Data Visualization

By : Ahmed Haytham

Intro

We will learn how to make an amazing plots according to **“Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks “ book by Jonathan Schwabish**

1. Some important questions
2. GUIDELINES FOR BETTER DATA VISUALIZATIONS
3. Kind of plots and its use with examples
4. There is v2 from this presentation focus more in type of plots and its use

Important Questions

1 photo = 1K Words

What is data Visualization ?

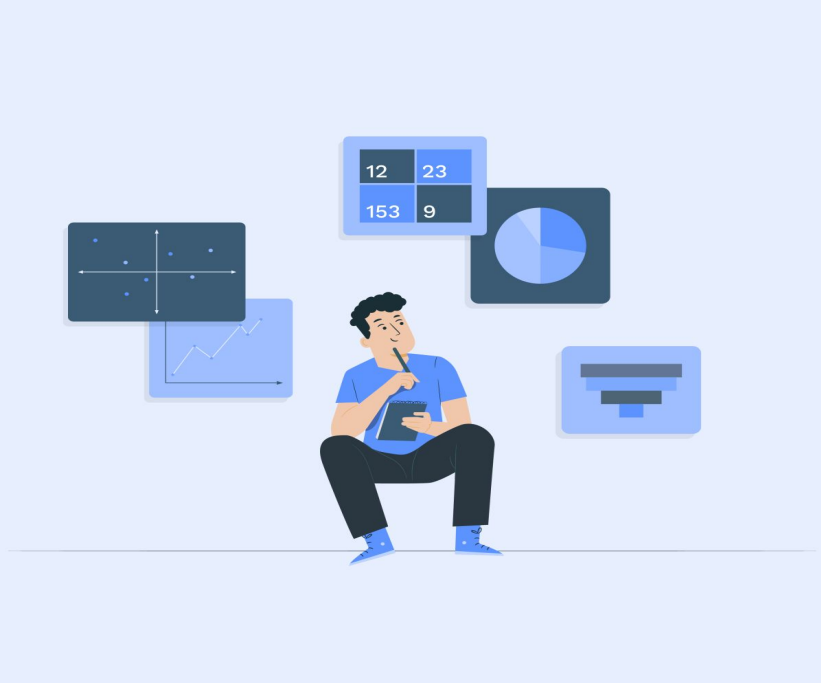
A way to represent the data in charts to get information

What is the Challenge ?

The biggest challenge in data visualization is to make the data clean

GUIDELINES FOR BETTER DATA VISUALIZATIONS

1. Treat the reader like a child
2. Show the data
3. Reduce the clutter
4. Integrate the graphics and text
5. Avoid the spaghetti chart



Show the data

Your reader can only grasp your point, argument, or story if they see the data. This doesn’t mean that all the data must be shown

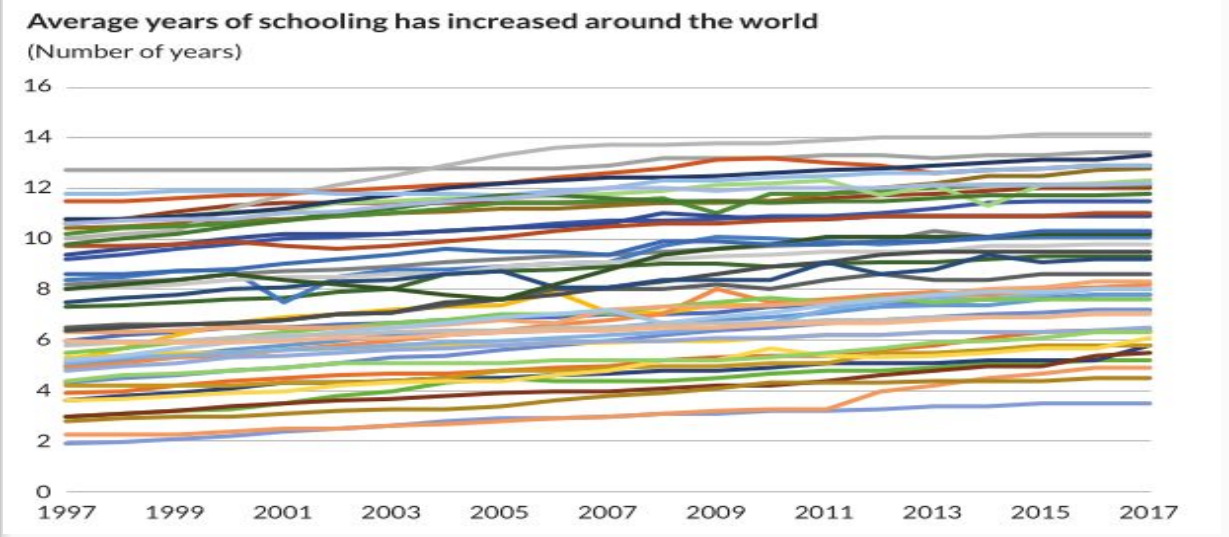
As chart creators, our challenge is deciding how much data to show and the best way to show it.

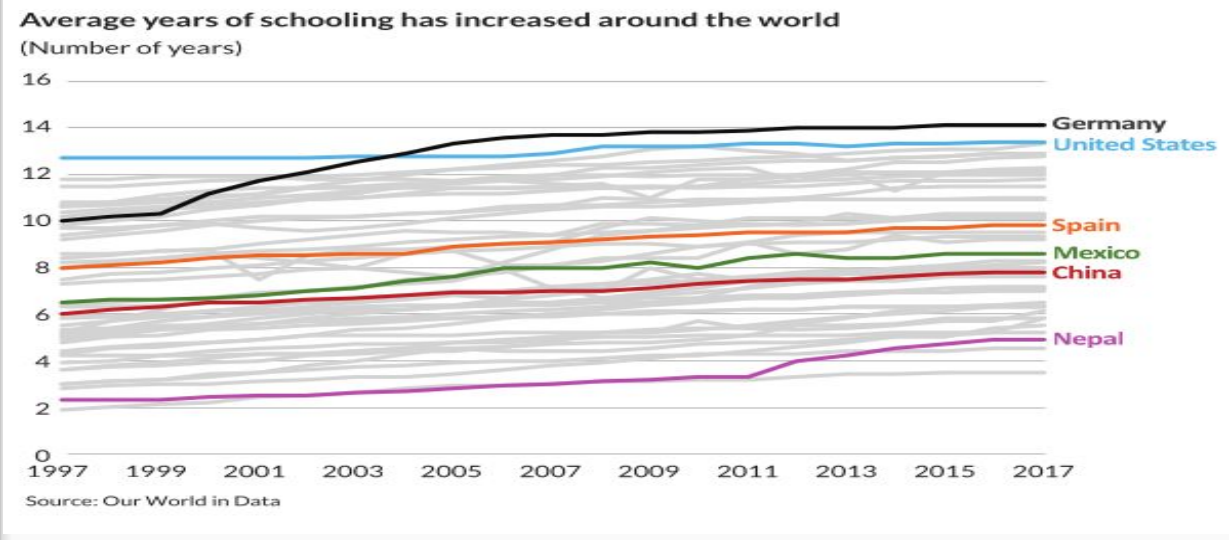
This is bad plot. You can't extract a single piece of information other than it's for the United States

Show the data

First plot has a lot of informations but it is very hard to be read by anyone need a lot of focus to get one information

Second plot unlike the first it has less data shown in it so you can get more information and it is very simple to be read





REDUCE THE CLUTTER

The use of unnecessary visual elements distracts your reader from the central data and clutters the page. There are lots of different types of chart clutter we might want to avoid.

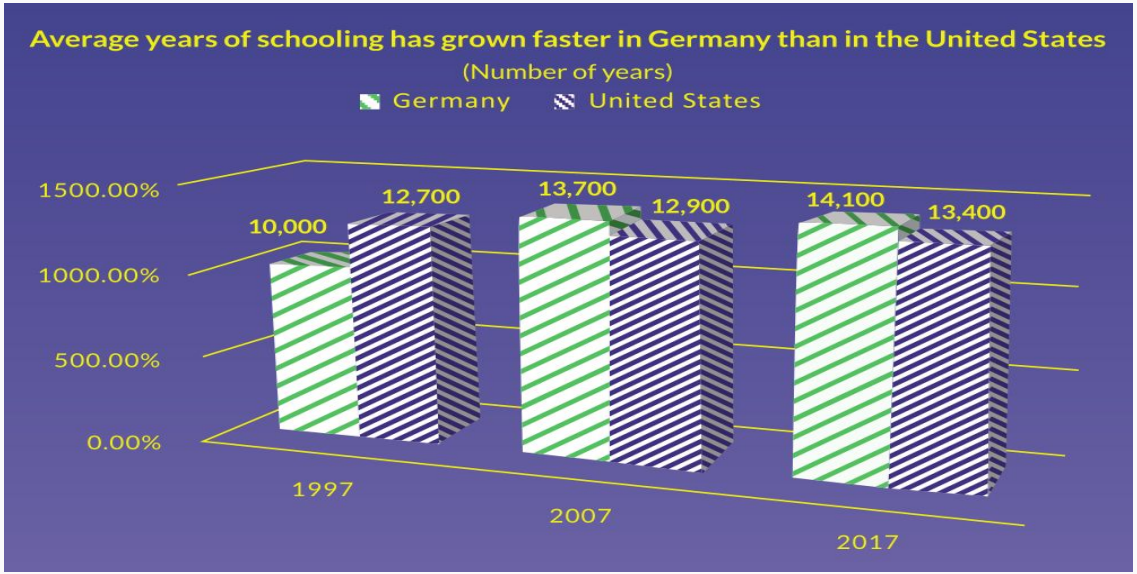
There are lots of different types of chart clutter we might want to avoid. There are basic elements like heavy tick marks and gridlines, which we should remove in almost every case

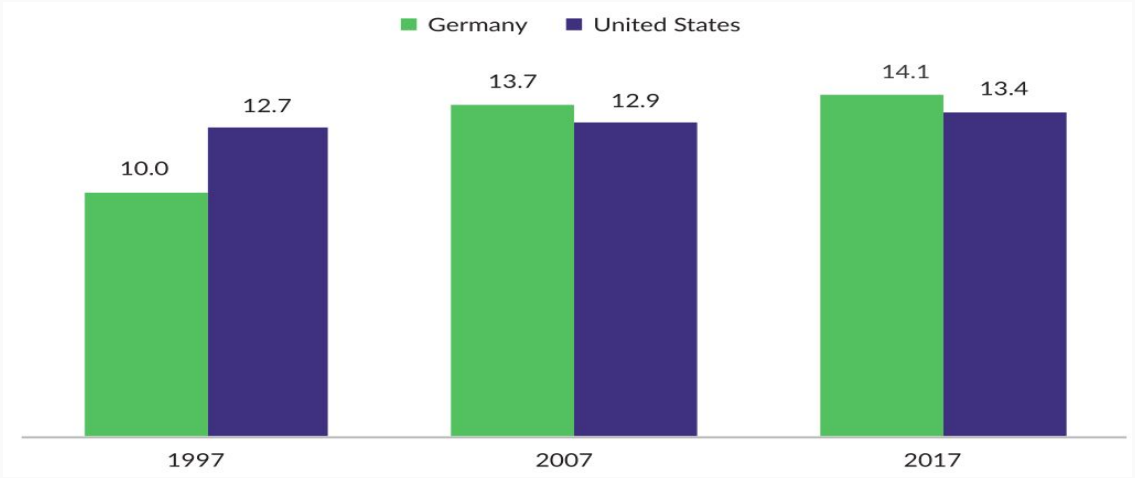
Some use unnecessary dimensions that distort the data. And others contain too much text and too many labels, cluttering the space and crowding out the data.

REDUCE THE CLUTTER

First plot has a lot of unimportant content with 3D it really need a lot of effort to understand it and get information

Second plot unlike the first it has just important components to get the same information with less effort



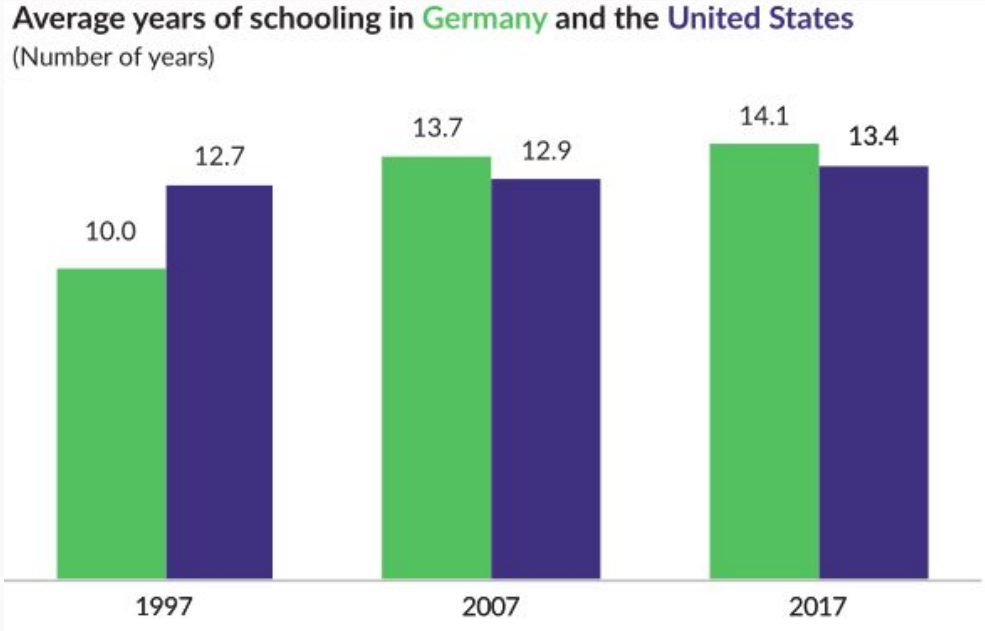


INTEGRATE THE GRAPHICS AND TEXT

The use of unnecessary visual elements distracts your reader from the central data and clutters the page. There are lots of different types of chart clutter we might want to avoid.

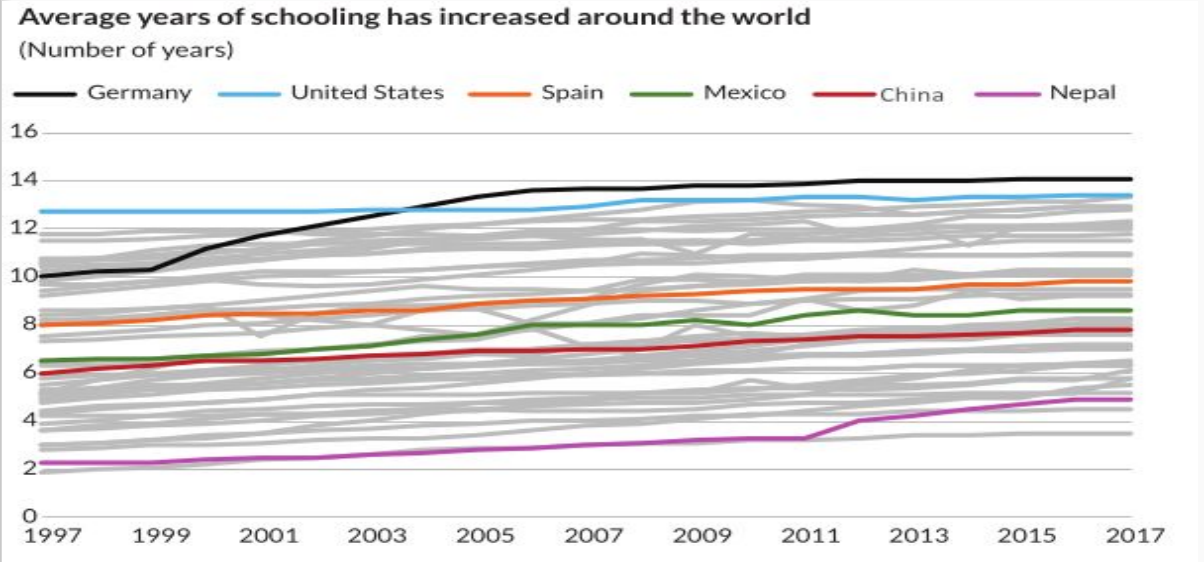
There are lots of different types of chart clutter we might want to avoid. There are basic elements like heavy tick marks and gridlines, which we should remove in almost every case

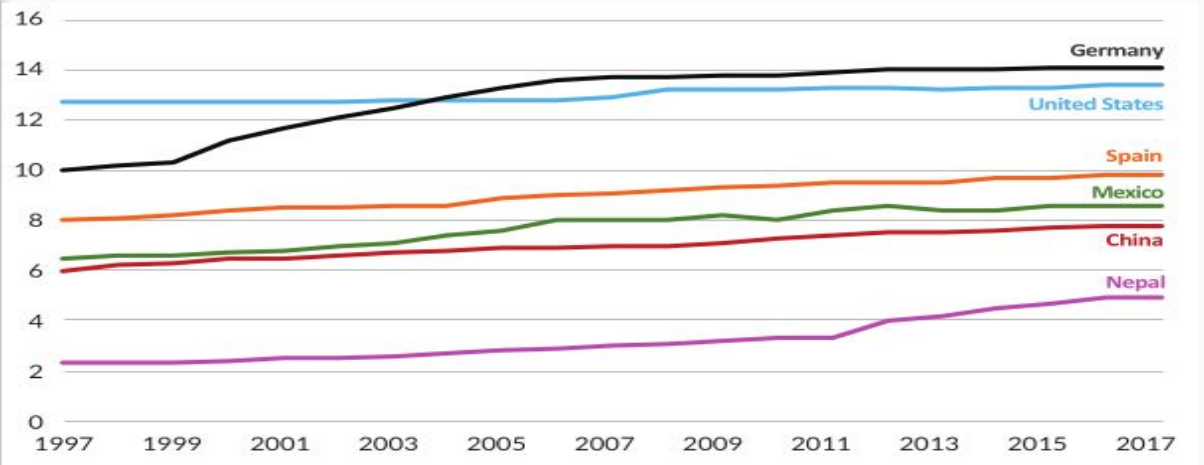
Some use unnecessary dimensions that distort the data. And others contain too much text and too many labels, cluttering the space and crowding out the data.



1_REMOVE LEGENDS WHEN POSSIBLE AND LABEL DATA DIRECTLY

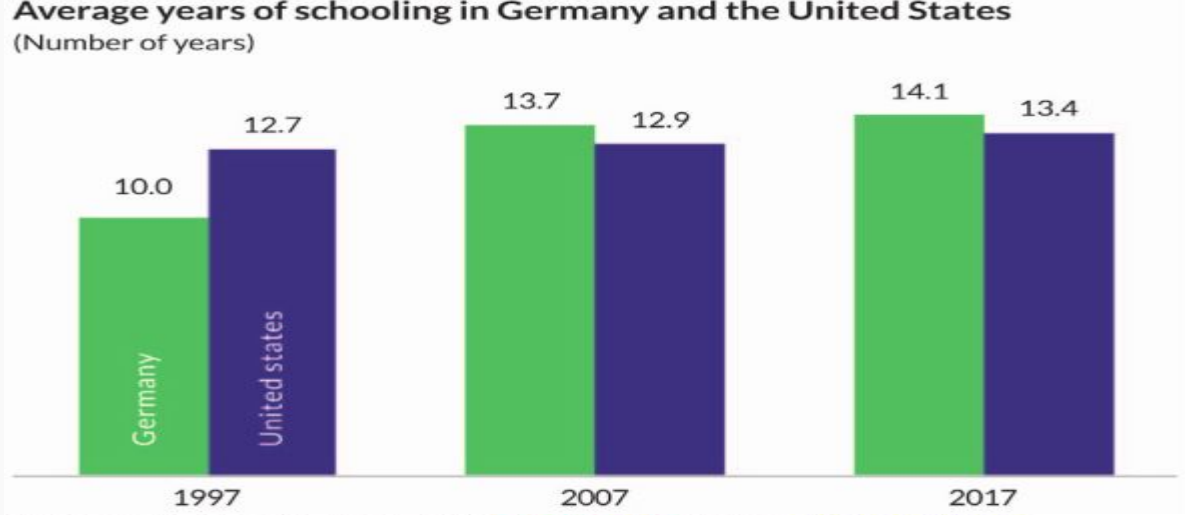
No explanation needed

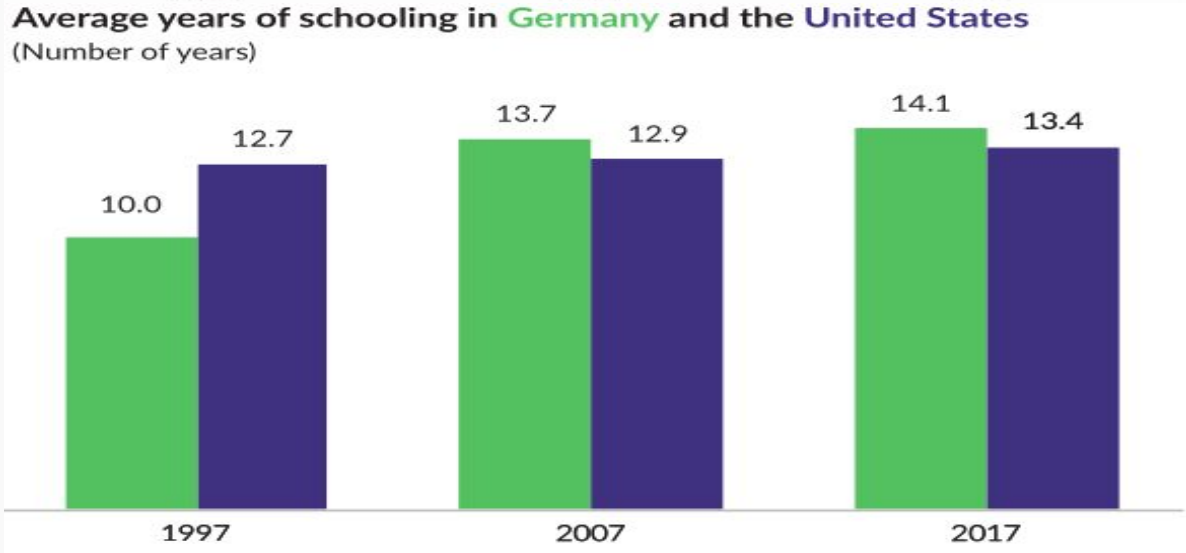




1_REMOVE LEGENDS WHEN POSSIBLE AND LABEL DATA DIRECTLY

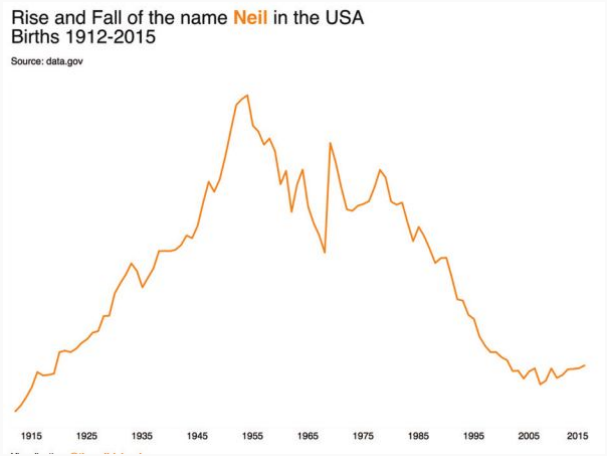
No explanation needed

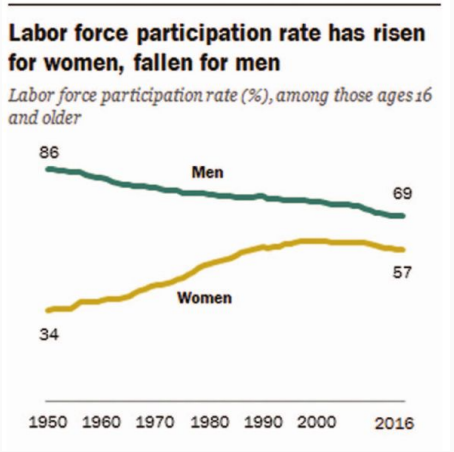


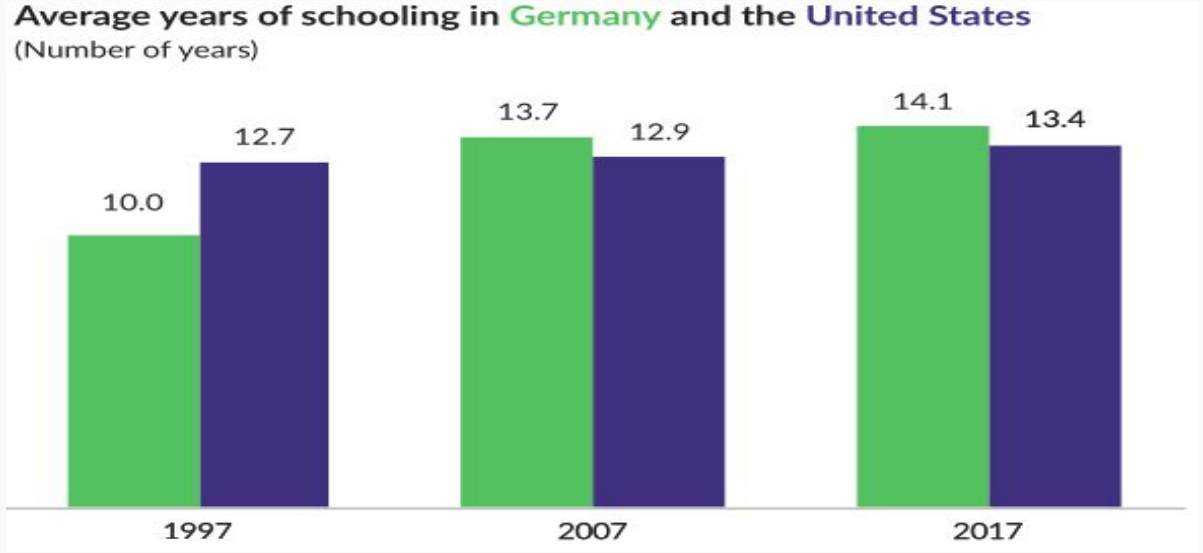


2_WRITE THE TITLE LIKE A NEWSPAPER HEADLINE

Most titles are neutral descriptions of the data



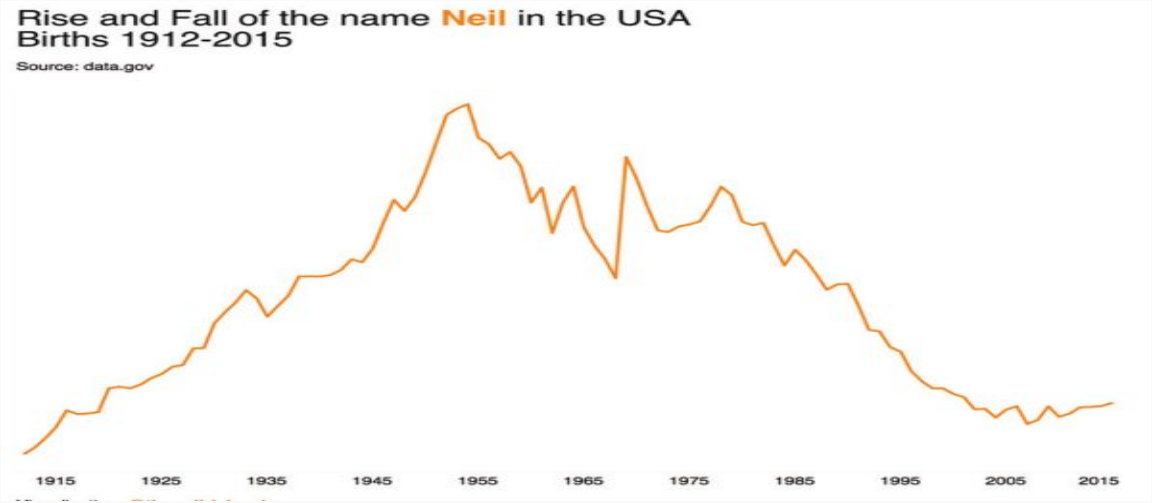


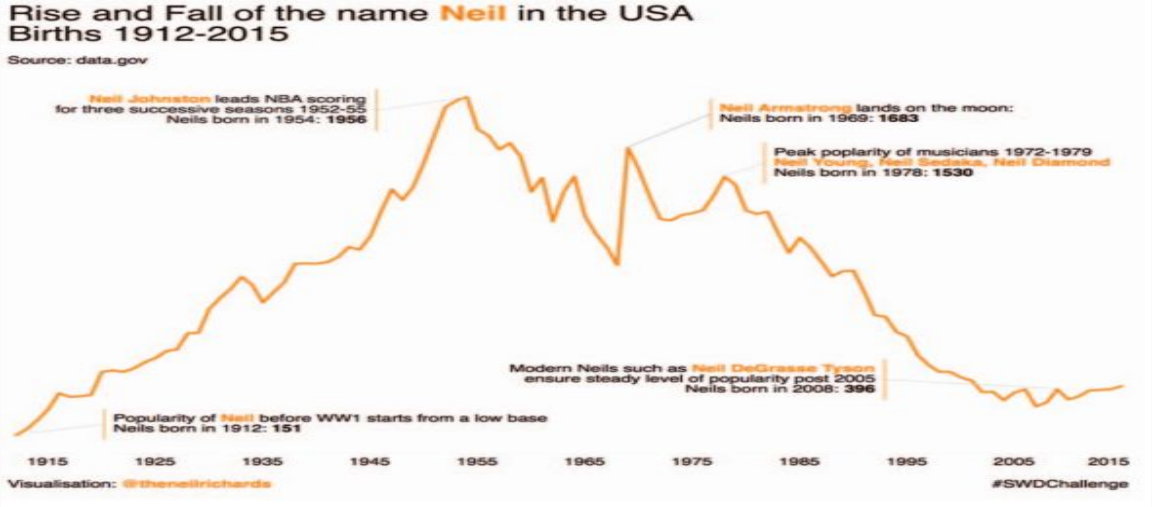


3_ADD EXPLAINERS

Once the chart is made and the title is settled, ask yourself, Would this chart benefit from more text?

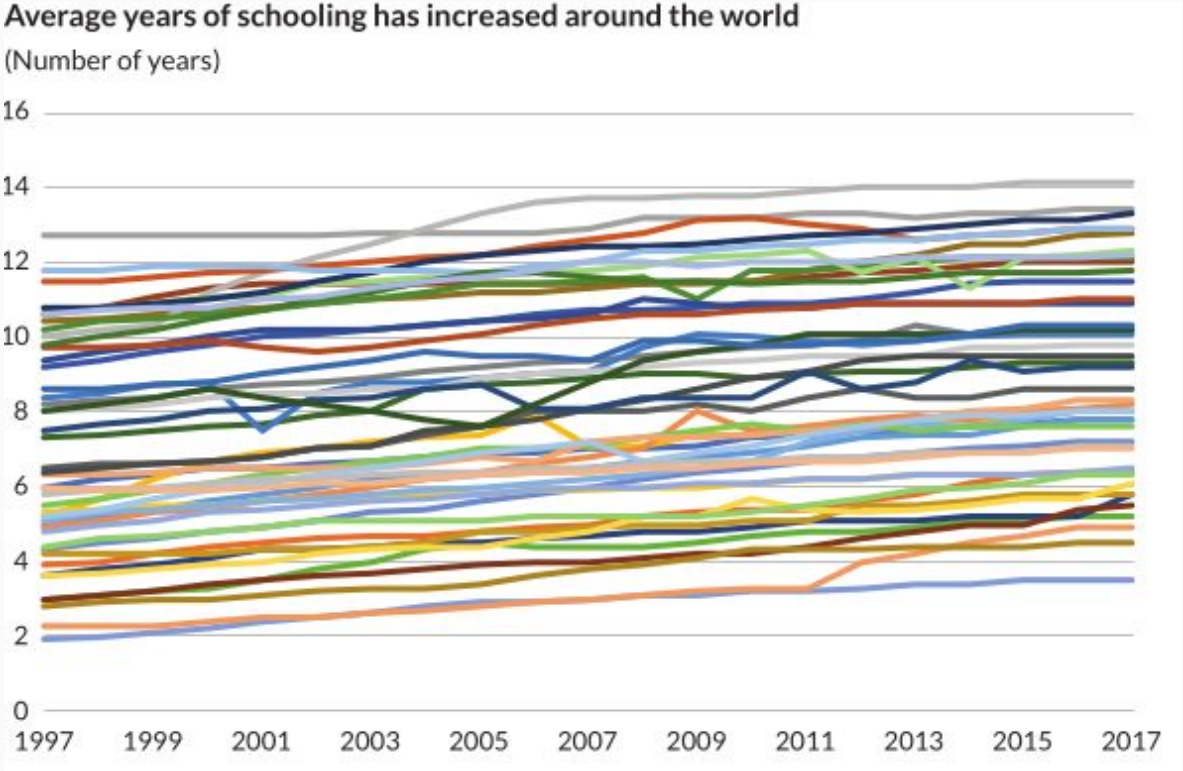
Sometimes data sets have peaks or valleys, outliers or variations that bear explanation. Adding detail in graphs can push your argument, highlight points of interest, or (in cases of nonstandard graphs) even explain how to read it.





4_AVOID THE SPAGHETTI CHART

It's obvious when a certain graph contains too much information—line charts that look like spaghetti, maps with dozens of colors and icons, or bar after bar after bar littering a chart.



Data Types

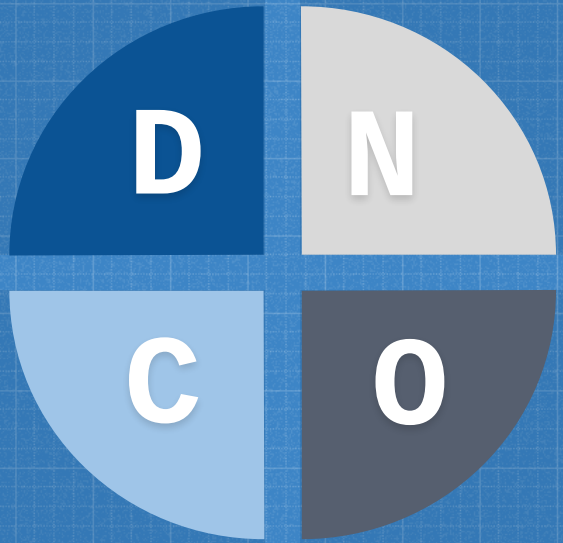
Numerical

Discrete Data

numerical data with countable elements. I.e they have a one-to-one mapping with natural numbers **countably**

They are represented as a set of intervals on a real number line **uncountable**

Continuous Data



Categorical

Nominal

Sometimes called naming data, it has characteristics similar to that of a noun. Not range able

categorical data includes elements that are ranked,

Ordinal

Data Types

The reason for the diversity in the types of data and information to be extracted that is to be clarified and highlighted. There are many and many that cannot be counted from the data visualizations that can be used.

Plots Types

1. COMPARING CATEGORIES
2. TIME
3. DISTRIBUTION
4. GEOSPATIAL
5. RELATIONSHIP
6. PART TO WHOLE
7. QUALITATIVE
8. TABLES

COMPARING CATEGORIES

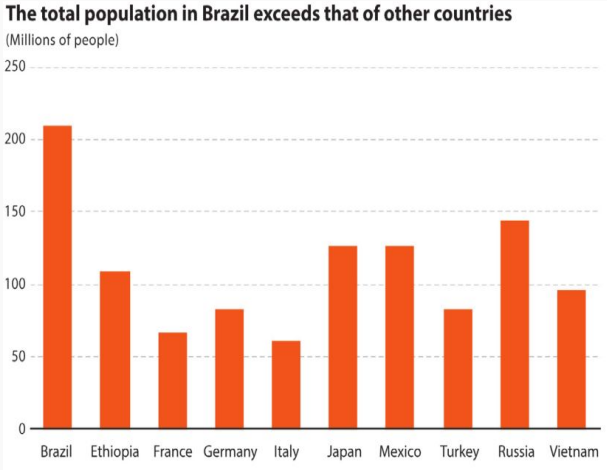
The graphs in this chapter are intended to help our readers compare values across categories. Bars, lines, and dots can all let our readers compare within and between groups. In some cases, we want our reader to see both levels and change, or some other variable combination; in other cases, we want to focus their attention on one comparison or another

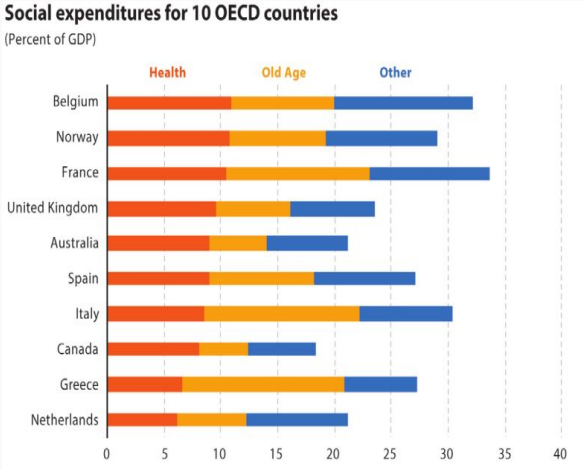
The challenge when comparing categorical data is deciding what we want the chart to convey. Is there a primary argument or story? Is there something you can identify as the most important comparison you want the reader to make? As chart creators, we need to prioritize what we want our charts to do. By putting every bar or dot in the graph, we can obscure the point we wish to convey

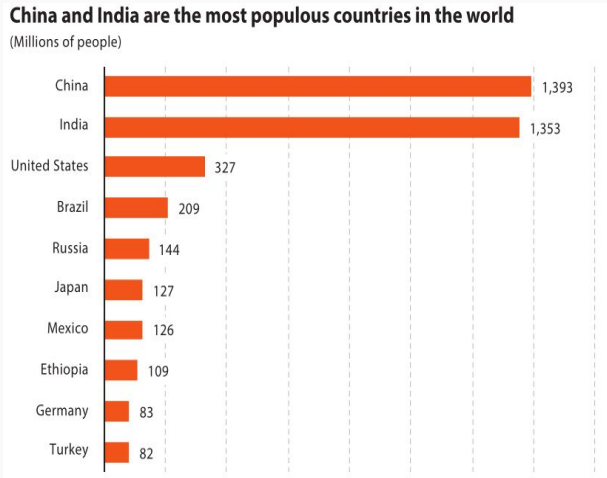
1_ Bar Chart

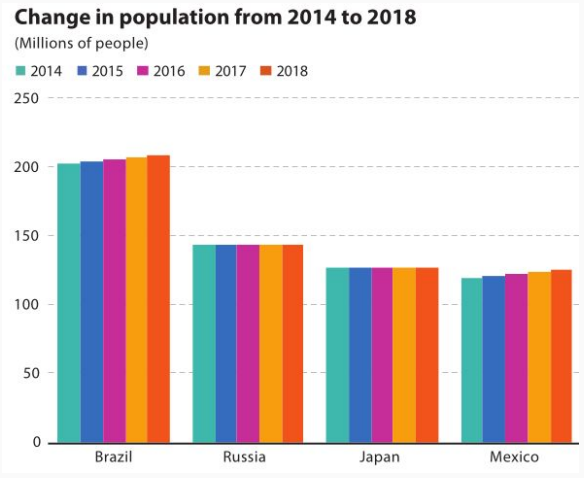
One of the most familiar data visualizations, the length or height of the rectangular bars in bar and column charts depict the value of your data. The rectangles can be arranged along the vertical axis so that the bars lie horizontally (often called a bar chart) or vertically on the horizontal axis (often called a column chart)

bar Has a lot of types in next presentations DV 2 i will explain more with more plots









TIME

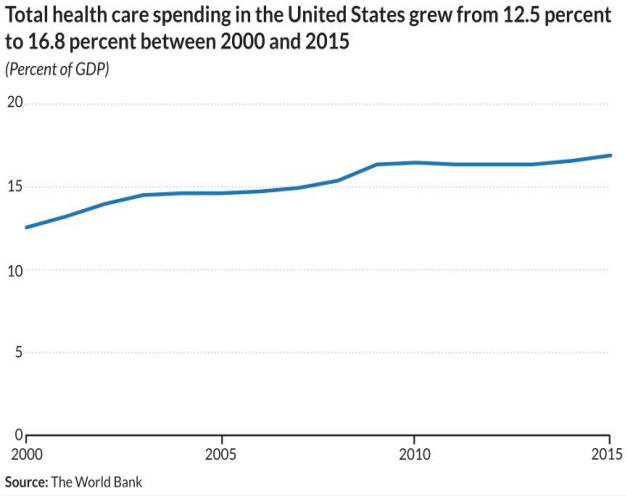
Many of the visuals in this chapter are variations on the line or area chart. Some let us include more data on the page than usual, while others allow us to combine changes over time with some other view of the data. With horizon charts and streamgraphs, for example, we can include more data in a single visual, but they are probably not best suited for detailed comparisons

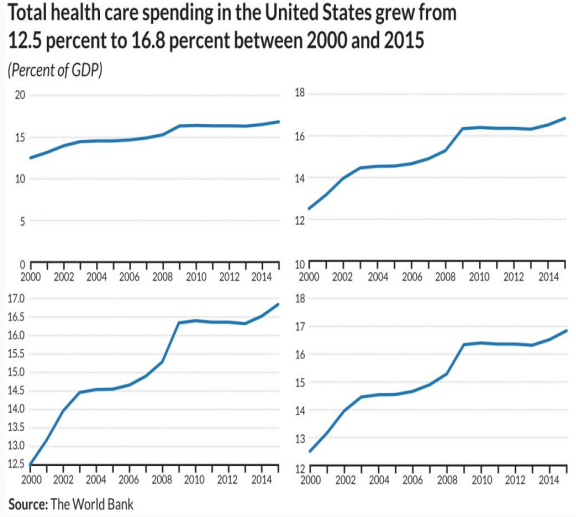
1_ LINE CHART

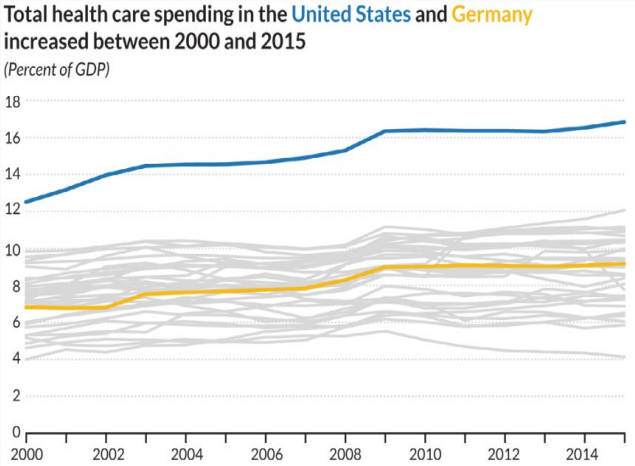
The line chart and the bar chart may be the most common charts in the world. The line chart is easy to read, clear in its representation, and easily drawn with pen and paper. Data values are connected by lines to show values over a continuous period, tracking trends and patterns.

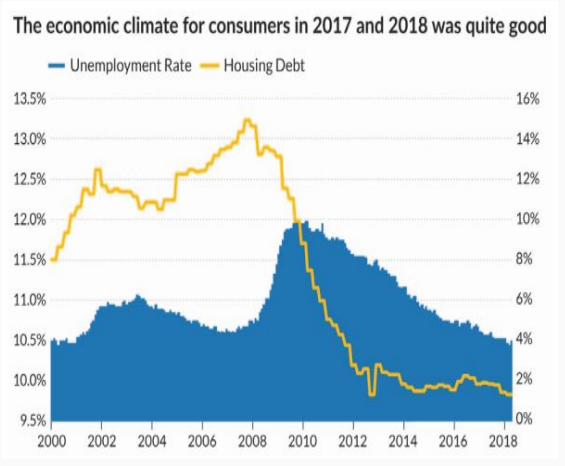
how a trend emerged over time.

Has a lot of types in next presentations DV 2 i will explain more





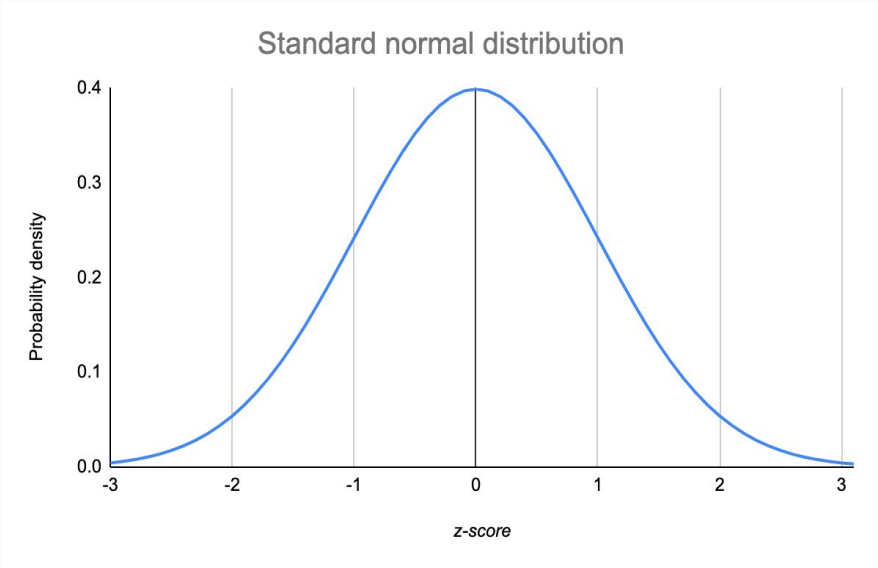




DISTRIBUTION

This chapter covers visualizations of data distributions and statistical uncertainties. These may be inherently difficult for many readers because they may not be as familiar with the statistical terminology or the graphs themselves, which may look quite different from the standard graphs they are used to seeing.

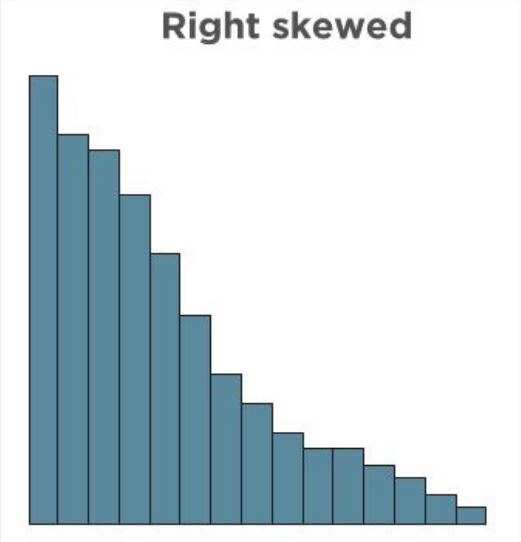
Charts like the fan chart and the box-and-whisker plot show statistical measures like confidence intervals and percentiles. Violin plots, which depict entire distributions, may look so foreign that your reader will need detailed explanations to understand them

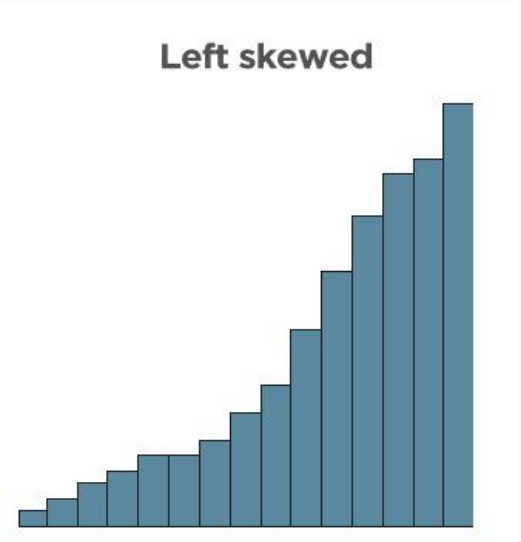


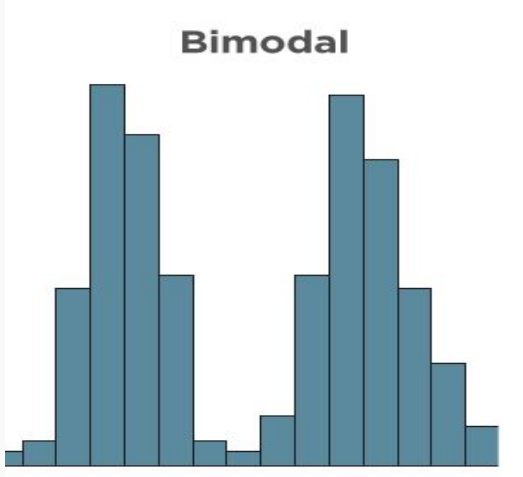
1_ HISTOGRAM

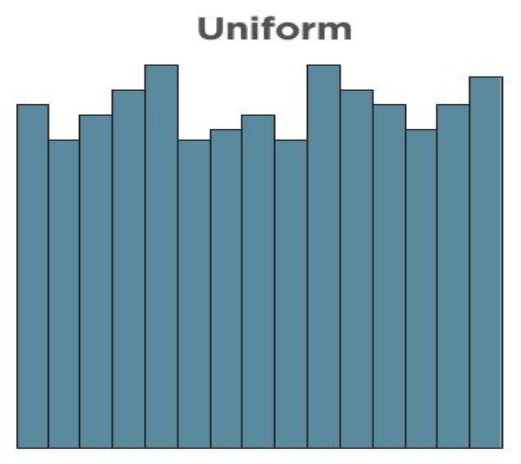
The histogram is the most basic graph type for visualizing a distribution. It is a specific kind of bar chart that presents the tabulated frequency of data over distinct intervals, “called bins”

The entire sample is divided into these bins, and the height of each bar shows the number of observations within each interval. Histograms can show where values are concentrated within a distribution, where extreme values are, and whether there are any gaps or unusual values.



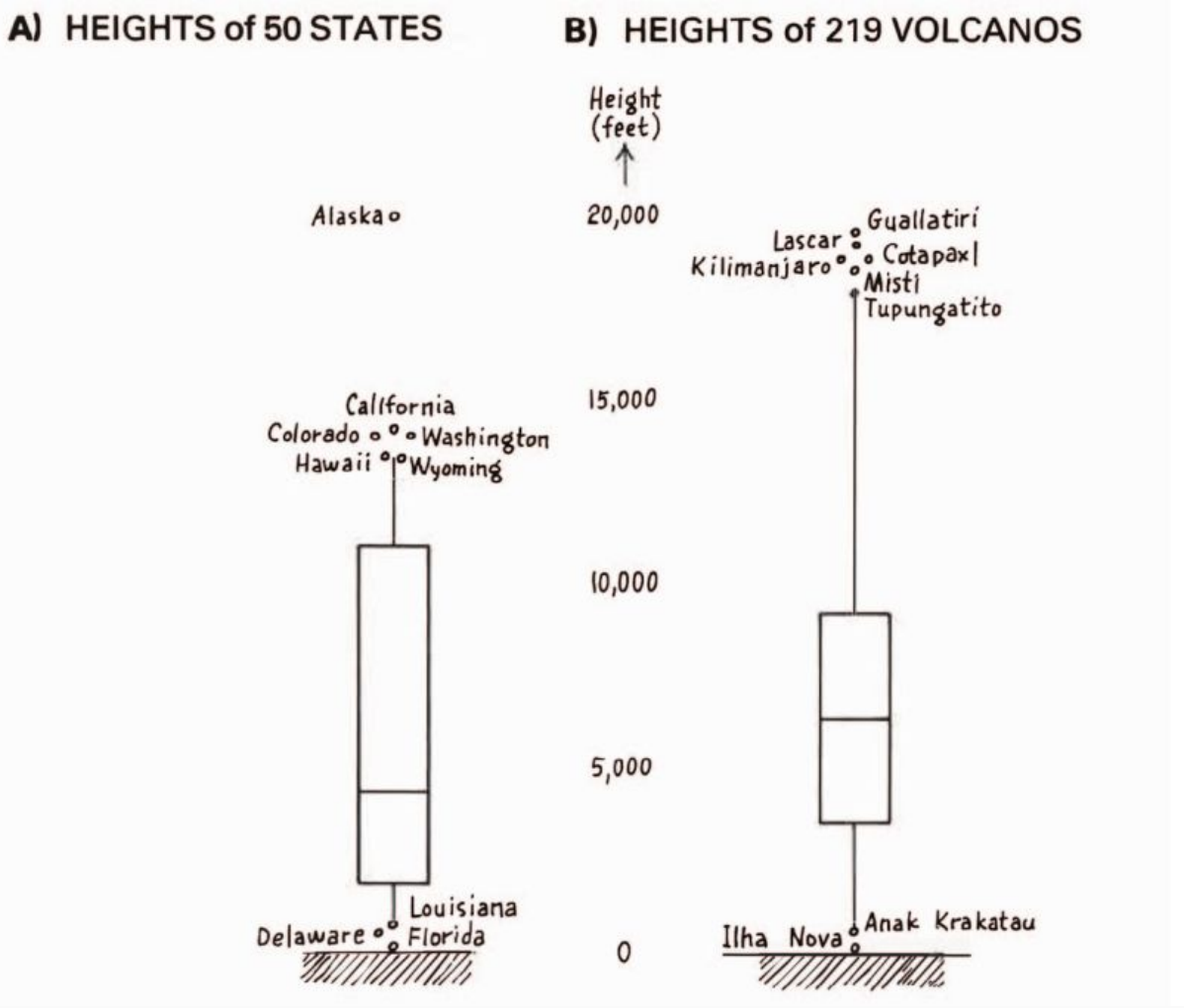






2.A_box and whisker

When you visualize the distribution of your data, you can show the entire distribution or just specific points within it. The box-and-whisker plot (or boxplot), originally called a schematic plot by its inventor John W. Tukey, uses a box and line markers to show specific percentile values within a distribution. You can also add markers to show outliers or other interesting data points or values. It is a compact summary of the data distribution, though it displays less detail than a histogram or violin chart.



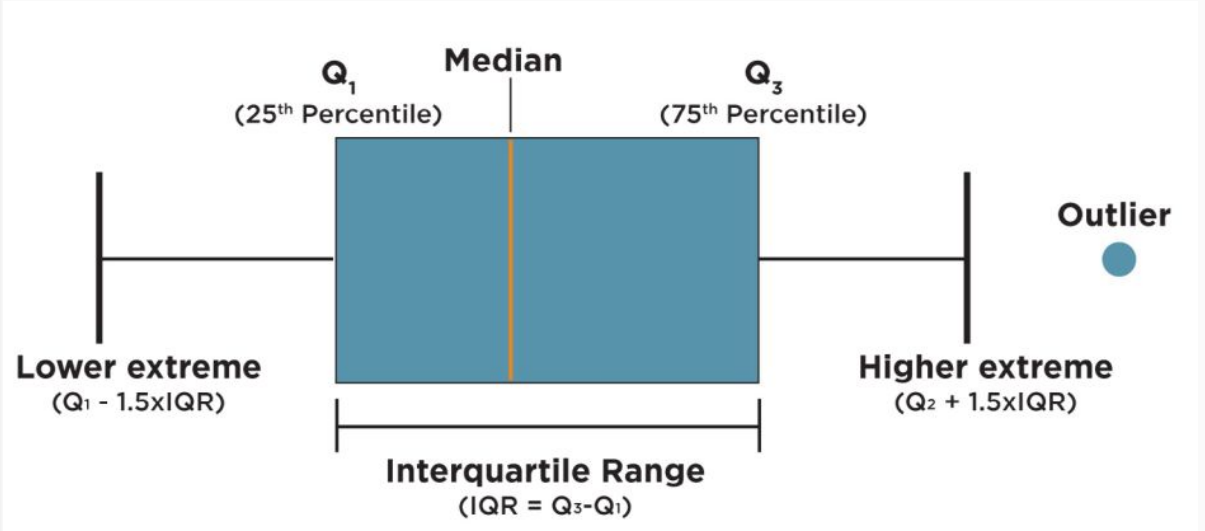
2.B_box and whisker

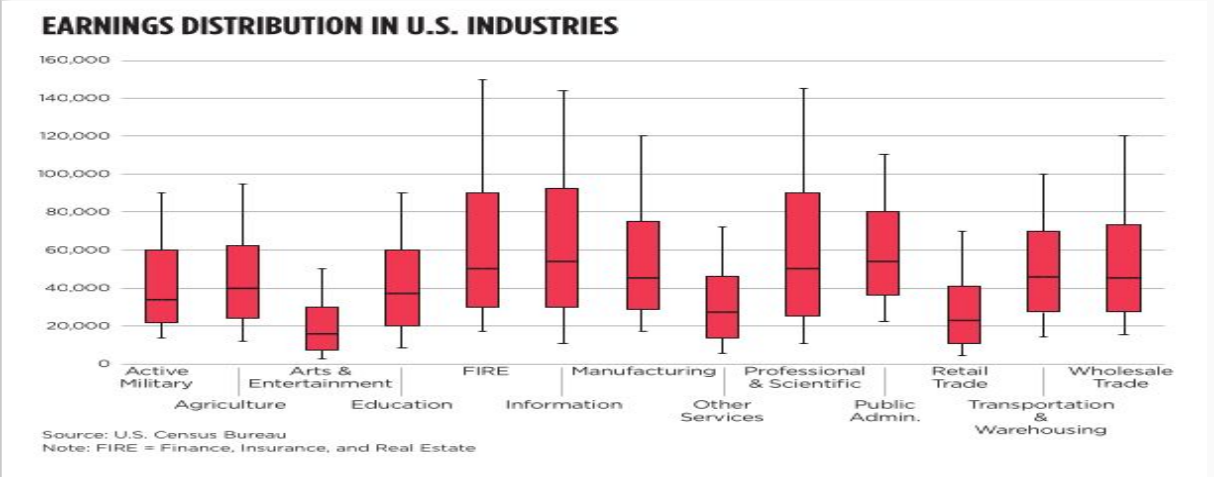
The basic box-and-whisker plot consists of a rectangle (the box), two lines (the whiskers) that emanate from the top and bottom of the box, and dots for outliers or other specific data points. Most standard box-and-whisker plots have five major components

1. The median, encoded by a single horizontal line inside the box
2. Two hinges, which are the upper and lower edges of the box and typically correspond to the first quartile (or the 25th percentile) and third quartile (75th percentile). The difference between these two points is called the Interquartile Range or IQR.
3. The higher and lower extremes (sometimes the maximum and minimum) are placed at a position 1.5 times the IQR (recall the Box on page 74).
4. Two whiskers (the lines) connect the hinges to a specific observation or percentile.
5. Outliers are individual data points that are further away from the median than the edges of the whiskers.

2.C_box and whisker

Each of these components can vary depending on which parts of the distribution we wish to show. Some creators replace outliers with fixed quantiles such as the minimum and maximum values or the 1st and 99th percentiles. Some use the semi-interquartile range (Q3-Q1)/2, which can generate asymmetric whiskers





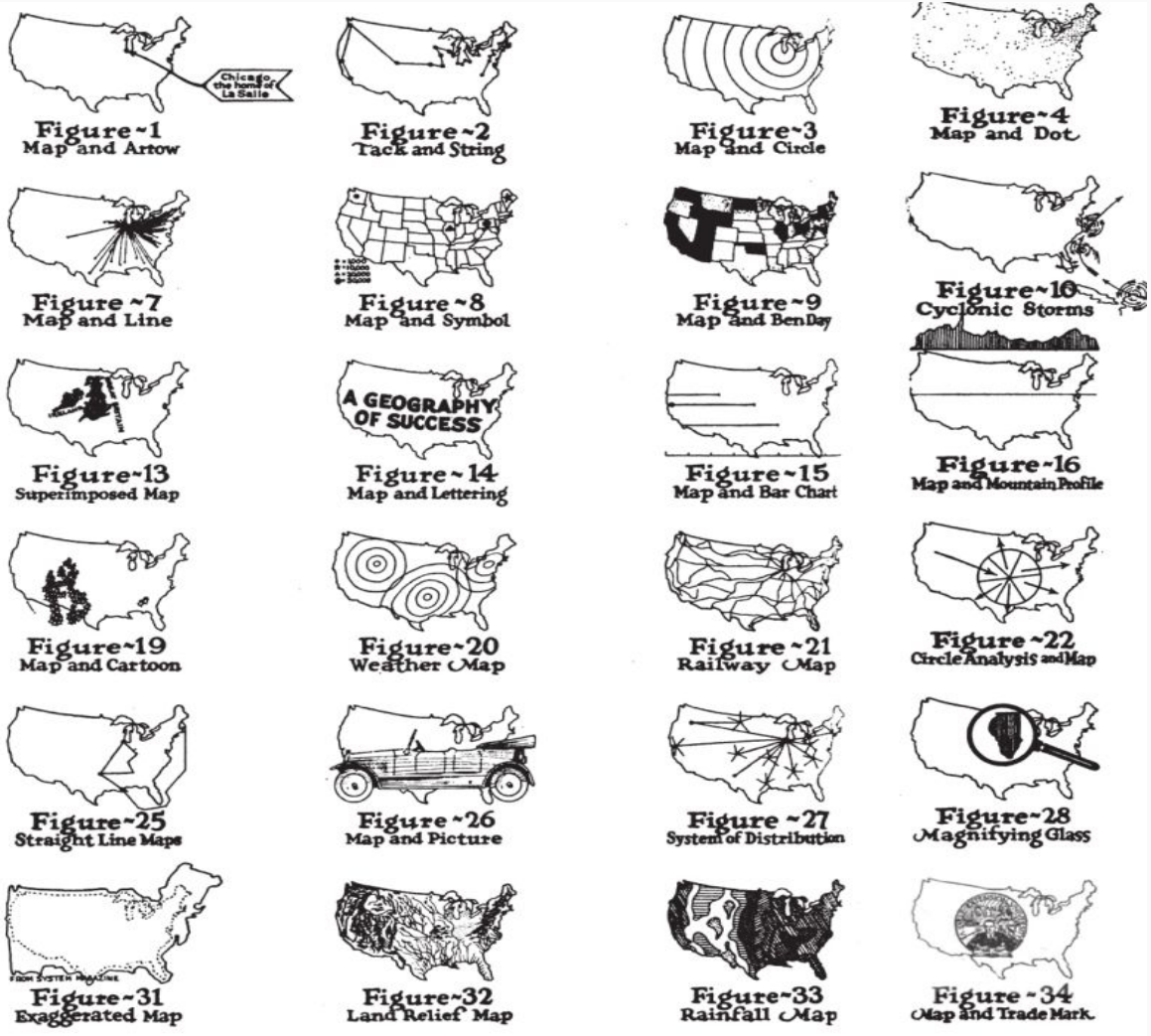
GEOSPATIAL

There is an obvious advantage to plotting geographic data on a map—people can find themselves in the data. They can literally see themselves in the data, a connection with the subject matter that other visualizations cannot muster. Plotting geographic data may mean adding color to geographic areas likes states or countries, or adding circles, squares, lines, or other shapes on top of a geographic map.

Data-driven maps are not new. In 1922, the “Maps and Sales Visualization” on the next page shows the reader thirty-six different ways to place data on a map

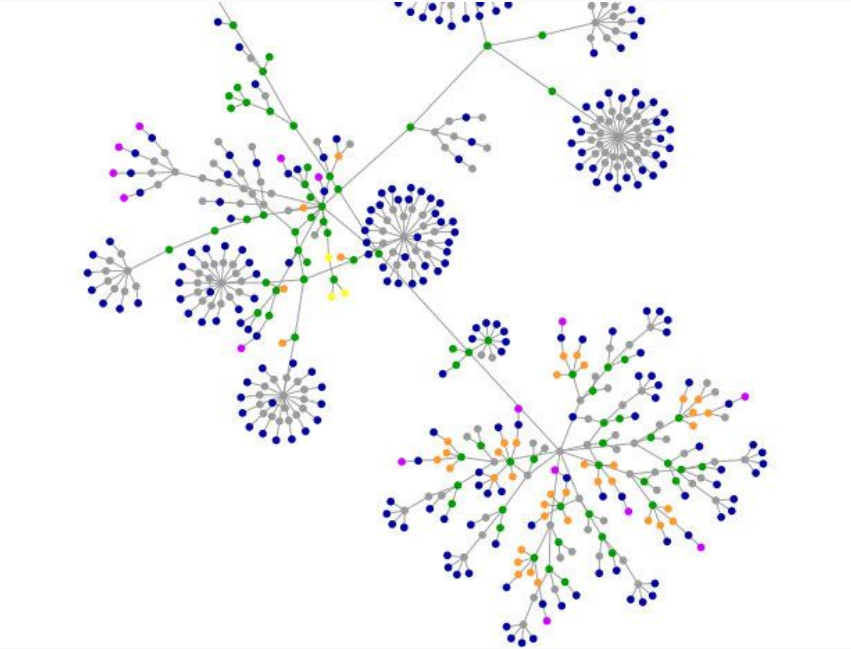
1.CHALLENGES OF MAPS

Due to the many problems in maps, especially converting the globe into two dimensions, and how to determine places, all of this will be in more detail in the other presentation in more detail.



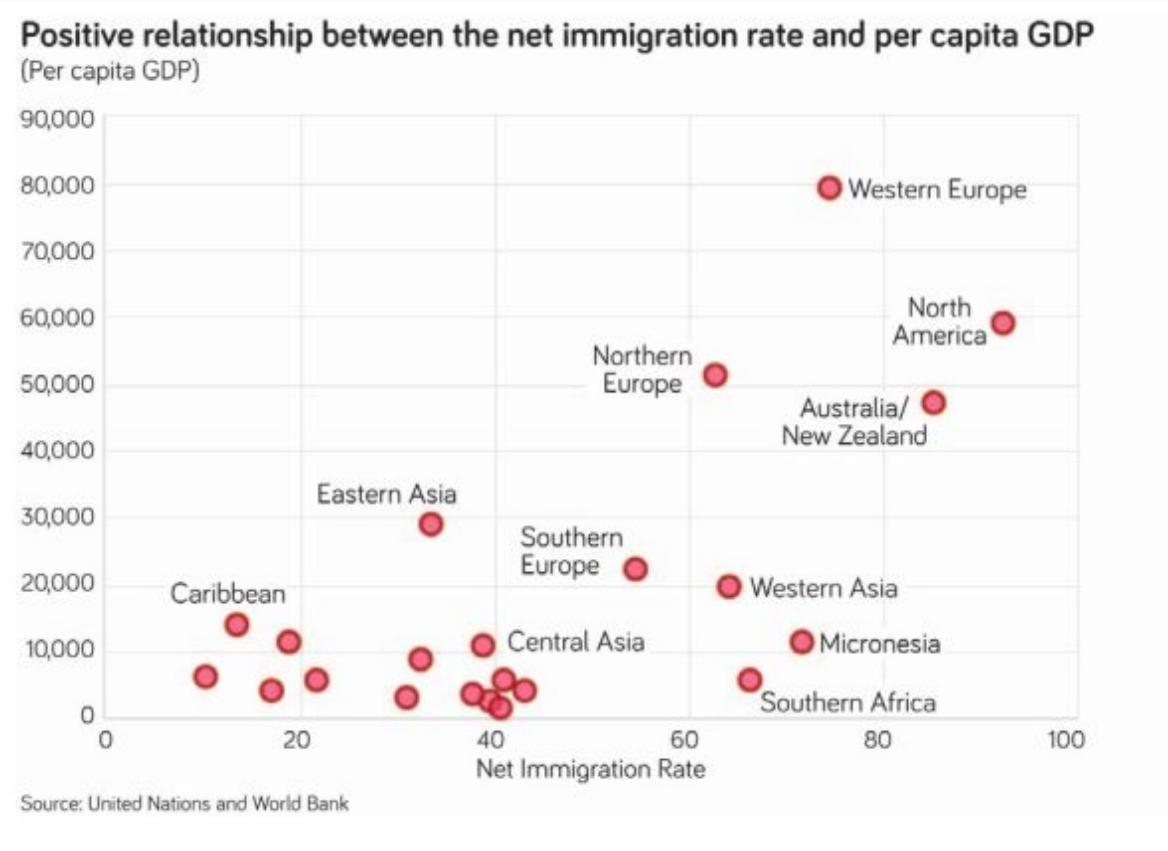
RELATIONSHIP

The charts in this chapter show relationships and correlations between two or more variables. Perhaps the most familiar chart type in this class is the scatterplot, a chart in which the data are encoded to a single horizontal and vertical axis. Other shapes and objects can also be used to visualize the relationship between two or more variables—a parallel coordinates plot uses lines, while a chord diagram uses arcs within a circle. These charts can show the reader correlations and even causal relationships.



1_ SCATTERPLOT

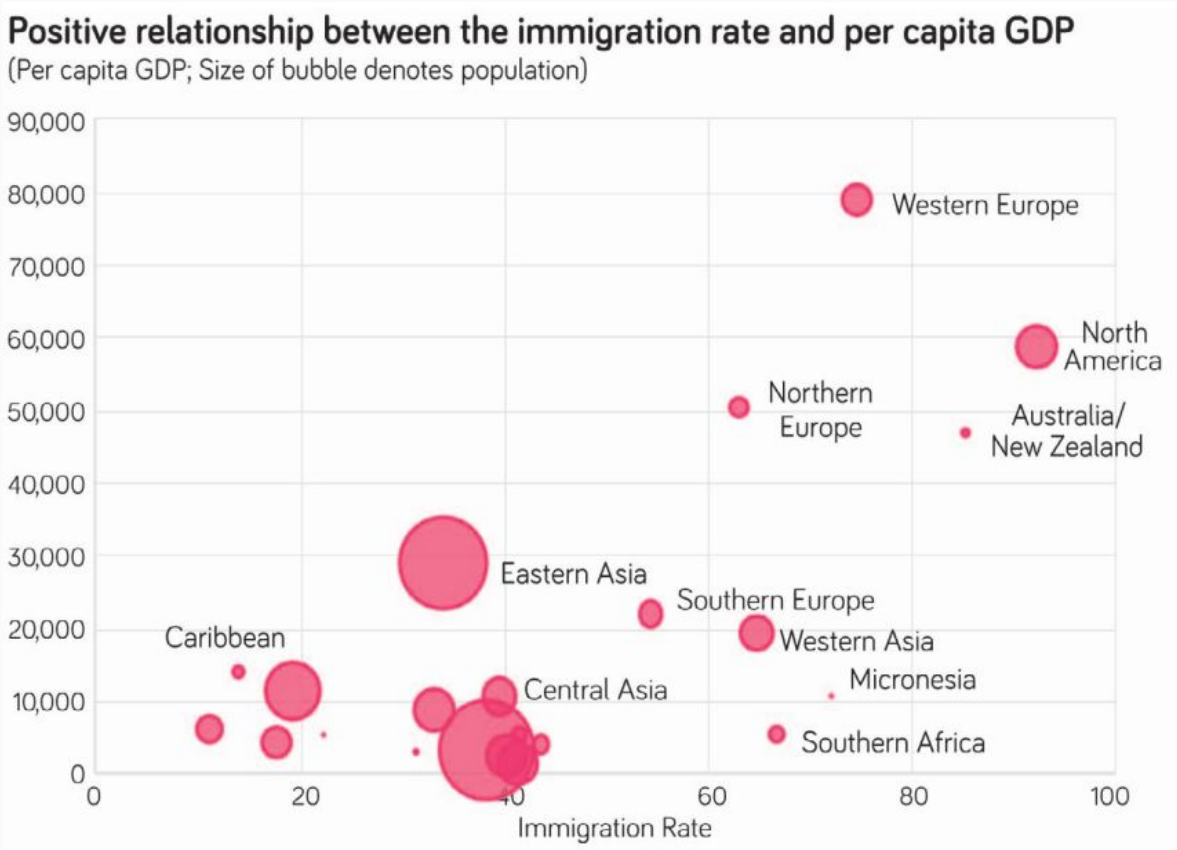
The scatterplot is perhaps the most common visualization to illustrate correlations (or lack thereof) between two variables one variable is plotted along a horizontal axis, and the other along a vertical axis. The specific observations are plotted in the created space. Unlike a bar chart, the scatter plot axes do not necessarily need to start at zero, especially if zero is not a possible value for the data series.



2_ BUBBLE PLOT

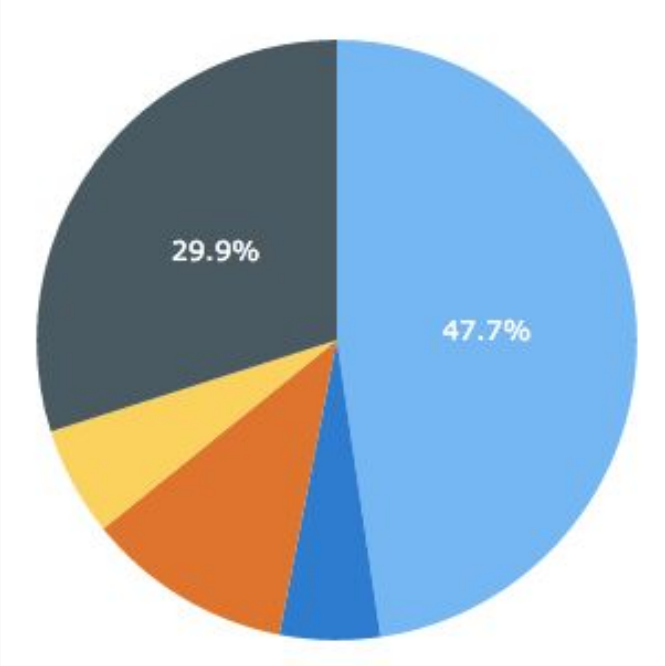
The scatterplot can be transformed into a bubble plot (or bubble scatterplot) by varying the sizes of the circles according to a third variable. The data points don't have to be circles, they can be any other shape that doesn't distort our perception of the data

As mentioned in the section on bubble charts, the circles should be sized by area, not radius . Color can help group or highlight certain points or direct the reader's attention to different parts of the graph.



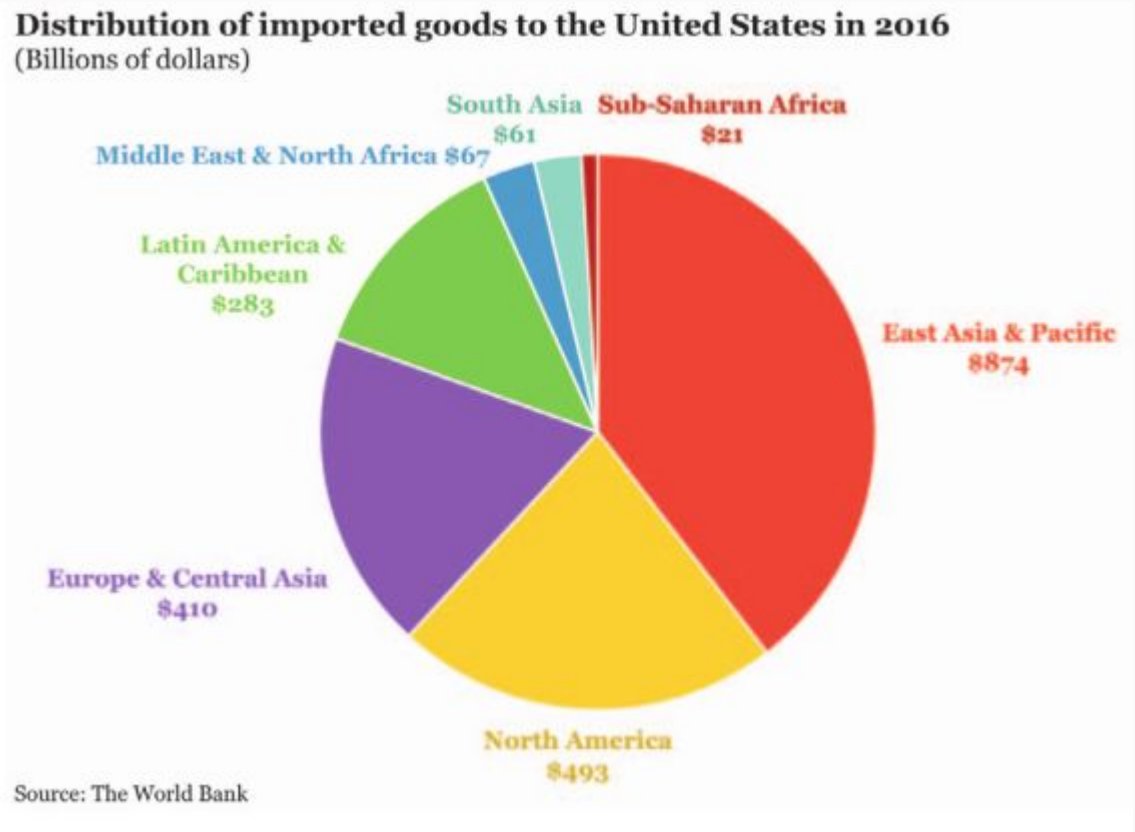
PART OF WHOLE

This class of charts shows how the shares of some amount relate to the total. The most popular and familiar graph in this class is the pie chart, which introduces a variety of perceptual challenges, as we'll see. Other charts in this class like the treemap and sunburst have different perceptual issues and, as always, we must ask ourselves whether we must show all of the components and how they sum to the total. Graphs in this chapter can also be used to visualize hierarchical data—data that can be grouped into layers where natural groups exist which we have already seen in cases such as the tree diagram



1_ PIE CHARTS

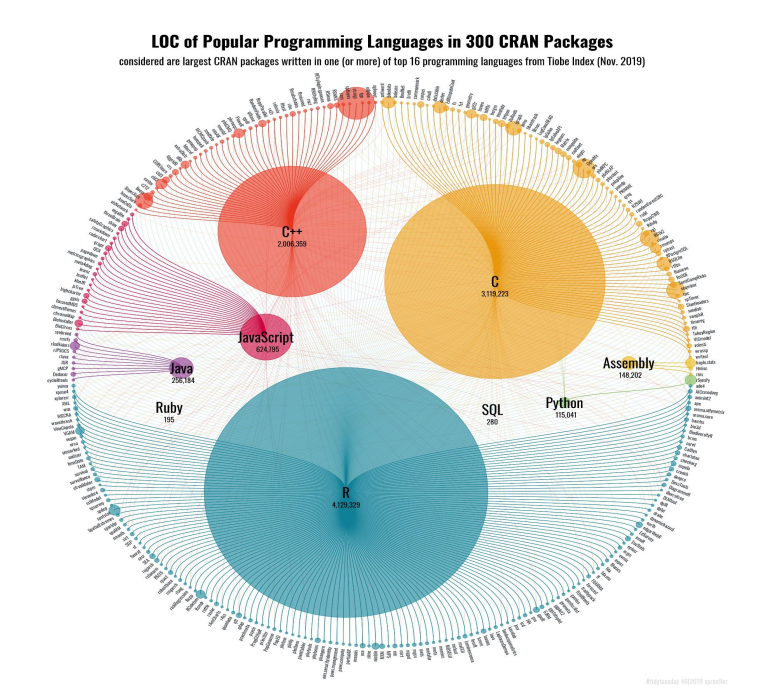
Disdain for pie charts pervades the entire field of data visualization. The most often-cited reason is that pie charts are a poor visualization choice because we have a hard time discerning exact quantities when they are visualized as slices of the pie. If we return to the perceptual ranking chart, we'll see that pie charts fall below the middle of the ranking not good choice more details in second slides



QUALITATIVE

Until now, we've explored charts that mostly communicate quantitative data. But there are also graphs that communicate qualitative data, non-numerical information collected through means of observations, interviews, focus groups, surveys, and other methods. The charts in this chapter primarily communicate words and phrases.

We can build narratives and tell stories around qualitative results in ways that can be more difficult with quantitative data. Downloading a big data set, running regressions, and creating tables may provide more generalizable results, but readers don't connect to them the way they connect to a story. Qualitative data can help tell those stories.



1_ ICONS

Iconography, especially, can help visualize your qualitative data. Icons can be purely decorative, they can represent data , or they can guide the reader from one phase of a visual to the next. Icons (including emojis) are themselves a visual language, so they can simplify and communicate ideas and feelings that are otherwise difficult to express. They may also help readers with certain intellectual or cognitive disabilities engage with your work. A body of research has shown that these kinds of visual-graph forms of communication have helped advance successful language development

Improved infant and maternal health: Researchers have found links between increased EITCs and improvements in infant health indicators such as birth weight and premature birth. Research also suggests receiving an expanded EITC may improve maternal health.

Better school performance: Elementary and middle-school students whose families receive larger refundable credits (such as the EITC and CTC) tend to have higher test scores in the year of receipt.

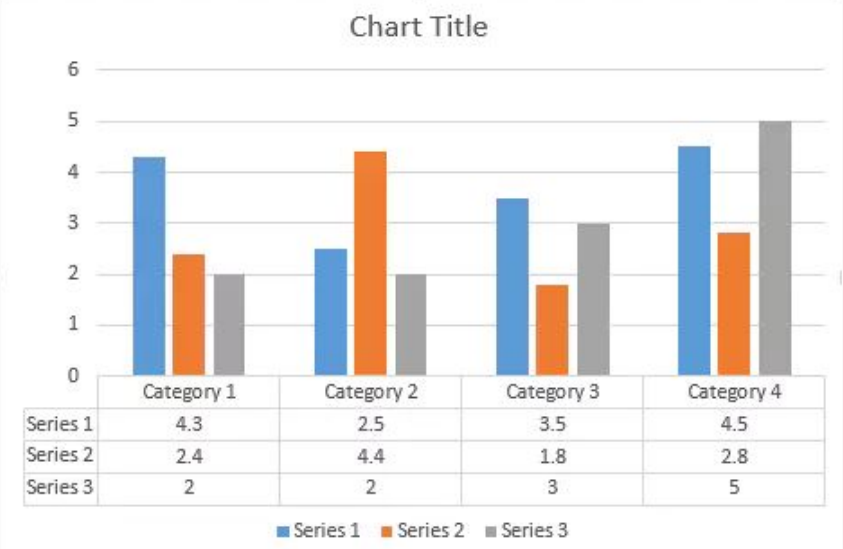
Greater college enrollment: Young children in low-income families that benefit from expanded state or federal EITCs are more likely to go to college, research finds. Researchers attribute this to lasting academic gains from higher EITCs in middle school and earlier. Increased tax refunds also boost college attendance by making college more affordable for families with high-school seniors, research finds.

2_ WORD CLOUDS

Word clouds are perhaps the most popular and familiar way to visualize qualitative data, but they are really a way to display quantitative data: the number of times a word appears in a text. In a word cloud, the size of each word is adjusted according to its frequency

TABLES

Yes, tables are a form of data visualization. If you want to show the exact amount of every value in your data, a table might be your best solution. They are not the best solution if you want to show a lot of data or if you want to show the data in a compact space—but still, a well-designed table can help your reader find specific numbers and discover patterns and outliers.

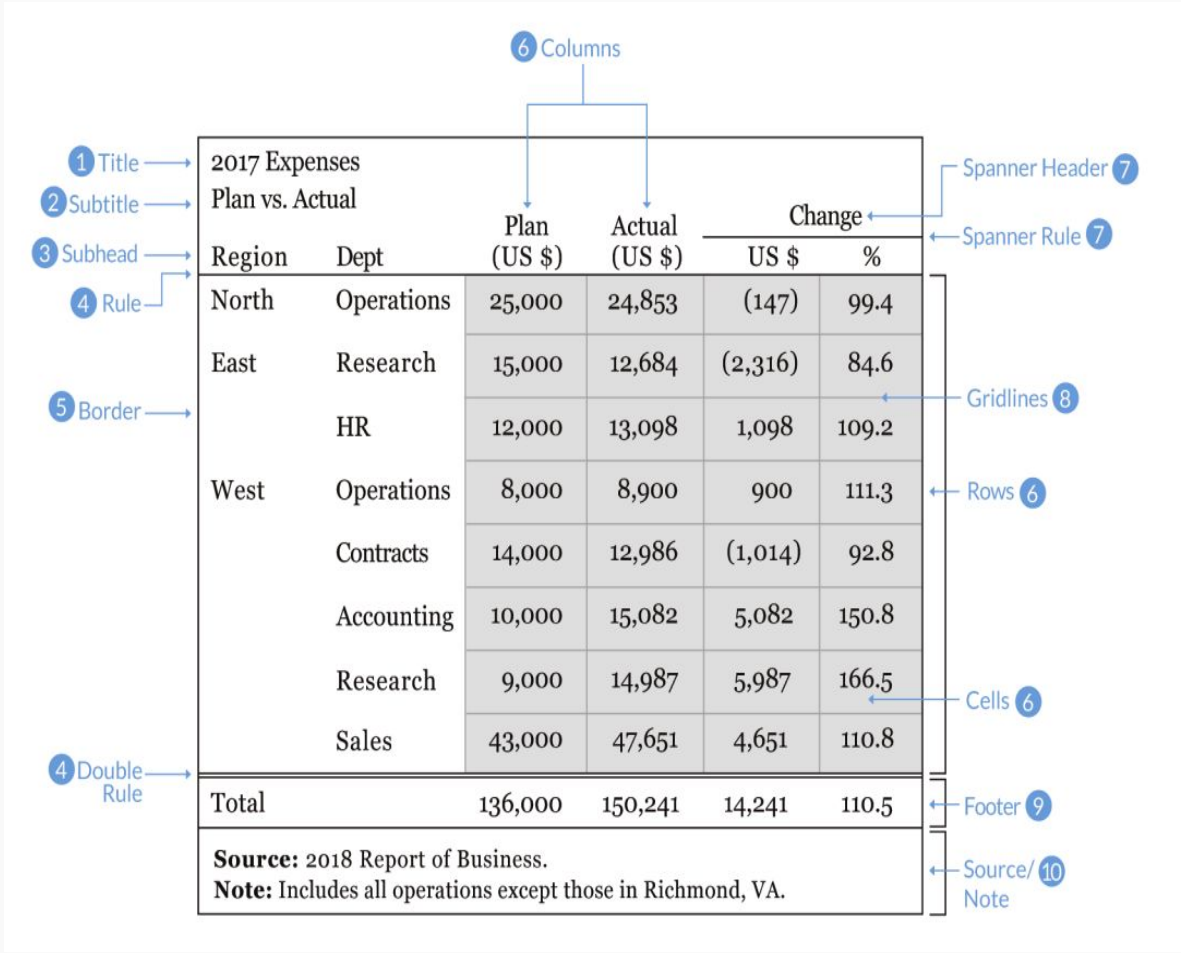


	Category 1	Category 2	Category 3	Category 4
Series 1	4.3	2.5	3.5	4.5
Series 2	2.4	4.4	1.8	2.8
Series 3	2	2	3	5

1_ PROPER ANATOMY OF A TABLE

We must first understand the components of a table before we can understand how and when to adjust them to improve our data presentation. This diagram shows the ten primary components of a table

More in second slide



Thanks!

Contact us:

ahpro001@gmail.com

@Ahmed Haytham