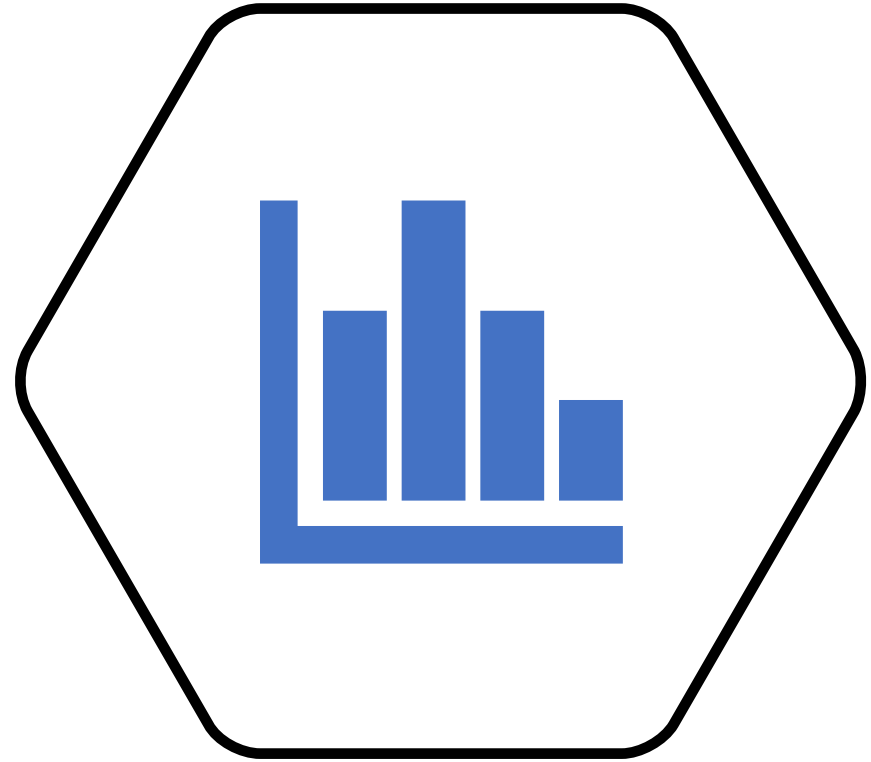
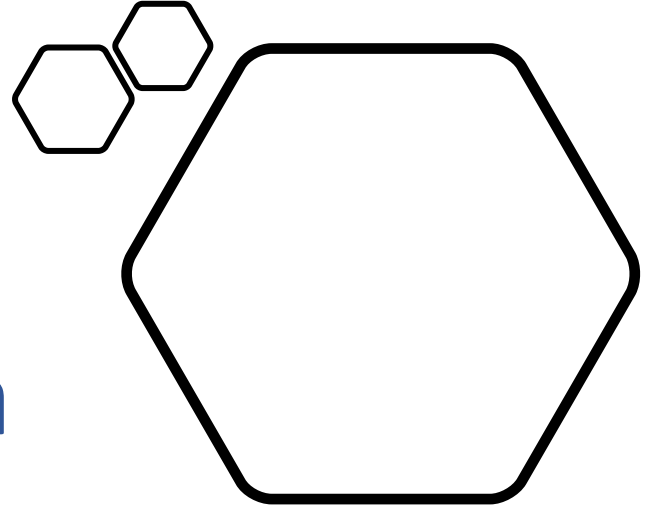


Introduction to Data Visualization



Why do we need data
visualization?



Tesla Stock Price – 2020&2021

Date	Stock Price
1/2/2020	86.052
1/3/2020	88.602
1/6/2020	90.308
1/7/2020	93.812
1/8/2020	98.428
1/9/2020	96.268
...	...
12/28/2021	1088.47
12/29/2021	1086.19
12/30/2021	1070.34
12/31/2021	1056.78

506 rows of data total.

Humans can't read so many numbers and make sense of the data. We need to be able to observe the trends – the price going up or down.

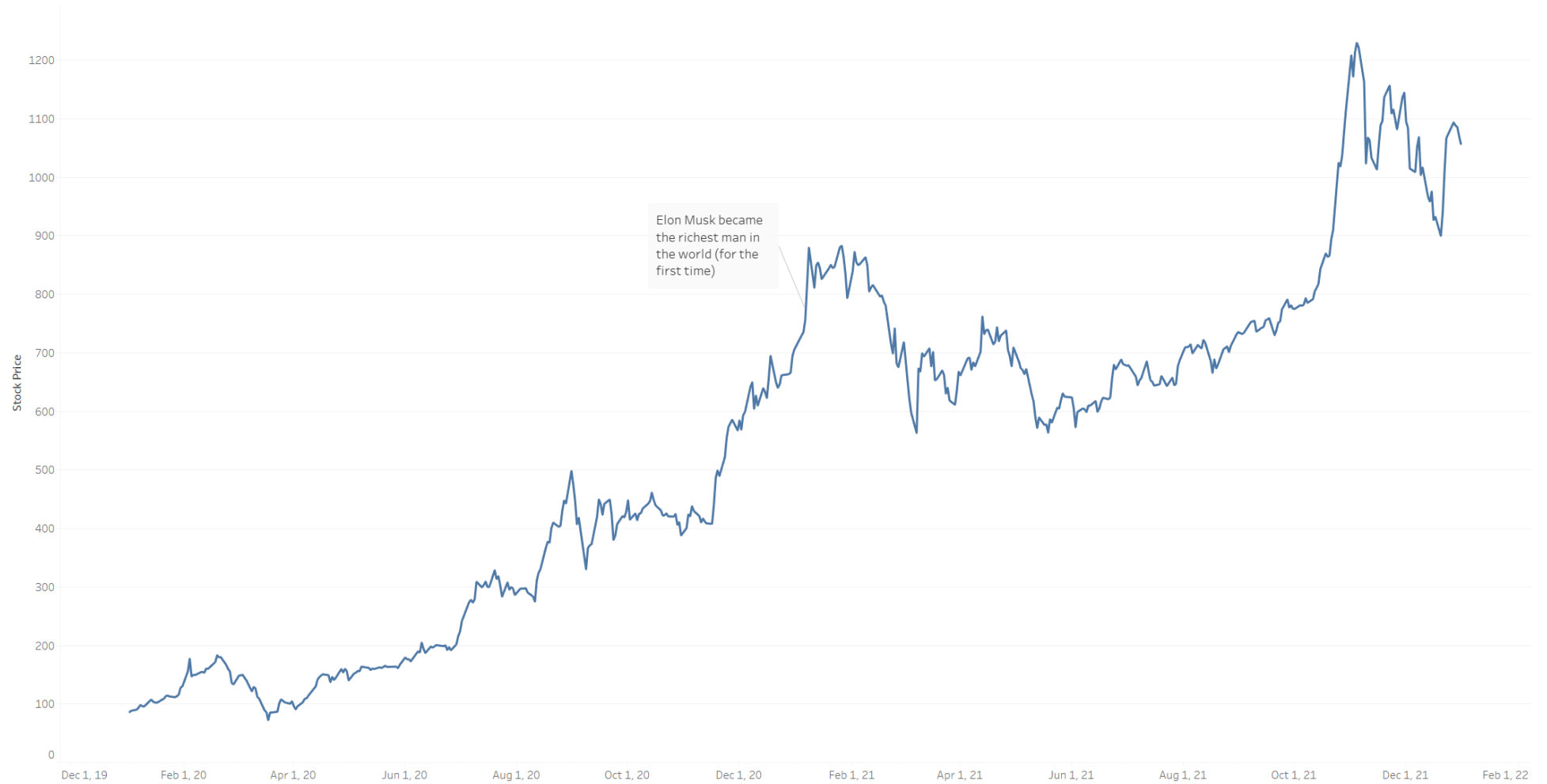
Maybe we want to associate the prices with historic/economic events.

The best way to understand this data is to plot it.

Tesla Stock Prices (2020-2021)



Tesla Stock Prices (2020-2021)



Data visualization gives us a summary of what is important

In the chart examples we are not focusing on the individual stock price values. What is important are the trends – the price going up or down.

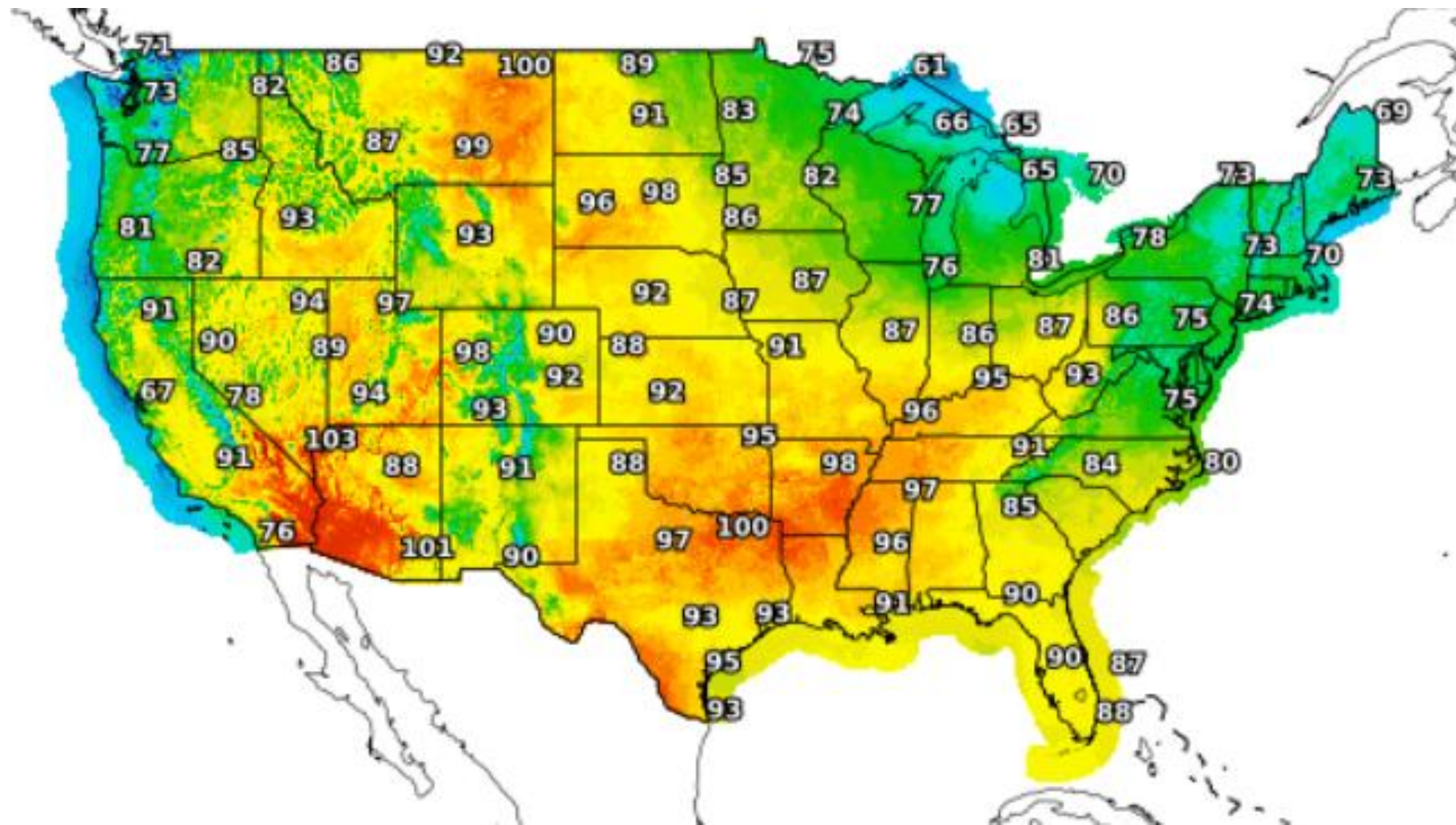
The charts allow us to focus on what we really need to see about the data.

“If I had more time, I would have written a shorter letter.”

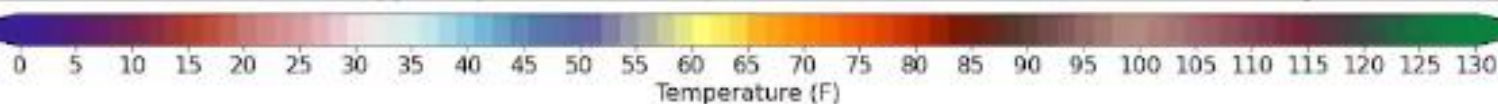
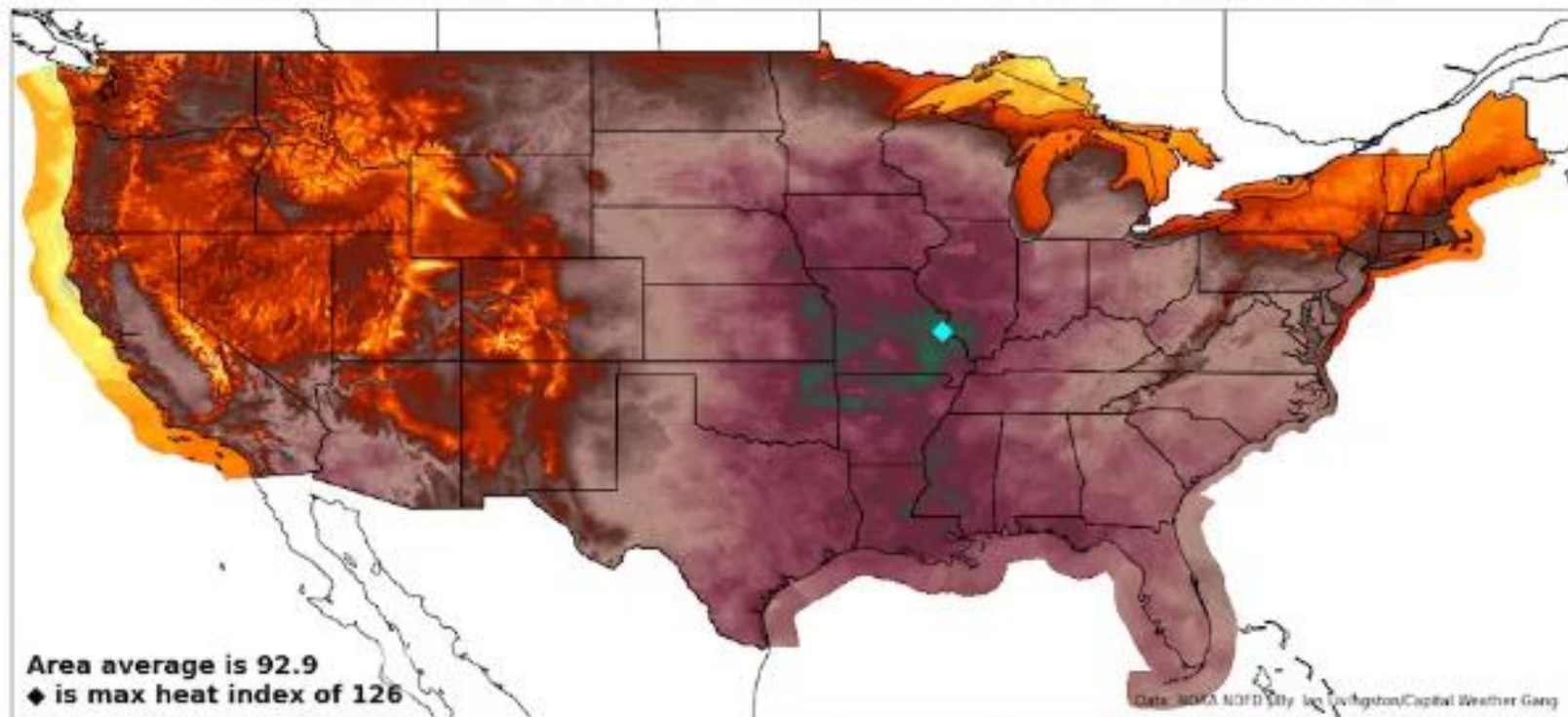
~Blaise Pascal (most likely)

Data visualization applications

- Data visualization is not only important for businesses. In this class we will focus on business questions but there are endless applications for data visualization.
- In fact, everything you see displayed on a screen (even PowerPoint and this slide) is some sort of data visualization.
- The next few pages show us different examples of everyday charts.





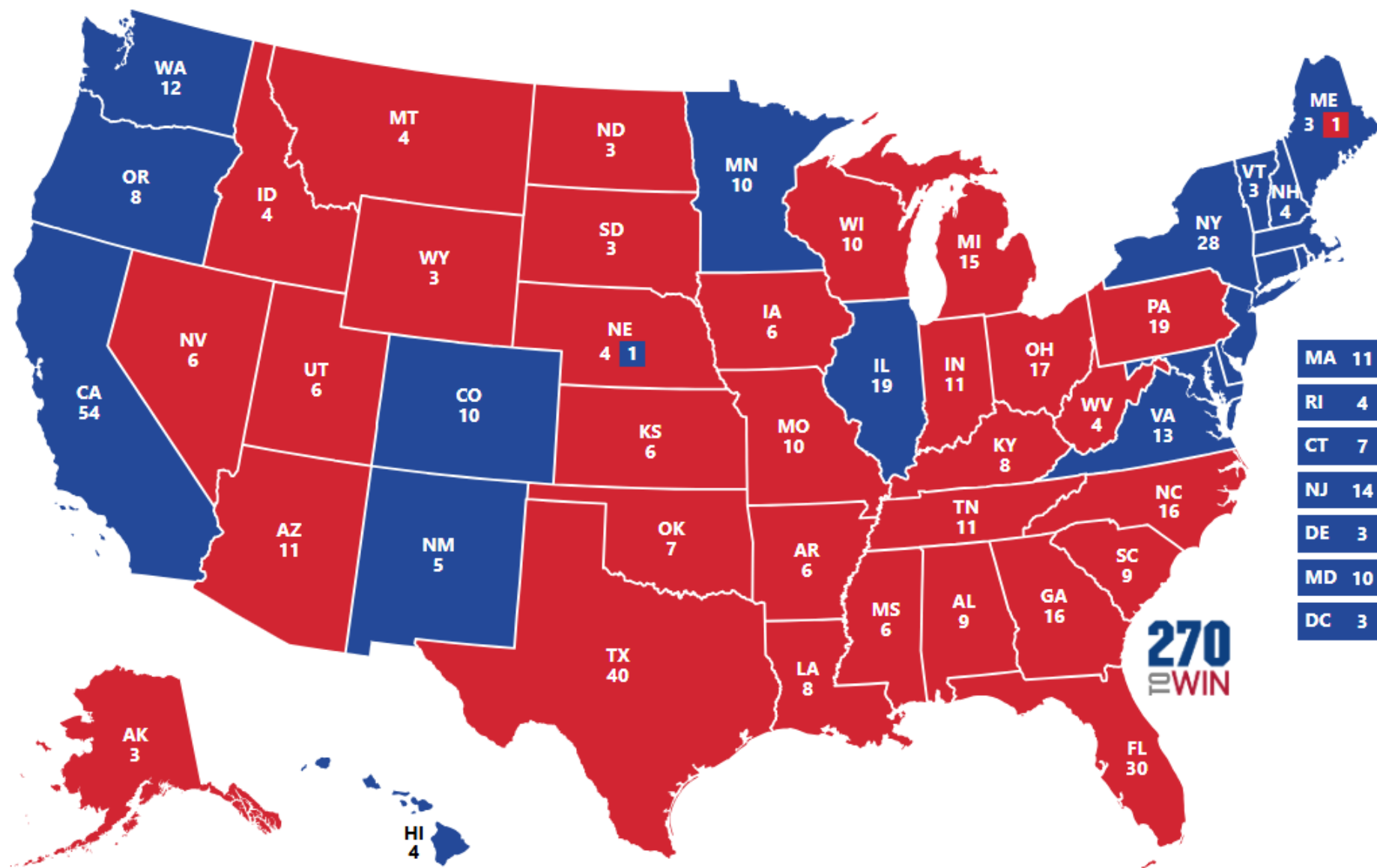
Forecast maximum heat index over the next week



Team

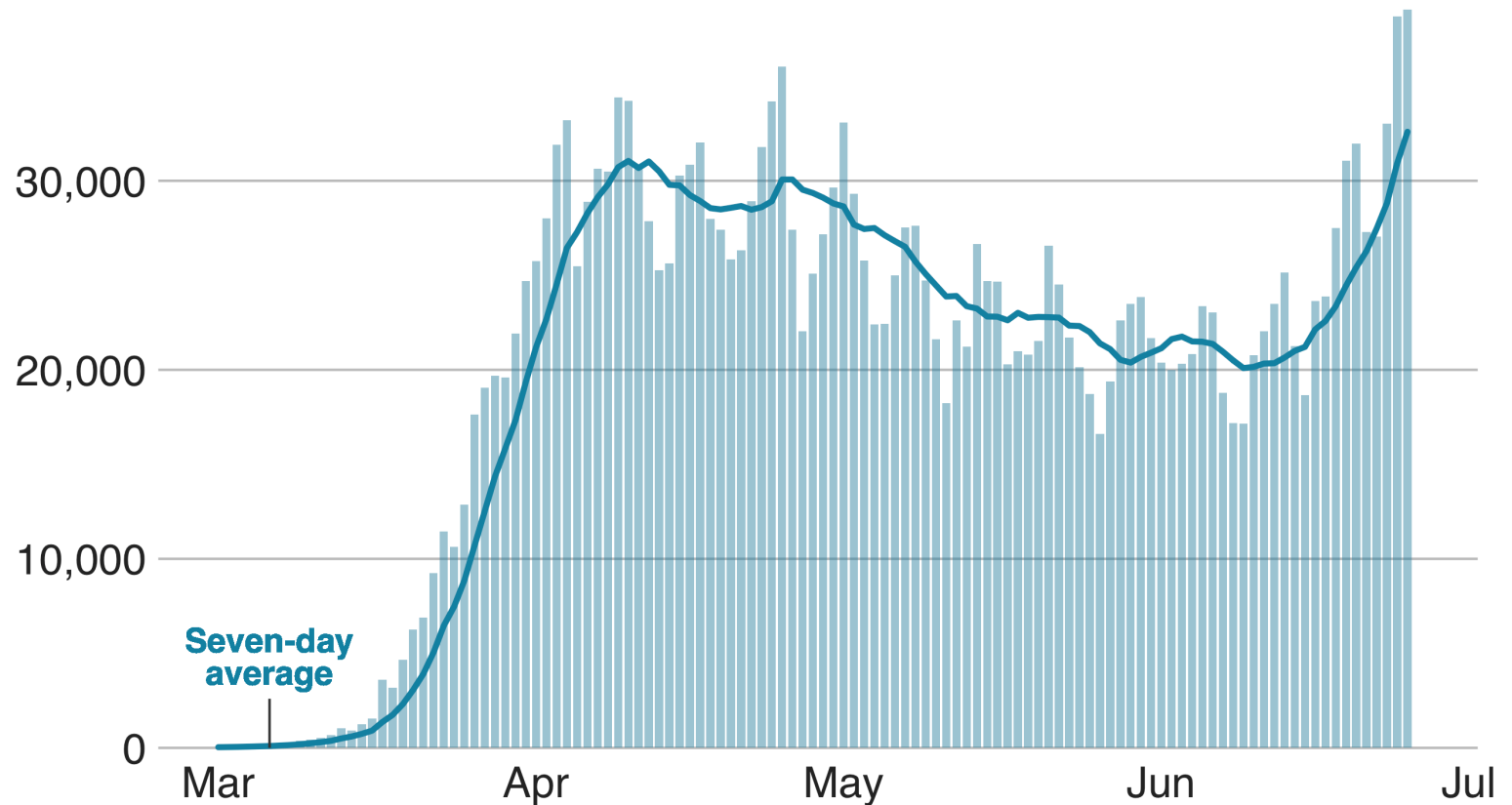
① ② ③

1	 United States	40	44	41	125
2	 China	40	27	24	91
3	 Japan	20	12	13	45
4	 Australia	17	19	16	52
5	 France	16	26	22	64
6	 Netherlands	15	7	12	34
7	 Great Britain	14	21	29	64
8	 South Korea	13	9	10	32
9	 Italy	12	13	15	40
10	 Germany	12	13	8	33



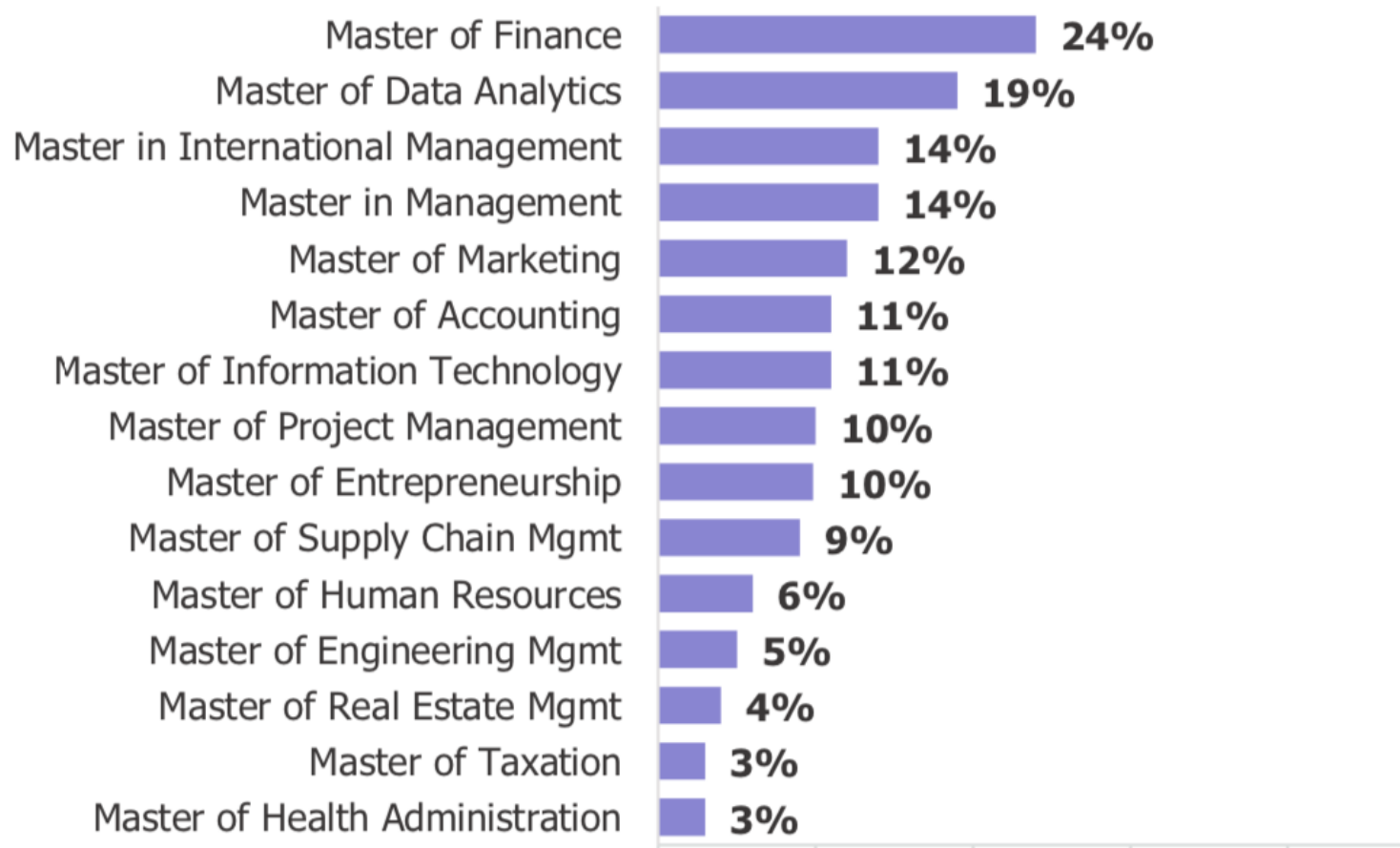
Cases are rising again in the US

Number of daily confirmed coronavirus cases



Source: COVID Tracking Project

Business Master's Program Types Considered



2D



3D



4D



These visualizations use the same data but are targeted at different audiences.

Data visualization for Business

Data visualization for business includes charts (a.k.a graphs or plots), dashboards and stories.

Charts and Dashboard are the focus of this class.

When do business analysts need to make visualizations

1. For formal presentations where the audience is your boss, business partners or other stakeholders at the company.
2. To produce reports about company metrics that update regularly – they can be documents such as PDF files or dashboards.
3. To create a dashboard to monitor certain aspect(s) of the business.

In these 3 situations you are creating a product for someone else (an external audience). Showing them good visualizations will make you look good.

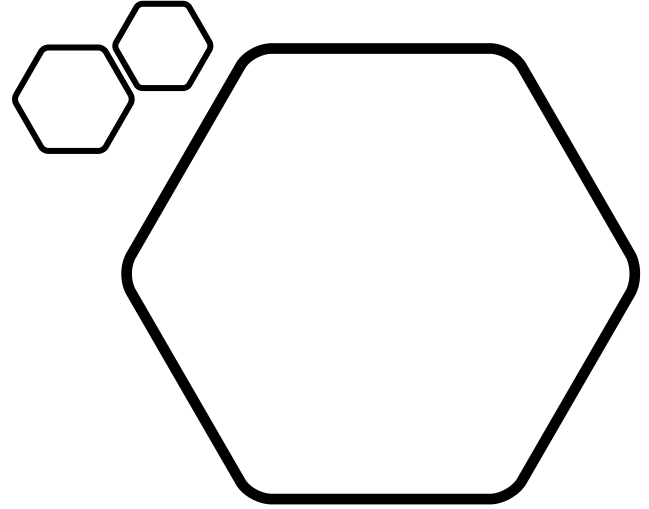
When do analysts need data visualization skills

4. As part of the process of analyzing the data.
 - to produce a summary of the data so you can strategize a solution.
 - to look for patterns, outliers and trends.
 - to understand the results of some algorithm.

Steps for visualizing data

1. Know what data you want to show.
2. Based on the data, choose the best type of chart to show the data.
The type of data determines the type of charts you can use.
3. Style the visualization.

Types of Data



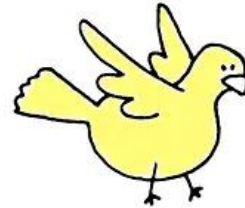
Main types of data for visualization

- **Categorical data** - text (string) values (sometimes called dimensions).
- **Numeric data** – integers or floating points #s
- **Dates** (or date and time)
- **Boolean values** – True/False or 0/1
- **Geo data** – used to plot on a map

Text (string) data

Usually used to describe categories or categorical features. Categories (or categorical data) is data that cannot be quantified. Depending on what you want to show sometimes geo locations, dates or even numbers can be treated as categories.

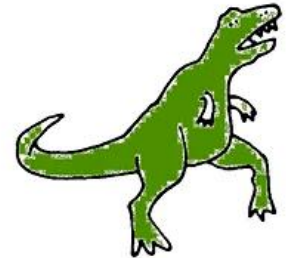
CATEGORICAL DATA:



I am a bird.
I am yellow.
I am awesome.



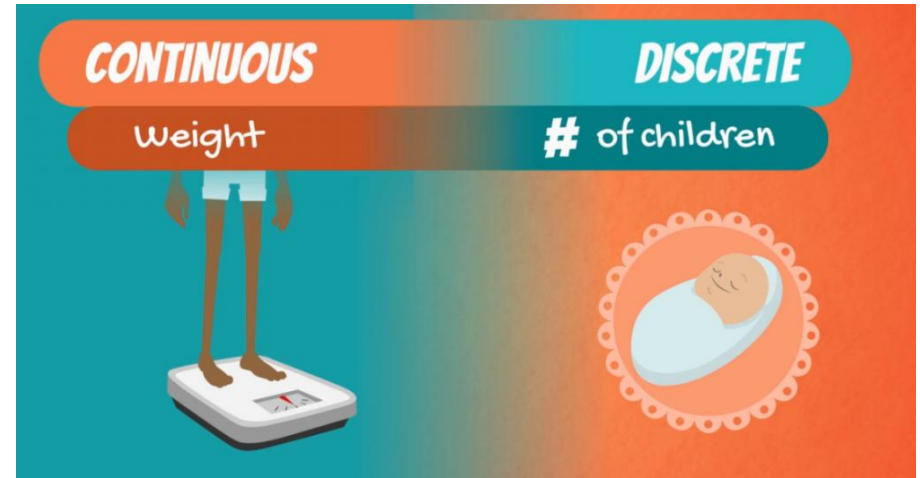
I am a seahorse.
I am orange.
I am super awesome.



I am a T-rex.
I am green.
I am extinct.

Numeric data

- Values that you can arrange in increasing or decreasing order. They can be integers or floating-point numbers. The numeric values are meant to be compared or included in calculations/aggregations.
- Most often it makes a difference if the numerical data is **discrete** or **continuous**.



Dates

Dates are very commonly used because we are often interested in seeing how data changes over time. Dates can be in any format – year, mm/yy, mm/dd/yyyy, quarters, etc.

In some cases, adding specific time to the date (hour, minutes, seconds) is also important.



Boolean values (True = 1, False = 0)

Boolean values are yes/no values. They are flags usually used for filtering or counting number of occurrences.

Geo data



Geo data allows us to create map charts.

Geo data can be latitude and longitude coordinates. They essentially act as x-y values on a flat map.

Alternatively, geo data can be country names, states (if within the US usually), cities, zip codes ... It can go all the way down to a street address.

What types of data do we see here

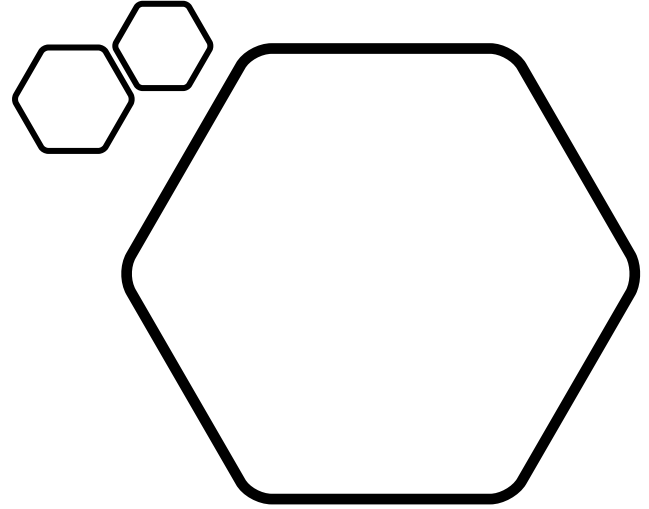
Result Table					
customer_numb	first_name	last_name	order_numb	order_date	order_total
002	John	Doe	1001	10/10/09	250.65
002	John	Doe	1002	2/21/10	125.89
003	Jane	Smith	1003	11/15/09	1597.99
004	John	Smith	1004	11/22/09	180.92
004	John	Smith	1005	12/15/09	565.00
006	John	Jones	1006	11/22/09	25.00
006	John	Jones	1007	10/8/09	85.00
006	John	Jones	1008	12/29/09	109.12

What types of data do we see here

X.locale	Address.CountryName	Address.Latitude	Address.Longitude	Address.Subdivision	ID	IsDaylightSavingsTimeRecognized	LocationMilestones.OpenDate	PhoneNumber
en-US	United States	33.608825	-85.783182	AL	2153	TRUE	2006-07-19T12:00:00	(256) 231-2900
en-US	United States	32.618108	-85.410636	AL	1499	TRUE	2005-07-20T12:00:00	(334) 705-0152
en-US	United States	33.55892	-85.076473	GA	1073	TRUE	1997-07-23T12:00:00	(770) 838-9221
en-US	United States	33.392907	-84.758875	GA	1223	TRUE	1999-10-06T12:00:00	(770) 502-0294
en-US	United States	32.534649	-84.971073	GA	1179	TRUE	1999-03-03T12:00:00	(706) 321-0831
en-US	United States	33.436865	-84.590146	GA	2129	TRUE	2006-07-19T12:00:00	(770) 282-2165

Short Break

Types of Charts



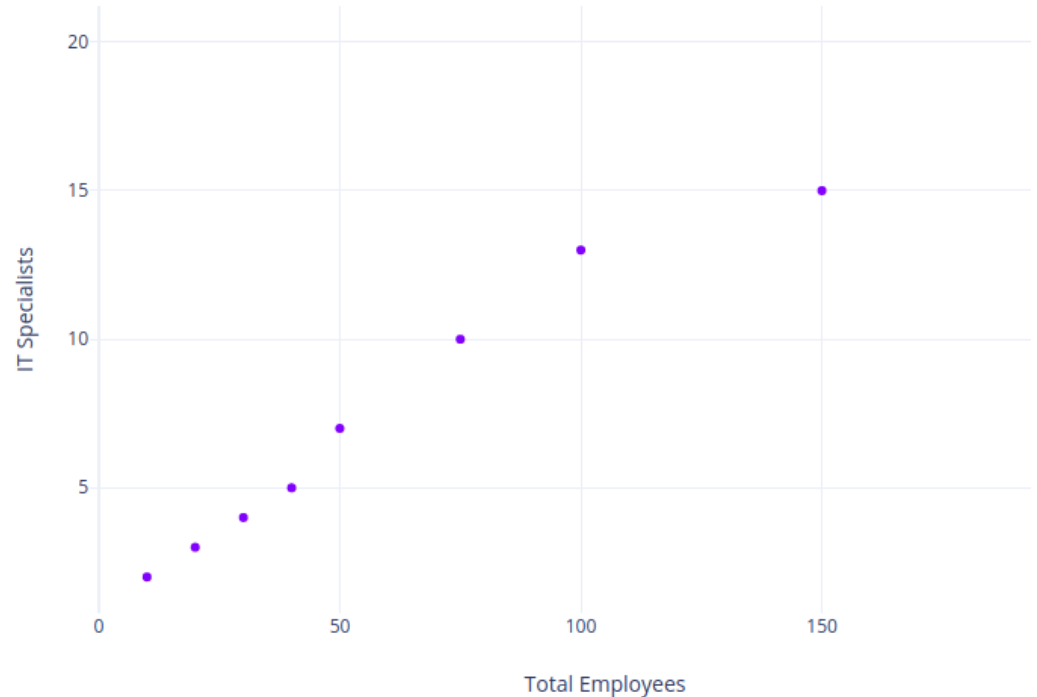
Most popular charts

- Scatter plot
- Line plot (line graph)
- Bar chart – vertical, horizontal, grouped, stacked, 100% stacked
- Area chart
- Histogram
- Heatmap
- Bubble plot
- Map charts – there are different types of map charts
- Pie chart

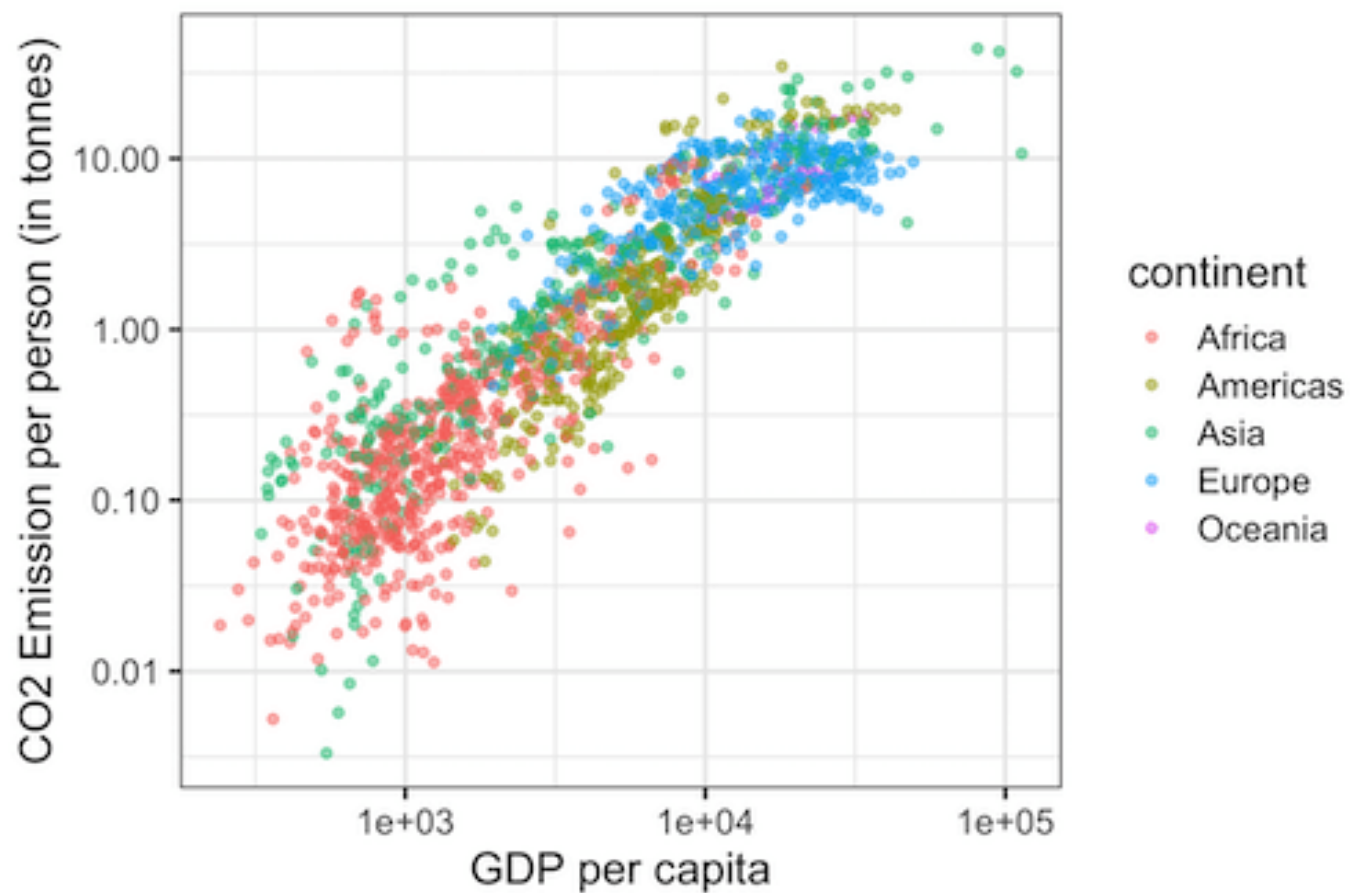
Scatter Plot

- Best shows the relationship between **two numeric variables**.
- Very useful when we want to see if two variable are correlated – if one goes up does the other go up, down or there is no pattern.

Optimal number of IT specialists



CO2 emission per person vs GDP per capita



Line Graph

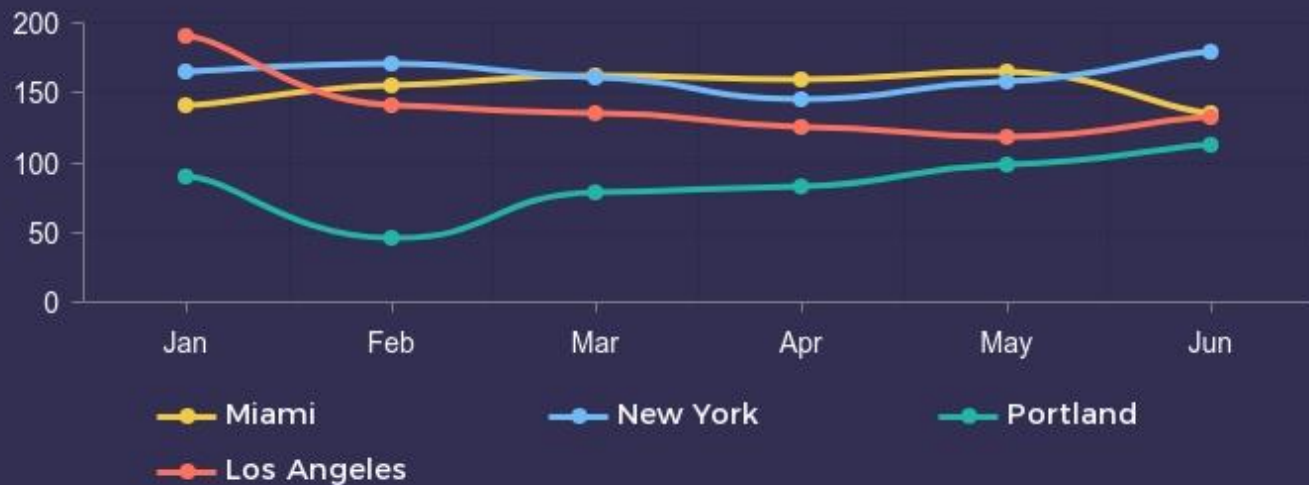
Tesla Stock Prices (2020-2021)



- Most often used to plot a time series – a numeric value that changes over time. So, typically we need a **date field** and a **numeric field**.
- Can show single or multiple time series

Sales by store

TECHNOLOGY

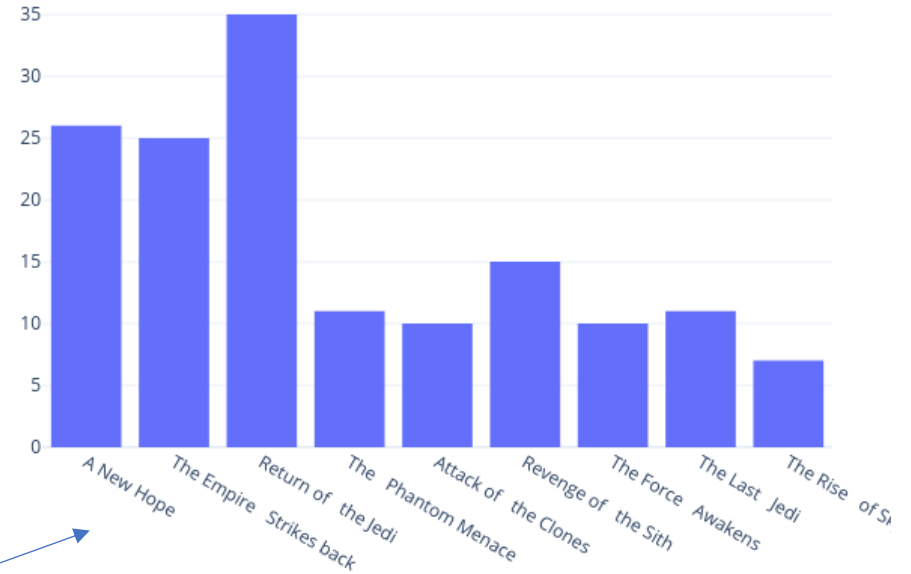


Bar Chart

Used for visualizing a **numeric value associated with a category**.

Shows relationships between groups.

Number of people who rate each movie as the best Star Wars movie ever

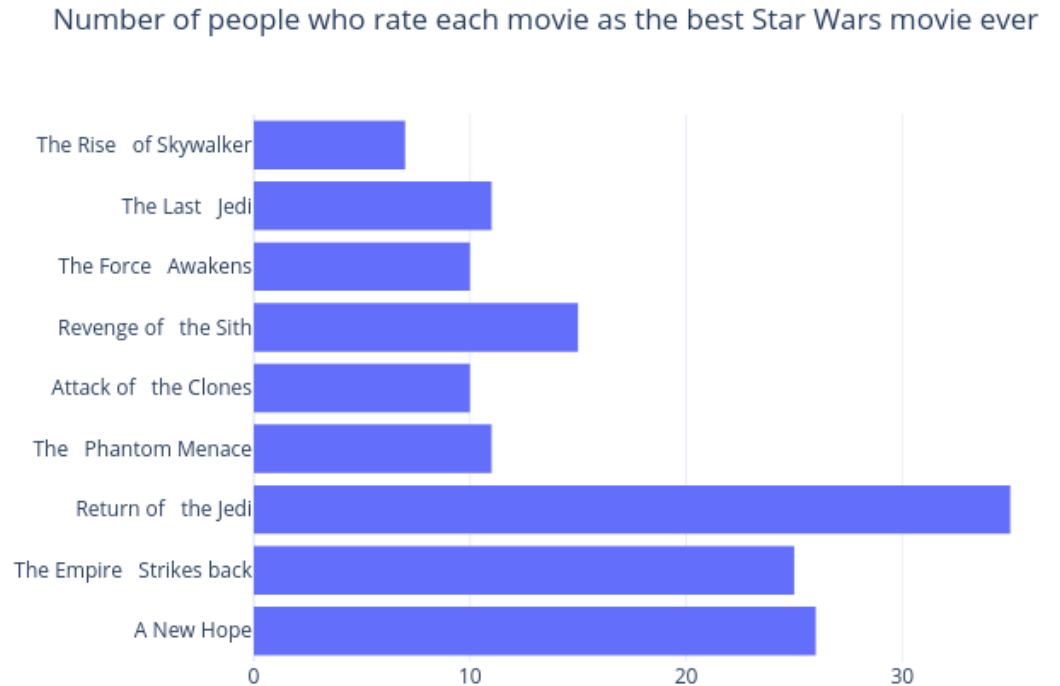


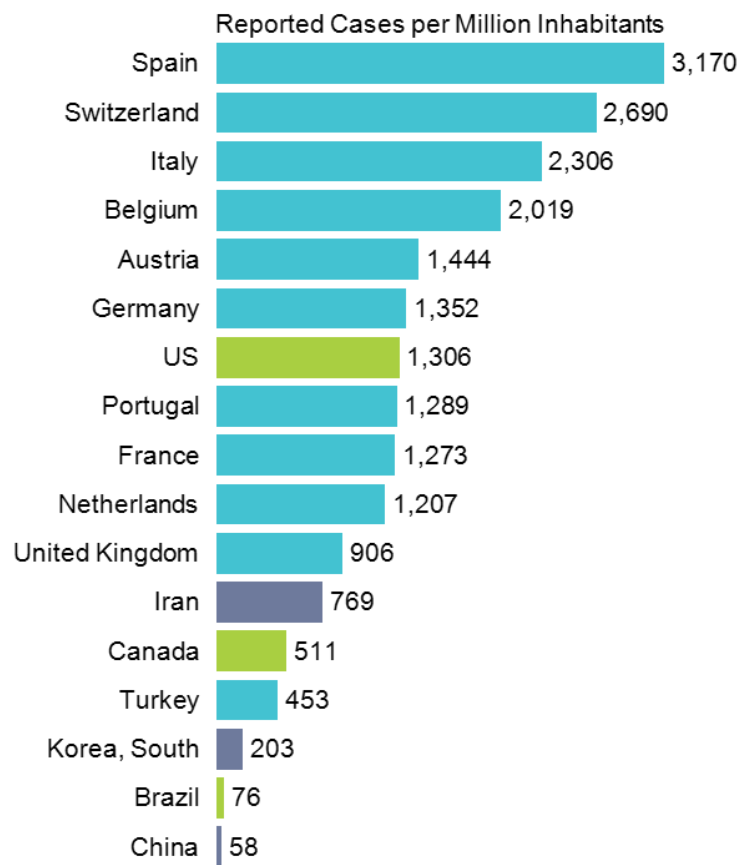
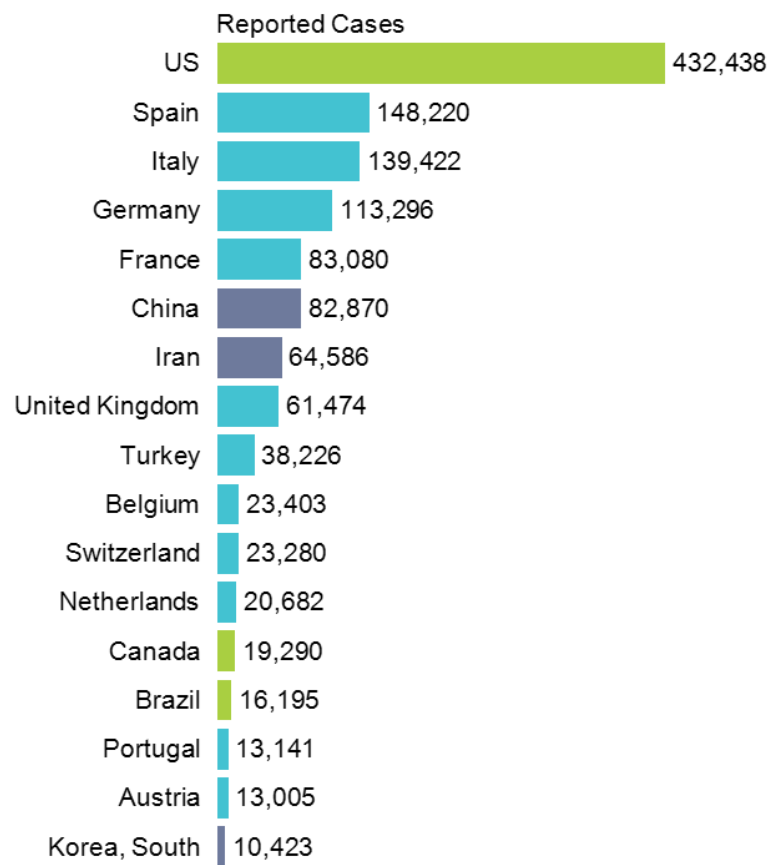
Avoid angled text

Horizontal Bar Chart

Same as the bar chart before but the axes (x-y) are flipped.

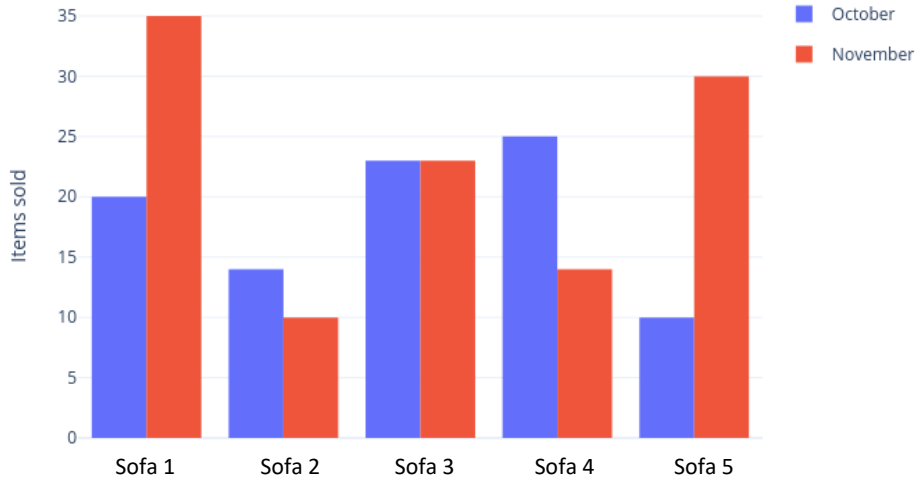
Best used for data with long category labels.





Grouped Bar Chart

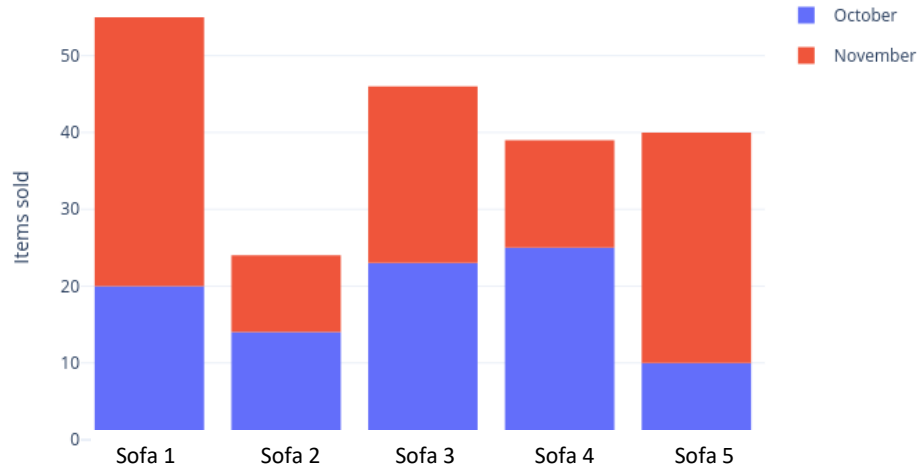
Sofa Sales per Model for October and November



Shows data for different sub-groups.

Stacked Bar Chart

Sofa Sales per Model for October and November

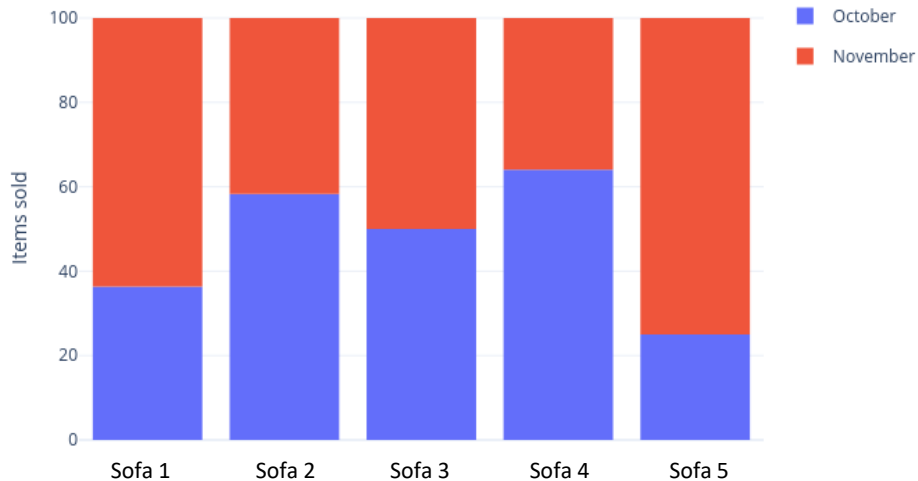


When you want to see the data as part of the total.

Here we see the total sales for sofas A, B, C, D and F.

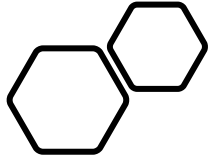
100% Stacked Bar Chart

Sofa Sales per Model for October and November



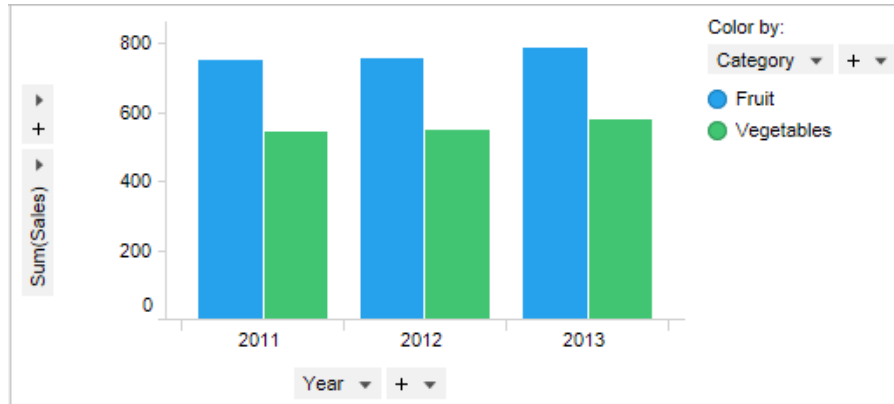
When you do not need to show the actual values, just the ratios.

Here we cannot tell that sofa A sells better than B. This information is lost but the comparison between October and November sales is emphasized.

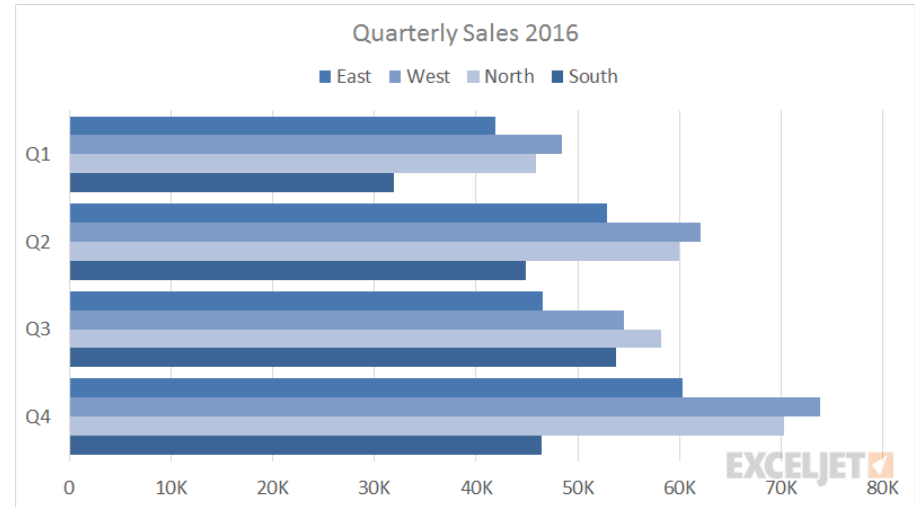


You can convert all versions of
vertical bar charts into
horizontal bar charts

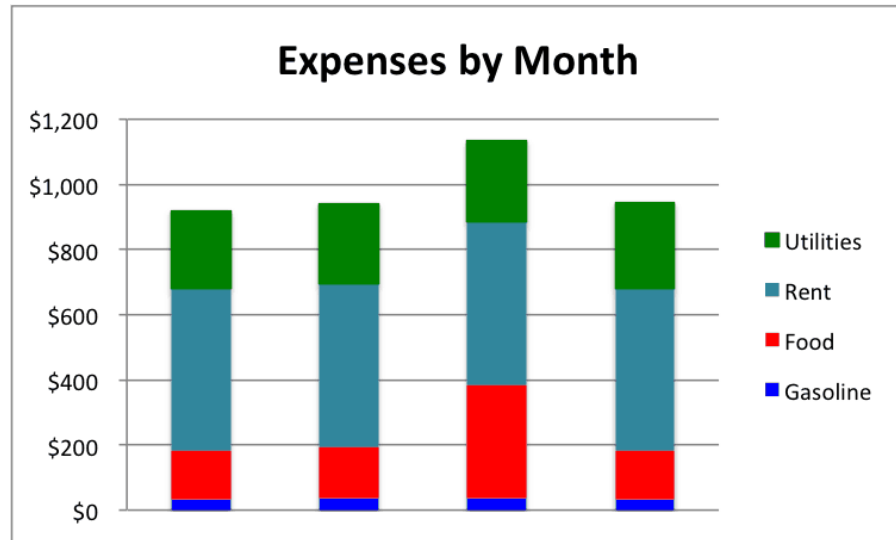
Vertical



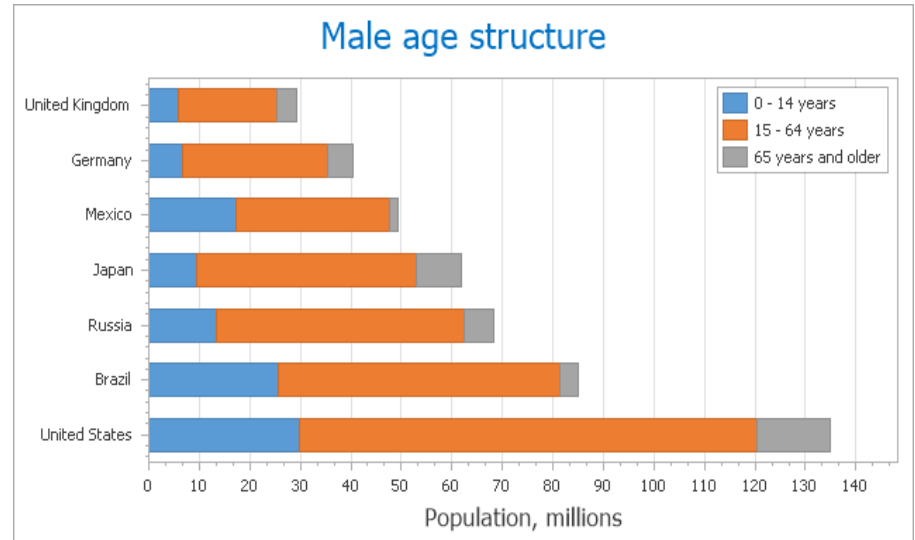
Horizontal



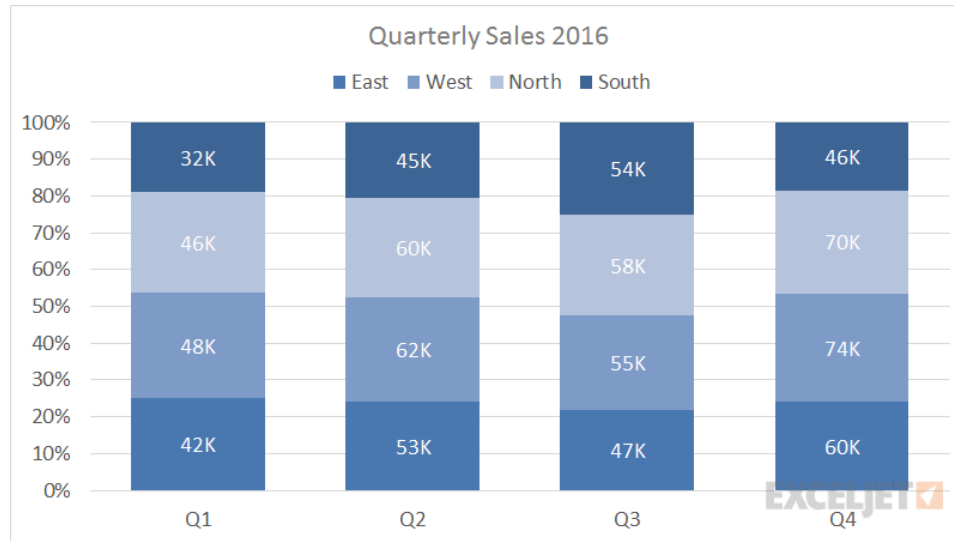
Vertical



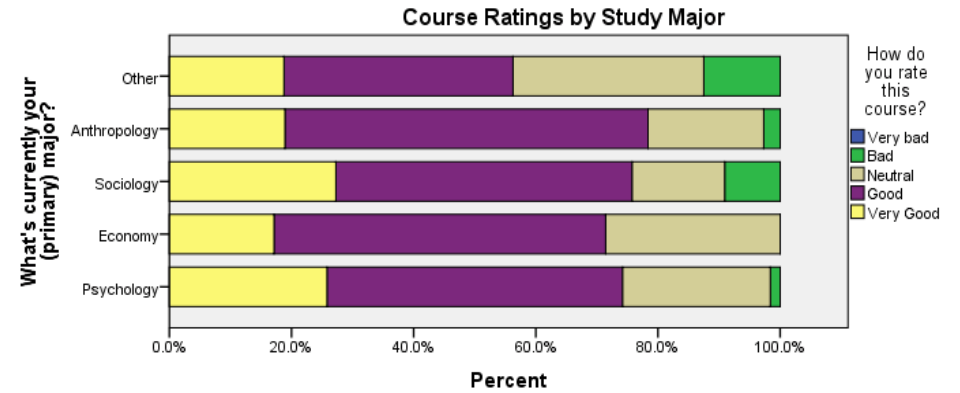
Horizontal



Vertical



Horizontal



Area Charts

A combination of a line and a bar chart. Usually tracks the change of a **numeric variable over time** but adds color to the area under the line.

Not a popular choice when using a single time series. All that color is noise.

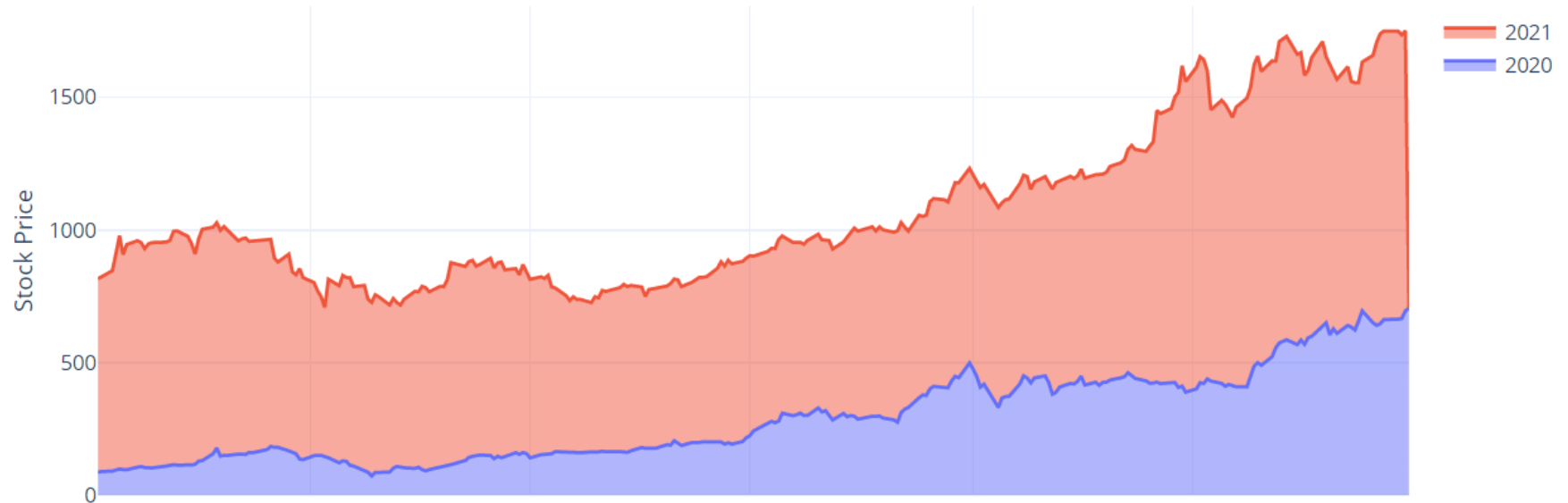
Tesla Stock Prices (2020-2021)

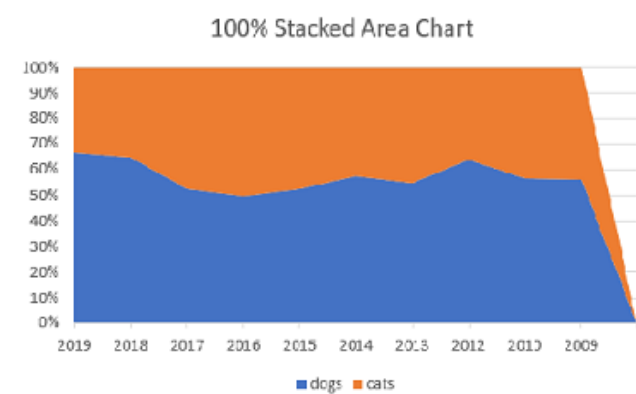
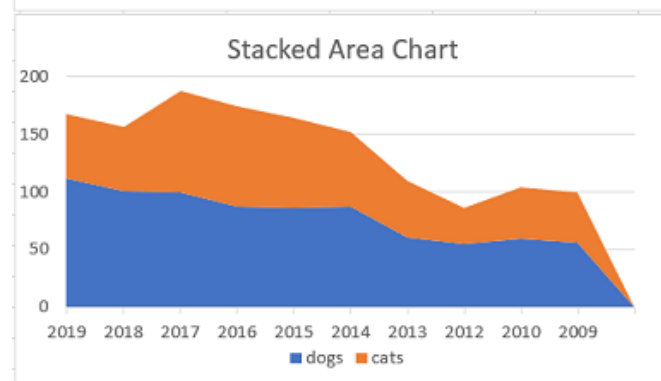
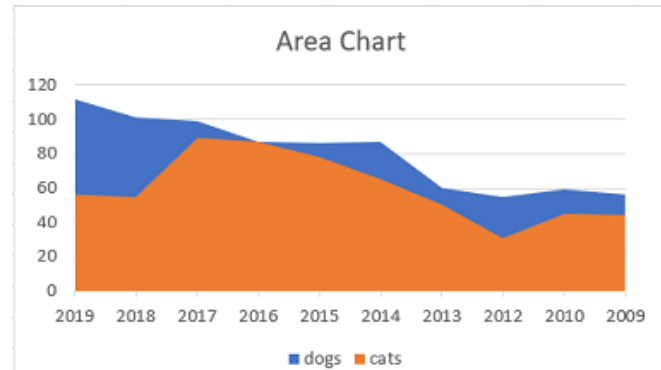


Area Charts

Area charts are designed to show multiple time series.

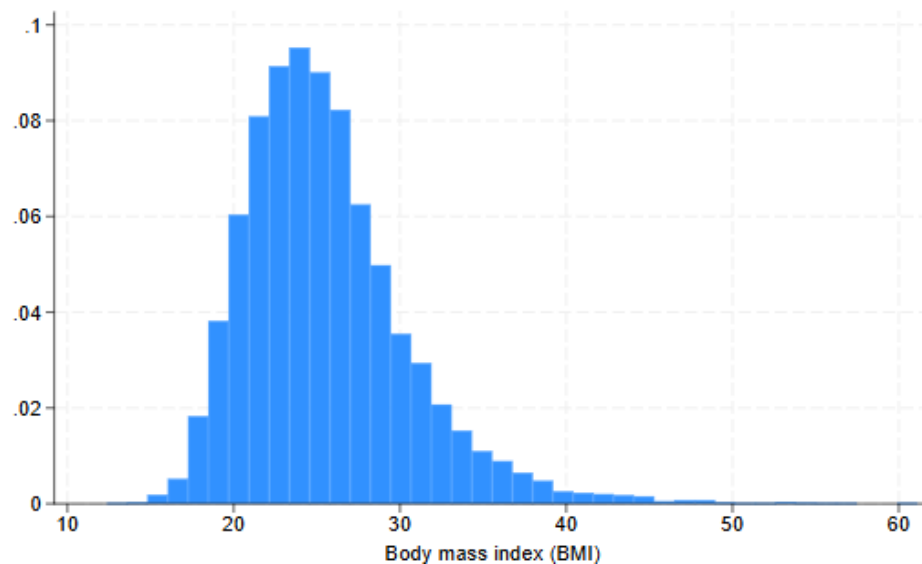
Tesla Stock Price 2020 vs 2021



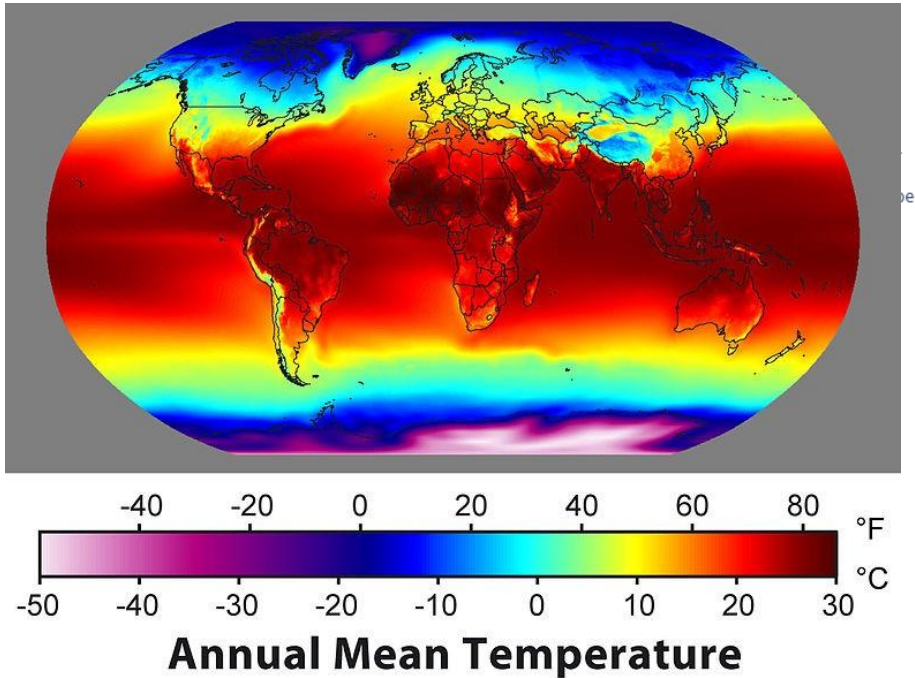


Histograms

A histogram always shows the distribution of a numeric variable. The variable values are binned into different groups and the height of the bars shows the number of observations in the bin.

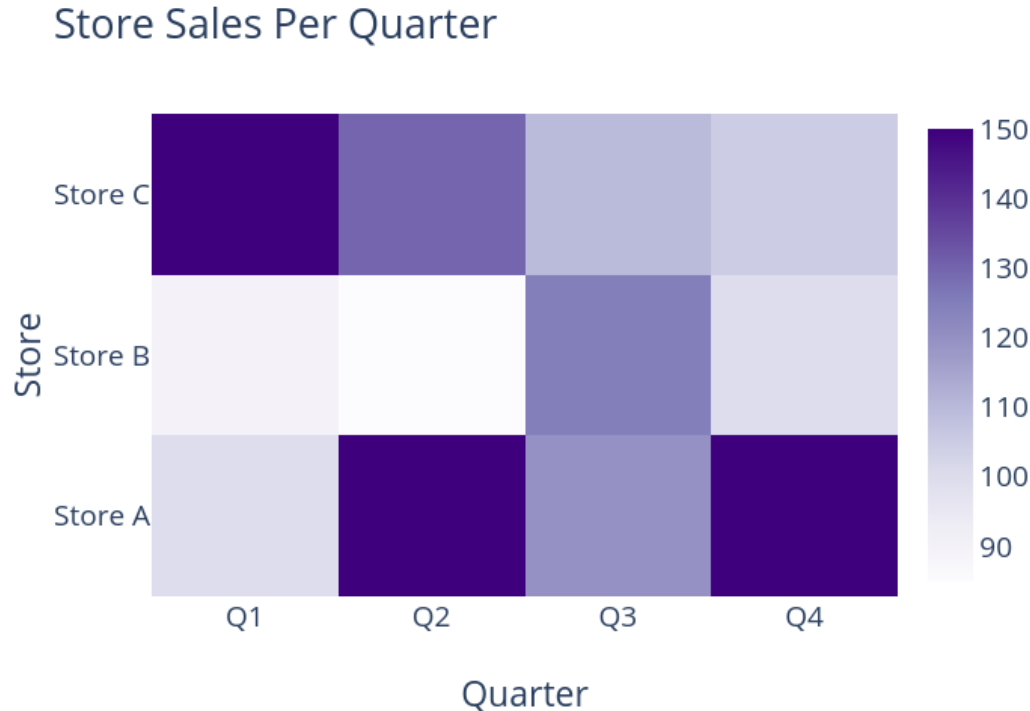


Heatmap Origin



It is more than a map with temperature measures.

Heatmaps today



A way to display 3D data as a 2D plot.

The axes can be **categories, dates or a combination of both.**

The colors come from **numeric data.**

In Tableau they are called “highlight tables”

		Ship Date				
		2021	2022	2023	2024	2025
Canada	Central			\$1,175	\$755	
	East	\$6,391	\$1,085	\$2,064	\$3,316	\$190
	West	\$3,033	\$1,721	\$1,489	\$8,113	
United States	Central	\$102,900	\$103,223	\$147,071	\$147,146	\$900
	East	\$119,992	\$164,992	\$180,090	\$213,106	\$601
	South	\$103,285	\$71,325	\$94,020	\$121,213	\$1,879
	West	\$144,206	\$139,904	\$190,144	\$249,425	\$1,779

Heat Map

Comparison: Overall

Breakout: **Department** ▾

Comparison: Overall

Breakout: Department ▾

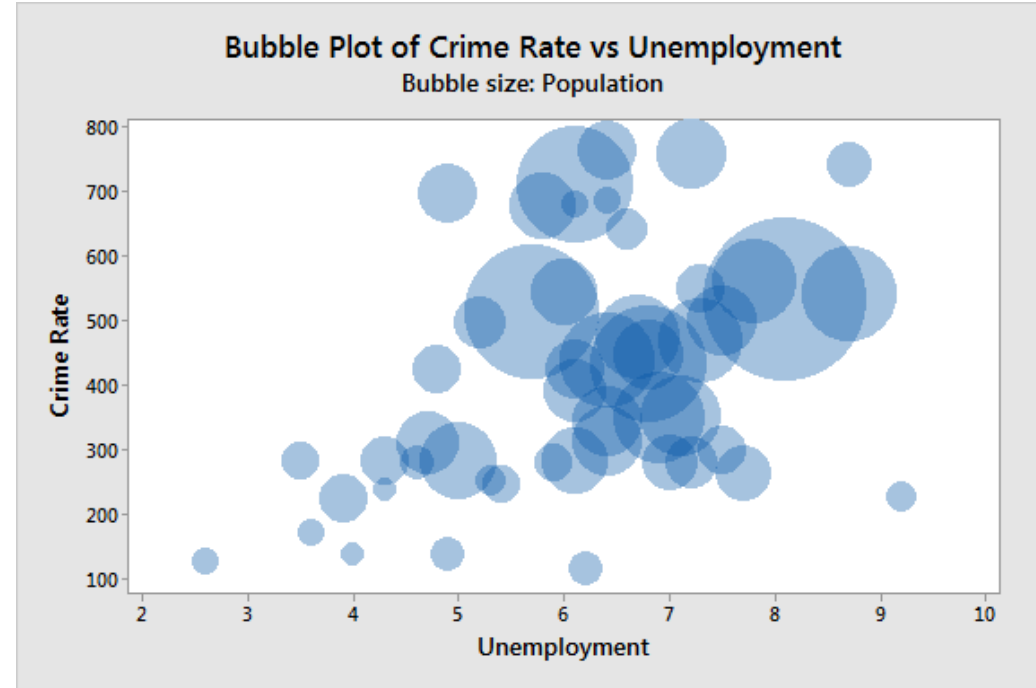
		Overall	Engineering	Finance	Human Resources	Marketing	Operations	Sales	Support
I feel proud to tell people where I work	54%	-9	-2	-11	-15	+18	+6	-19	
This company is good at directing resources (hum...ial or other) to the right places when necessary	72%	-4	+5	+9	+2	-4	+4	-7	
There are enough people to do the work on my team	56%	-1	+9	-9	-5	+2	+9	-11	
I have the tools / equipment / technology I need to do my job	76%	-14	+7	+12	-1	-2	+8	-2	
I have a good understanding of this company's strategy and goals	78%	+2	+3	-4	-7	+2	+3	-10	
I agree with the strategy and goals of this company	68%	-2	+17	+1	+1	+7	-5	+16	

Bubble Plot

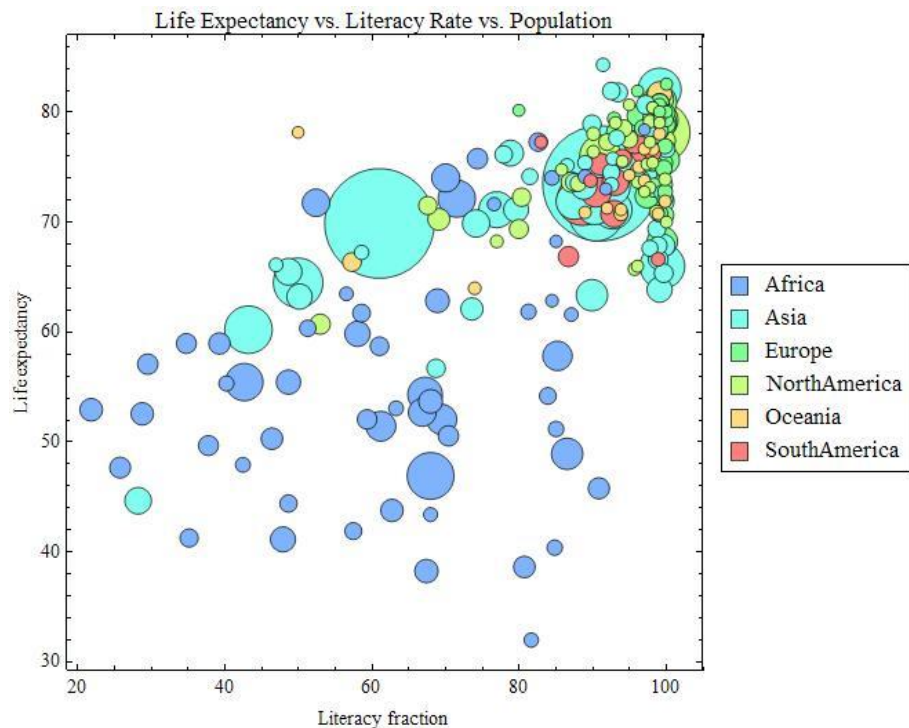
Another way of displaying 3D data on 2D plot.

A type of scatter plot where the size of the markers varies. The size data comes from a numeric value column. So, for a **bubble plot** requires 3 numeric fields.

Transparency is usually added for readability.



4D plot ???



You can keep adding shapes, sizes and colors to show more dimensions/data on the same chart but this is usually not a good idea.

Map Chart

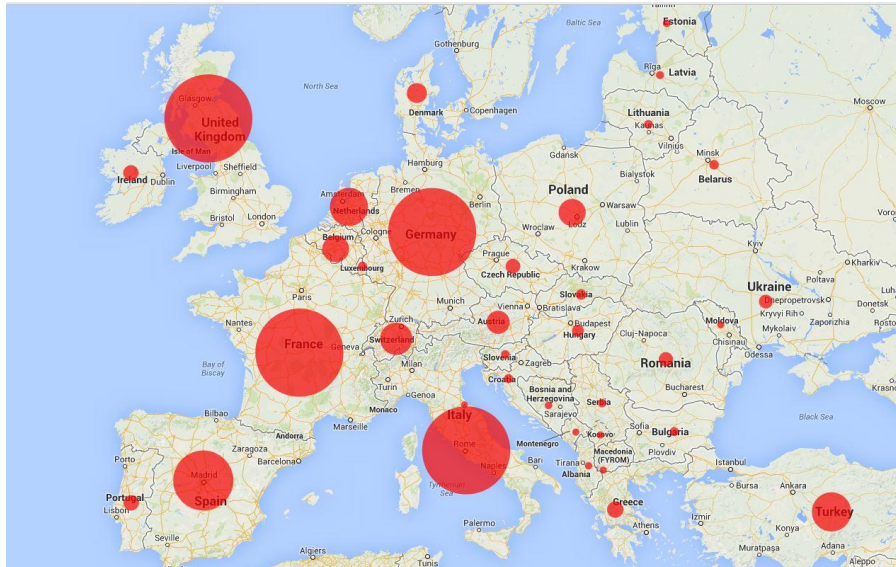


Anything plotted against a map is a map chart.

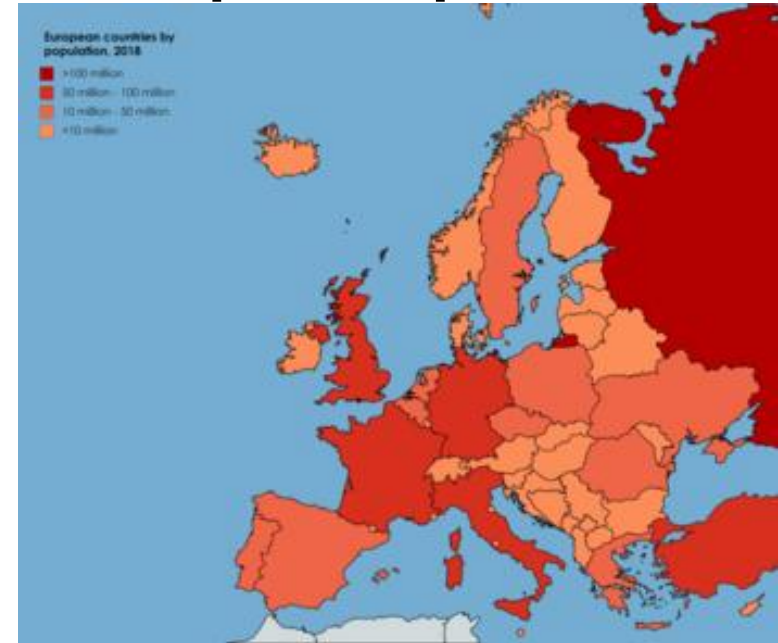
This is a bubble plot showing the relevant population size for the 25 largest cities in the US.

You need geo data to create map charts.

Bubble map



Choropleth map



Pie Chart



The pie slices represent a category. The size of the angle comes from a numeric column.

Pie charts use the same data as bar charts.

Pie charts are controversial

Pros:

- Pie charts are still very popular
- People love circles (really)

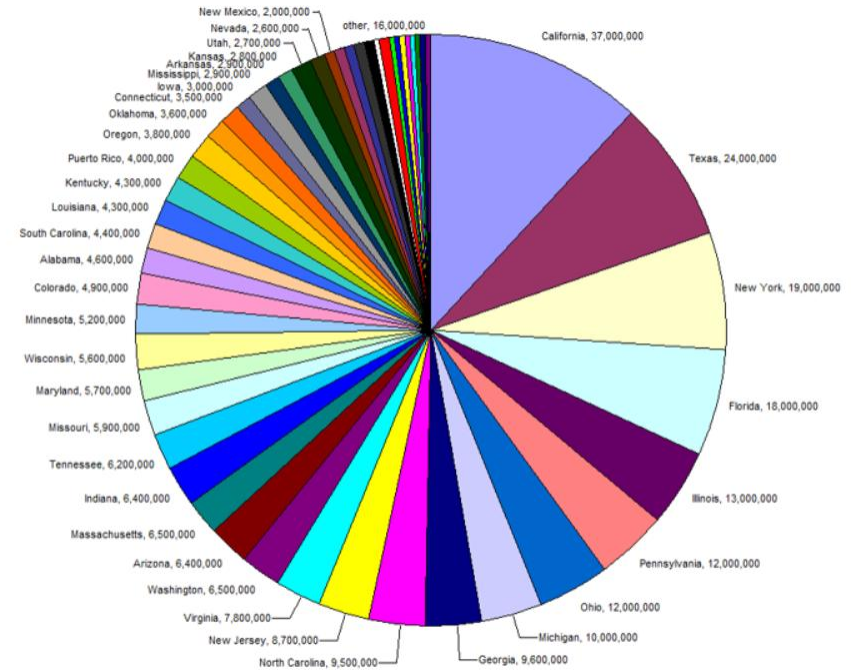
Cons:

- People are not good at spotting differences in angle size

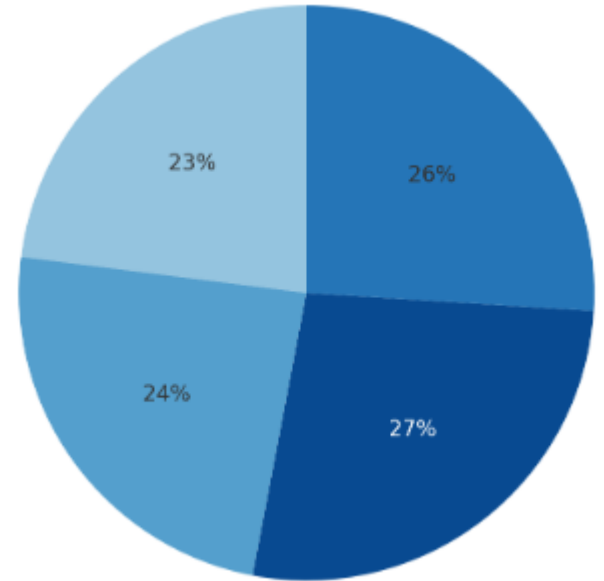
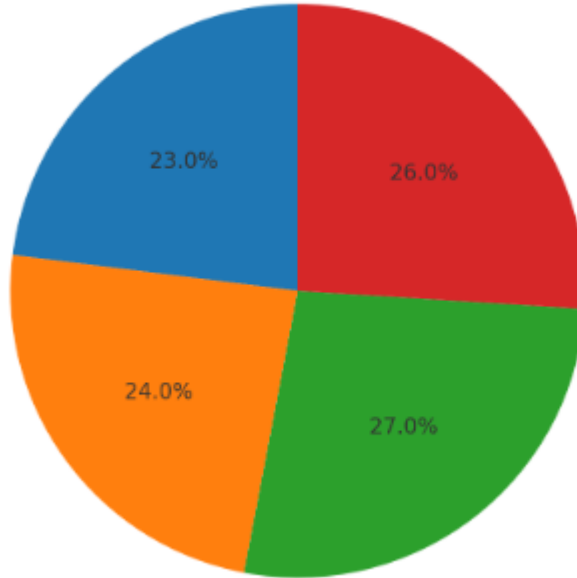
World's Most Accurate Pie Chart



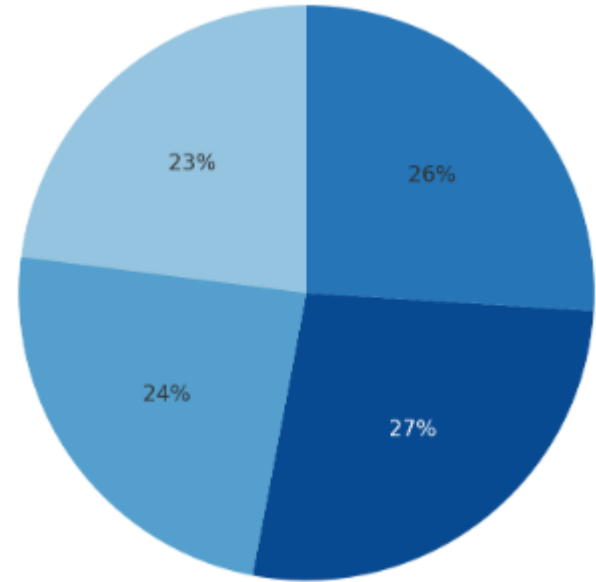
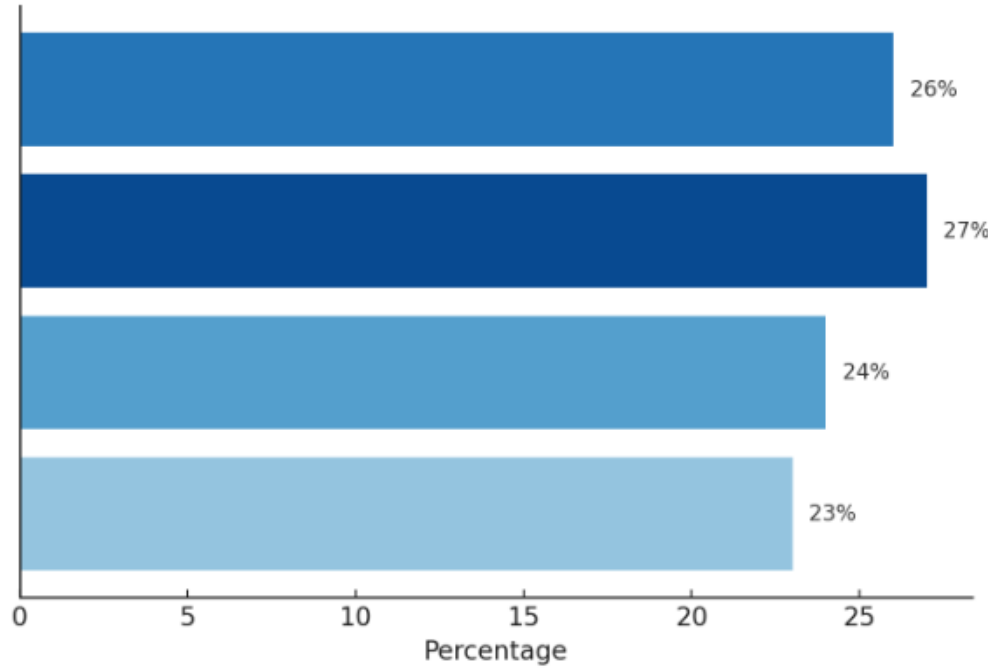
There are many ways to make a bad pie chart



But many pie charts can be improved



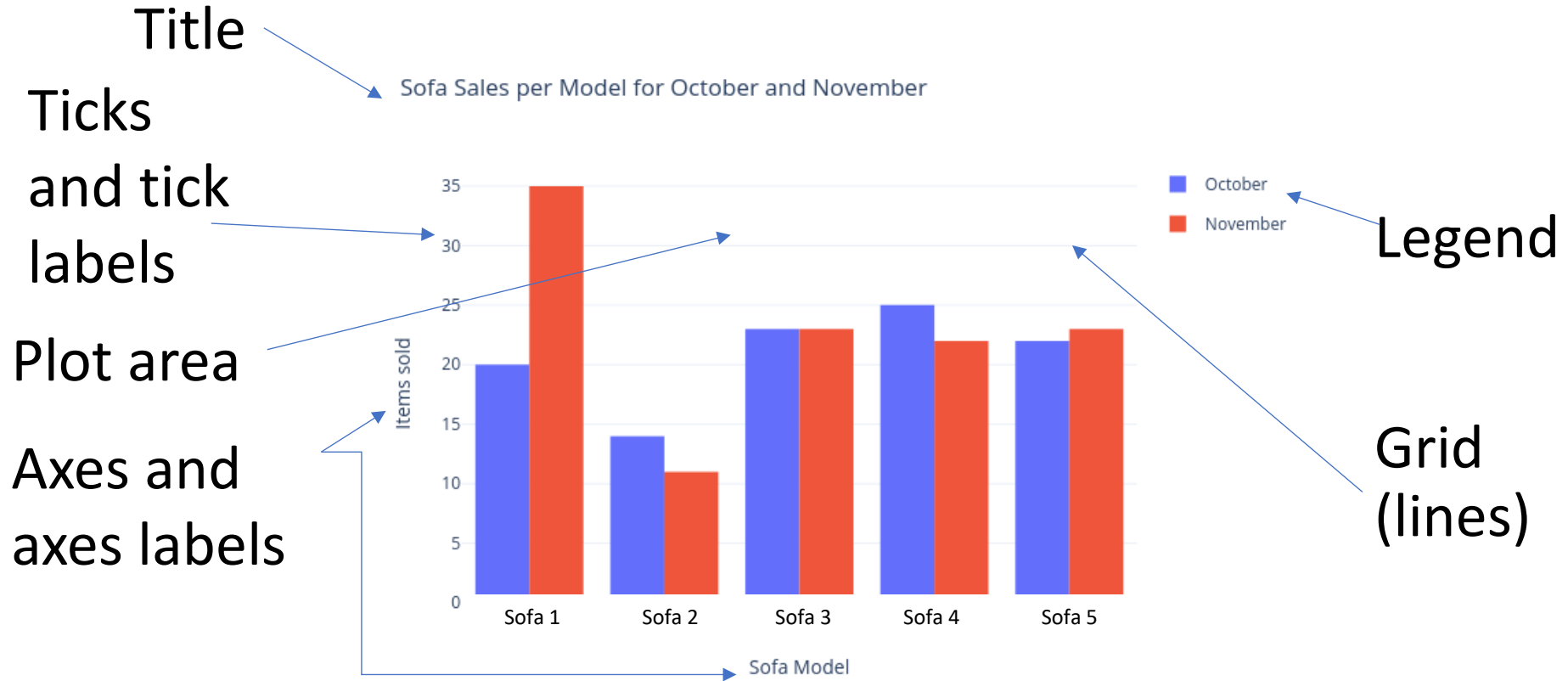
When in doubt use a bar chart instead



Conclusions about pie charts

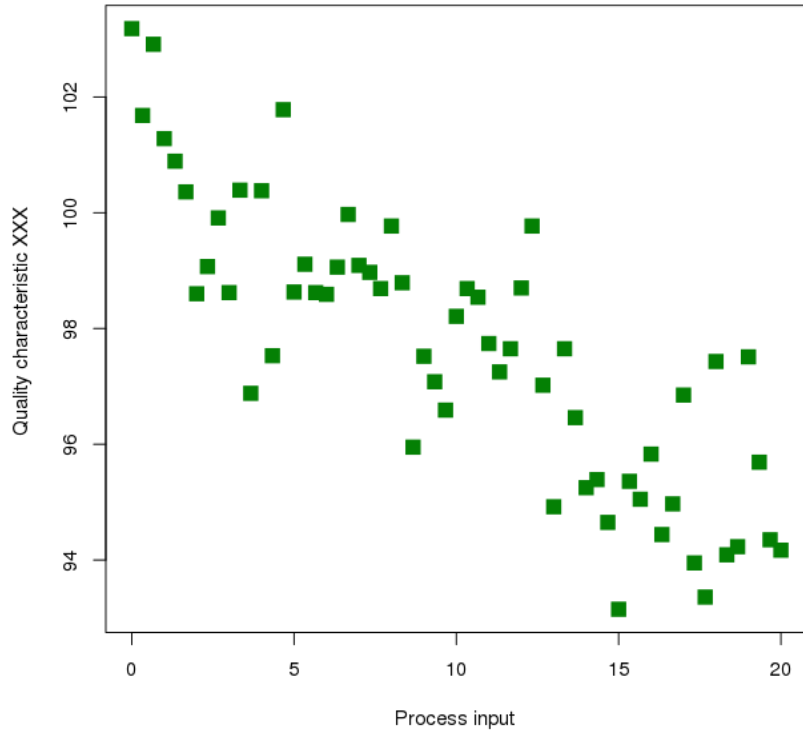
1. You can replace a pie chart with a bar chart. It's a safe option that always works.
2. If you insist on a pie chart, consider your audience and be extra careful about styling it.
3. **Don't use pie charts for more than 5 categories.**

Main elements of a chart



Identify the elements on this chart

Scatterplot for quality characteristic XXX



Title

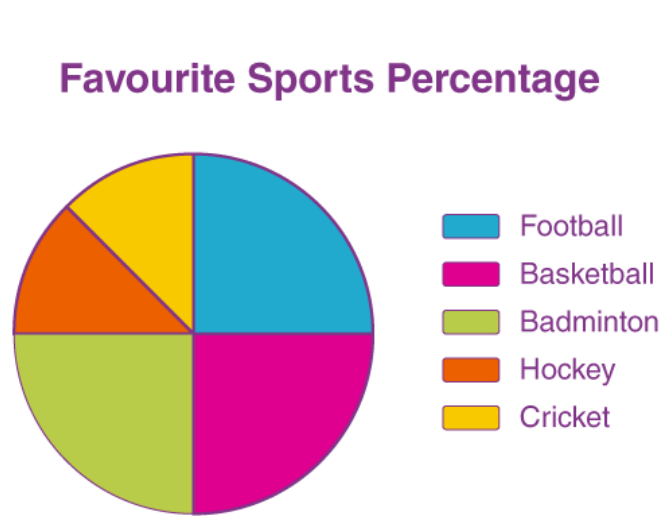
Axes and axes labels

Ticks and tick labels

Grid

Legend

Identify the elements on this chart



Title

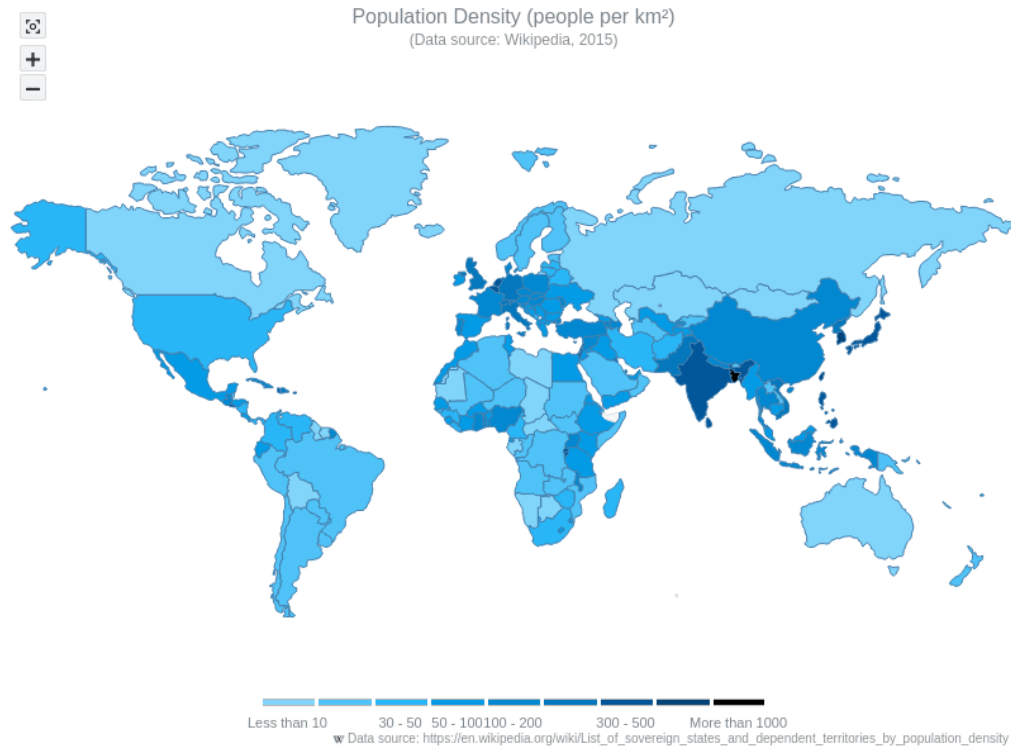
Axes and axes labels

Ticks and tick labels

Grid

Legend

Identify the elements on this chart



Title

Axes and axes labels

Ticks and tick labels

Grid

Legend

Tableau Desktop

Next week we'll start visualizing data with Tableau Desktop.

Make sure to install it and to activate it with the activation code posted on Canvas.