

# DATA VISUALISATION



# LEARNING OBJECTIVES

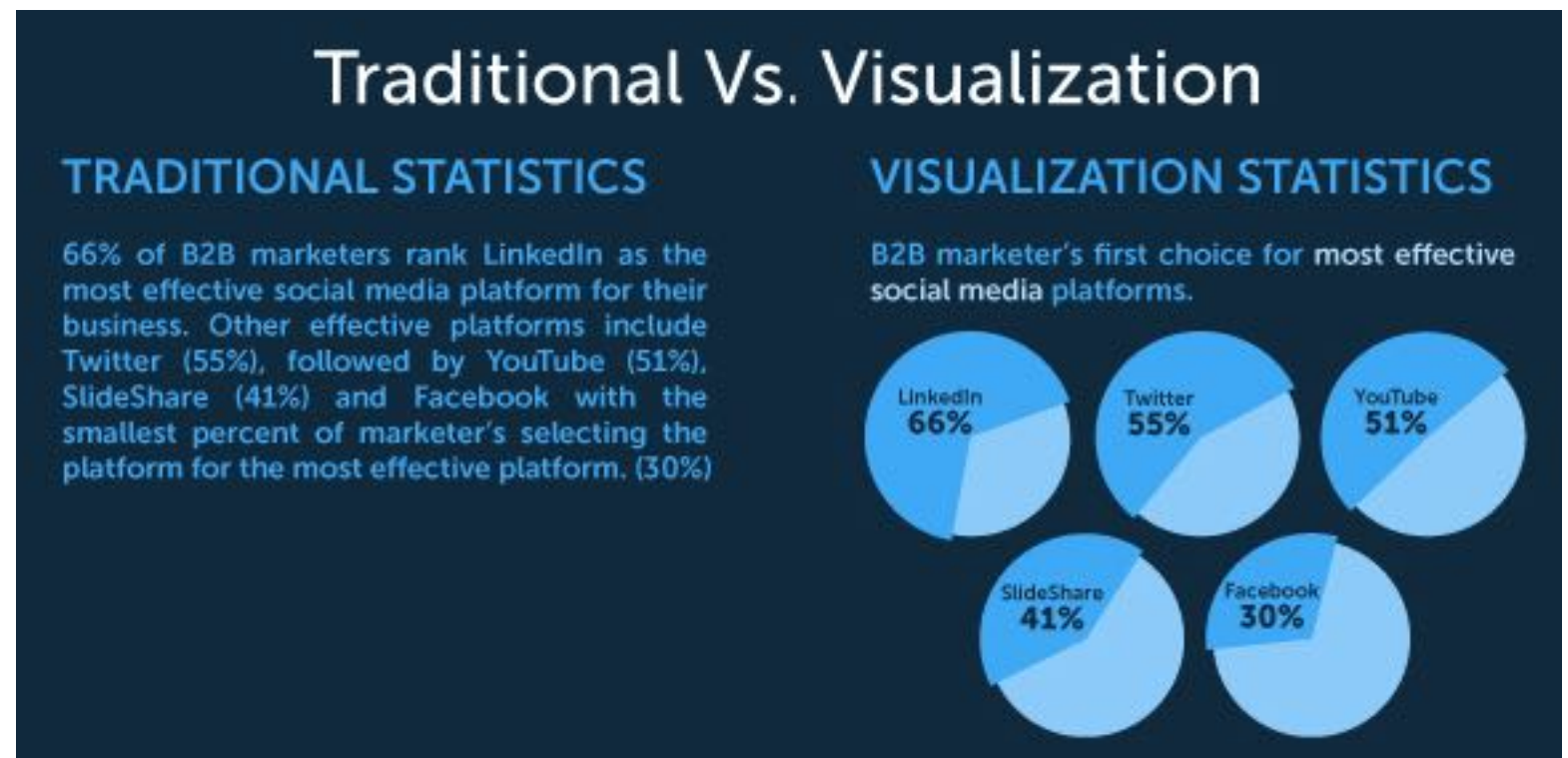
---

- Describe why data visualization is important.
- Identify the characteristics of a great data visualization.
- Describe when you would use a bar chart, pie chart, scatter plot, and histogram.
- Introduction to python libs. For data visualization

# WHY DATA VISUALIZATION

---

- Charts or graphs that are used visualize large amounts of complex data are easier to understand than traditional spreadsheets or reports.
- Data visualization is a quick and easy way to convey concepts in a universal manner that builds upon human cognition
- Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats
  - <https://www.youtube.com/watch?v=jbkSRLYSojo>



# WHY DATA VISUALIZATION

Statistical measures for data summarisation

vs.

Data visualization.

- Highlights the failures of summary statistics.
- Shows the effect of outliers on statistical properties.
- Attacks the impression among statisticians that "numerical calculations are exact, but graphs are rough."

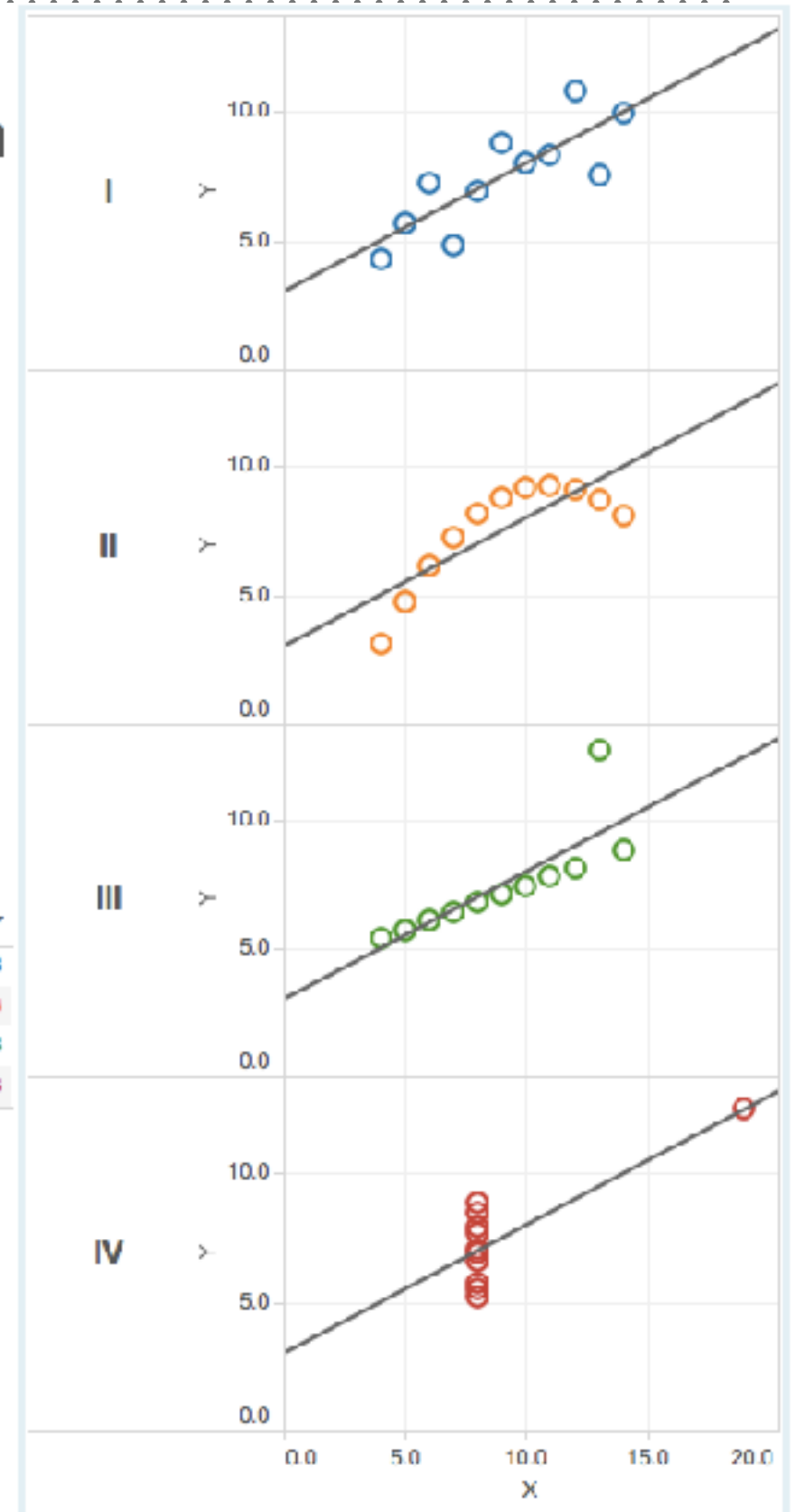
## Anscombe's Quartet: The power of visualization

These four data sets have identical summary statistics, yet the plots show vastly different stories

I	II	III	IV
(4, 4.3)	(4, 3.1)	(4, 5.4)	(8, 5.3)
(7, 4.8)	(5, 4.7)	(5, 5.7)	(8, 5.0)
(5, 5.7)	(6, 6.1)	(6, 5.1)	(8, 5.8)
(8, 7.0)	(7, 7.3)	(7, 5.4)	(8, 6.6)
(6, 7.2)	(14, 8.1)	(8, 5.8)	(8, 6.9)
(13, 7.5)	(8, 8.1)	(9, 7.1)	(8, 7.0)
(10, 8.0)	(13, 8.7)	(10, 7.5)	(8, 7.7)
(11, 8.3)	(9, 8.8)	(11, 7.8)	(8, 7.9)
(9, 8.8)	(12, 9.1)	(12, 8.2)	(8, 8.5)
(14, 10)	(10, 9.1)	(14, 8.8)	(8, 8.8)
(12, 10.8)	(11, 9.3)	(13, 12.7)	(19, 12.5)

Summary Statistics

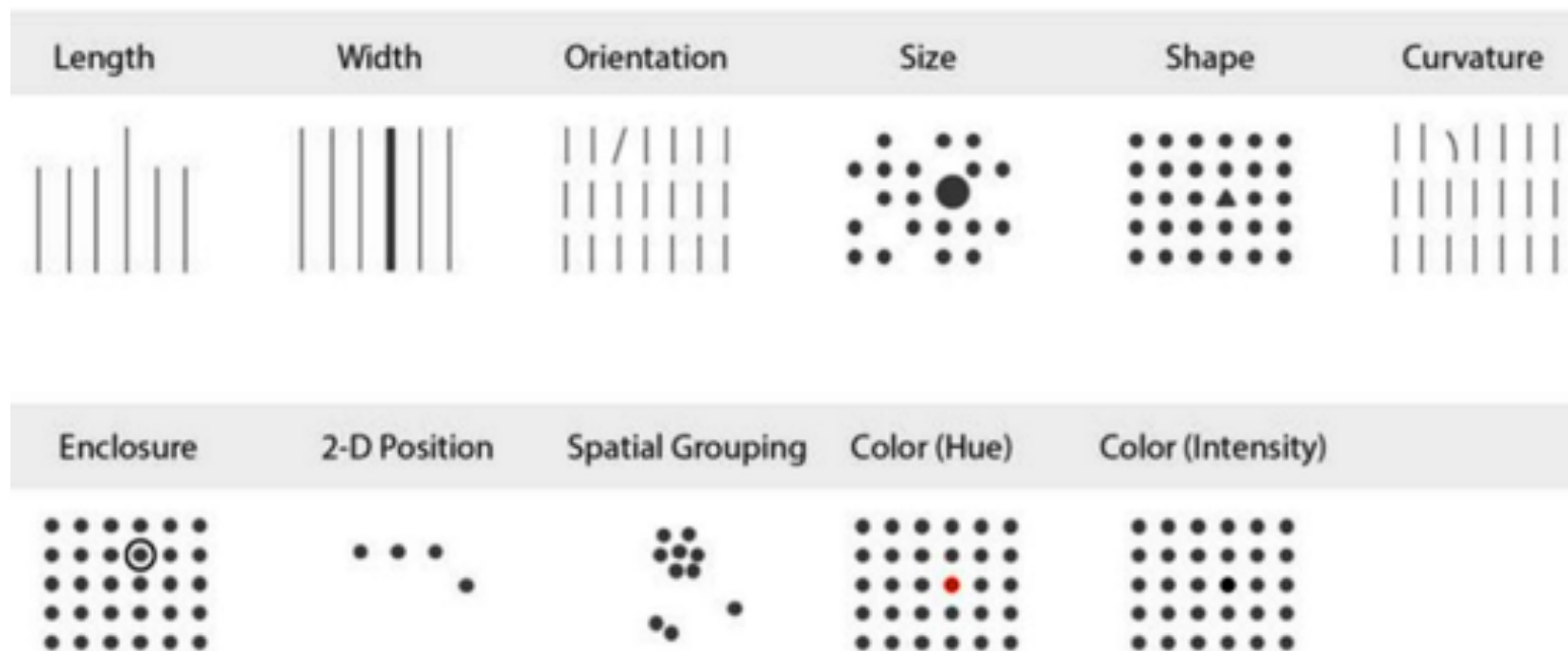
Plot	sum X	sum Y	avg X	avg Y	stdev X	stdev Y
I	99.0	82.5	9.00	7.50	3.32	2.03
II	99.0	82.5	9.00	7.50	3.32	2.03
III	99.0	82.5	9.00	7.50	3.32	2.03
IV	99.0	82.5	9.00	7.50	3.32	2.03



# WHAT IS A “GOOD” VISUALIZATION

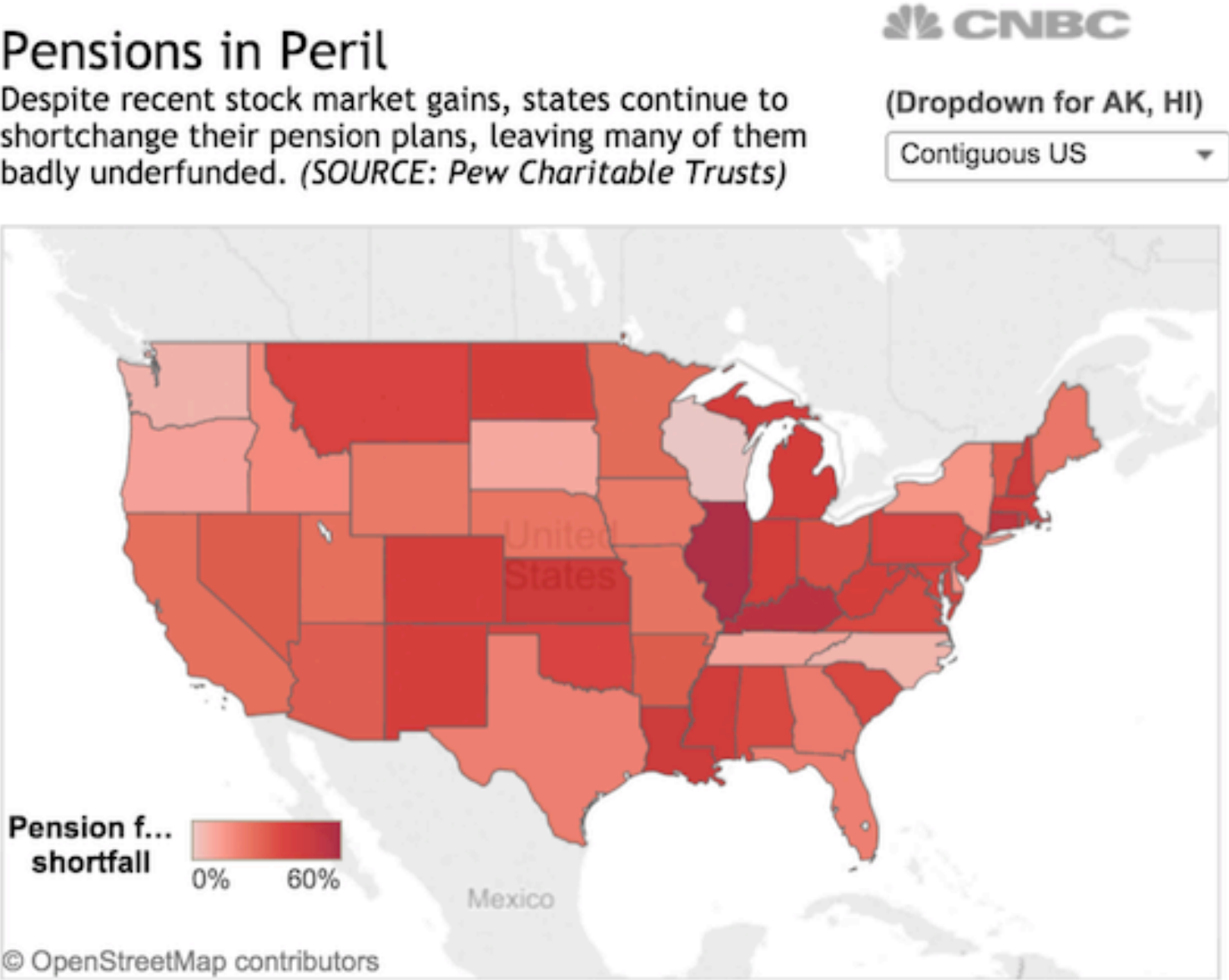
---

- Some visual cues have more of an effect on our brains than others. The ones we tend to focus on most are position, then color, then size.



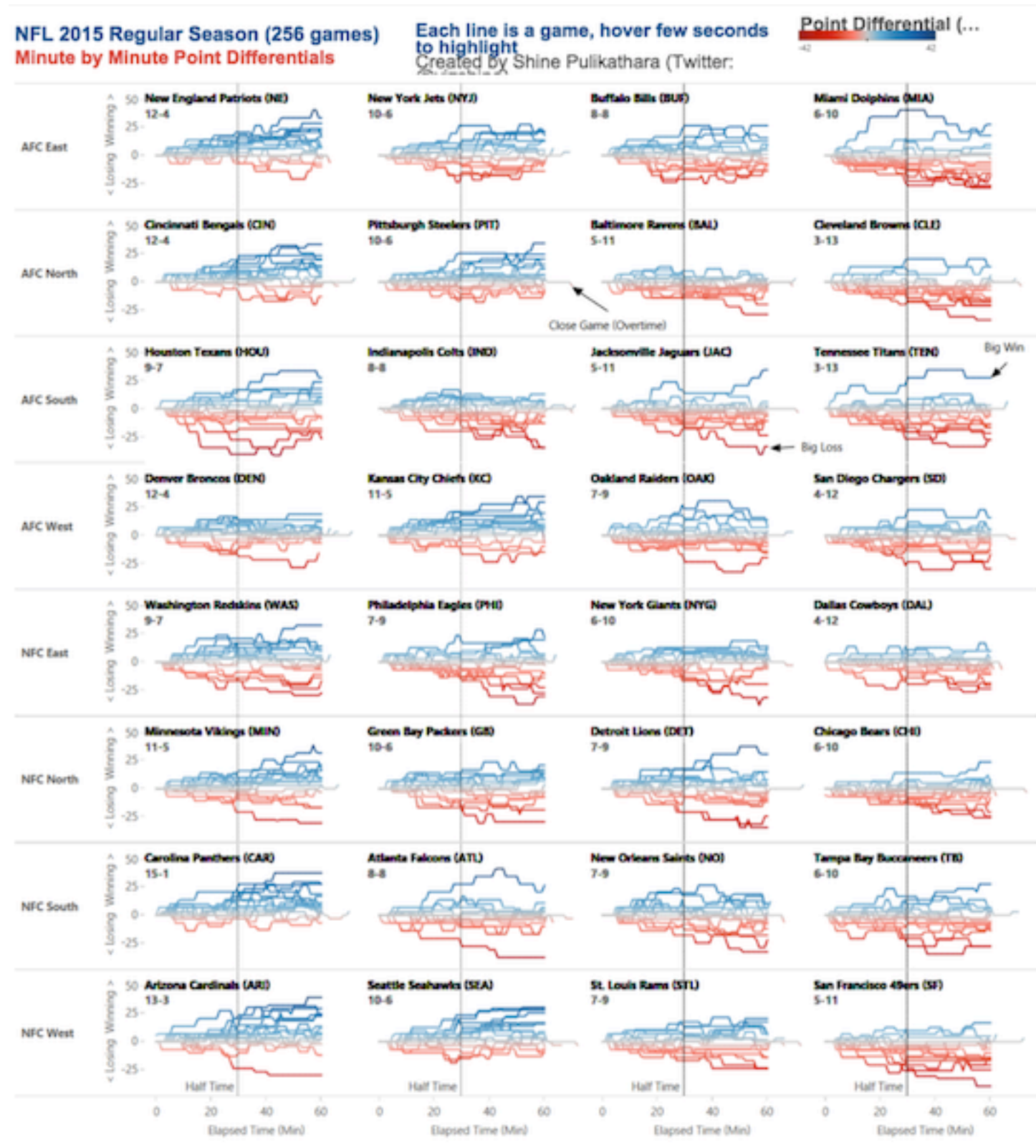
# SEQUENTIAL COLOURS FOR ORDER (LOW TO HIGH)

.....





# DIVERGENT COLOURS FOR VALUES WITH CRITICAL MIDPOINT

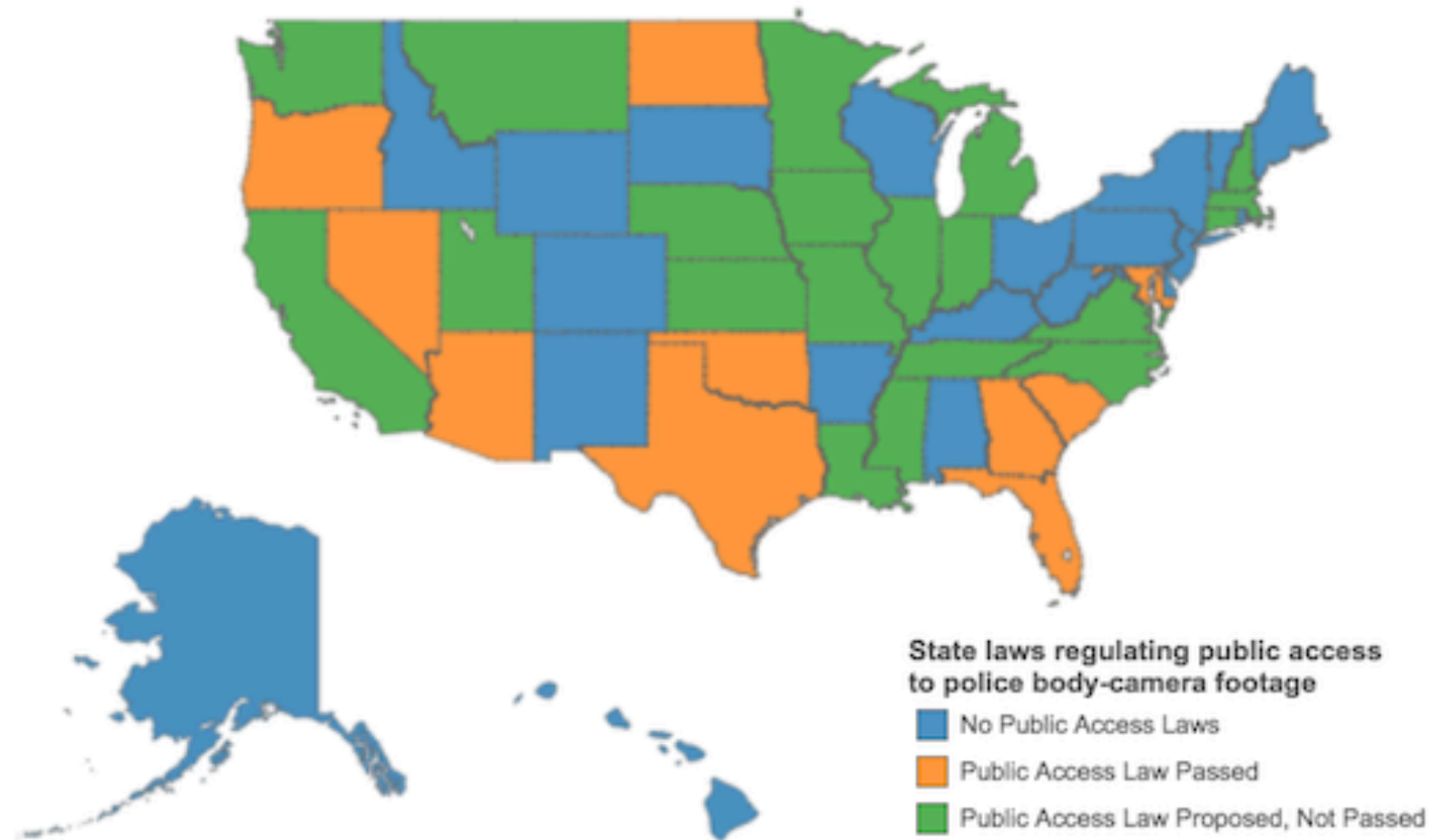


# CATEGORICAL COLOURS FOR GROUPING

.....

## 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.*



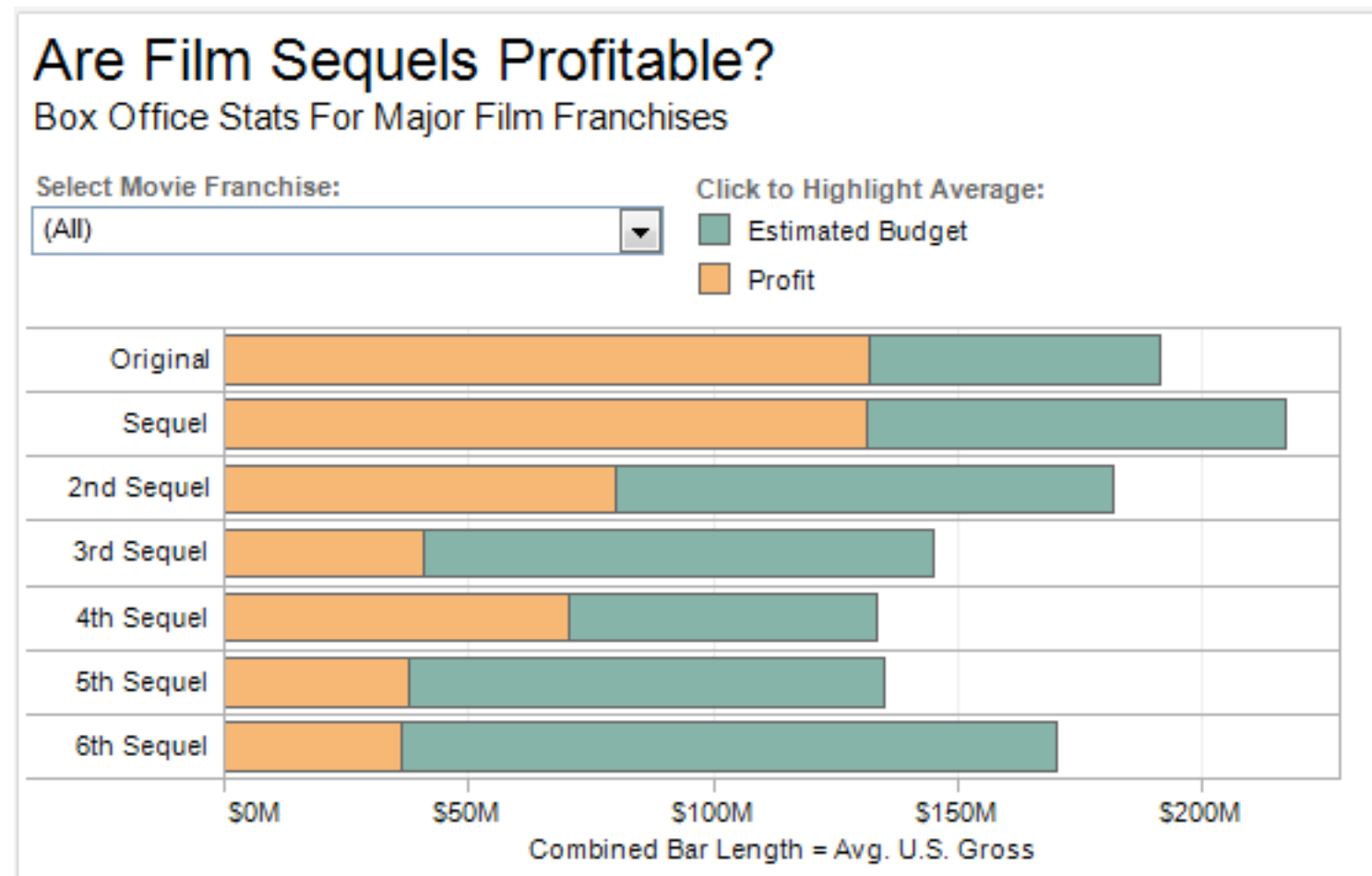
Source: Reporters Committee for Freedom of the Press



# BAR CHARTS

---

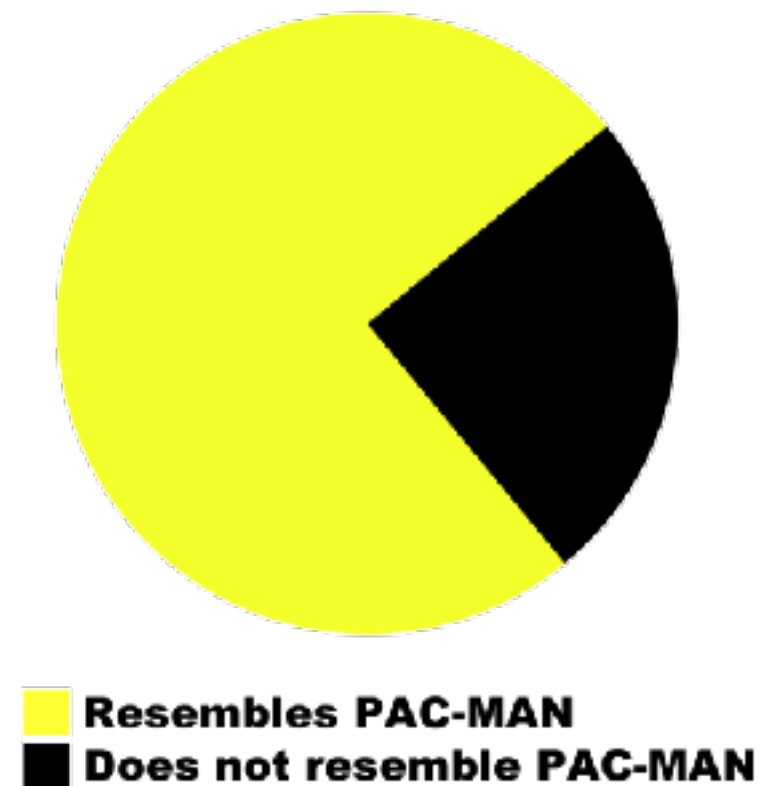
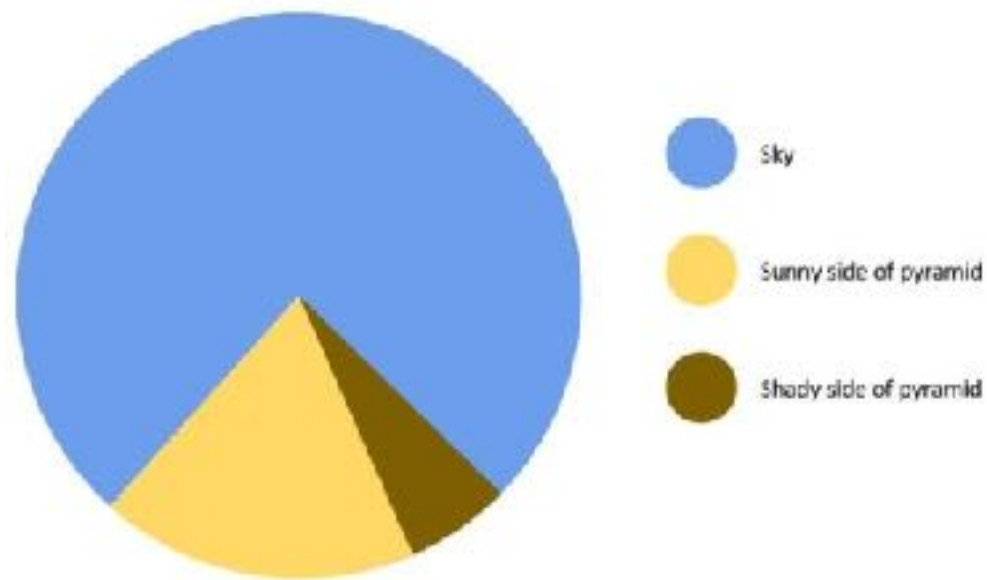
- Most common visualization
- Compare information
- Retrieve highs and lows
- Categorical  
Numerical



# PIE CHARTS

---

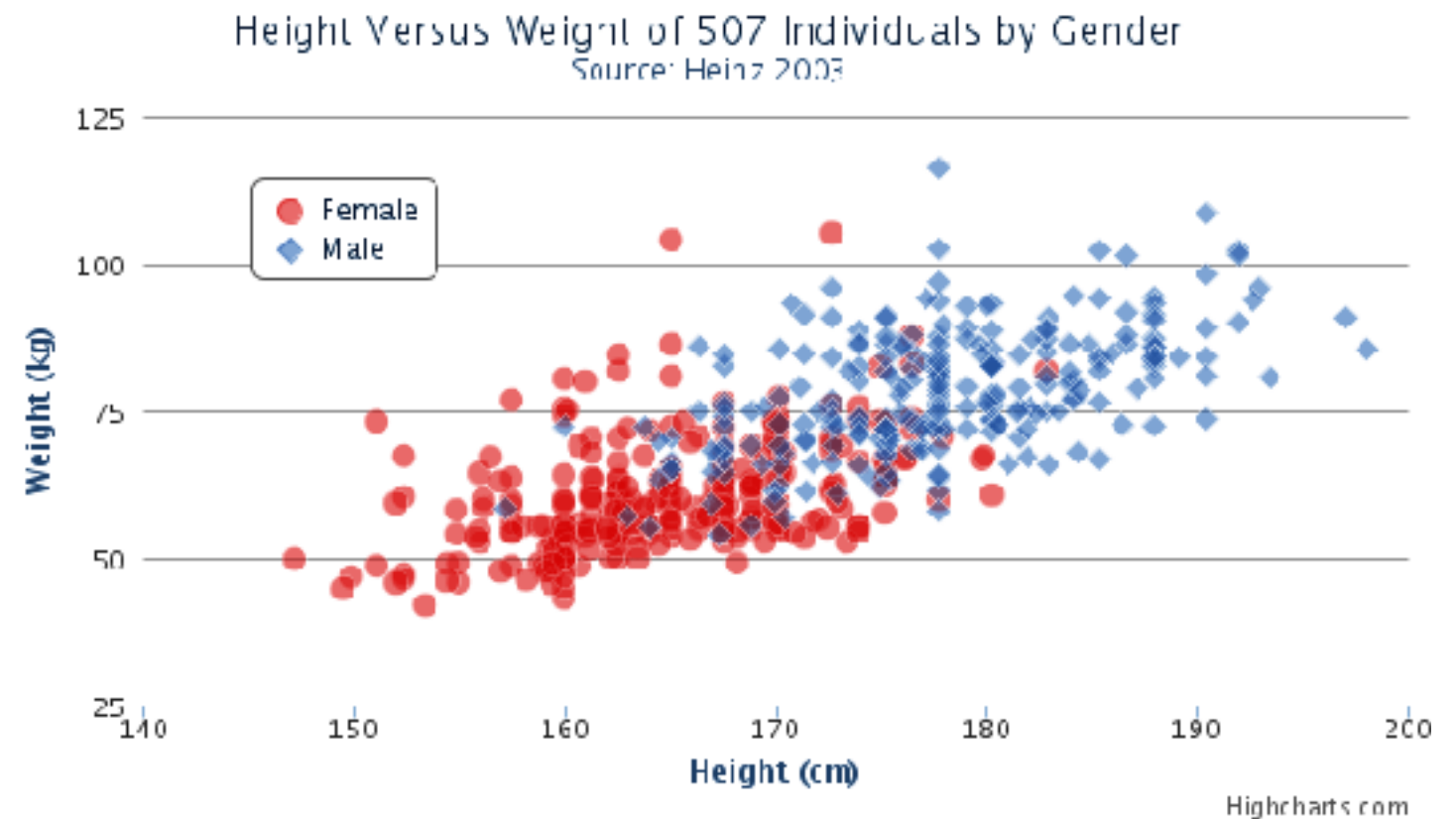
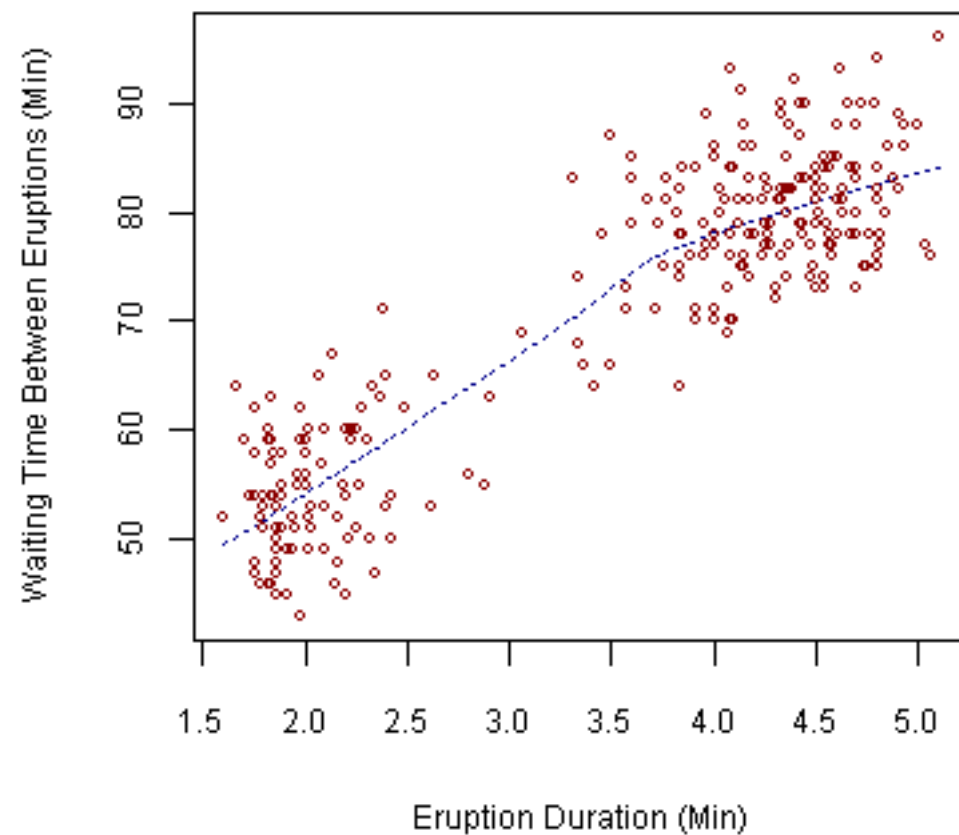
- Relative proportions or percentages of information.
- Can be easily mis-used (comparing multiple pie charts)



# SCATTER PLOTS

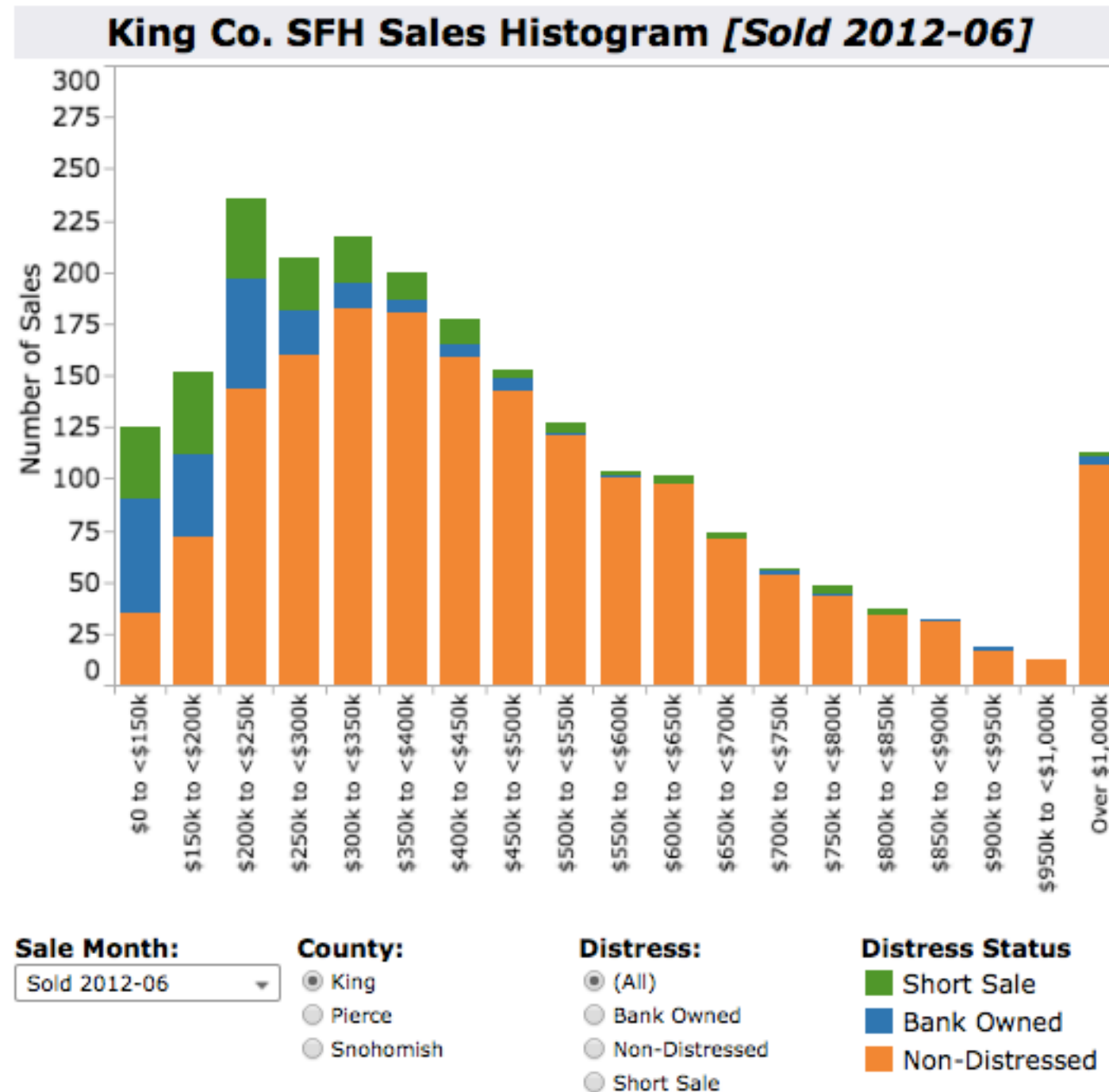
---

- Heavily used in data analysis
- Provide a sense of trends, concentrations, and outliers.



# HISTOGRAMS

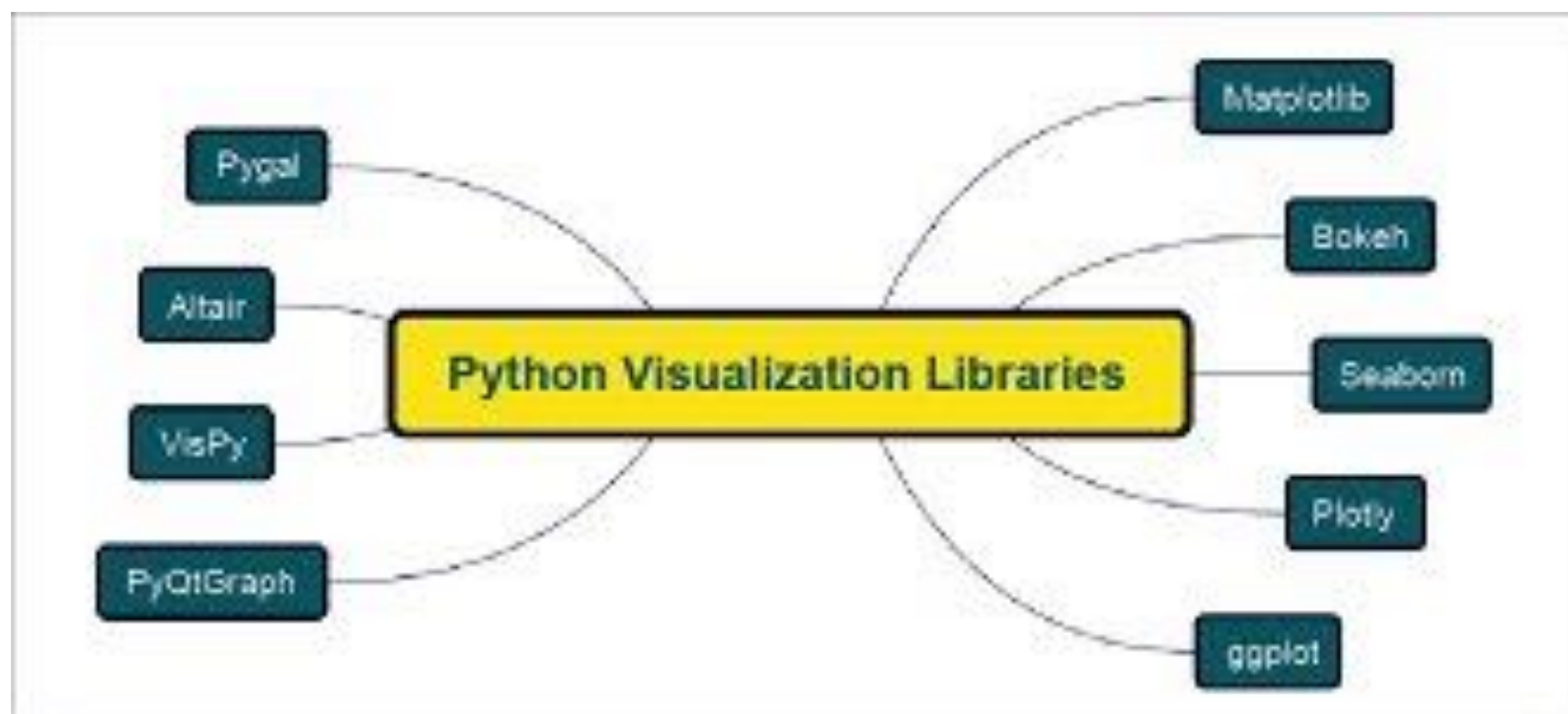
- Histograms are useful when you want to see how your data are distributed across groups.



# PYTHON VISUALIZATION LIBRARIES

---

- We shall use matplotlib and seaborn libs. For in class exercises and demos.
- Plotly library for high end interactive graphics



# USEFUL LINKS

---

- SAS data visualization
- ([http://www.sas.com/en\\_us/insights/big-data/data-visualization.html](http://www.sas.com/en_us/insights/big-data/data-visualization.html))
- Dear data
- <https://www.youtube.com/watch?v=iqaVe1MCTlA>
- Big data and data viz
- [https://www.youtube.com/watch?v=1\\_c18g\\_Lpu8](https://www.youtube.com/watch?v=1_c18g_Lpu8)
- London moves
- <https://www.youtube.com/watch?v=FaRBUnO5PZI>