

DATA VISUALIZATION

“The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey

INTRODUCTION:

Data Visualization makes data easier to understand. It is the representation of data in graphs, charts, images, and video. Software applications provide a means to create data visualizations and data dashboards—interfaces that allow us to visually explore and analyze data.

Visualizing data helps us view and understand something that is inherently not visual and potentially difficult to read. A single image can provide a way to explore and make sense of small to big data and open the possibility to “notice what we never expected to see.”

Data Visualization has become an indispensable tool to understand data. Computers and networks have transformed data from something that was once difficult to collect and use, to something that is abundant and easy to access. We are in the era of Big and Open Data, and researchers and data analysts help to make this data visible and therefore widely accessible.

Some fields closely related to Data Visualization include:

- Information Visualization (InfoViz)
- Information Design
- Data Storytelling.

These fields build upon the analysis of data and Data Visualization with respect to taking visual representations of data and crafting them into more easily digested visual formats for professionals and the public at large.

VISUALIZING DATA:

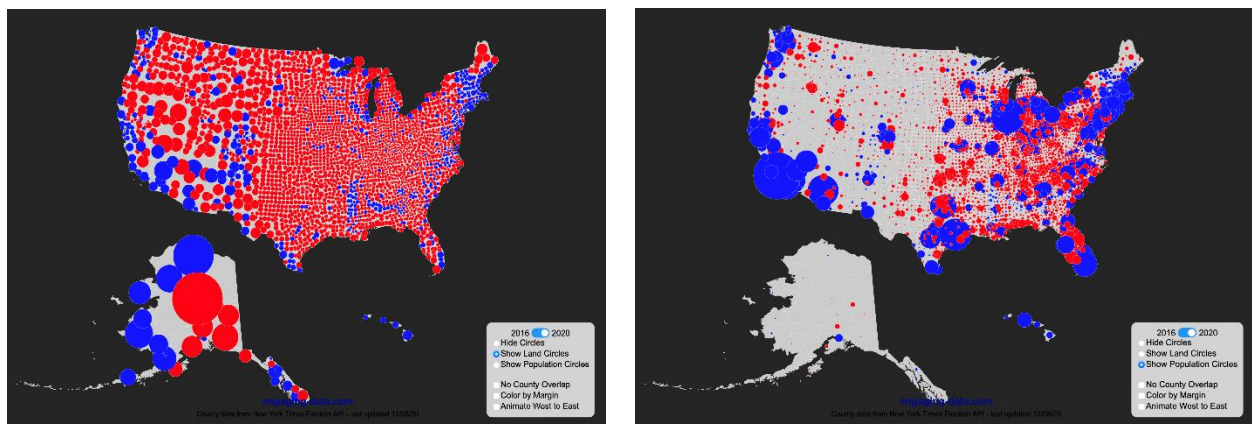
The process of creating Data Visualizations involves transforming numerical data into a visual form. In its numerical form, data is often saved and shared using spreadsheets—digital tables with rows and columns. These spreadsheets can be in a proprietary format such as Microsoft Excel, or

in a plain text format such as CSV (Comma Separated Values). This raw data often needs to be formatted before being imported to a software application for visualization. This process ranges from simple to extremely complex depending on the nature of the data, and the type of visualization. For example, Microsoft Excel is an application that allows for the storage and organization of data as a spreadsheet, but also includes a range of tools to view the data in graphs and charts. Other applications, such as Tableau, integrate tools that allow users to import data and create data dashboards to explore and compare data using different types of visualizations.

DATA LITERACY:

We are frequently presented with Data and Information Visualizations online and in different publications, and it is important to understand how these visual representations can simplify the complexity of the data, or only depict a partial story. As such they can intentionally or unintentionally tell stories that are either incomplete or incorrect. In our data-driven world, it is important for just about everyone to have some degree of data literacy with respect to being able to read, understand, and critique visual representations of data. Full data literacy is the ability to read, work with, analyze and communicate with data—skills that also involve using software tools to transform raw numerical data into different visual forms.

There are many choices available when creating data visualizations—the type of chart or graph, the number of variables shown, the scales used and the use of colour. All of these factors can work together to tell dramatically different stories. Take the following example of a visualization of the 2020 US presidential election results:



Source: US County Electoral Map – Land Area vs Population, Engaging Data.

While these two visualizations show the same data, they tell very different stories. The visualization on the left uses coloured circles, sized according to the geographical area to each voting region, whereas the map on the right uses coloured circles that are sized according to population of the voting region

Imagine description: US county electoral map. A dynamic way to engage data. This interactive map shows the election results by county and you can display the size of counties based on their land area or population size.

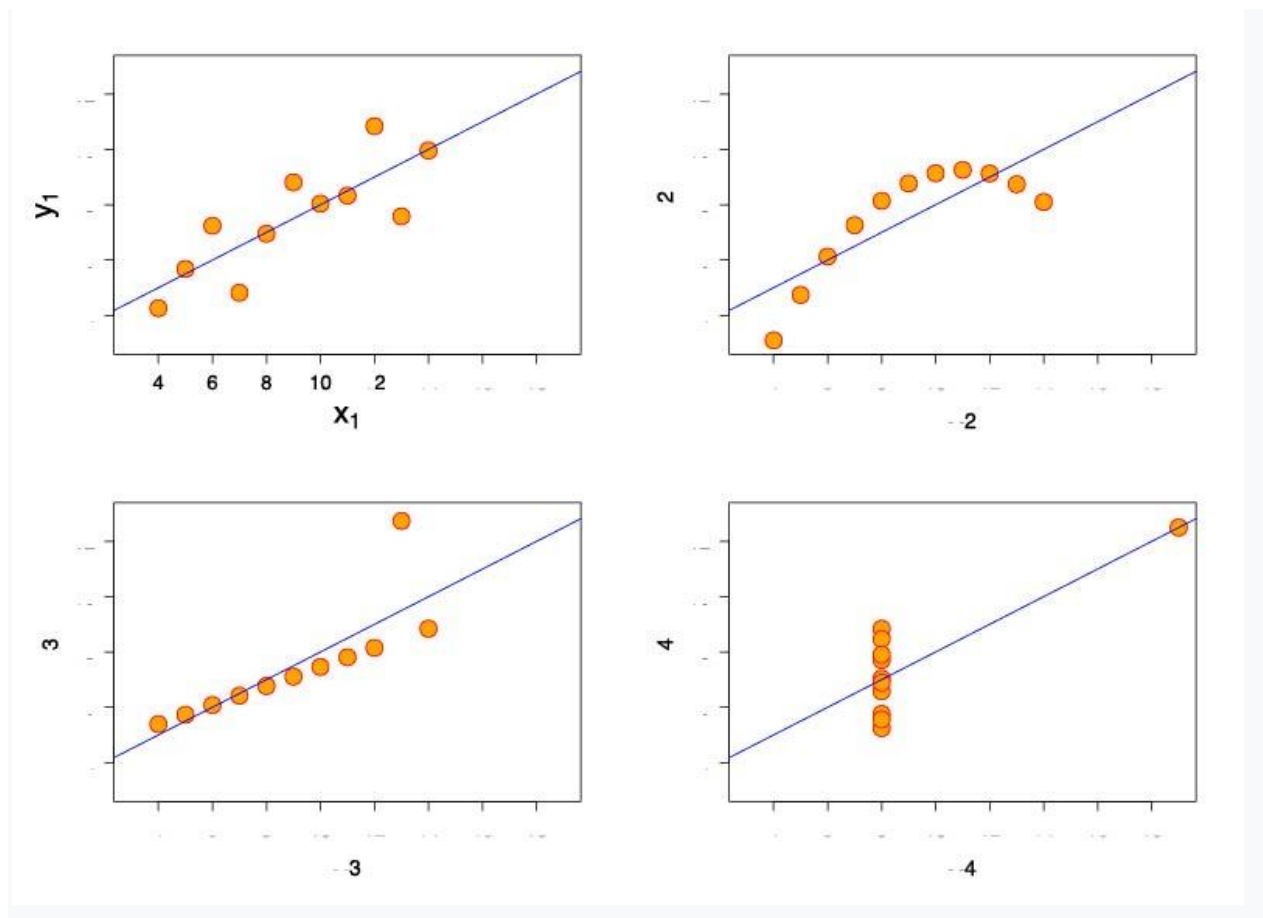
EXAMPLE: WHAT'S IN THE DATA?

In this example, we'll look at Anscombe's quartet, an example created by the statistician Francis Anscombe in 1973 to demonstrate the importance of graphic data when analyzing it.

Look at the chart on the right. There are four sets of data that show a relationship between two variables, x and y .

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

While these four sets of data look numerically similar, graphing them demonstrates that the relationship between the x and y variables is very different for each data set...



Source: Graphic from [Anscombe.svg](#) by [Wikipedia user Schutz](#). [Creative Commons Attribution-ShareAlike 3.0 Unported license](#).

This graphic represents the four datasets defined by Francis Anscombe for which some of the usual statistical properties (mean, variance, correlation and regression line) are the same, even though the datasets are different. Reference: Anscombe, Francis J. (1973) Graphs in statistical analysis. *American Statistician*, 27, 17–21.

QUIZ

1. What is Data Visualization?
 - a. The representation of data in graphs, charts, images and video.
 - b. The transformation of numerical data into a visual form.
 - c. A process used to analyze and study data.
 - d. All the above.**

Correct answer: d

Feedback: Data visualization is the visualization of numerical data in charts, images and video, a process used to study and analyze data.

2. What is (arguably) the most important benefit of Data Visualization?
 - a. To communicate complex numerical data in publications and media.
 - b. To make data easier to store and share.
 - c. To learn and discover things about the data not evident in its numerical form.**
 - d. To make data more visually appealing.

Correct answer: c

Feedback: Data visualization is a tool to analyze data and discover things not evident in its numerical form. While Data visualizations can be spun into Information Visualizations in publications and media, its core value is as a tool to understand data and not merely to make data more visually appealing.

3. What is a data dashboard?
 - a. A software application to help with the visualization of network data.
 - b. An interface that allows users to visualize and explore data using different tools.**
 - c. A way of visualizing statistical data.
 - d. A mobile platform for data visualization.

Correct answer: b

Feedback: A data dashboard is a software interface where data can be explored using different types of visualizations and analysis tools.

4. What's the difference between Data Visualization and Information Visualization?
 - a. There is no difference, they are the same.
 - b. Data Visualization is an indispensable tool when analyzing data whereas Information Visualization is used to communicate key elements of the data in publications and media.**

- c. Data Visualization is the visual representation of numerical data by statisticians, and Information Visualization is the visual representation of numerical data by designers.
- d. While Information Visualization is an accurate representation of data, Data Visualization can be misleading.

Correct answer: b

Feedback: While Data Visualization is used to understand and analyze data, Information Visualization involves communicating features of the data or patterns that the data reveals.

5. What does it mean to be data literate?
- a. It is the ability to read and understand numerical data without the use of data visualization.
 - b. Always knowing the best way to represent different types of data, graphically.
 - c. **It involves the ability to read, understand and critique the visual representation of data as information.**
 - d. To have studied in the field of information design.

Correct answer: c

Feedback: While data literacy is the ability to read, work with and communicate with data, on a basic level it involves being able to read and critique data visualizations in news and media.

ACTIVITY



Purpose: In this activity, we'll take a critical look at some Information Visualizations communicating Covid-19 statistics in the media.



Task: Head to the following website and have a look at the article on “[How Bad Covid-19 Visualizations Mislead the Public.](#)”

Now think back on the steps involved with transforming data into Data and Information Visualizations and consider the following questions. Hopefully they will provide you with food-for-thought in your journey towards data literacy!

1. Do you think that any of the examples in the article demonstrate an intentional misuse of Data and Information Visualization?
2. Can you think of any Data or Information Visualizations that you have seen lately? Did the visualizations seem to draw conclusions by design?
3. Do you think that data storytellers and information designers should provide more transparency with respect to their data sources and choices with respect to how they have chosen to represent data?

 **Time:** 10 mins

FURTHER RESOURCES

Journal: [The Good, the Bad, and the Biased: Five Ways Visualizations Can Mislead \(and How to Fix Them\)](#)

Web page: [Visual Capitalist](#)

Udemy course: [Data Storytelling and Data Visualization](#)

REFERENCES

Tableau. [What Is Data Visualization? Definition, Examples, And Learning Resources](#)

Wikipedia. [Data and Information Visualization](#). 2022

McCormick, B. H., DeFanti, T. A. & Brown, M. D. [Visualization in Scientific Computing](#). Vol 21, no 6, ACM SIGGRAPH, 1987.

Wikipedia. [Data Literacy](#). 2021

Stobierski, T. [Bad Data Visualization: E5 Examples of Misleading Data](#). Harvard Business School, January 2021.

Wikipedia. [John Tukey](#) 2022

Robert Falkowitz. [Information Visualization or Data Visualization?](#) Concentric Circle Consulting, 2019.

Alice Park, Charlie Smart, Rumsey Taylor, and Miles Watkins. An Extremely Detailed Map of the 2020 Election. New York Times, 2021.

Engaging Data. US County Electoral Map – Land Area vs Population. 2020.

Nora Mulvaney, Audrey Wubbenhorst, and Amtoj Kaur. Critical Data Literacy: Strategies to Effectively Interpret and Evaluate Data Visualizations. Toronto Metropolitan University, 2022.

Wikipedia. Anscombe's Quartet. 2020

VIDEO TRANSCRIPT

Things for Digital Knowledge. Concordia Library.

David Clark, Technology Analyst

Intro slide: Concordia library logo

Interviewer: What is data visualization? [question displayed on the screen]

David Clark: Data visualization is essentially making data easier to understand. So it's the representation of numerical data in charts, graphs, images, and sometimes even video.

Visual image description: how to visualize data, graph/chart, image, video.

Visual image description: examples of data visualization in charts, forms, graphs.

Visual image description: example of a data dashboard.

David Clark: There are many software applications to facilitate the process of creating data visualizations, and sometimes what we call data dashboards, which are interfaces that allow us to see different visual representations of data to try different things. It's very much a discovery process.

David Clark: There's a quotation by an American statistician and mathematician named John Tukey, and that is “the greatest value of a picture is that it forces us to notice what we never expected to see” [text displayed on the screen]. And I really like that because that's for me, and I think for many people, what characterizes data visualization, is that it's a discovery process. You take large, unwieldy datasets that are hard to work with and you can, in a single picture, begin to look and explore and analyze that data. But at the same time, it's very much an intentional

manipulation. And so in trying to represent that data in different visual forms, you might see different things or maybe even obscure things. I think one thing that's important to think about when we talk about data visualization is some other fields that are closely related to data visualization, like information visualization or info viz and data storytelling and also data design. These are sort of other fields.

Visual image description: example of different visual forms, such as timeline, time series, Gantt chart, different type of maps, alluvial diagram.

David Clark: If we think about taking numerical data, working it into visualizations, and then taking those visual representations and perhaps creating stories with them or making inferences or information visualizations for users, for news or for media or for papers. But we have to understand that this is one story. There may be many stories in the data. At one point in history data visualization was something that was done often by statisticians and mathematicians. But we live in an era of big and open data. There's so much data around us that's generated by everything and every industry. And so data visualization has become an indispensable tool to understand that data. Data was once really hard to access. And, of course, stats was a place where it was generated and collected, but now it's collected everywhere.

Interviewer: Can you give us some examples? [text displayed on the screen]

David Clark: Well, an early example that I really like is by an epidemiologist from London in the 1800s, John Snow. And in this one example, there was a cholera outbreak. And I believe at the time that they thought that this was airborne. But what John Snow did was map in something that was very much an early geospatial visualization.

Visual image description: map of cholera outbreak in London in 1800s by John Snow

David Clark: He mapped all the cases of cholera onto a map of London. And by doing that, it was very easy to see that most of the cases were all happening around their water supply, a water pump that was in the central part of that area. So it became very easy to see through this geospatial visualization what was causing the outbreak, that it wasn't airborne, that it was the water supply.

Interviewer: Can you explain the process of data visualization? [text displayed on the screen]

David Clark: So, in a nutshell, the process of creating a visualization involves transforming numerical data into a visual form.

Visual image description: Microsoft Excel logo labeled “numerical data” with an arrow to a chart icon labeled “visual form.”

David Clark: Often numerical data or data that's stored is stored in the form of spreadsheets. So columns and rows of data. So the steps involved are basically taking that numerical data, editing it, wrangling it, working it into a form in which it can be imported into a software application. And depending on the software application, it might require different formats.

Visual image description: Microsoft Excel logo labeled “numerical data” with an arrow to a chart icon labeled “visual form.”

David Clark: For example, a very basic or rudimentary visualization tool, but still very powerful, is Microsoft Excel, which is a spreadsheet. But it provides a way of taking that data that's in the rows and columns and looking at it in the form of graphs and charts.

Visual video clip: example of generating a chart from data stored in an Excel file

David Clark: But every different visualization tool or application might require the data to be in different forms and treating of the data from its raw spreadsheet form into the form that it can be imported could be quite simple, as in the case with Excel, or it could be quite complicated.

Visual video clip: examples of more complicated forms of data and their related visualization methods.

David Clark: For example, working with network-based visualization tools like Gephi, there might be actually some coding involved in getting it from its raw state into the state that it can be visualized. So once it's imported into that tool, then you can begin looking at different visualizations of it. Another application called Tableau provides a way of creating more like data dashboards that I mentioned before. And when it imports data, it imports from a raw spreadsheet form such as Excel or something called comma separated values into a database, which is what enables the tool Tableau to be able to create dashboards in different ways of seeing the data.

Interviewer: Can you explain what data literacy is? [text displayed on the screen]

David Clark: Sure. Yeah. So as citizens, we are frequently presented with data and information visualizations online and in different publications. And it's important for us to understand how these visual representations of data can simplify the complexity of the data, or only tell a partial story.

Visual slide text: data literacy to understand the data

David Clark: Because of that, data visualization and infographics can either intentionally or unintentionally tell stories that are either incomplete or incorrect. In the data driven world that we live in today, it's important for just about everyone to be able to read, understand and even critique visual representations of data as information. Being data literate can take this one step further and involve the ability to create our own data visualizations. And that might be learning a tool such as Tableau or even Microsoft Excel.

Interviewer: Can you give an example? [text displayed on the screen]

David Clark: Sure. A great example is by a Belgian designer, Karim Douïeb, who works at a company called Jetpack. And what Karim did was create a visualization of the U.S. 2020 U.S. election results. There were a number of election maps, so geospatial visualizations that were showing the results of the election but coloring a map according to the geographic size of the voting region. And of course, the geographic size of the voting region doesn't correspond to the number of voters that are in that region.

Visual video clip: dynamic data showing US election results based on land size and population size

David Clark: So what Karim did is creating a visualization that essentially morphs geospatial visualization of land size to population size. Different representations of those election results are very different and tell very different stories. So, this is a great example of how a representation of data in a visual form can tell a very different story depending on how it's done.

Visual: thank you and credit. CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik. Music: Lofi Summer Background by Vladislav Kurnikov on Pixabay.