibility, as they are an additional visual cue to differentiate between categories. By incorporating patterns alongside colors, the visualization becomes more accessible to individuals with visual disabilities. These patterns provide an alternative means of distinguishing data points or categories, ensuring that the information is effectively communicated to a wider range of viewers.

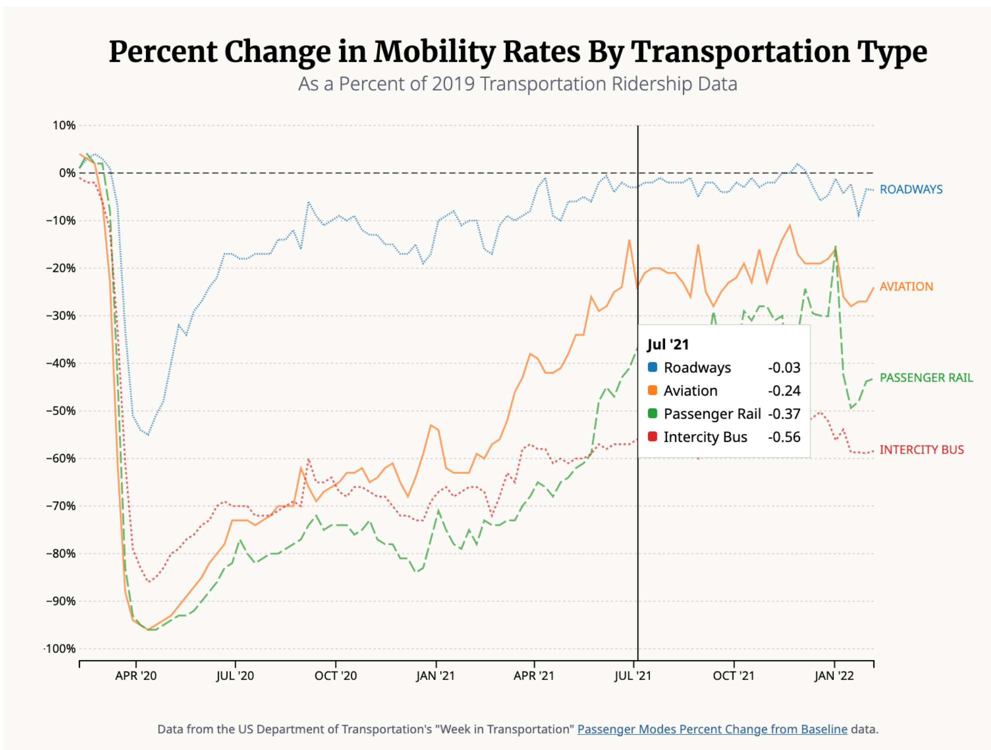


Figure 2: Different colors are used to denote categorical variables. The chart also employs a tooltip, allowing for user exploration of specific dates.

Clarity is also promoted in the inclusion of interactive elements, which empowers users to actively engage with the information. These interactive features enable users to manipulate the visualization based on their specific interests and explore different facets of the data. State filtering, for example, allows users to select and focus on specific states of interest, providing a more localized view of transportation patterns. Similarly, date selection functionality enables users to explore the data over different time periods, facilitating the analysis of temporal patterns and trends. This interactive feature allows users to investigate how transportation patterns have evolved over time, identify significant events or changes, and make comparisons between different periods. The introduction of these interactive elements not only enhances clarity but also promotes a more personalized and engaging user experience. This interactivity encourages a deeper level of understanding and exploration, enabling users to grasp the complex relationships within the data more effectively.

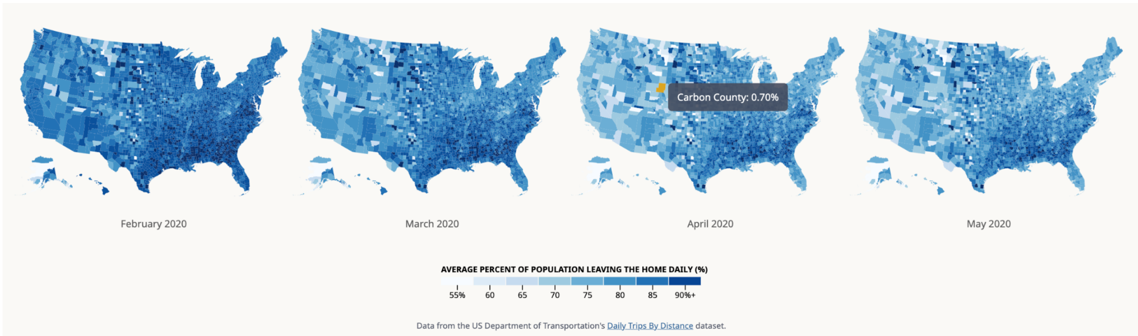


Figure 3: Interactive map elements allow the users to explore the data for particular counties at their discretion.

Moreover, the article prioritizes accuracy, ensuring that the data presented is reliably represented. The information is thoroughly vetted to minimize errors and inaccuracies, enhancing the credibility of the article's findings. This commitment to accuracy helps to build trust with the audience and reinforces the article's value as a reliable source of information. The data sources used are all public information produced by the United States Federal Government, enhancing the reliability of the article. The public sources of the data used are linked in the charts themselves so as to allow for individual exploration of the data and replication of the charts.

The use of familiar and unidimensional graphs in the data visualization portfolio aims to make the article accessible and understandable for a general audience. The charts, along with the accompanying text, strive to present the patterns and trends of transportation in a clear, concise, visually appealing, and easily comprehensible manner. The synergy between the visual and textual components ensures that the information is communicated in a comprehensive and easily digestible manner for general audiences, striving for more intelligible graphical representation.

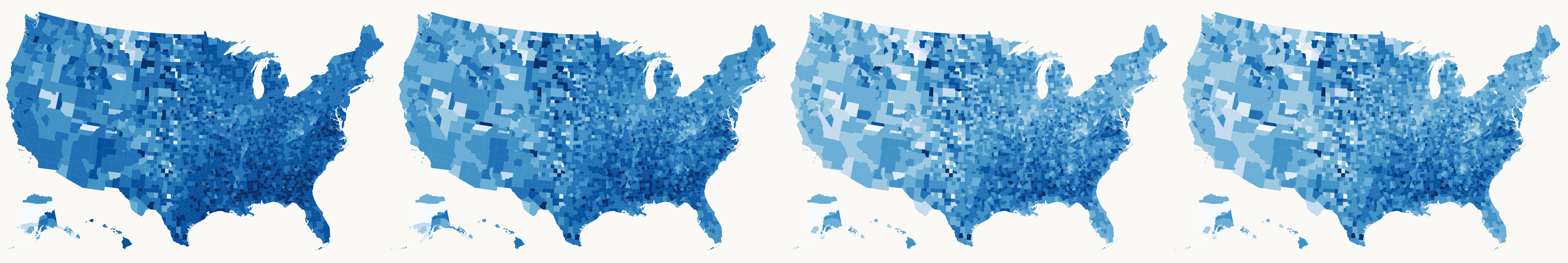
COVID-19 AND TRANSPORTATION:

How the Pandemic Affected Where We Went and How We Got There

A look into how the COVID-19 pandemic impacted—and, in some ways, still impacts—transportation and mobility patterns across the United States.

BY ALLISON TOWEY

The University of Chicago



February 2020 March 2020 April 2020 May 2020

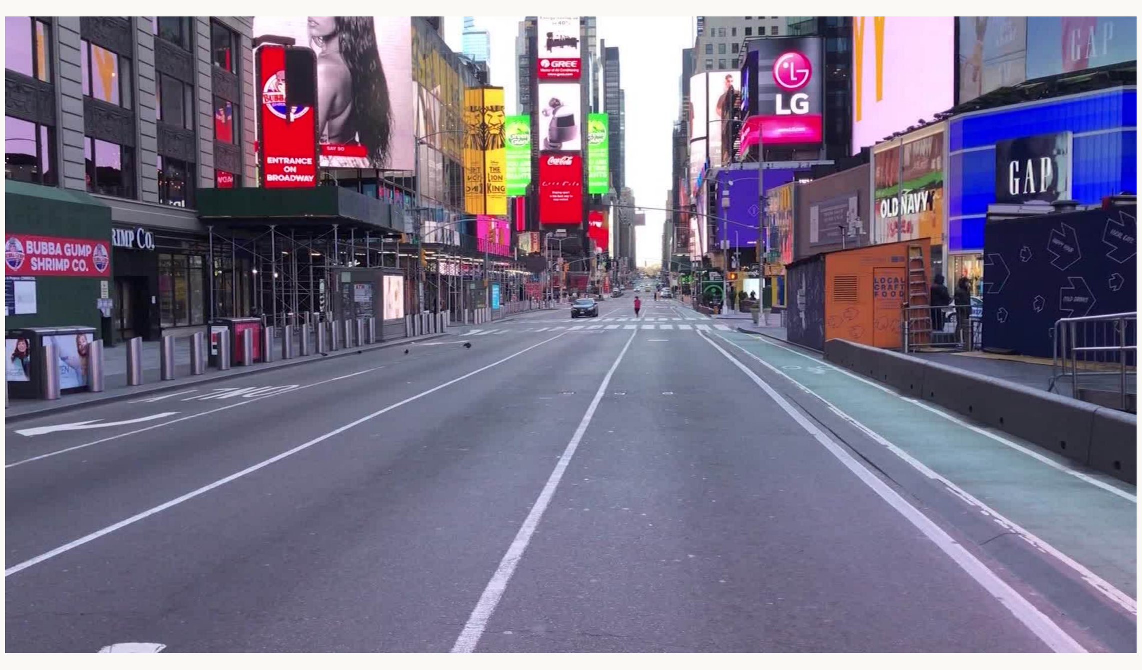
AVERAGE PERCENT OF POPULATION LEAVING THE HOME DAILY (%)

55% 60 65 70 75 80 85 90%+

Data from the US Department of Transportation's Daily Trips By Distance dataset.

Undoubtedly, the coronavirus pandemic has caused major, and at times lasting, effects on our lives. From how we work, to where we go to school, to how we shop, to how we socialize, our habits and routines have adapted to a new normal of increased remote access opportunities. The rise of telework, teleschool, online shopping, and video-chatting have decreased our need to leave our homes. In the immediate and uncertain days of the pandemic, these changes allowed for us to continue with our lives without in-person connection, slowing the transmission of the disease in the hopes of saving lives.

With the worst days of the pandemic now (hopefully) in our rearview mirror, this "new normal" has had some lasting effects on our routines. In looking at American transportation patterns, we can see the effects of the pandemic on how often, how far, and with which transportation methods we travel, from the early days of the pandemic to now, nearly two years later.



An empty scene in Times Square in New York City, April 2, 2020. Source: CNN

The pandemic's effects on travel varied by location.

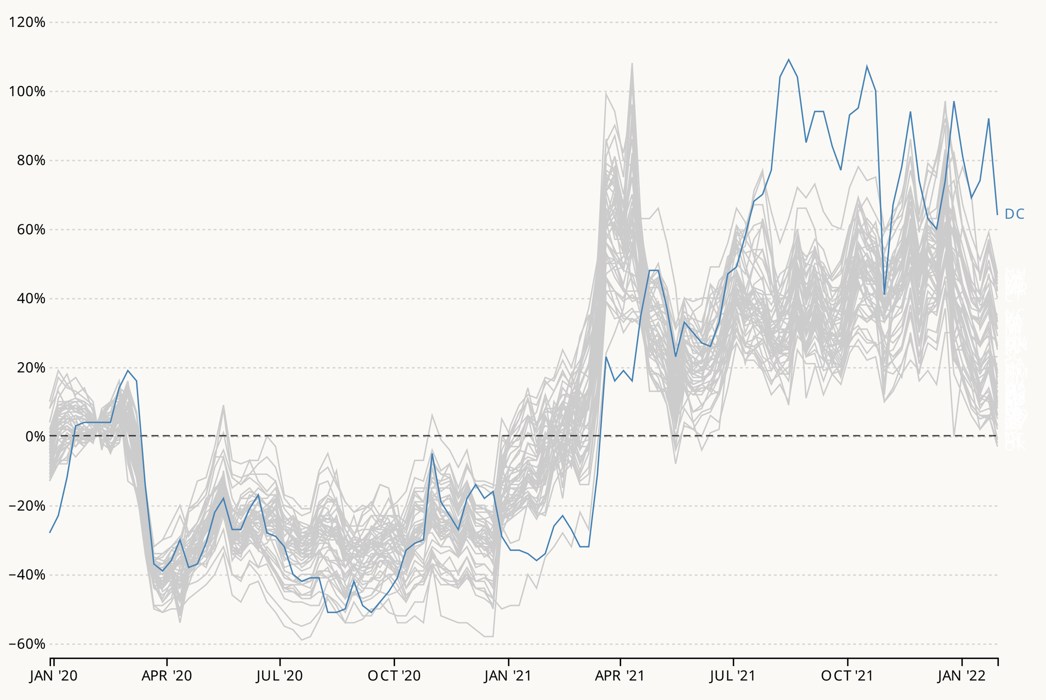
As shown in the maps above, regional differences in the United States impact travel patterns. The Southeast, both before and during the early days of the pandemic, has a higher mobility rate, or more people, on average, leave their homes on a given day.

Across the country, however, travel decreased significantly in April as the pandemic spread through the country and lockdowns took effect. April had the lowest rate of people leaving their homes. Nationally, April 2, 2020 had a mobility rate of just over 63%. For context, the rate on April 2, 2019, just a year prior, was 89%. Such a change in the amount of travel in the United States likely has a variety of causes: school cancellations, telework opportunities, restaurant and bar closures, and more reduced the need for Americans to risk their health and leave home.

Percent Change in Mobility Rates By State

As a Percent of 2019 Mobility Data

Selected state: DC



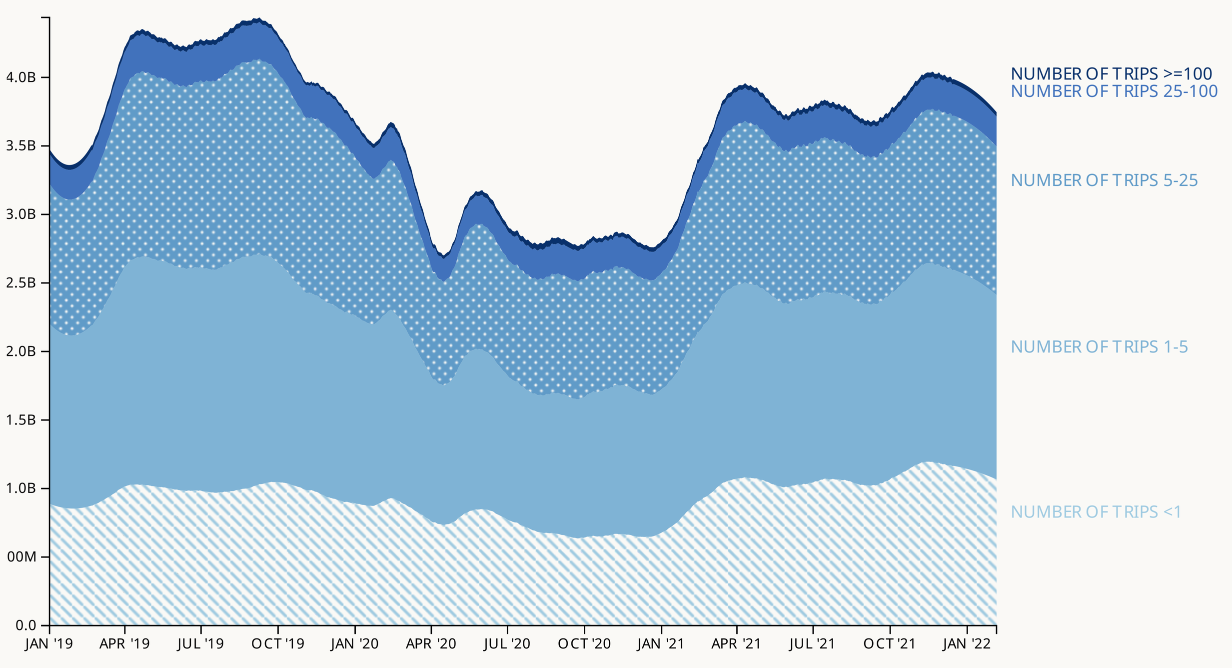
Data from the US Department of Transportation's Daily Trips By Distance dataset.

Interestingly, however, after the initial few months of the pandemic, the second half of 2020 and beginning of 2021 saw a resurgence of people traveling outside of the home. In fact, the spring and summer months of 2021 had a higher rate of travel year-over-year, showing a return, at least in part, to travel. Each state had a different path of mobility patterns through the pandemic, which different state laws permitting varying levels of travel. To explore these trends, hover over various states above.

The pandemic not only affected if people would leave their homes, but also how far they traveled; in fact, local travel of less than 1 mile, was only somewhat decreased during the pandemic. Further travel was more severely impacted by the pandemic. Flight and travel cancellations undoubtedly had a role in the lower number of long trips taken. In 2020, over 4,000 flights, constituting over 1% of total flights, were cancelled entirely, with another 6% rescheduled to a later time. (Source: The Federal Aviation Administration)

National Number of Trips Taken By Distance From Home

Daily Data (in miles)



Data from the US Department of Transportation's Daily Trips By Distance dataset.

While travel, both near and far, has rebounded since the early days of the pandemic in 2020, travel is still below its 2019 numbers. Though we are only a couple years removed, it appears that there has been a longer-term effect of the pandemic on our travel patterns, even as lockdowns, flight cancellations, and public health concerns have subsided.

Bus drivers in masks in Chicago, IL, April 2020. Source: Active Transportation Alliance

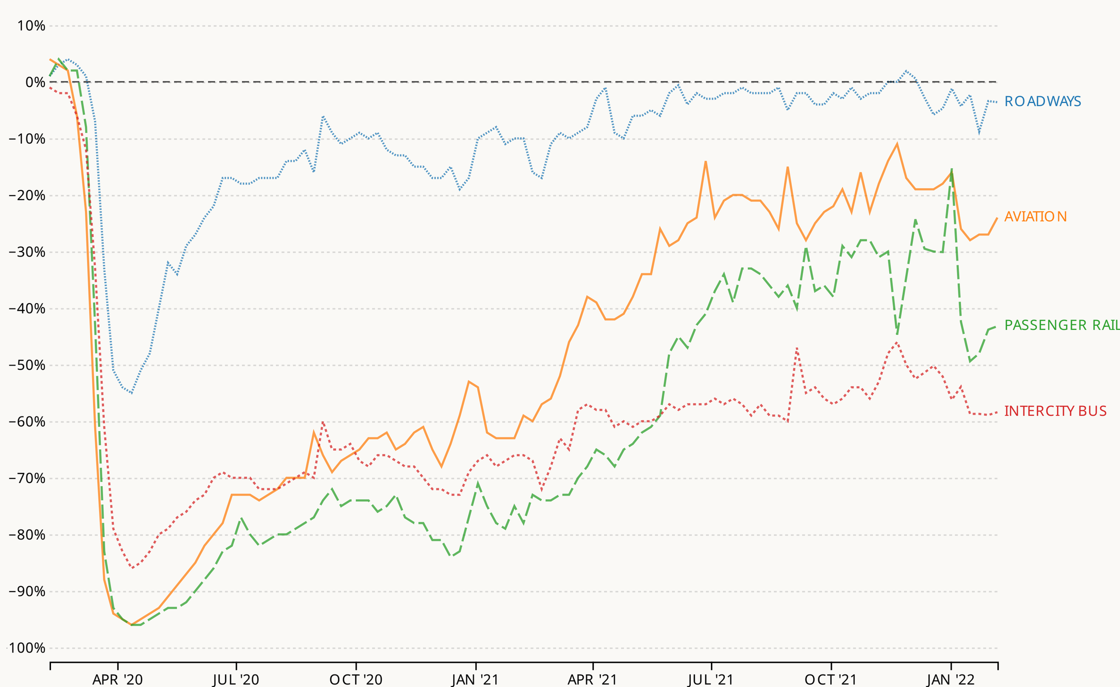
While car traffic has resumed, public transportation and aviation remain below their 2019 ridership numbers.

Every day, millions of commuters in the U.S. continue to rely on public forms of transportation, whether that is buses, rail, or aviation. Along with the decreased need to travel due to telework, teleschool, and other remote opportunities on the rise, the COVID-19 pandemic also forced riders to contend with social distancing and other safety measures. To adapt to the spread of the airborne virus, transportation agencies in major cities had to quickly change how they offered services. On planes, buses, and rail lines, social distancing measures, masking requirements, and decreased vehicles in service greatly impacted transportation as we know it.

In fact to this day, bus, rail, and aviation ridership numbers remain below their 2019 rates, with buses racing the most significant decrease in ridership. The only form of transportation near its initial rate (denoted by miles traveled nationally) is automobile traffic. As car usage has rebounded, other, public forms of transportation have failed to recover quickly. It is unknown if such travel will return. As more remote socialization and working options arise, perhaps public transportation's levels will remain at their decreased clip, though only time can tell.

Percent Change in Mobility Rates By Transportation Type

As a Percent of 2019 Transportation Ridership Data



Data from the US Department of Transportation's "Week in Transportation" Passenger Modes Percent Change from Baseline data.

While some pandemic-related changes to our transportation and travel have subsided, longer-term effects, such as decreased public transportation ridership, remain. In our new world of telework and video-conferencing with loved ones, where we go and how we get there remain an evolving process.