Austin Peay State University

Data Visualization

A Key to Human Understanding

Austin Kerr

CSCI 4800

Dr. Jiang Li

22 Apr. 2021

Austin Kerr

Dr. Jiang Li

CSCI 4800

22 April 2021

Data Visualization: A Key to Human Understanding

Data visualization has existed in many forms throughout the period of humans’ existence. Though there can be clear examples of visual data dating back to the stone age, the importance of data visualization and its many uses has been increasing as human society continues to innovate. This fact shines clearly in our modern world with the plethora of absorbable media each individual has access to. Data visualization has presented itself as a tool when humanity finds a new need. Visualizing data greatly helps the communication and understanding of new thoughts and ideas.

This research paper will observe examples of data visualization’s usage throughout history. We will comment on what makes a particular method of visualization effective and some issues faced with data visualization methods. We will describe the growing dependence on visualized data in our modern world. We will also discuss possible modern issues in which data visualization may assist in public understanding and discourse. Observations will be based on data and research collected from an assortment of other scholarly papers, editorials, and research columns.

Data visualization is and always will be a staple need for human innovation. Visualization of data also has its uses in the simple exchange of daily discourse whether it is for scientific, empathetic, or philosophical reasons. Data visualization has been imperative in how humans have excelled throughout time and how they communicate with one another. As such, it will remain a cornerstone in human communication and transformation for all eternity.

Historical references of data visualization date all the way back well into the Stone Age, nearly 40,000 years ago. For quite some time, researchers saw cave paintings and etchings as simply a form of art. Through continued study and even more new findings, researchers have started to realize these markings may have been the beginning of some sort of communication system. This attempt at communication may have been used to communicate hunting routes, warn of dangers in the nearby area, or for groups of people to communicate the stories of their own specific groups. Though much simpler than current examples, these etchings and blotches scrawled across cave walls show us that communication resonates through visual imagery even back in ancient times.



Fig. 1. “Rock paintings from the Cave of Beasts (Gilf Kebir, Libyan Desert) Estimated 7000 BP” By Clemens Schmillen - Own work, *CC BY-SA 3.0*, https://commons.wikimedia.org/w/index.php?curid=31399425

Slightly more recent examples, such as the Egyptians and Maya people, show us a system of writing with hieroglyphs. These civilizations used their visual symbols not just to communicate words and feelings but also to record and relay numerical data. Both the Mayans and Egyptians have an extensive recorded history of their leaders, dates when drought and famine happened, casualty data from great wars, and much more. These examples mirror more closely our current usages for data visualization because they introduce the usage of numerical data, dates, and geographical data. Still used today, maps have been a visualization of geographical data as far back as 2500 BC. Many of these representations have allowed major historical events and the rise and fall of kingdoms be forever remembered and studied.

Ever since the boom of the Information Age, data visualization has been in integral part of nearly everyone’s daily lives. Traffic pattern reports, social media applications, and even the daily weather forecast all involve the use of some sort of multimedia data visualization methods.

Within the past 15 years specifically, data visualization technology has been the main form of communication of topics through mobile devices such as phones, tablets, and laptops. The most prevalent example of mobile devices using data visualization technology within the past two years has been the creation of applications used to track contact tracing throughout the Covid-19 pandemic. Mobile phone tracking and Covid-19 test results data are being used to give users an idea if they have come into contact with someone who may have been exposed to the virus. Most applications using this technology use histograms, charts, and other multimedia data to inform users what steps they can take to minimize risk in their everyday schedule.

The increase in use of mobiles devices is largely due to the ease of adaptation when users are first learning to use the devices. Due to the prevalence of mobile devices, companies, news agencies, scientists, and others wishing to share their data with the world have had to learn how to adapt their data for the mobile platforms. Great efforts have been made to package data into a form accepted by and easily understandable to the common person. A scientist may very well be able to look at a table of values and understand the significance in the variance of the numbers, but that is not the case for everyone. Examples such as this require a different approach to how data is presented. Largely imagery and interactive user interfaces has become the choice media for formatting big data in order to deploy it and inform the general population. As an example, during the 2020 presidential election, Google made an interactive applet allowing users to track election data. Users could look at data on a national, statewide, or county scale. Users also had the ability to separate the data between the major presidential ticket, the house of representatives race, and the senate race. Formulating the data to be easily read and allowing users to view subdivisions of the data is what made that applet a successful example of data visualization deployed to the general public.

Sometimes, the way data is presented can become a persuasive aspect of how the presenter of the data wants consumers to feel about certain aspects of the data. This can be beneficial by highlighting important data points, coloring memorable phrase words in bright colors, or even by making words and phrases about a specific topic larger than the rest of the data. These methods can be useful in focusing the scope of data without removing any of it, but it also presents an issue when misused or used with somewhat devious intent. Those viewing data may fall into a confirmation bias, filtering data until they have a view which resembles wat they may have already been expecting or hoping for. In the same way, distributors of the data may present the data in such a way that draws attention to very specific aspects, creating a bias of its own. The key to simplifying data would be to make sure no data is removed or changed, only simplified to be understood by a wider audience.

A good example of data visualization engages the attention of consumers, but a great example can affect them psychologically and emotionally. A list of data on the amount of average marriages per month of the year could be colored in pinks, whites, and reds which are colors most people associate with romance. Statistics about drunk driving crashes may have warning slogans or images of skulls to warn readers of the dangers of drinking and driving. While the data itself is what is being analyzed, something as simple as a color change or an image can greatly change how a consumer perceives the data they are viewing. If you engage with the emotions of viewers, not only is their attention more likely to be held but they are more likely to remember how the data made them feel. That associated feeling is what could help the remember the data more clearly and the message behind the reason for presenting the data.

We have now discussed what data visualizations must be. Data visualizations must be simplified in some way, will normally have multiple visual or interactive elements, and are normally trying to inform or persuade consumers about a specific topic. To be successful, they cannot be a jumble of data such as would be encountered while conducting experiments. This is where many scientific and research based resources fall short when presenting data. The general public like to watch, play with, and be shown things. Data must be simplified and beautified in order to accommodate the literacy and attention of a wide audience regardless of the topic the data represents.

As an example, let us imagine there is data presented about car sales in the US over the year of 2020. Let us imagine the full data covers new cars, used cars, cars sold by owner, and cars sold by dealers. If the visualization of the data shows only used car sales, viewers may believe that data is representative of all sales for that year, even though dealer sales are not included. Even if all the data is included, if one side of the data is largely highlighted while the other is not, consumers may still make these same false assumptions. In some cases, distributors of data will maliciously leave out or misrepresent data to fit a specific narrative. As data visualization reaches more of the general populace, it will be important for the public to be able to make the distinction between good data and bad data.

There are still multiple uses for data visualization that are only just now being explored or have yet to be touched. Some visualizations are used so much that they have transformed many times into new representations. In ancient times maps were made. Maps became popular with all people as travel became a commodity, so atlases were made to assist travelers in planning their routes. Eventually, with technological advancement, we have technologies such as Google Maps where data is given to visually and auditorily guide users to their destinations. There is even consumable data which allows users to see traffic patterns and congestions rates on roads, presented both online and from many local news sources. As data such as roadway mapping and traffic evolves into what it has today, so too will these untouched data be used and adapted as needed.

Most of the modern advances in data visualization come from tracking data through social media, cellular companies, and internet companies. Cellular traffic in an area can give a lot of insights about the popularity of an area. As mentioned earlier, physical road traffic can be tracked through the cellular traffic on the roads. Companies such as Google, Facebook, and others use internet search data to keep analytics of the popularity of stores, restaurants, tourist attractions, and even conversation topics. Google has the ability to tell users with a visual representation how busy a business will be due to tracked cellular usage in that business tracked over a period of time. Google and Facebook both track internet search history to present advertisements, or data, relevant to what each person has been searching. The usefulness of these tactics is evident, but they do spark concern about the fine line between data tracking for the greater good and at what point it becomes an invasion of privacy. The most prominent solution has been to make these tracking services toggleable as to if they will function for each individual user.

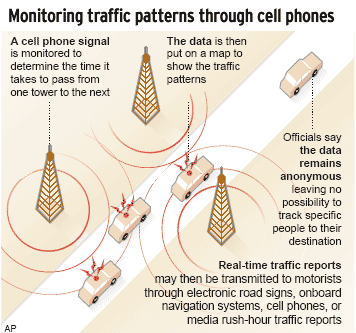


Fig. 2. Associated Press. “Untitled.” *NBC News,* https://www.nbcnews.com/id/wbna9698139

Let us look at possibly one of the most influential data visualizations of our modern era. With relative ease, any person with access to the internet can look up the values and live changes every day for the American Stock Market. While many may not participate in the buying and selling of stocks, the changes in the values of the market can sometimes affect the economy of the entire world, thereby affecting the lives of every individual in it. Many sites and providers, Google included, have graphs for each market or individual stock showing data as it changes. Normally, data can be viewed from as short as one day of change all the way back to the change of values from the data’s inception. Many Americans who do not participate in trading stocks still view or at least know some of the changes in the daily presentation of this data simply because of the impact the changes ay have on their lives. Some of the most common changes are with natural gas and oil companies, as changes with these companies can greatly affect the price for fuel, electricity, and other amenities Americans have come accustomed to having daily. The way this data is represented is extremely important because any misrepresentation could have dire consequences. If oil has gone down a small percentage in the last two hours, yet it is visualized as if it is dropping rapidly, the public may go out and panic buy gasoline. A panic buy such as this could cause prices to skyrocket and cause shortages much like what was seen during the Vietnam War era. With that realization, the way data is portrayed has a level of responsibility to be sure the data is presented in a genuine manner, not easily able to be misinterpreted.

An area of discourse which may benefit from more readily available data visualization is the discussion about firearms and their legality for qualified American citizens to own them.

Related more towards the fields of Science, Technology, Engineering, and Mathematics (STEM) there are multiple tools which analysts can use to format and visualize data. In my own personal experience, I have used simple programs such as Excel and more complex programs such as MatLab and LaTeX. These programs, and others just like them, are useful for those working in data sciences not only for the intent of organizing their own data but also for formatting data in a manner more pleasing to the general audience they may plan to have. One example of the uses of MatLab was for the visualization of a flight from an instrument cluster that had been attached to a weather balloon. Data was able to be saved directly into MatLab and was then put through a series of formulas to extrapolate the data into a image of a flight path. That image was then presented in the program in the form of an interactive three-dimensional graph. From this image, students working with this data at the time were able to identify the points in the flight path where the balloon stopped gaining altitude and the point at which parachutes deployed from the instrument cluster.

LaTeX can be used for similar interactions, but it is on of the programs of choice for statisticians. Data from many formats can be read directly into LaTeX, and modifications can be made to remove things outside of tolerances for the sake of controlling outliers. LaTeX also has the capabilities to form graphs with the data stored, allowing a multitude of different forms seen in general mathematics such as a bar graph, histograms, box-and-whisker plots, and even three-dimensional density charts. With these programs, there is little that cannot be done to and with the data in order to make it easier for consumers to understand what they are looking at. MatLab and LaTeX are not the only programs used for such modifications, but they are very popular in today’s STEM fields.

I myself am currently using data visualization techniques while designing a web-based application for my senior capstone project. The web application incorporates personal finance applets to allow users to track things like their monthly budget, savings planning goals, and net worth. Through the design process, my team and I have had to make many decisions regarding how we want certain data to be presented. The format of the presentation for output data could make a service highly functional or very confusing if done incorrectly. With this in mind, we have all data clearly labeled so users will not be confused. We also have items organized into a grid pattern. this makes each of the pages of our app much more legible. These techniques may seem small, but ensuring we deploy these techniques helps us make sure the app will function for users in a way beneficial to them.

Data visualization techniques must have certain qualities in order to be effective. Data must be simple enough to understand that a large audience can grasp it, but it most not trade accuracy for the sake of simplicity. Data can be formatted to stand out or be flashy to get and keep and audience’s attention. Imagery and interactivity can be used to allow divisions of large data to be shown in smaller subsections and to allow viewers to feel more engaged with the subject matter of the data. While very useful when used correctly, focusing on specific parts of data can be misused to push a narrative that may not actually be representative of the full data set. Even the format of how the data is presented and images, colors, or words associated with the data may be used to affect how viewers feel about the data and the message trying to be communicated.

Technology is here to stay, and as technology becomes more available to the general public so will data on nearly any subject. We already live in an age where nearly any question can be asked and answered purely online simply for the sake of gaining new knowledge. Creators and distributors of data now have to format their data in a way that is understandable and engaging to far wider audiences than ever before. Data publishers must try to fully realize the main goal of their data and reveal it in a way that is both mentally and emotionally engaging. As the quest for knowledge grows in the mind of everyone, so should the quest to teach grow in those that format and publish data of all subjects. As the human race continues to advance and evolve, the hunger to learn will only get stronger. The availability of subject knowledge will only become even more available than it currently is. Creators and distributors should learn how to properly take advantage of good data visualization techniques. Consumers of data should also learn how to divide good, whole data apart from bad, misleading, partial data. This issue will persist as more knowledge becomes abundant, as it already has been more of an issue in recent years. In conclusion, data visualization is and will always be an ever-increasing need for our society as a whole. Not only as an amenity, but as technology has advanced, it is needed simply to function.

Works Cited

Aparicio, Manuela, and Carlos J Acosta. “Data Visualization.” *ACM Digital Library*, 1 Nov. 2014, dl-acm-org.ezproxy.lib.apsu.edu/doi/pdf/10.1145/2721882.2721883.

Associated Press. “Monitoring Traffic Patterns Through Cell Phones.” *NBC News*, NBC News, 15 Oct. 2005, www.nbcnews.com/id/wbna9698139.

Hepworth, Katherine. “Big Data Visualization: Promises & Pitfalls.” *ACM Digital Library*, Communication Design Quarterly, dl-acm-org.ezproxy.lib.apsu.edu/doi/pdf/10.1145/3071088.3071090.

Lee, Bongshin, et al. “Data Visualization on Mobile Devices.” *ACM Digital Library*, 26 Apr. 2018, dl-acm-org.ezproxy.lib.apsu.edu/doi/pdf/10.1145/3170427.3170631.

Mani, Murali, and Si Fei. “Effective Big Data Visualization.” *ACM Digital Library*, 14 July 2017, dl-acm-org.ezproxy.lib.apsu.edu/doi/pdf/10.1145/3105831.3105857.

Rodriguez, Maria Teresa, et al. “Telling Stories with Data Visualization.” *ACM Digital Library*, dl-acm-org.ezproxy.lib.apsu.edu/doi/pdf/10.1145/2804565.2804567.

Schmillen, Clemens. *Rock Paintings from the Cave of Beasts (Gilf Kebir, Libyan Desert) Estimated 7000 BP*. 1 Mar. 2014, en.wikipedia.org/wiki/Cave\_painting#/media/File:Bestias11.JPG. Accessed 20 Apr. 2021.