





multiverse

Data Science Unit 2



Data Visualisation





In this session we will...

- 
1. Understand why Data Visualisation is Important
 2. Identify what makes a 'Good' Data Visualisation
 3. Identify the Most Appropriate Visualisation



Why use Data Visualisation

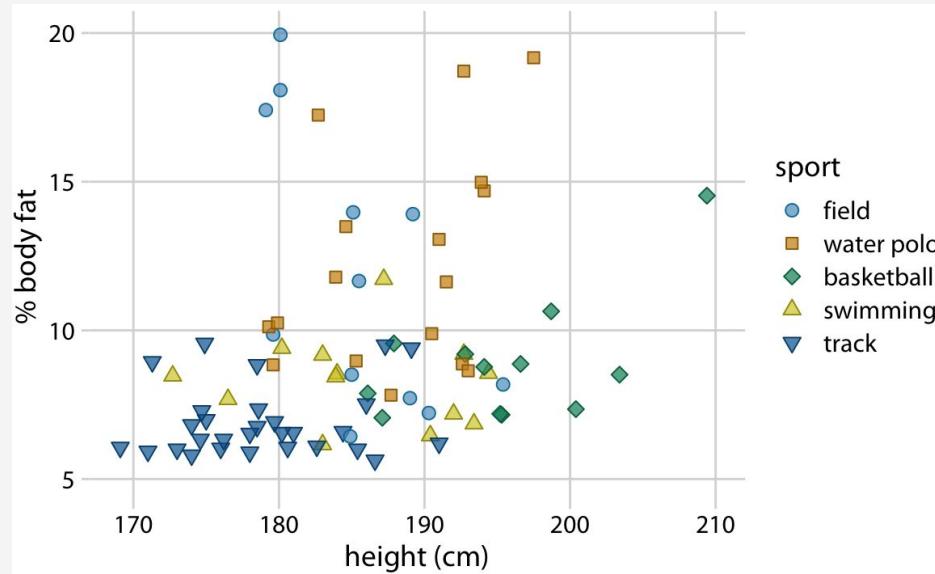




Why do you think
Data Visualization is
useful?



Humans are exceptional at
processing complex information
when it is presented in visual form





Anscombe's Quartet

Without visualization, we rely on statistics to interpret data

But...

What happens if our basic metrics are the same?

I

10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

II

10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.13
7	7.26
5	4.74

III

10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

IV

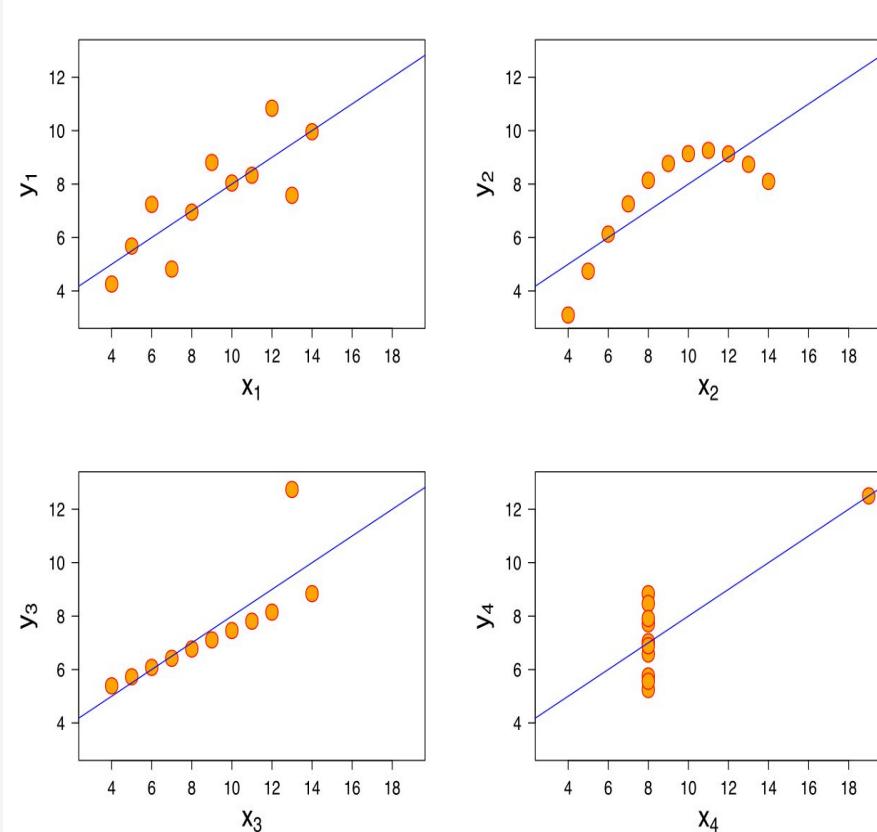
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89



Visualization complements statistics



Property	Value
x-mean in each case:	9 (exact)
x-variance in each case:	11 (exact)
y-mean in each case:	7.50
y-variance in each case:	4.122 or 4.127
Correlation between x & y:	0.816
Linear regression line:	$y = 3.00 + 0.500x$



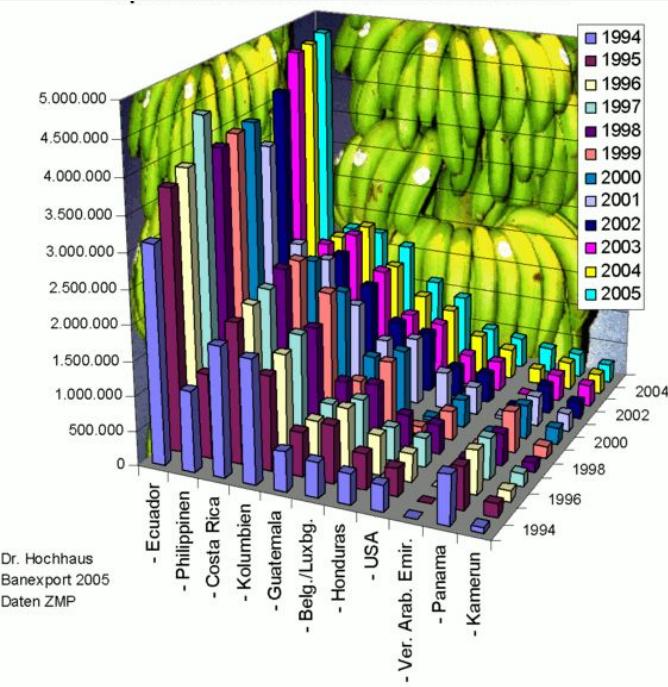


multiverse



Attributes of a good visualisation



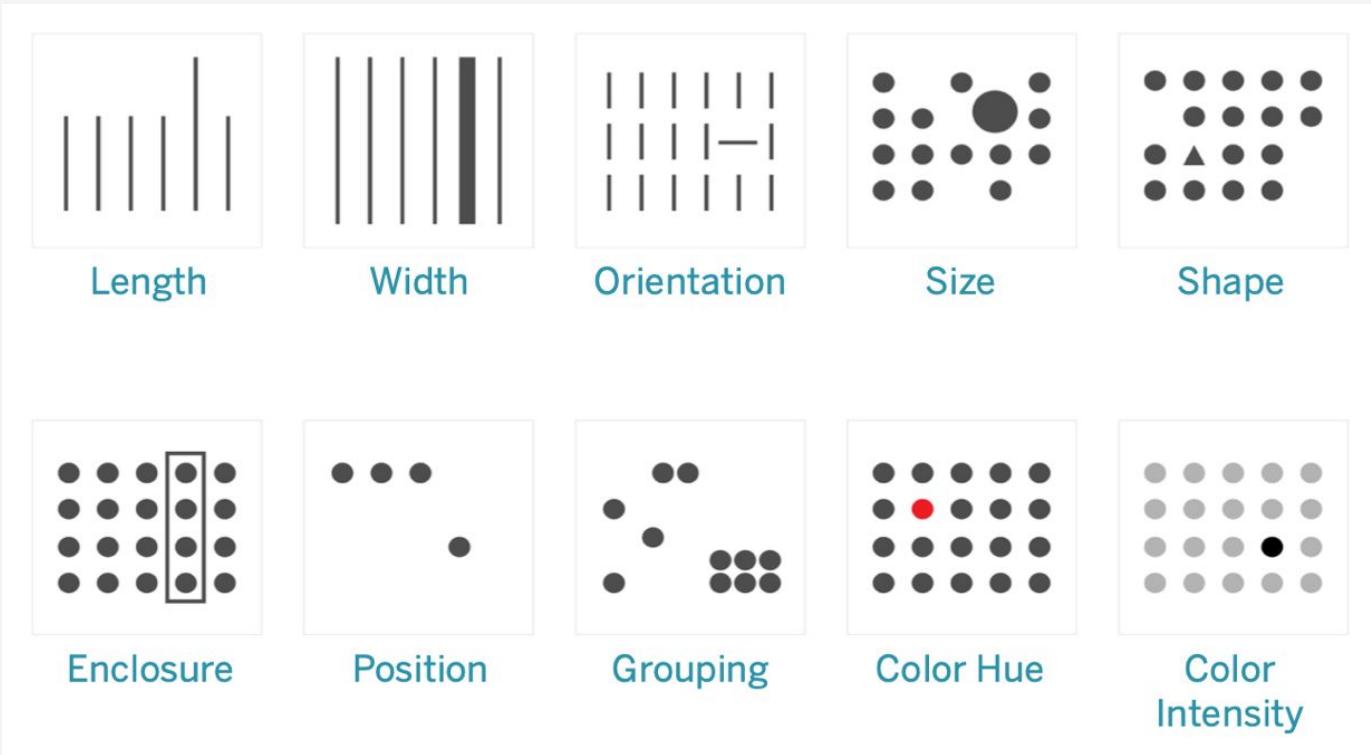


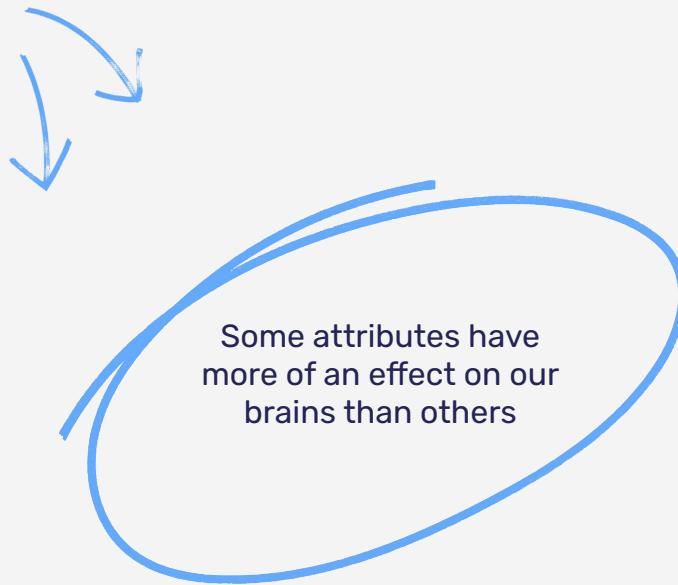
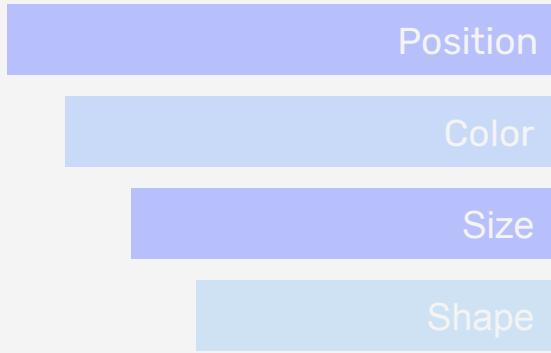
Human eyes good
at seeing visual
patterns!...

Sometimes.



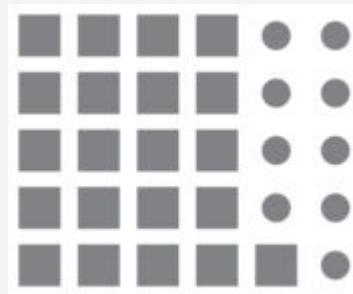
Preattentive attributes of visual perception







Let's look at three visualizations. Which one catches your attention most? Why?

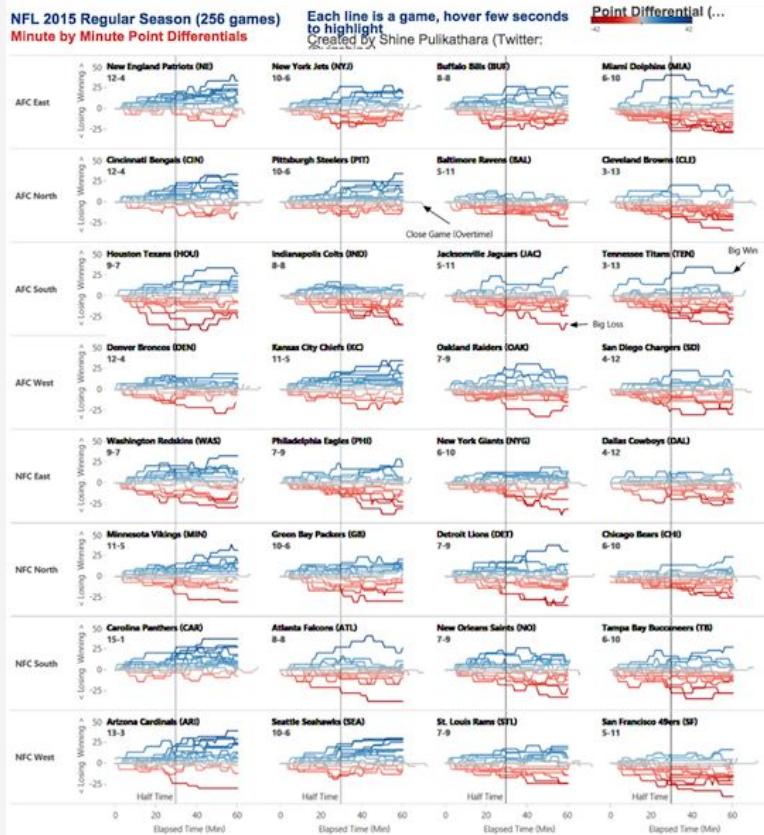




You can use color in one of three ways:

- Divergent
- Sequential
- Categorical

Divergent colors are used to show ordered values that have a critical midpoint



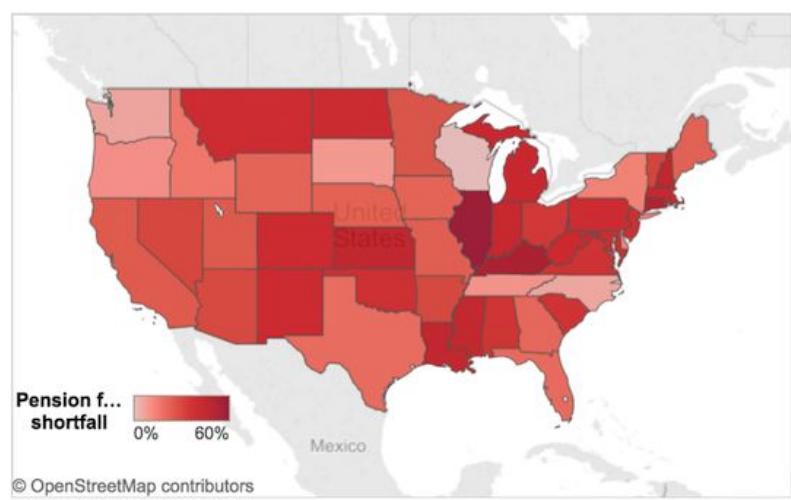


Sequential colors are used to show values ordered from low to high

Categorical colors are used to distinguish data that falls into distinct groups

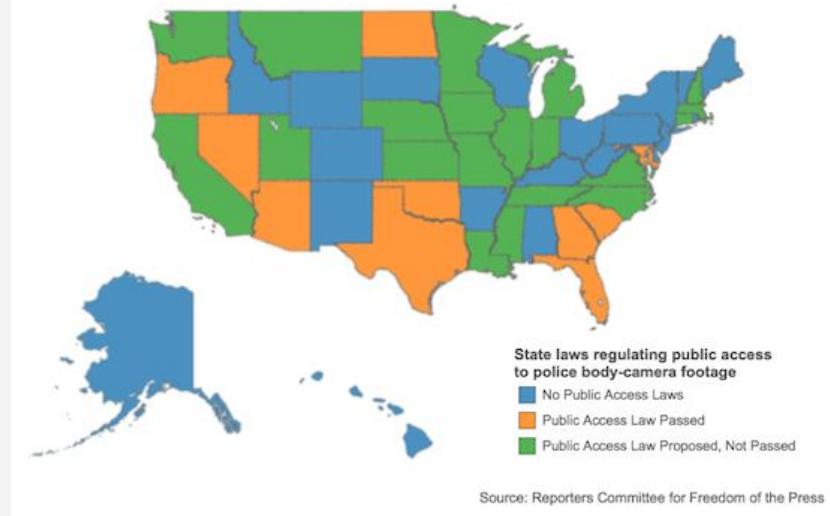
Pensions in Peril

Despite recent stock market gains, states continue to shortchange their pension plans, leaving many of them badly underfunded. (SOURCE: Pew Charitable Trusts)

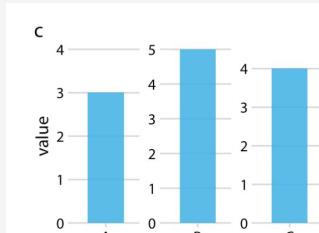
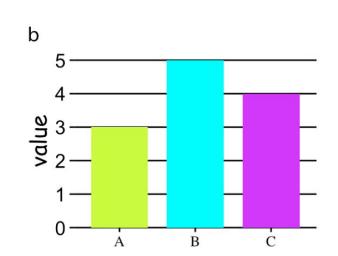
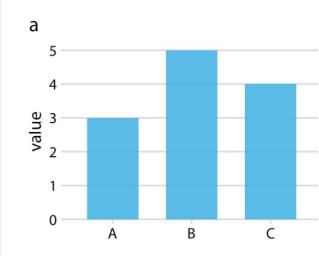


Body Camera Laws

Ten states have passed laws that control the public's access to footage from police body cameras. Hover over each state for more information.



Which chart follows good visualization principles? Why?





multiverse



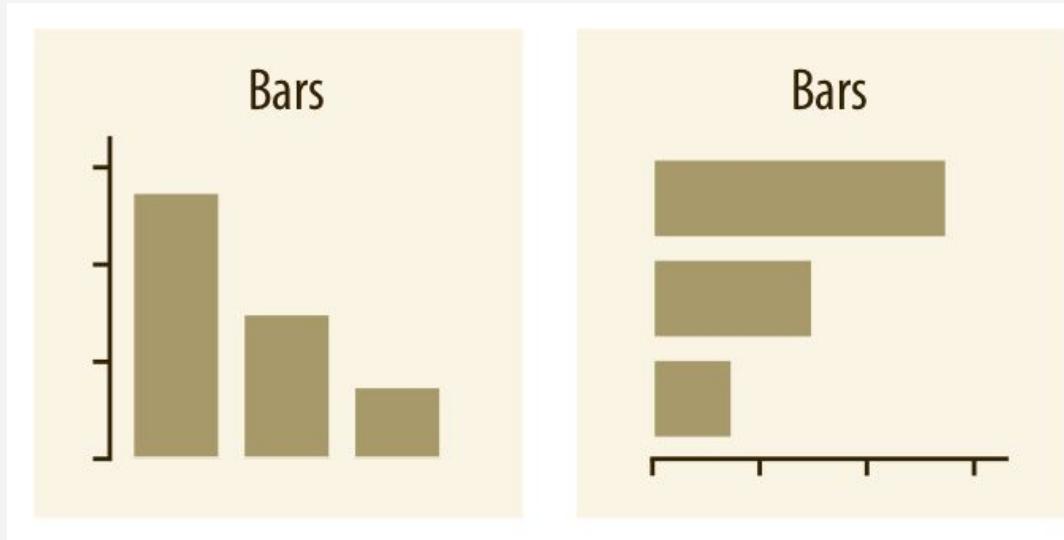
Choosing the right visualization





Visualising Amounts

Bar charts are one of the most common ways of visualizing data. Why?





Bar chart – Example



Are Film Sequels Profitable?

Box Office Stats For Major Film Franchises

Select Movie Franchise:

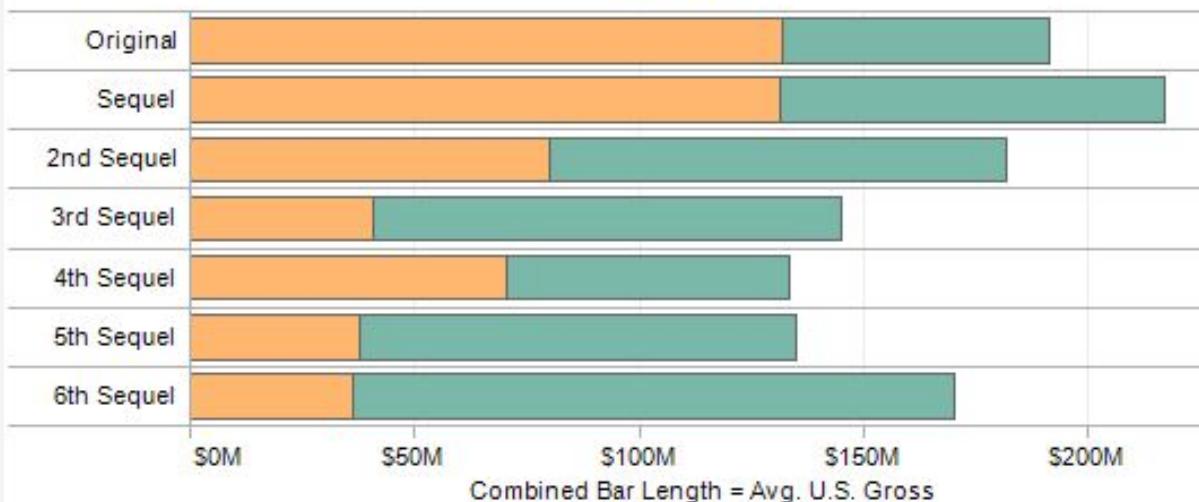
(All)



Click to Highlight Average:

Estimated Budget

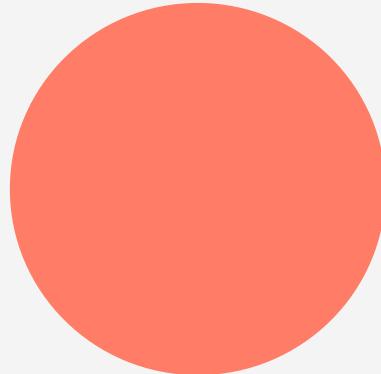
Profit





Visualising proportions

Pie charts can be used to visualize proportion, but they are the most commonly misused chart type.



Fun in pie charts

- fun
- No fun

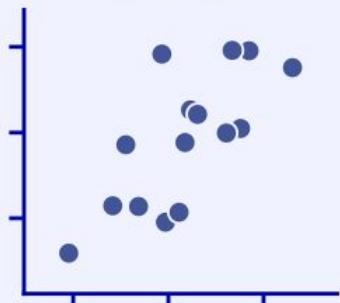


X-Y relationships

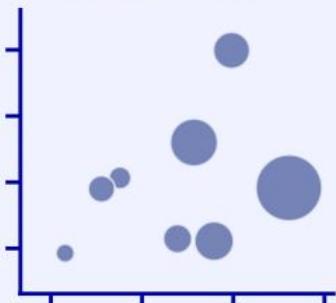


Scatter plots are a great way to give you a sense of trends, concentrations, and outliers, which provides a clear idea of what you may want to investigate further.

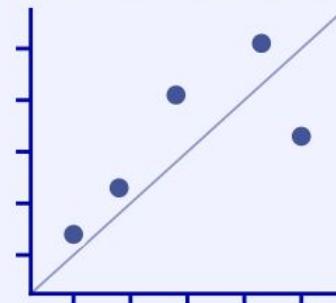
Scatterplot



Bubble Chart



Paired Scatterplot



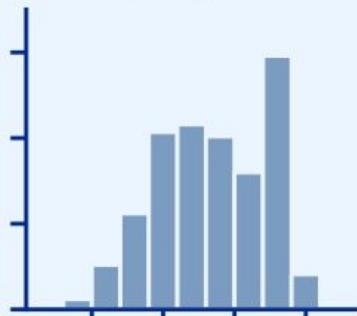


Visualising distributions

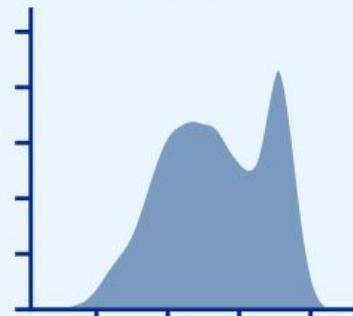
Scatter plots are a great way to give you a sense of trends, concentrations, and outliers. This will provide a clear idea of what you may want to investigate further.



Histogram



Density Plot

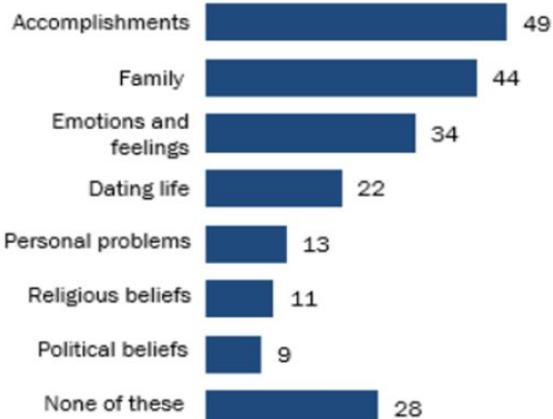




Which chart shows a distribution?



A



While about half of teens post their accomplishments on social media, few discuss their religious or political beliefs

% of U.S. teens who say they ever post about their __ on social media

B

SHARE OF AMERICAN ADULTS
IN EACH INCOME TIER

Upper 19%

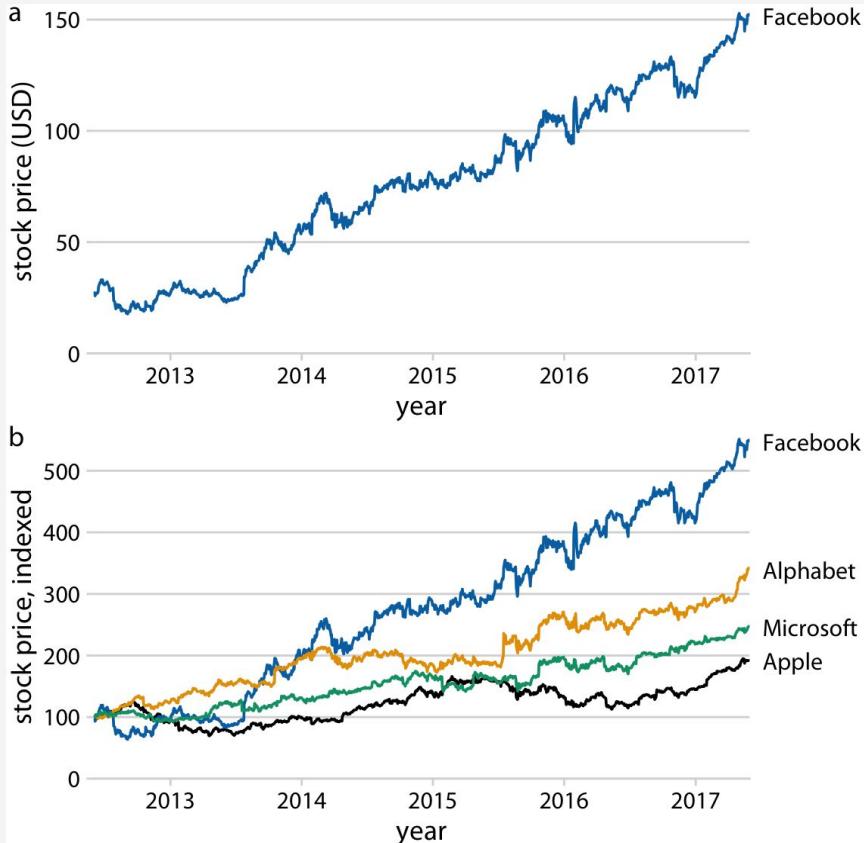
Middle 52%

Lower 29%

The figure shows growth of Facebook stock price over a five-year interval and comparison with other tech stocks?

What is wrong with this figure? Why?

Suggest which chart type we can use to make this a good visualization?

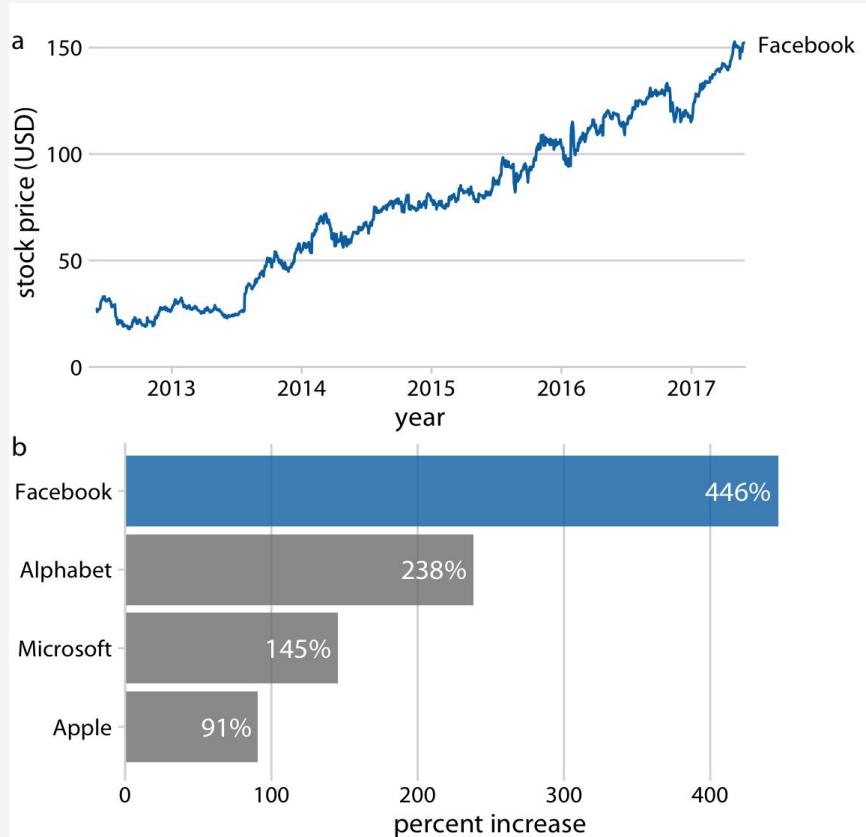




Possible suggestion:

Leave part (a) as is but replace part (b) with a bar plot showing percent increase

Now we have two distinct figures that each make a unique, clear point and that work well in combination.



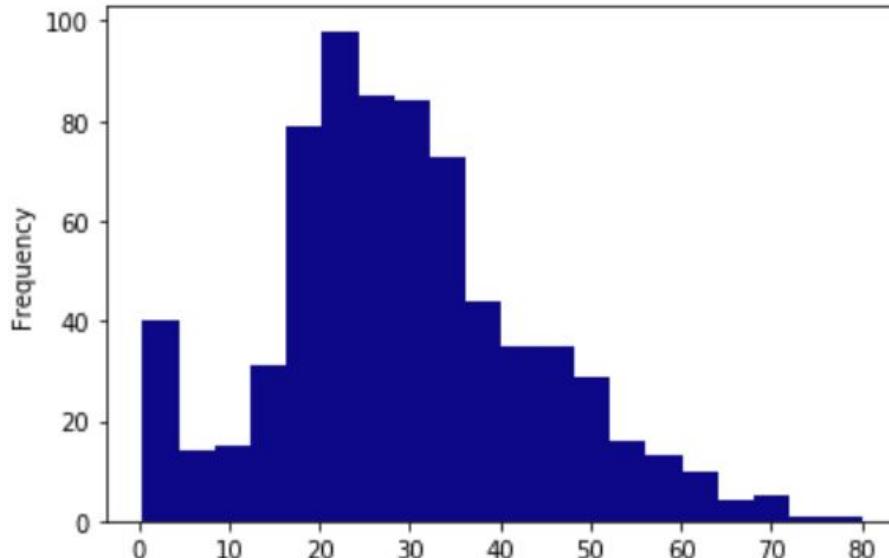
multiverse

Python Libraries





```
df.Age.plot(kind='hist',bins=20,colormap='plasma');
```





To effectively use data visualizations, you must be proficient with both the principles of visualization and the programming tools to generate plots.

In this lesson, we will use the Python libraries Matplotlib (Python plotting) and Seaborn (statistical data visualization).

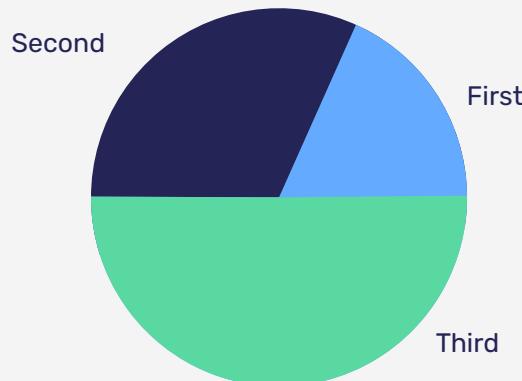
The Matplotlib logo consists of the word "matplotlib" in a lowercase sans-serif font, with the 'o' in "plot" replaced by a small sun-like icon with rays.

The Seaborn logo consists of the word "Seaborn" in a bold, dark teal sans-serif font.

matplotlib

```
import matplotlib.pyplot as plt
```

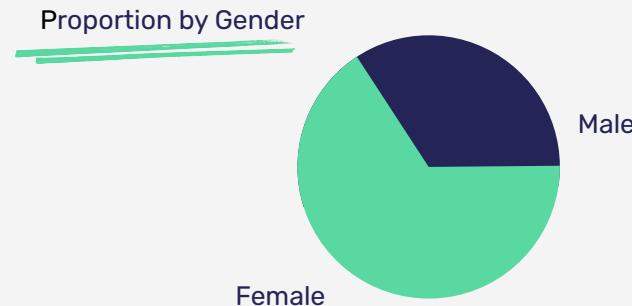
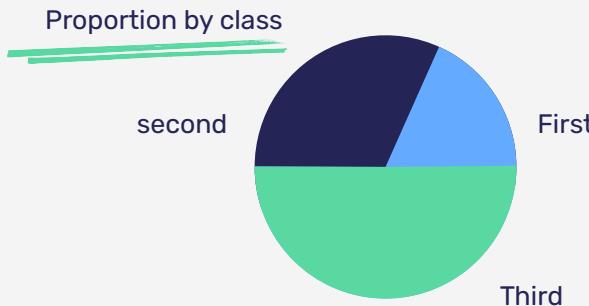
```
plt.pie([1,2,3],data=df.Pclass,labels=['First','Second','Third']);
```



matplotlib

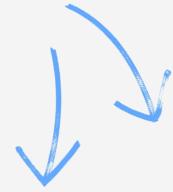
```
import matplotlib.pyplot as plt
```

```
fig,ax=plt.subplots(ncols=2)
ax[0].pie([1,2,3],data=df.Pclass,labels=['First','Second','Third'])
ax[0].set_title('Proportion by Class')
ax[1].pie([1,2],data=df.Gender,labels=['Male','Female'],colors=['Yellow','Green'])
ax[1].set_title('Proportion by Gender');
```

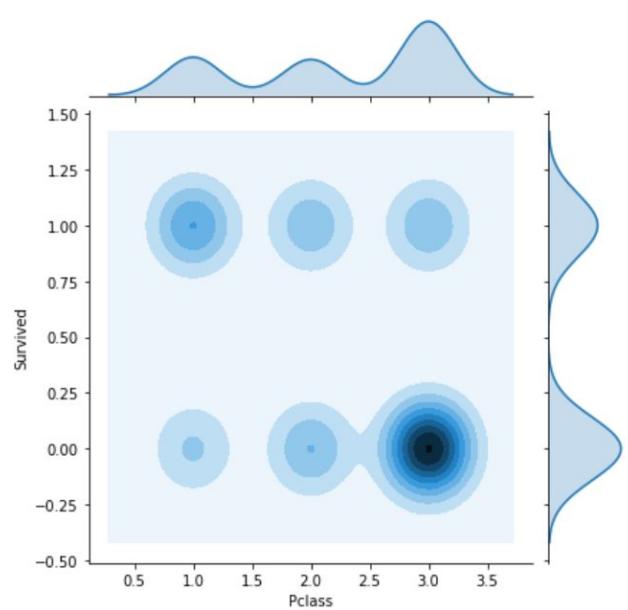




Seaborn



```
sns.jointplot(x='Pclass',y='Survived',data=df,kind='kde');
```



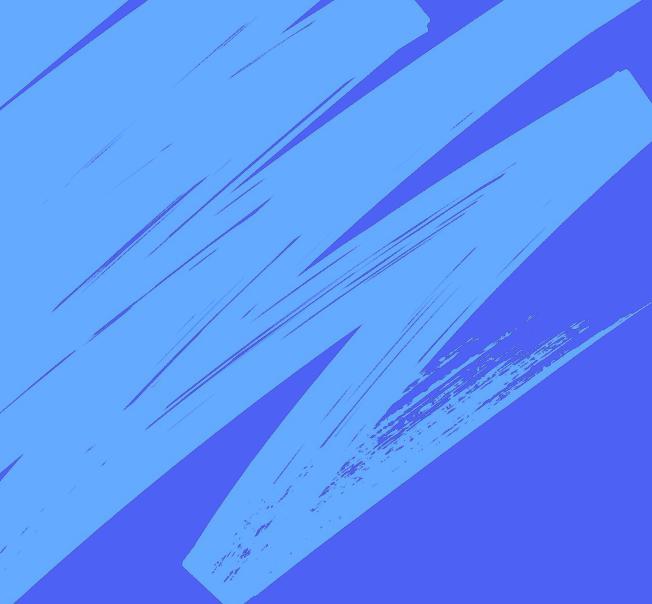
```
import seaborn as sns
```



Other libraries

Many other Python libraries exist for making visualizations. Some of the most popular include:

- [**Bokeh**](#): Python visualization library that targets the web browser (e.g., in Jupyter). Makes interactive plots, dashboards, data applications, etc.
- [**Graphviz**](#): Popular visualization library for graph data structures (e.g., edges, vertices, etc). Has Python extensions.
- [**Basemap**](#): Python Matplotlib extension for drawing static maps. There are many other Python libraries for plotting geographic data, including ones that might be easier to use, but many are not actively developed.
- [**D3.js**](#): JavaScript library for interactive web visualizations



multiverse



Other tools





Although this course emphasizes a Python approach to data science, a variety of non-programming tools are also used in industry. Often, these tools can be applied much more quickly than creating a custom Python solution.

For example:

- **Excel**: For quick data cleaning and simple graphs
- **Power BI**: A suite of business analytics tools
- **Tableau**: Business intelligence and analytics software
- **Periscope Data**: Data analysis platform
- **Plotly**: Create charts and dashboards

multiverse

Independent Practice





Python Plotting With Pandas and Seaborn

Open up the `python-data-viz-lab` in your practice folder to explore plotting the sales data with Python.





multiverse



Summary



Summary

- Why is data visualization so important?
- What are some considerations to keep in mind when creating a visualization?
- Describe when you would use the following types of charts or graphs:
 - Bar chart
 - Pie chart
 - Scatter plot
 - Histogram

Multiverse

Any Questions

multiverse

OTJ and SAL

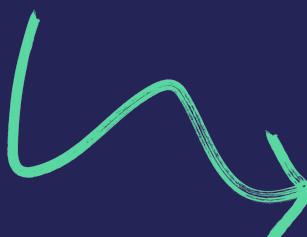


The text 'OTJ and SAL' is centered on a dark blue background. Above the text, there are two sets of stylized lines radiating outwards. The left set is composed of orange lines, and the right set is composed of blue lines. Both sets of lines have small, irregular segments at their ends, resembling stylized sunburst or starburst patterns.

The background features abstract teal brushstrokes on a dark navy blue surface. One large, horizontal brushstroke is positioned at the top left, while another large, curved brushstroke sweeps across the middle-left area. Smaller, more vertical and diagonal strokes are scattered throughout the upper portion of the slide.

multiverse

Thank you

A thick, teal brushstroke forms an arrow shape, pointing from the bottom center towards the right side of the slide.

Get in touch
info@multiverse.io