

Techniques

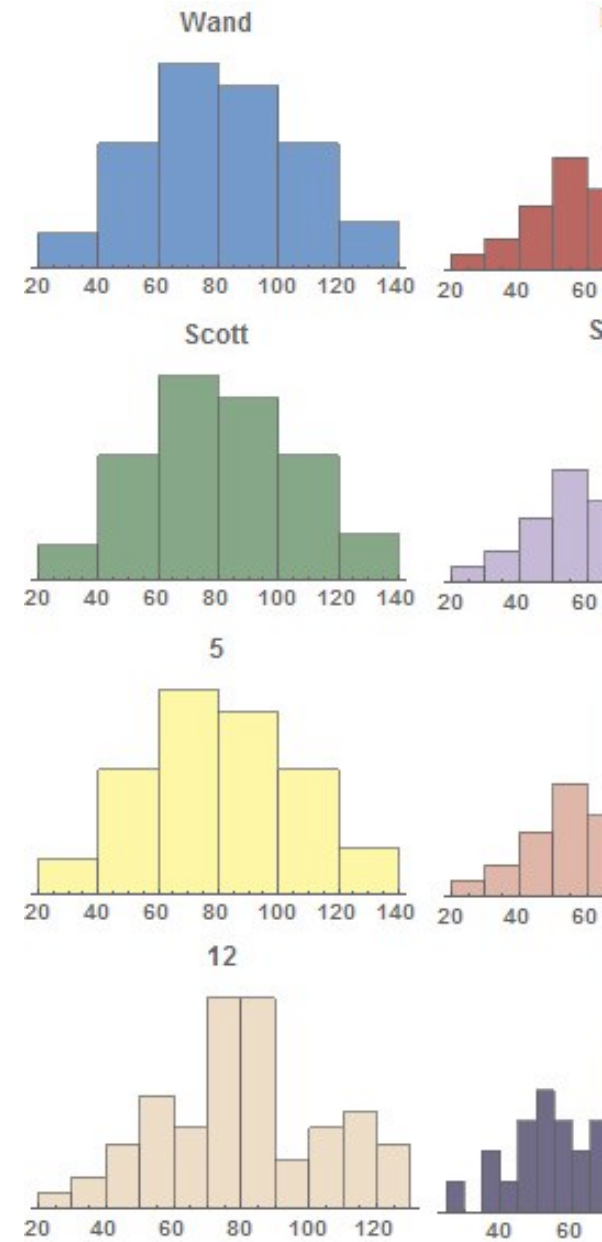
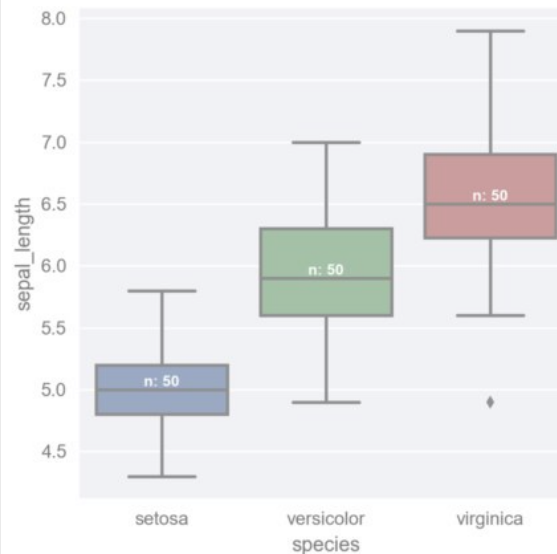
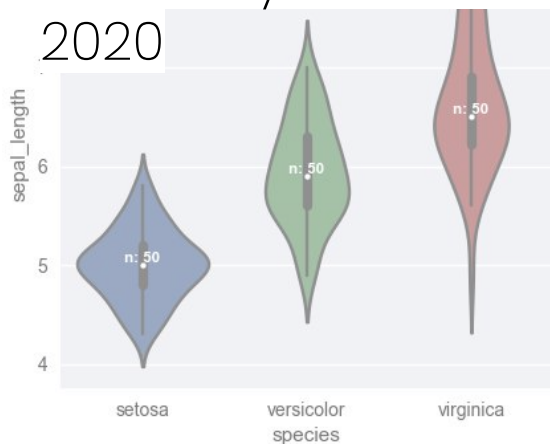
Basic Charts

Benjamin Bach

<http://benjbach.me>

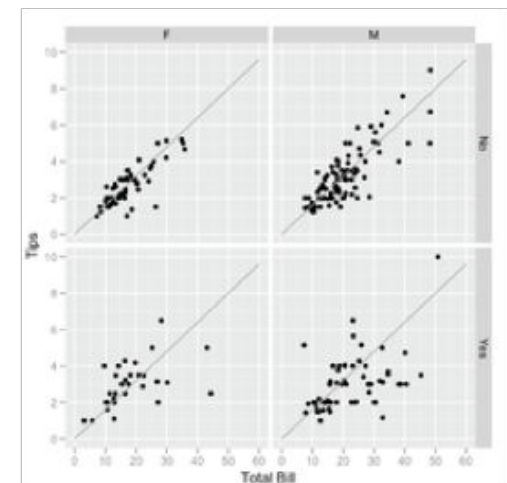
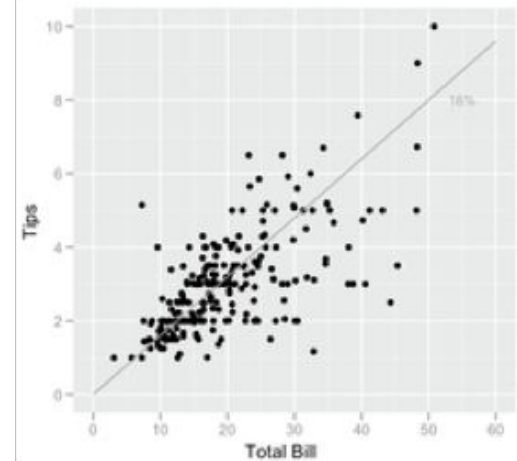
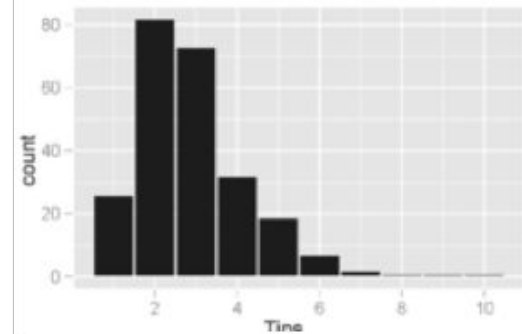
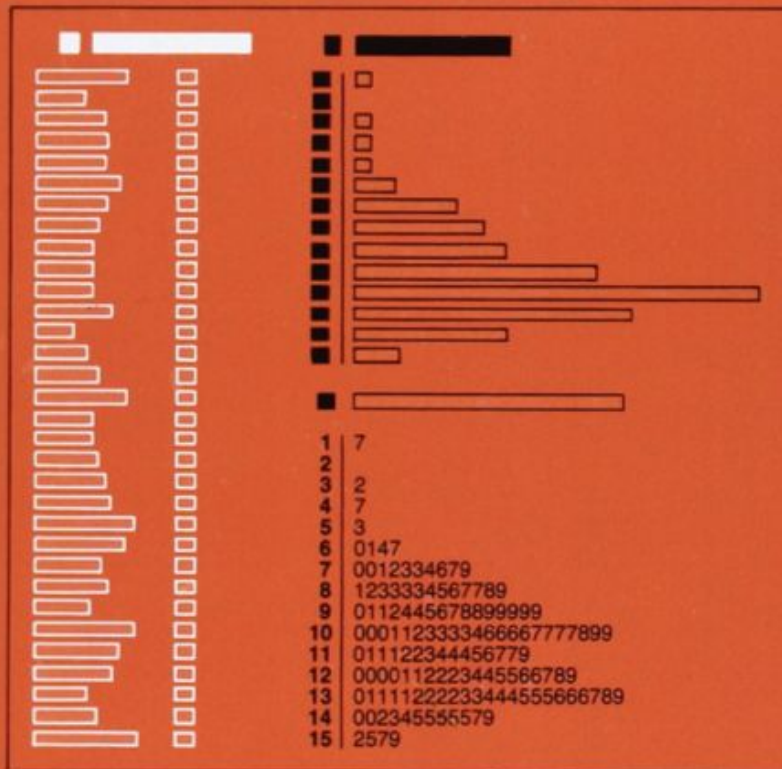
University of Edinburgh

2020



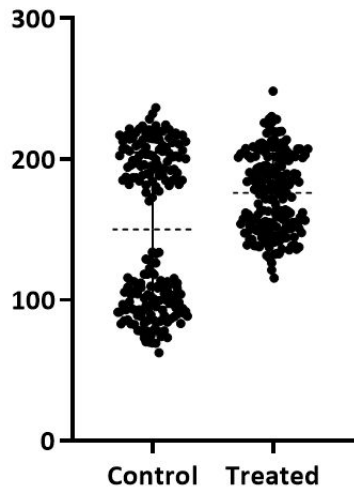
John W. Tukey

EXPLORATORY DATA ANALYSIS

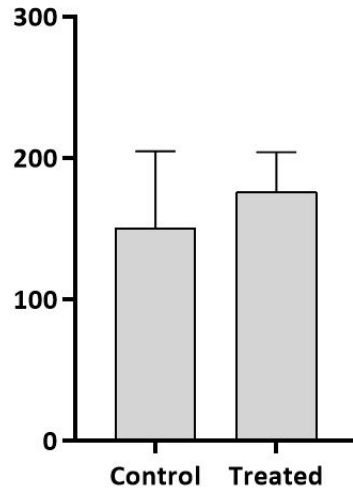


Basic Charts

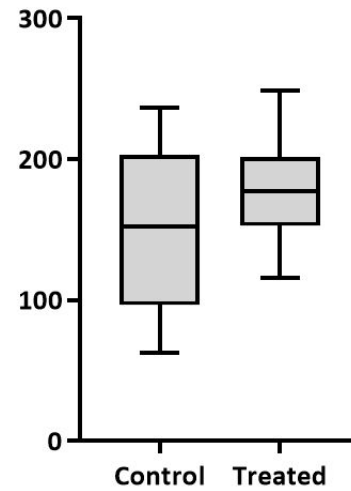
?



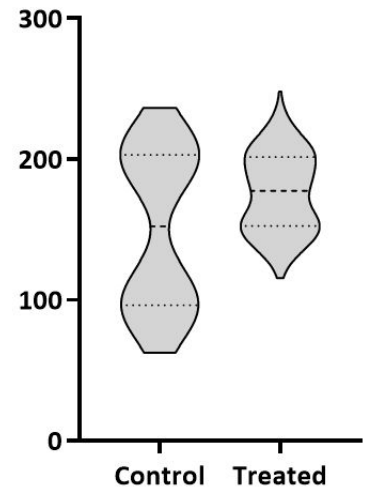
?



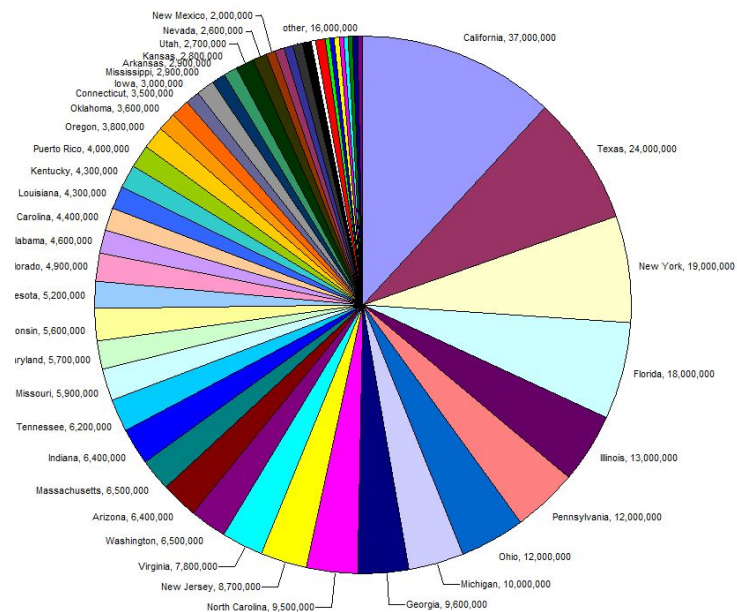
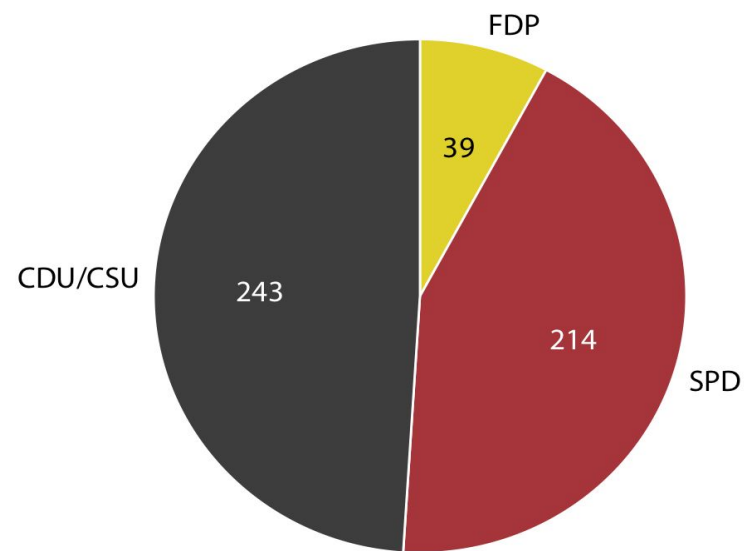
?



?



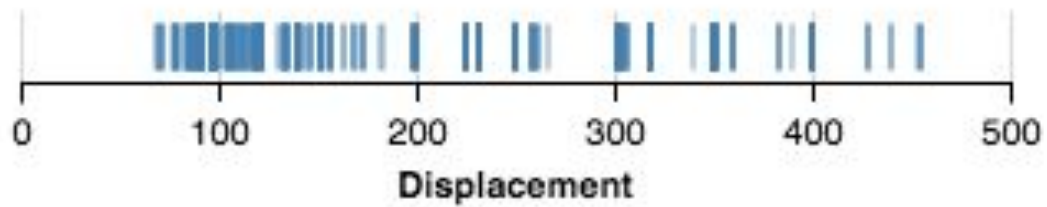
Basic Charts



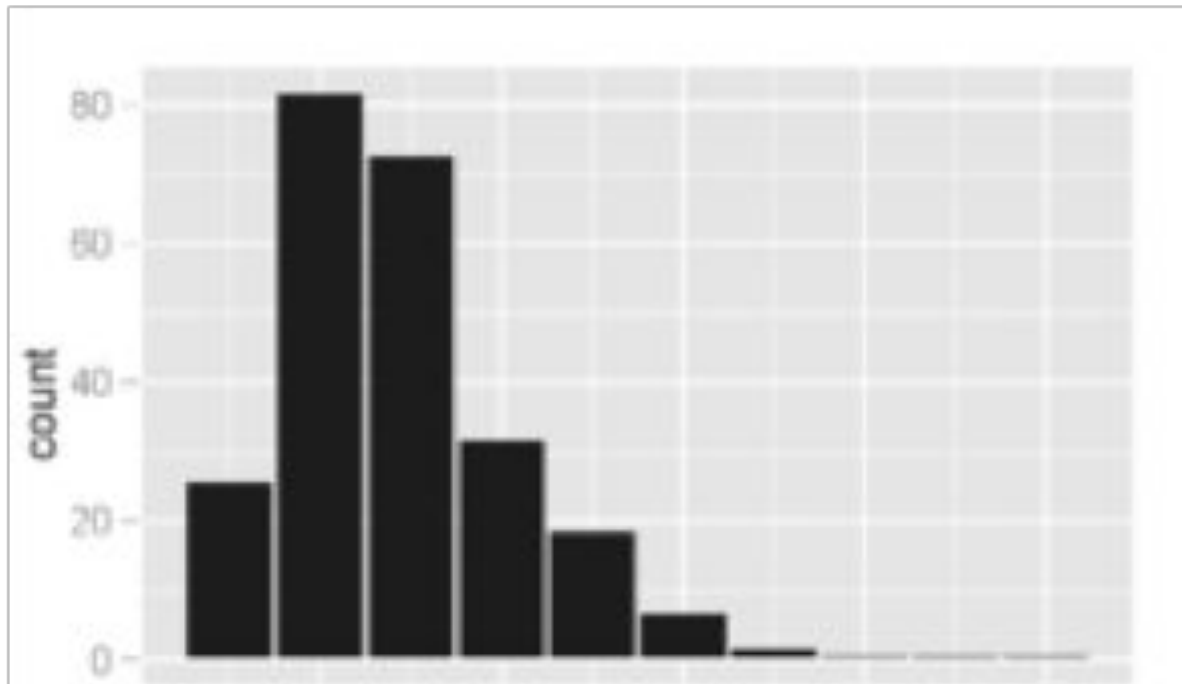
Outline

- Distributions & Histograms
- Mean, Mode, Average
- Chart types
- Uncertainty
- Pie Charts

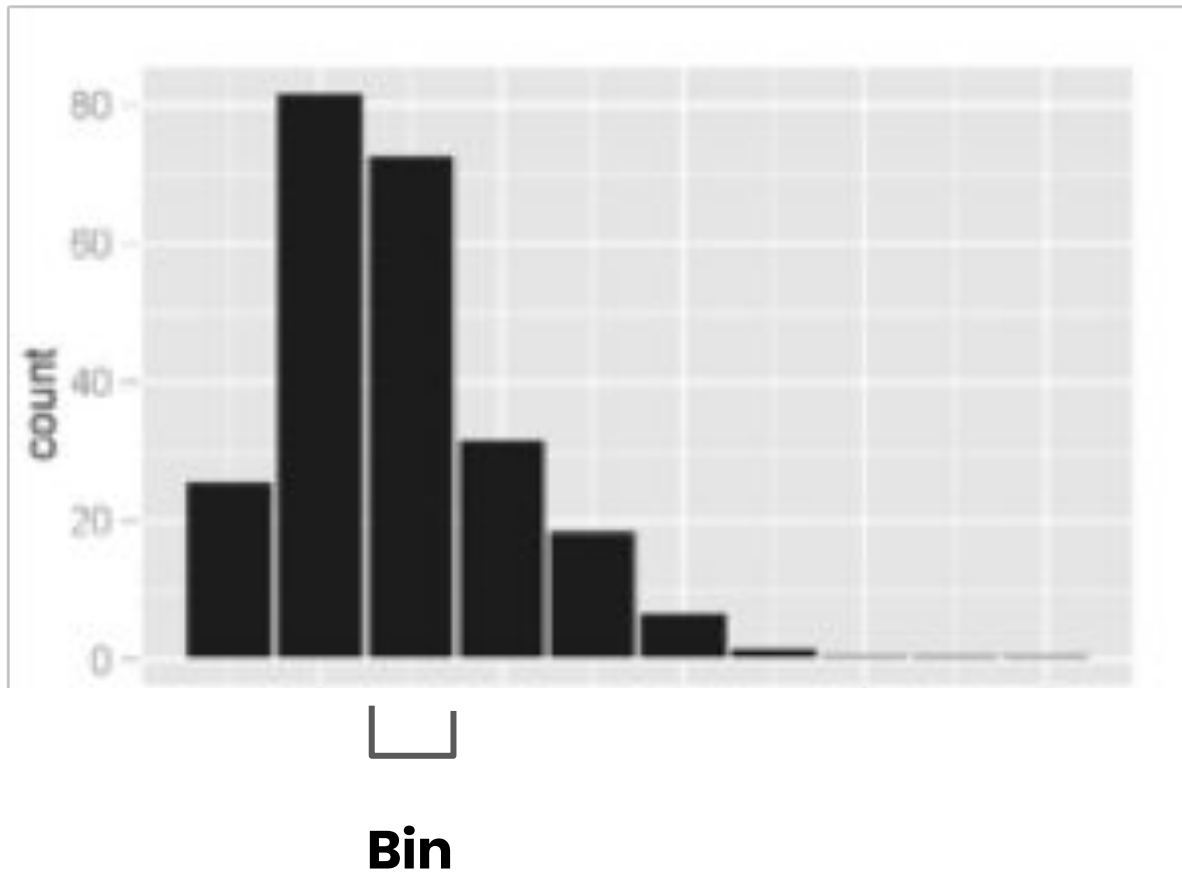
Distributions



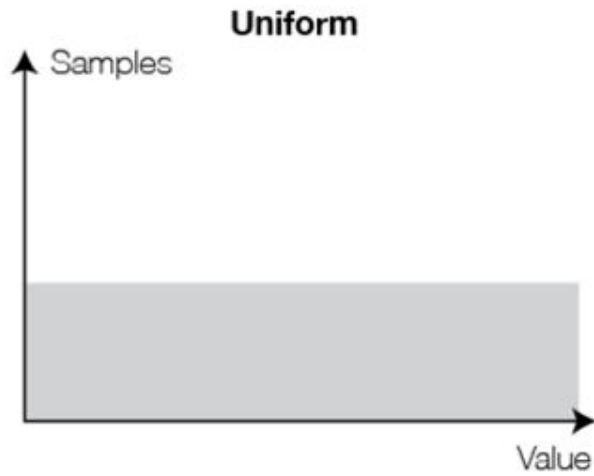
Distributions: Binning



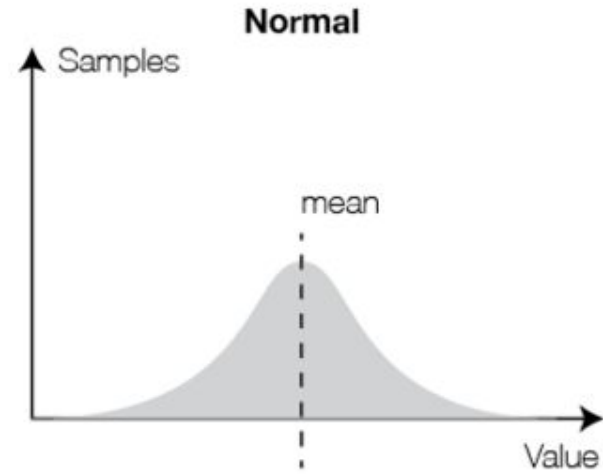
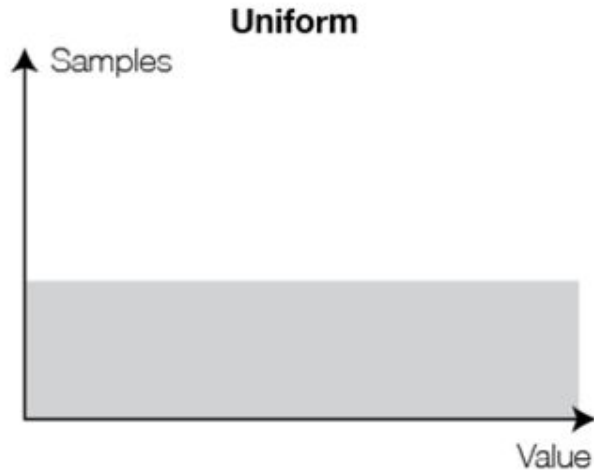
Distributions: Binning



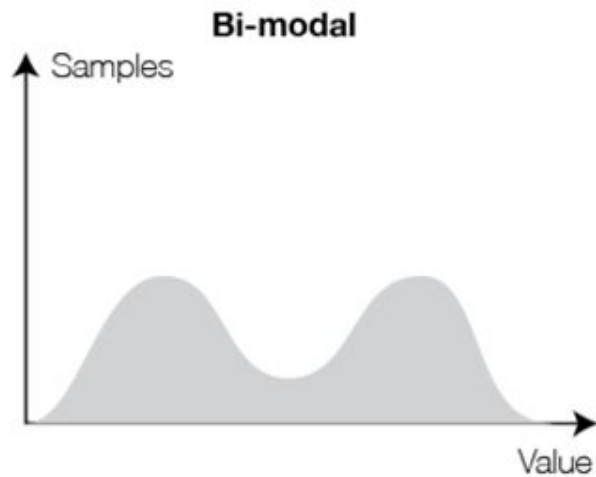
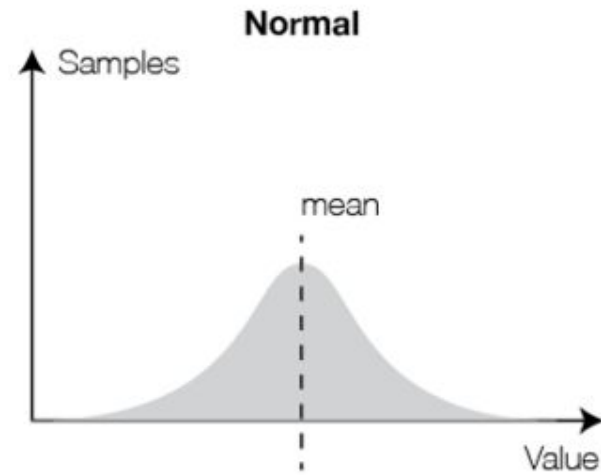
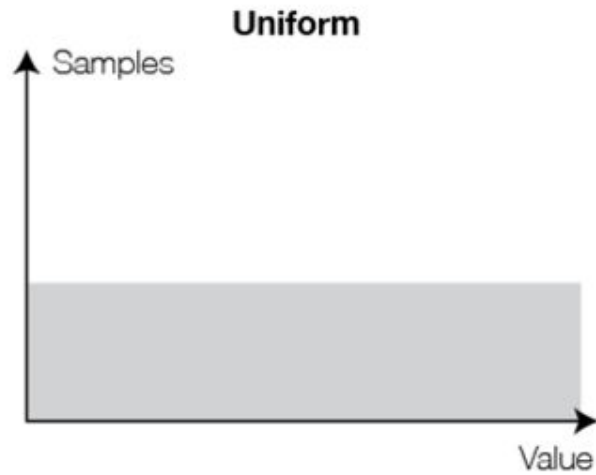
Distributions



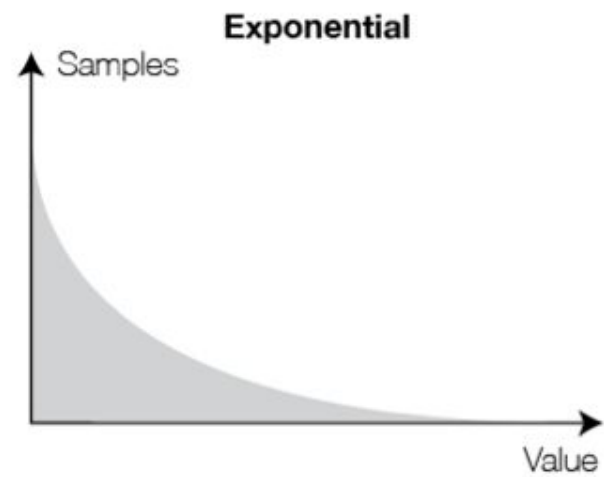
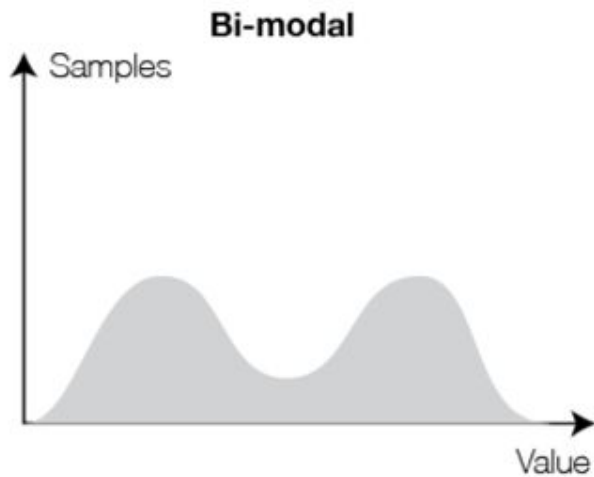
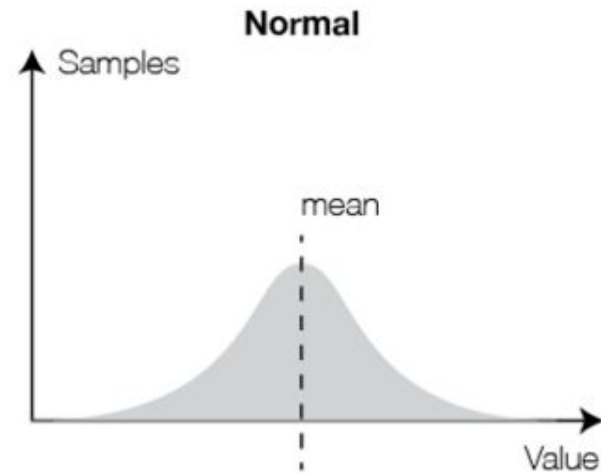
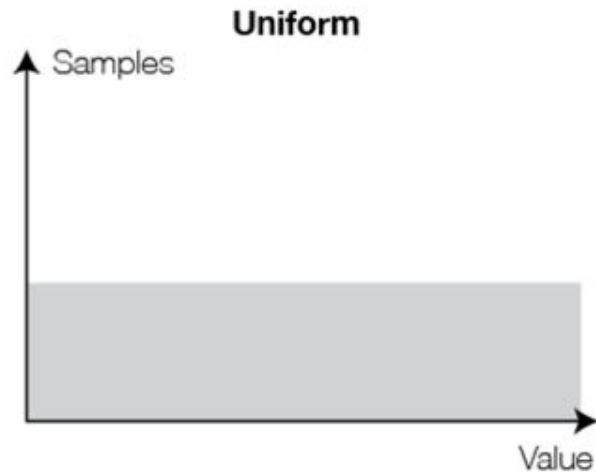
Distributions



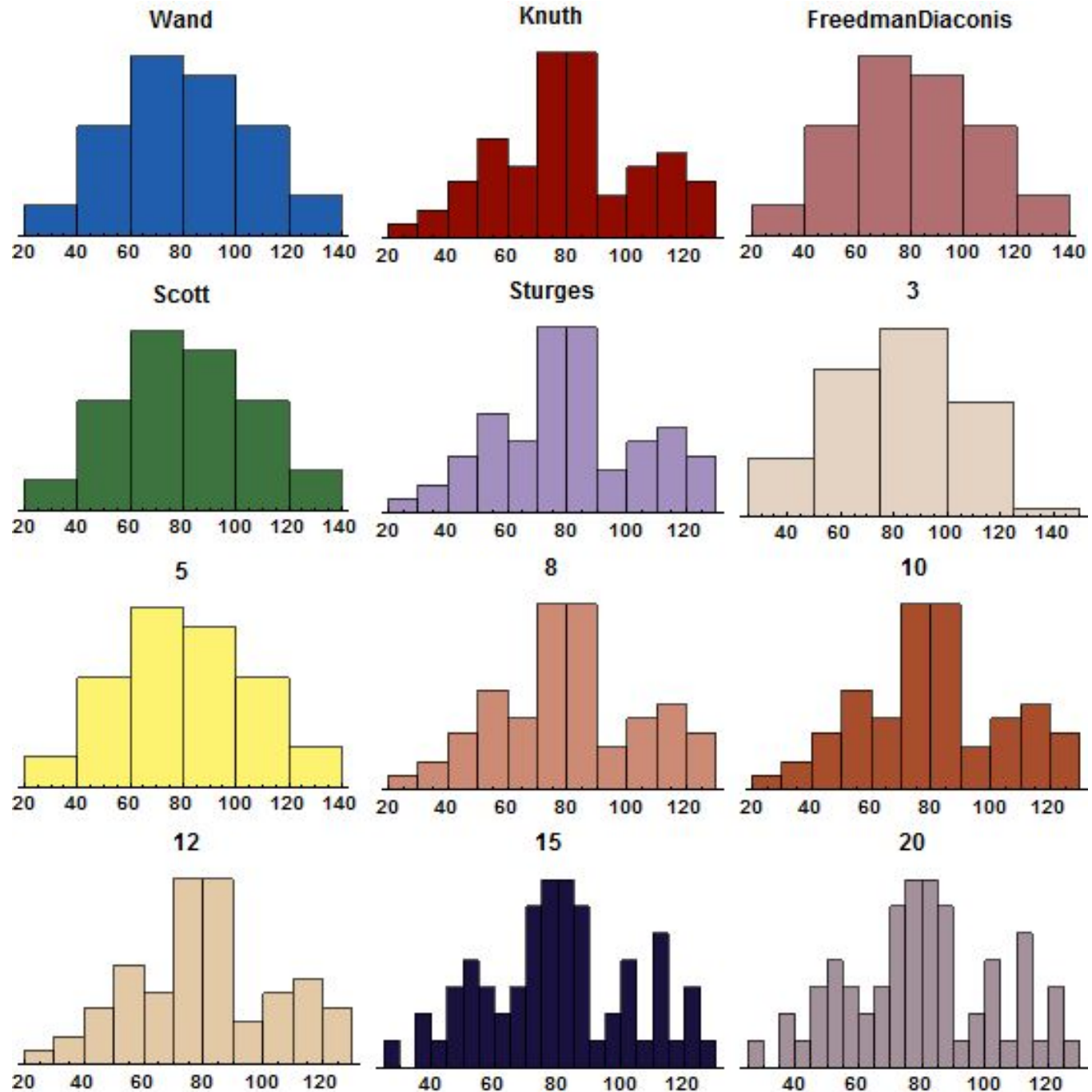
Distributions



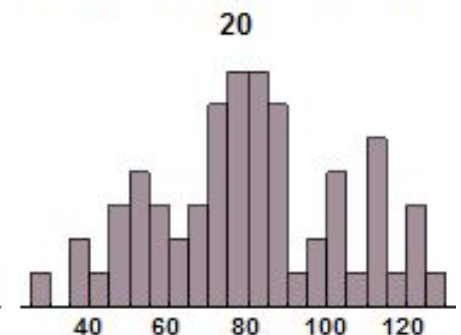
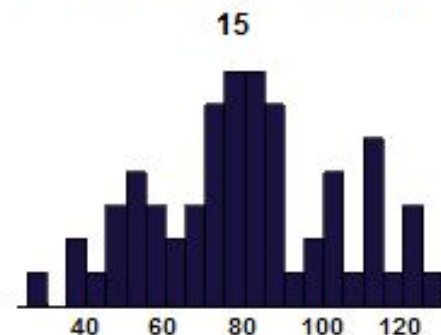
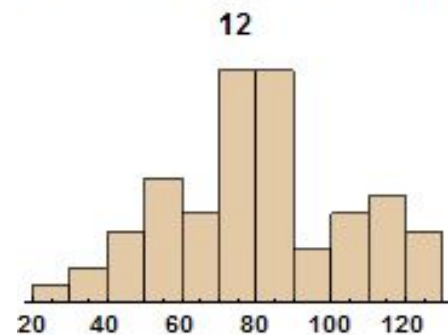
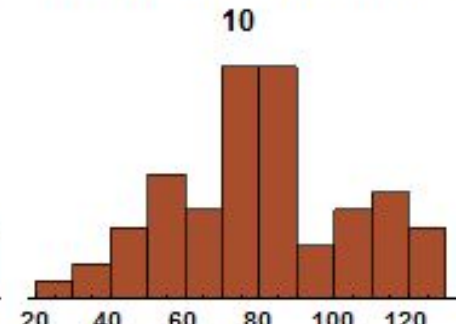
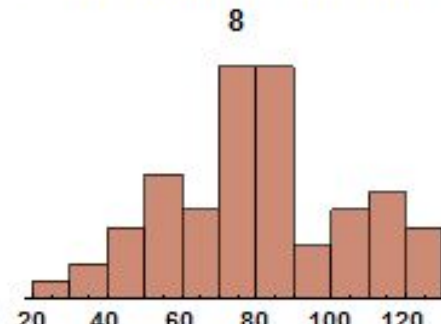
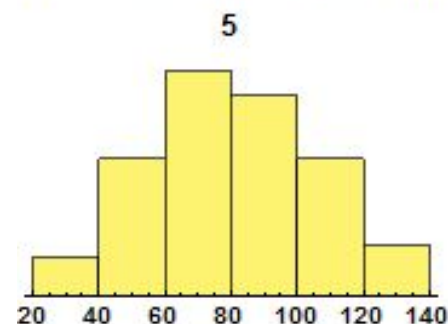
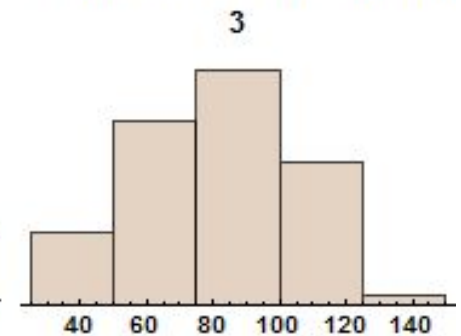
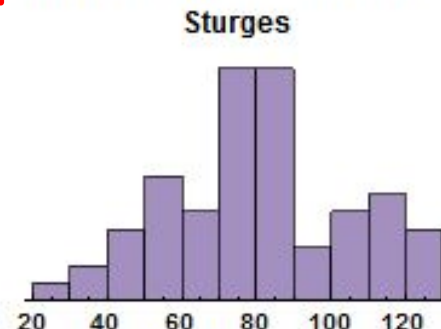
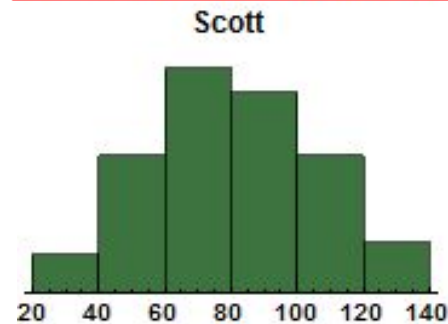
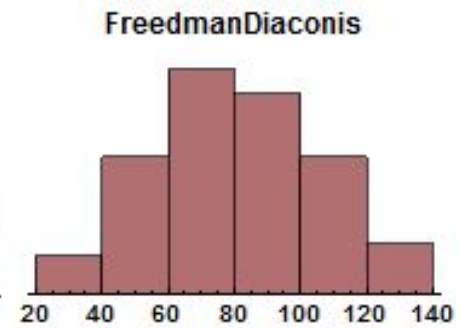
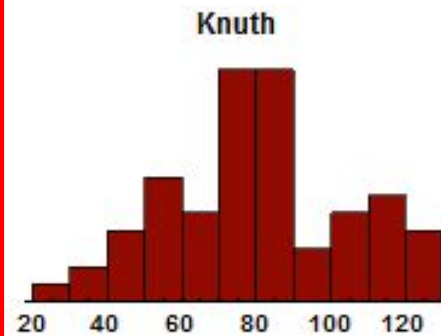
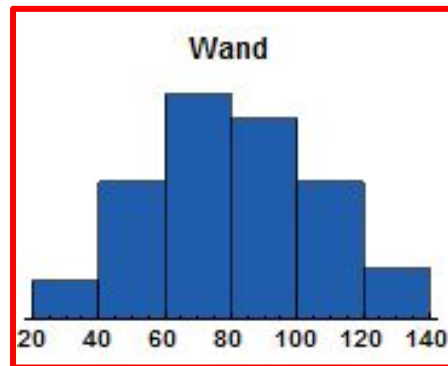
Distributions



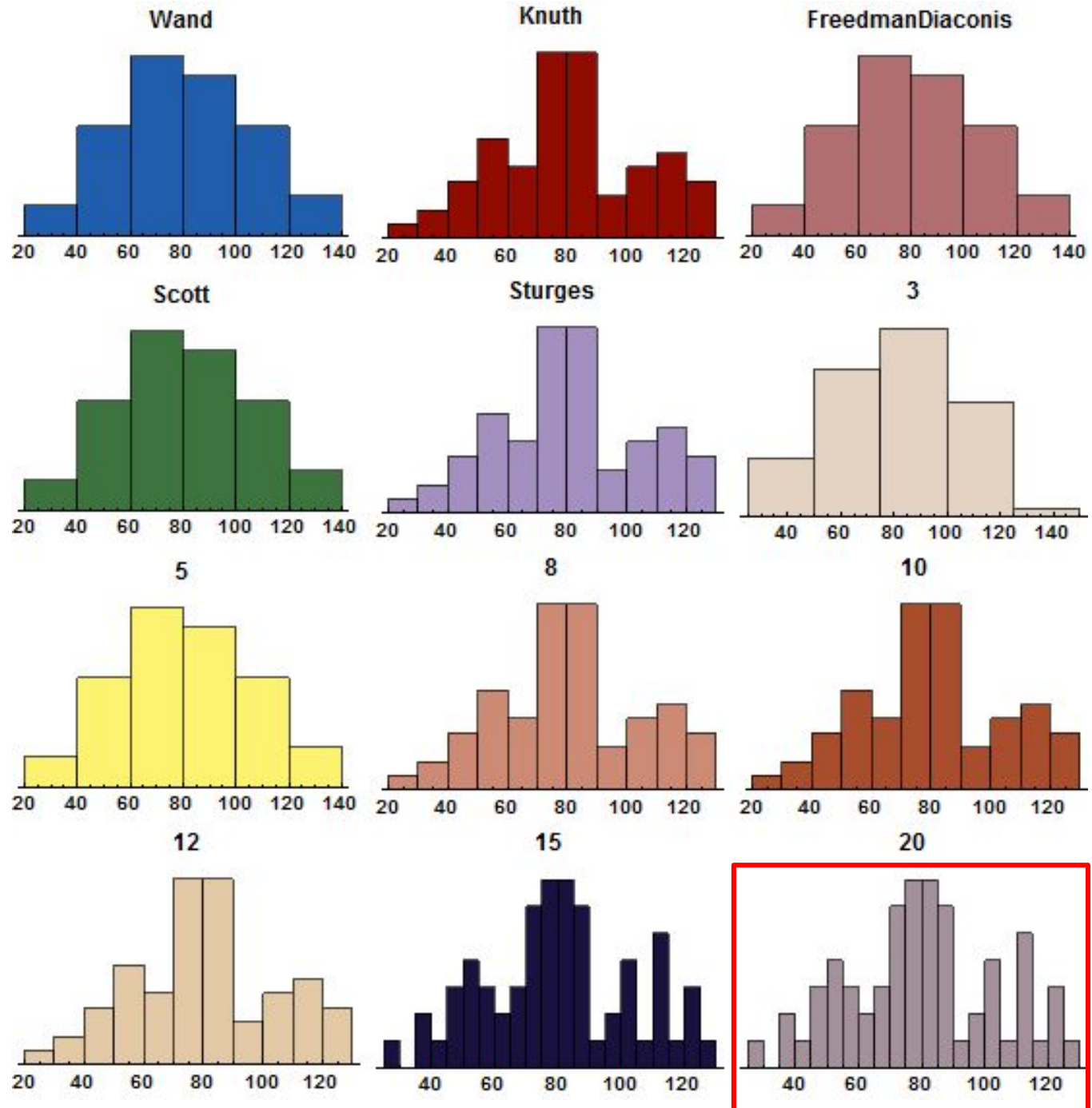
Histogram Binning



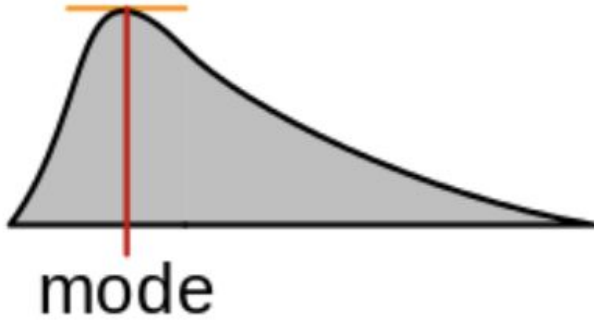
Histogram Binning



Histogram Binning

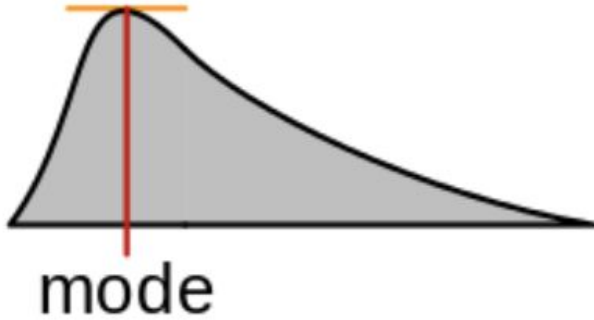


Descriptive Statistics

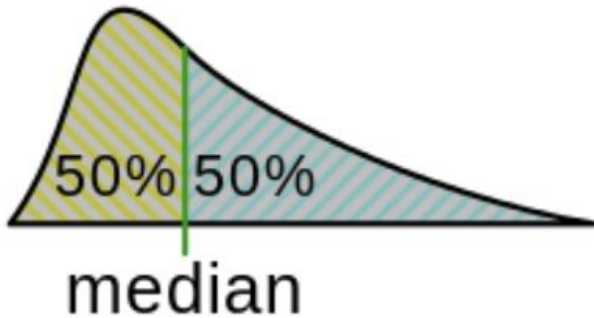


Mode: Most frequent value in data

Descriptive Statistics

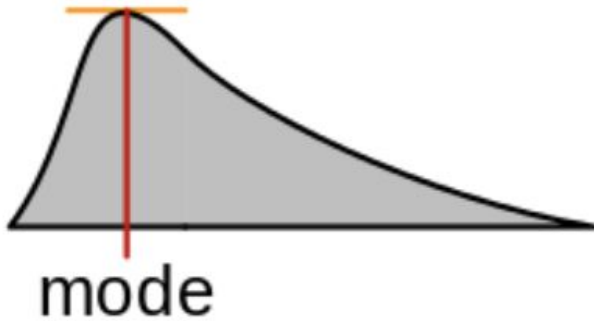


Mode: Most frequent value in data

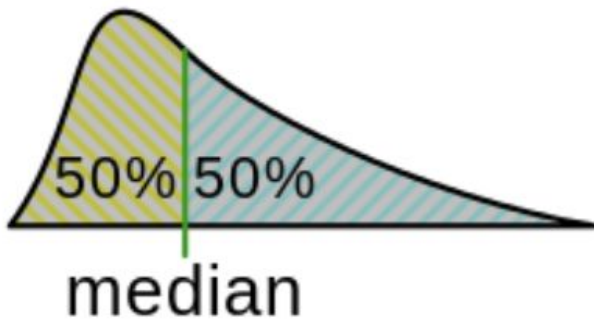


Median: splits samples in two equally sized sets (50:50)

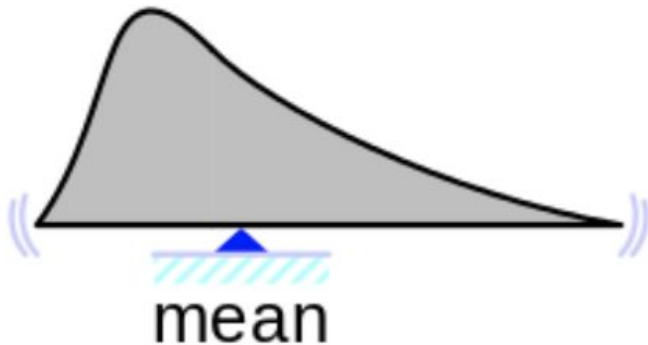
Descriptive Statistics



Mode: Most frequent value in data

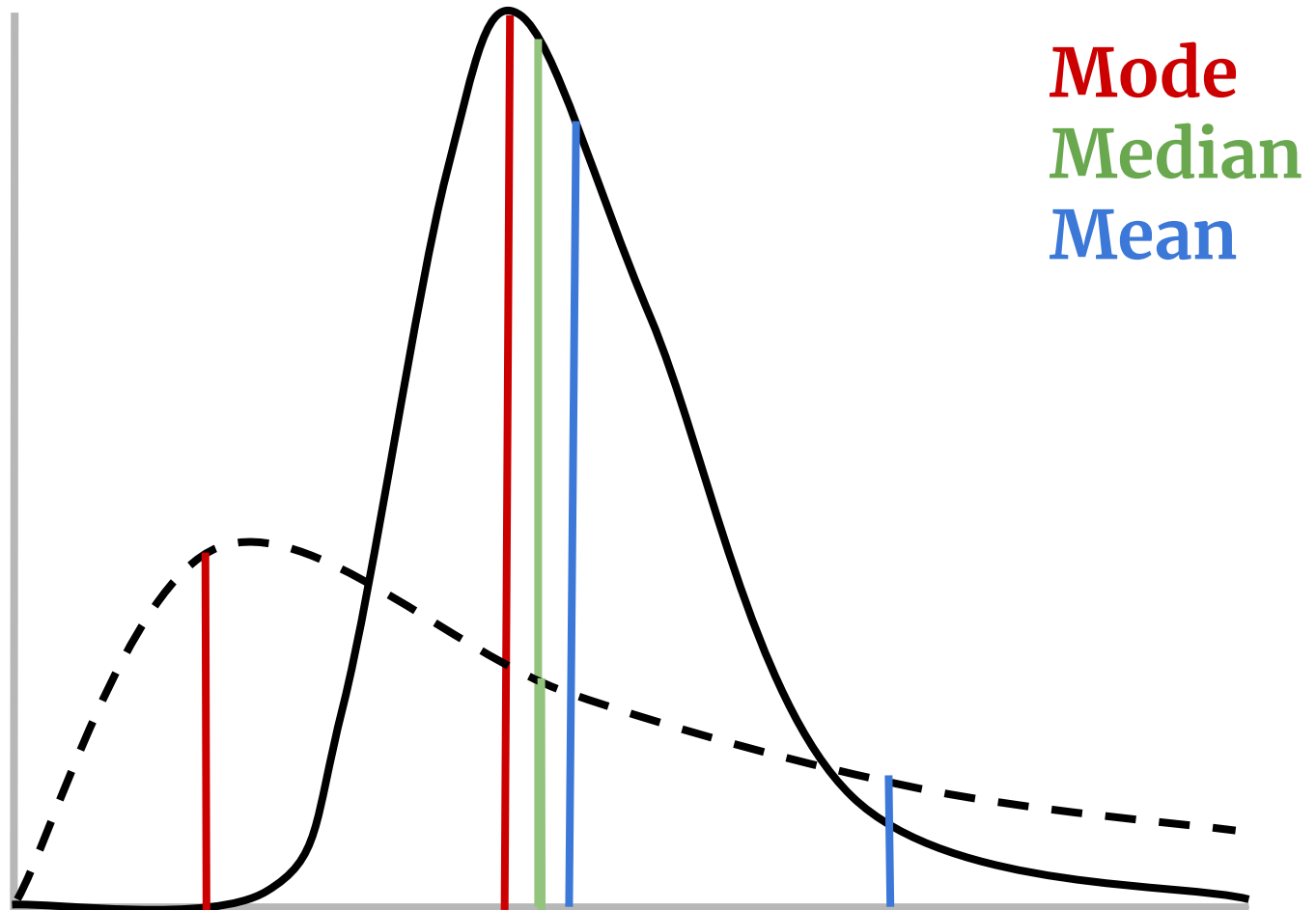


Median: splits samples in two equally sized sets (50:50)

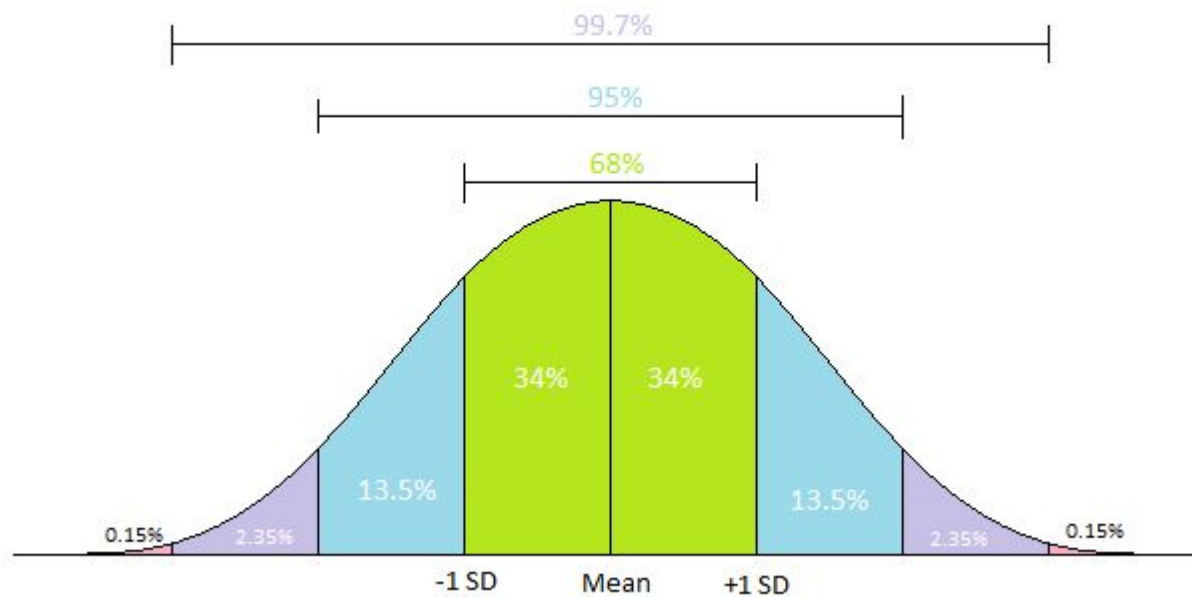
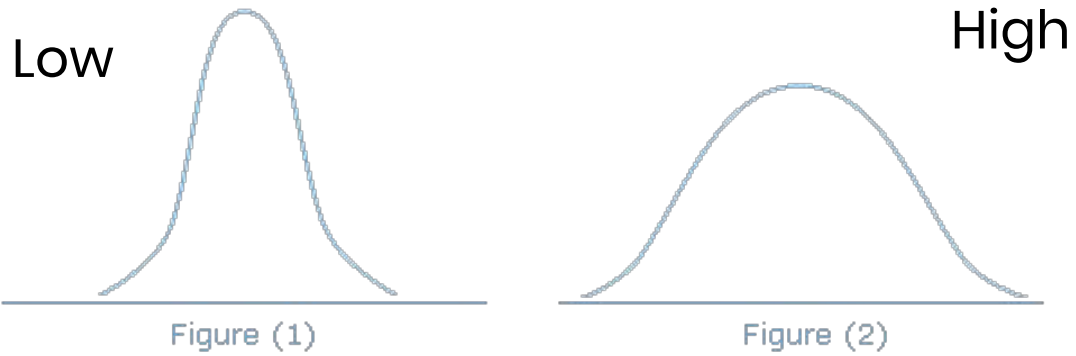


Mean: average value

Descriptive Statistics

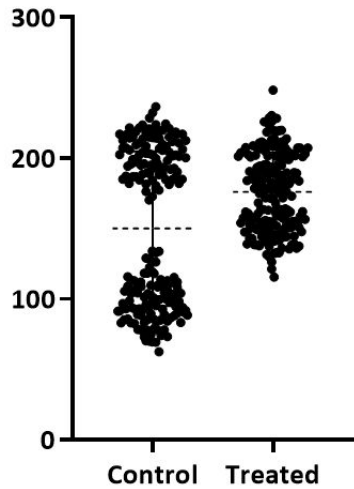


Standard deviation

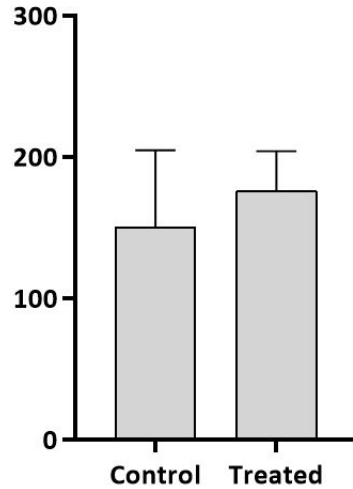


Visualizing Distributions

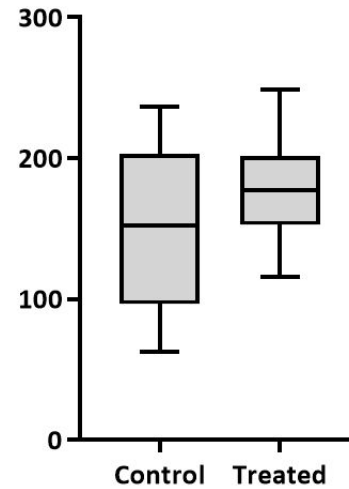
Plotting Distributions



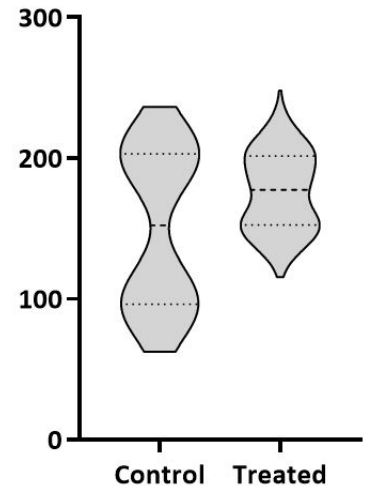
Dot plot/
Beeswarm plot



Bar charts +
error bars

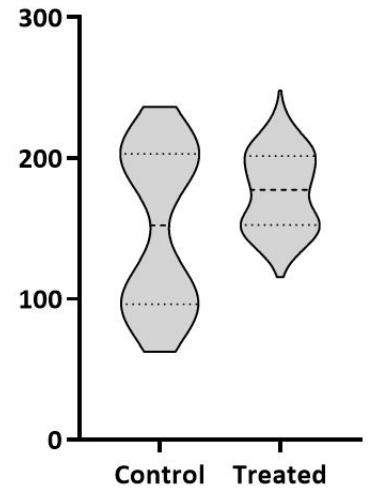
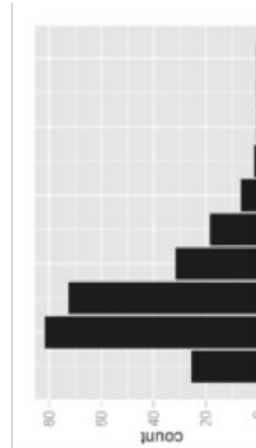
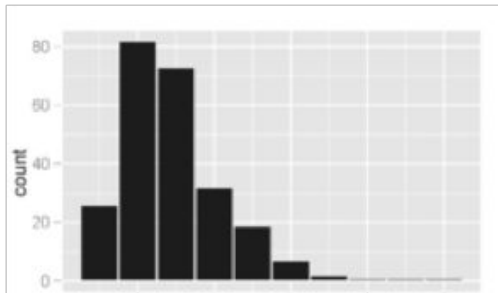


Box plots

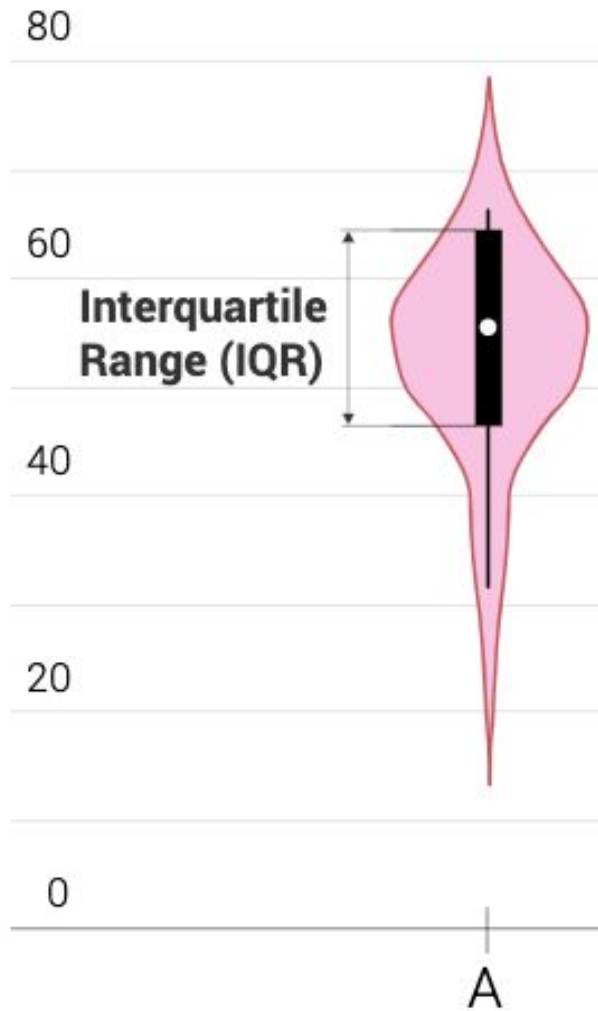


Violin plots

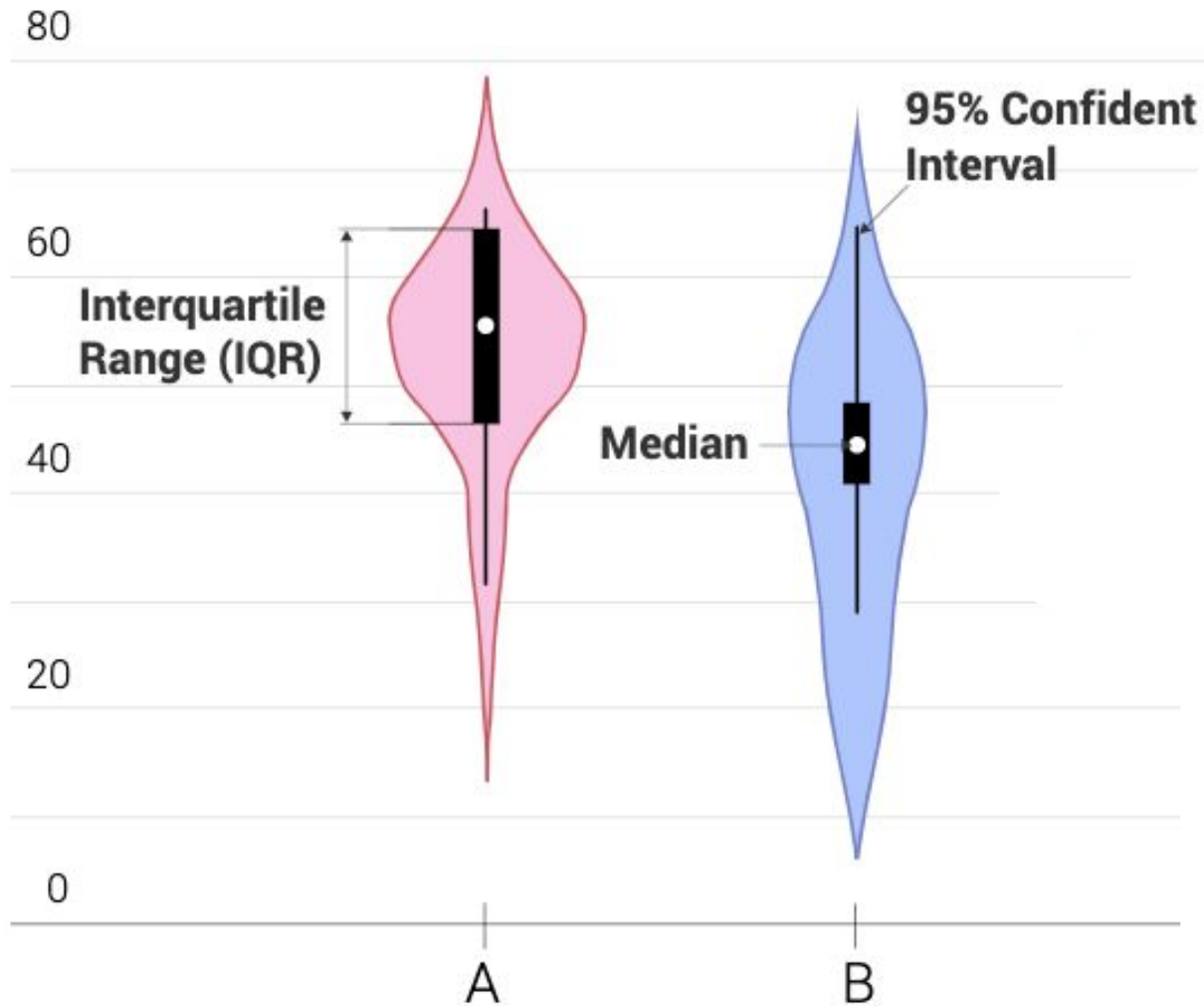
Violin plots



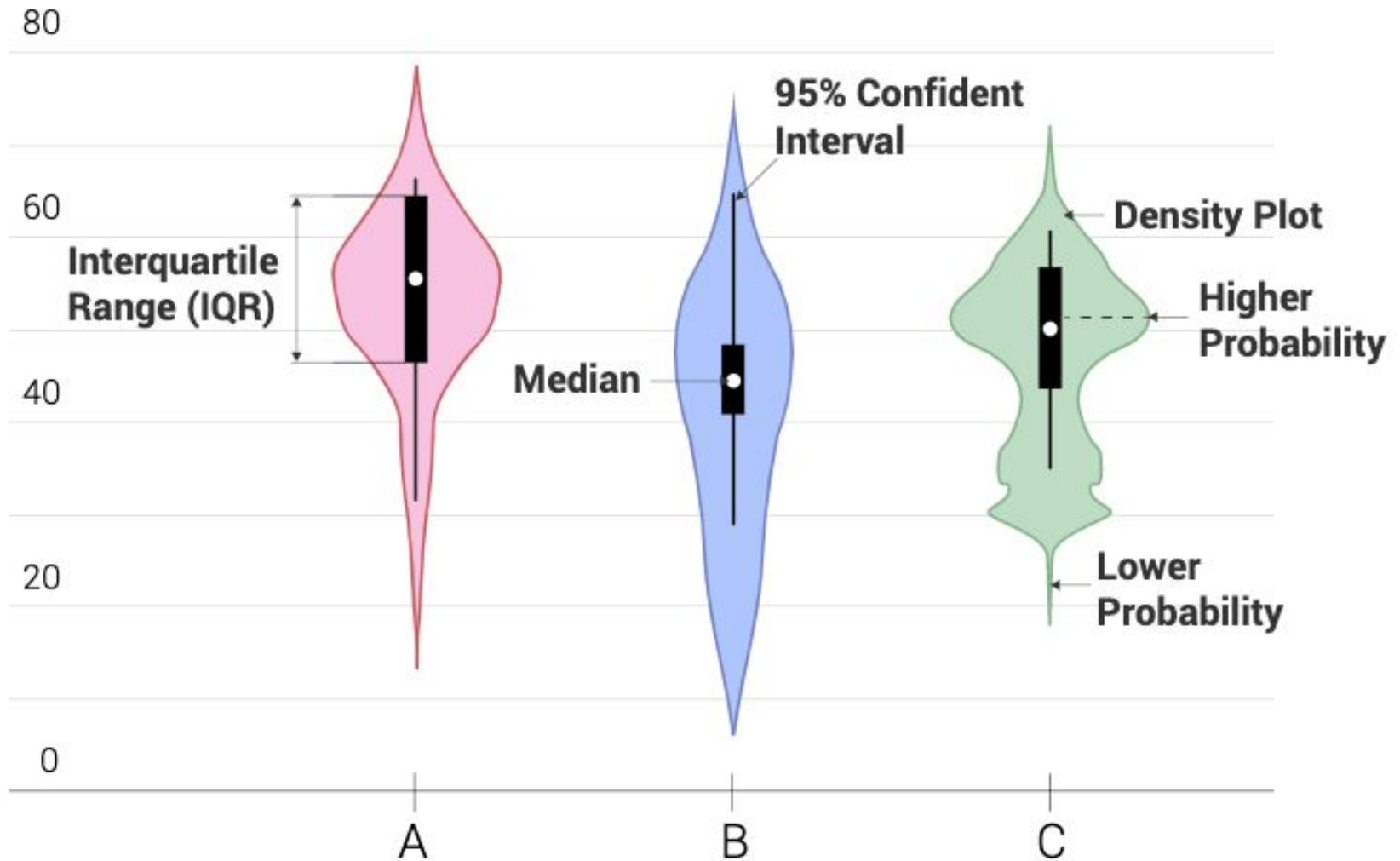
Violin plots: measures



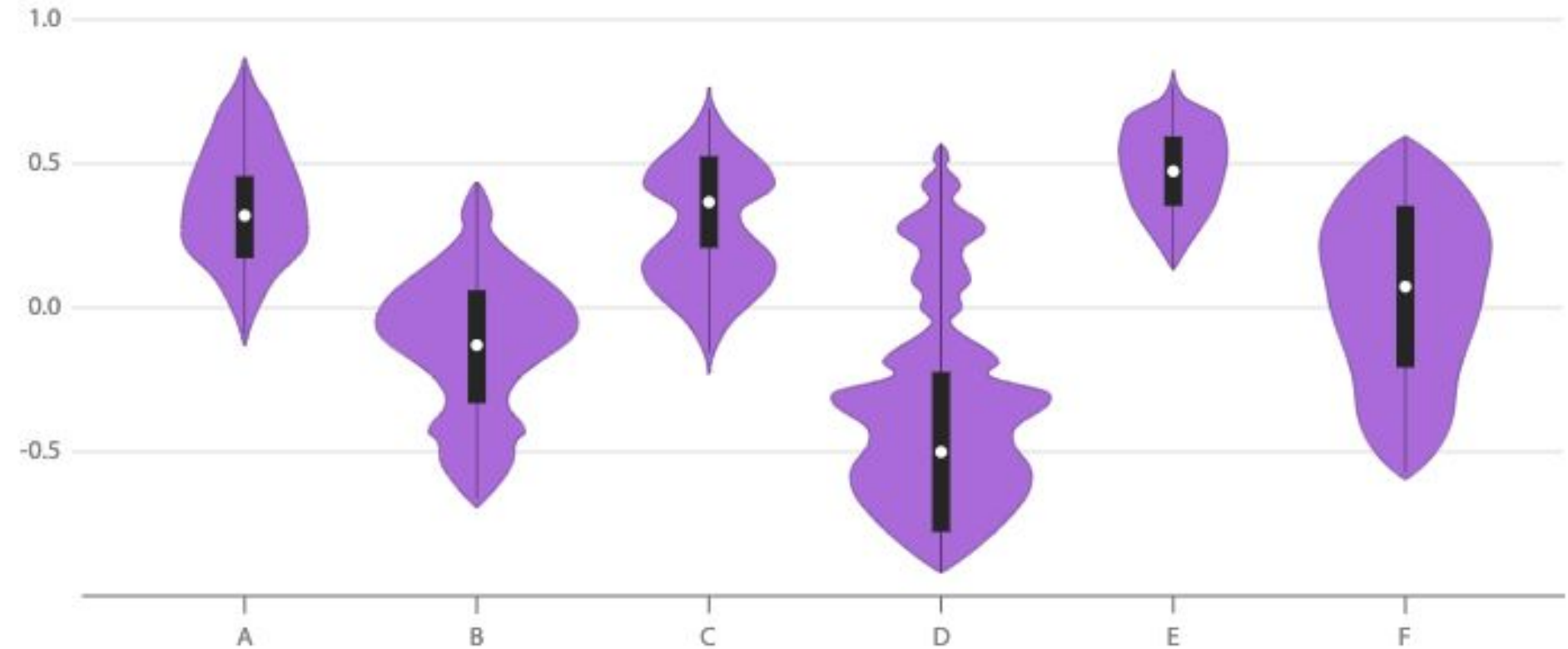
Violin plots: measures



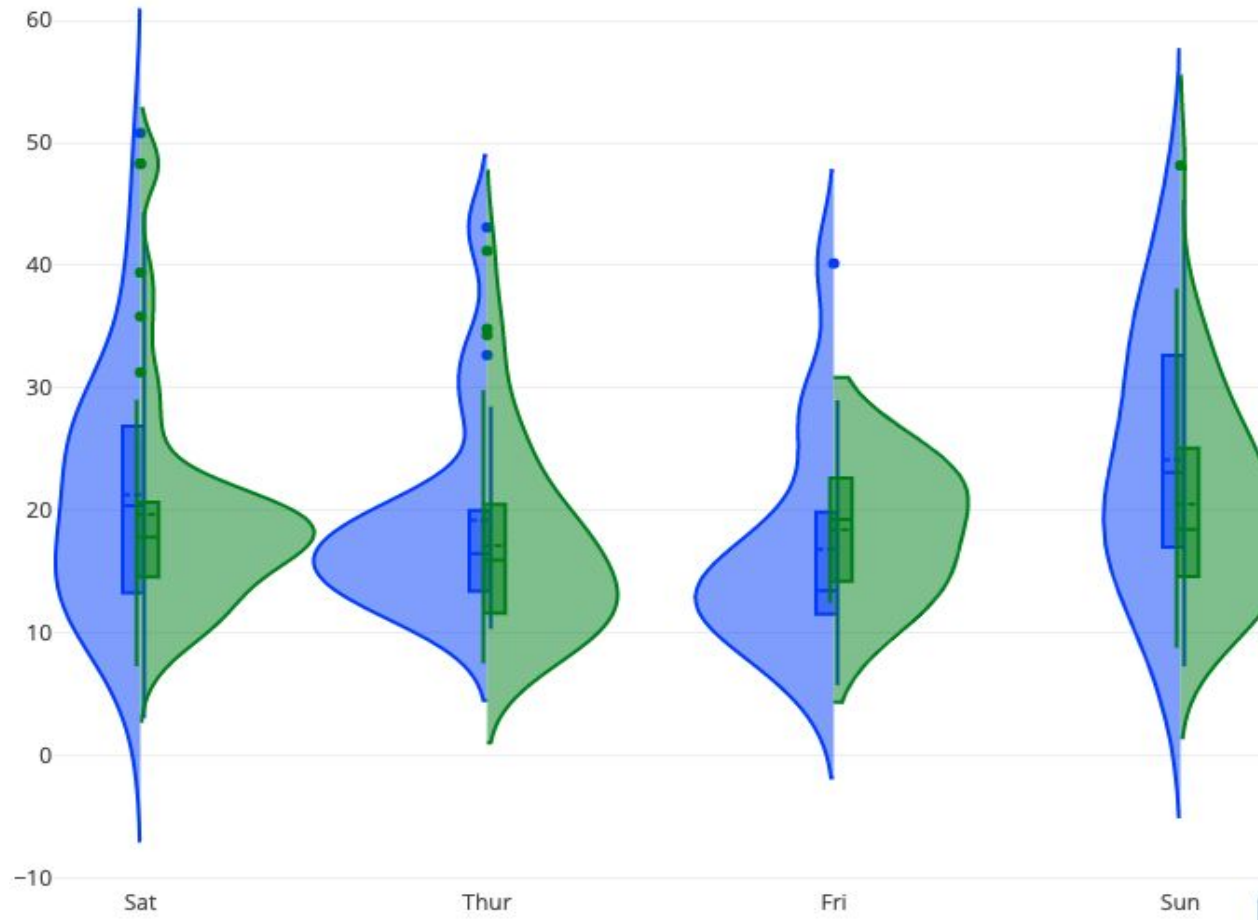
Violin plots: measures



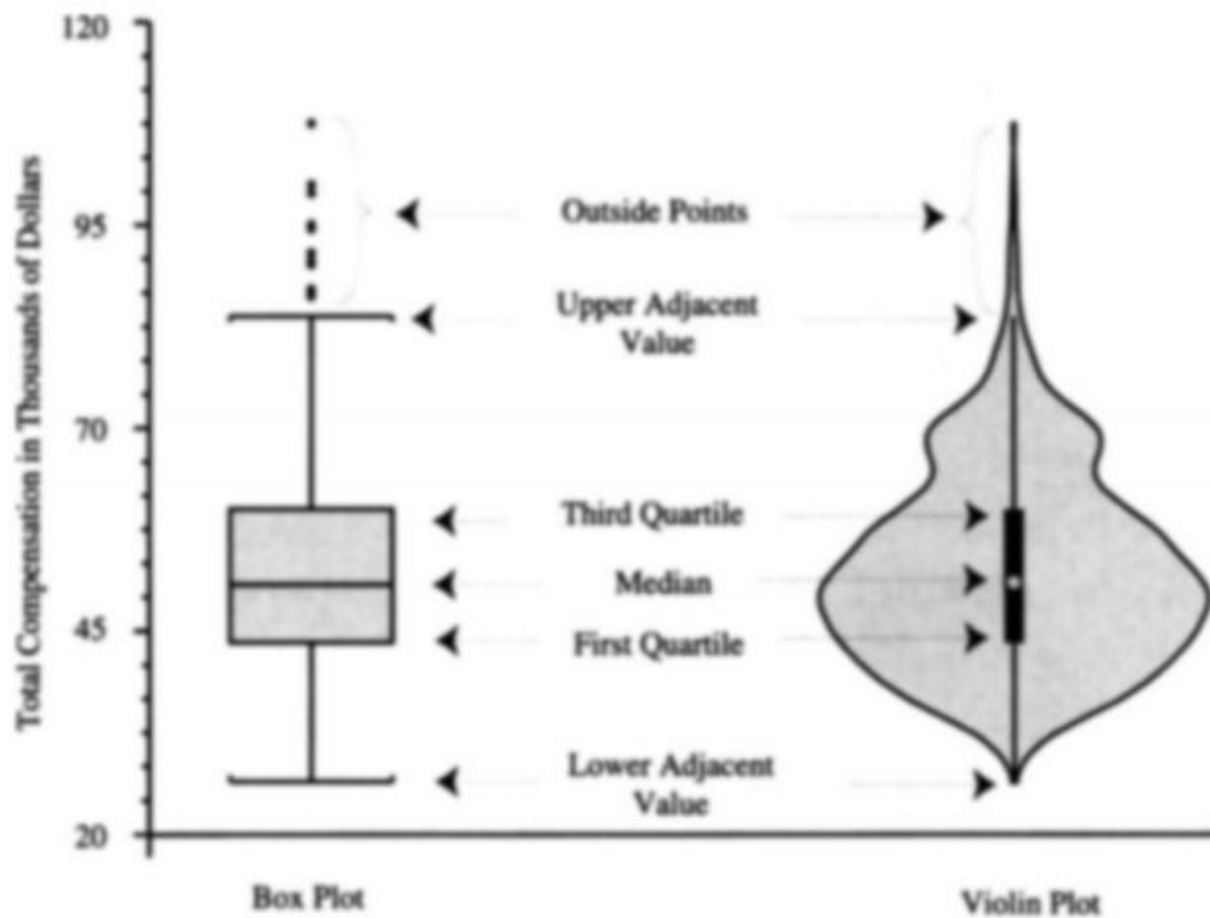
Violin plots: example shapes

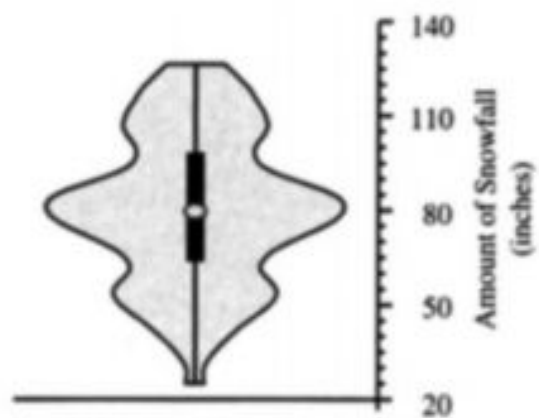
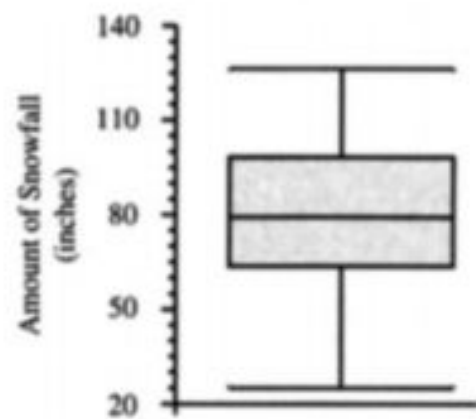


Violin plots for comparison

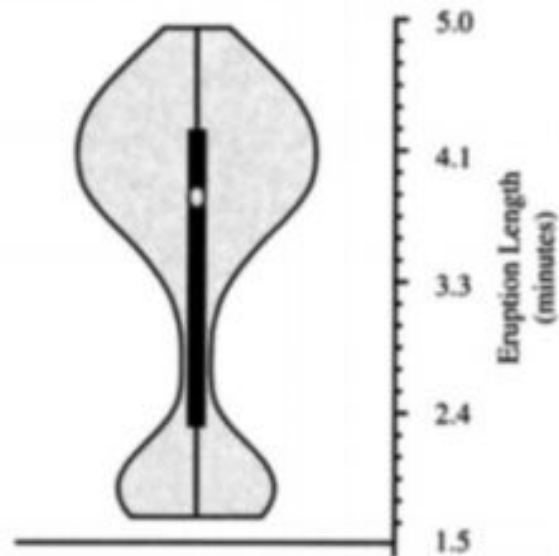
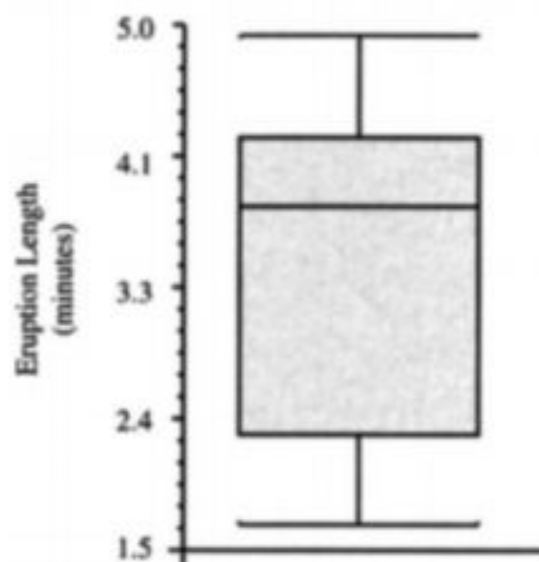


Box plots





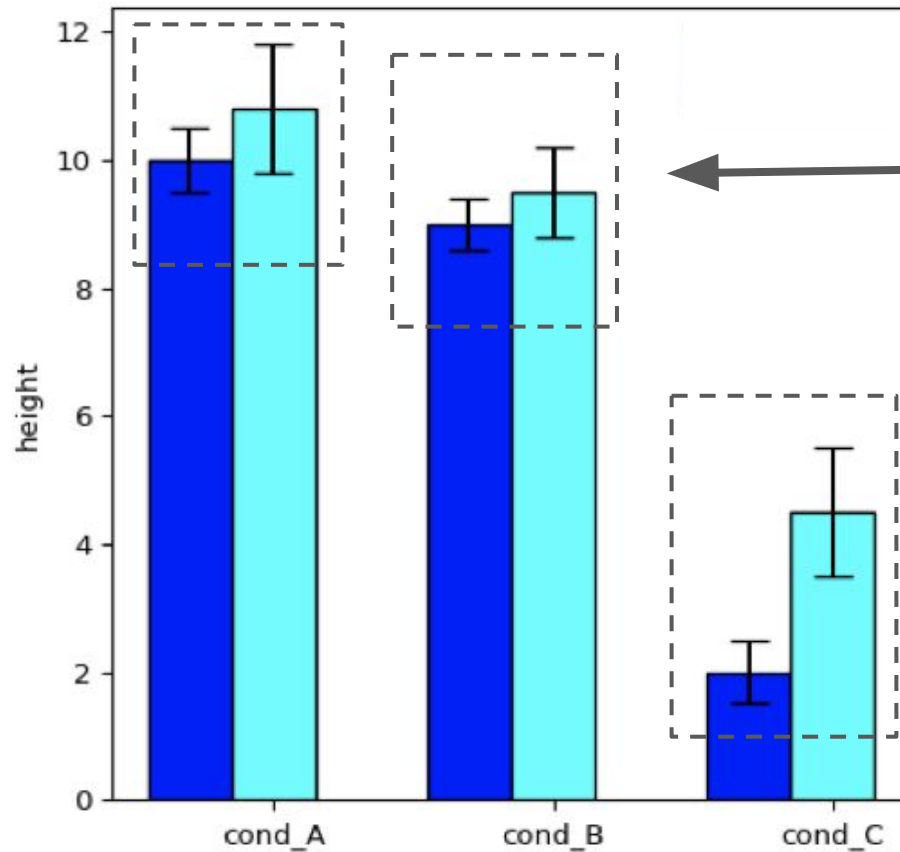
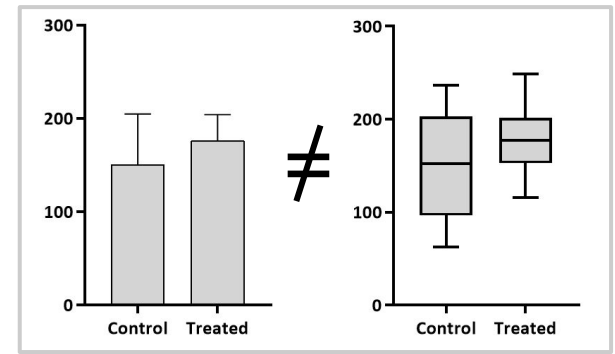
(a)



(b)

Bar plots + Error bars

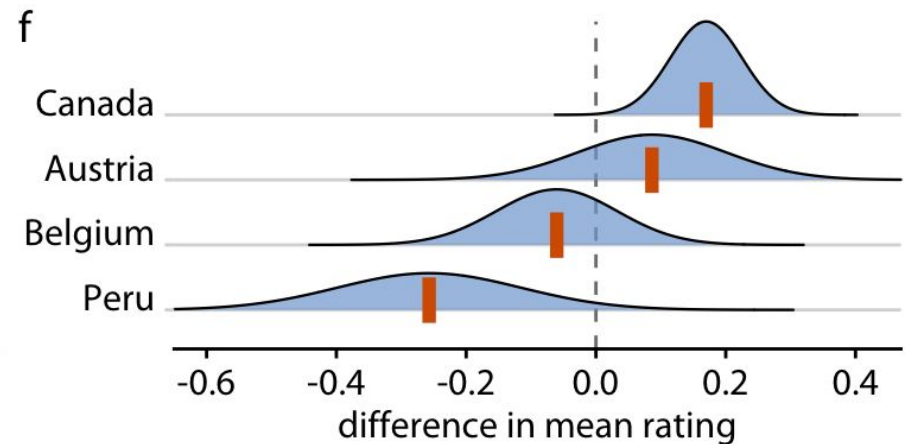
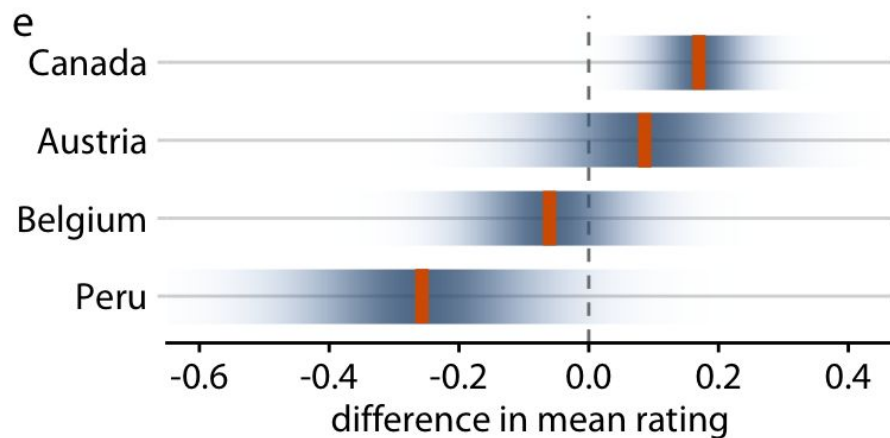
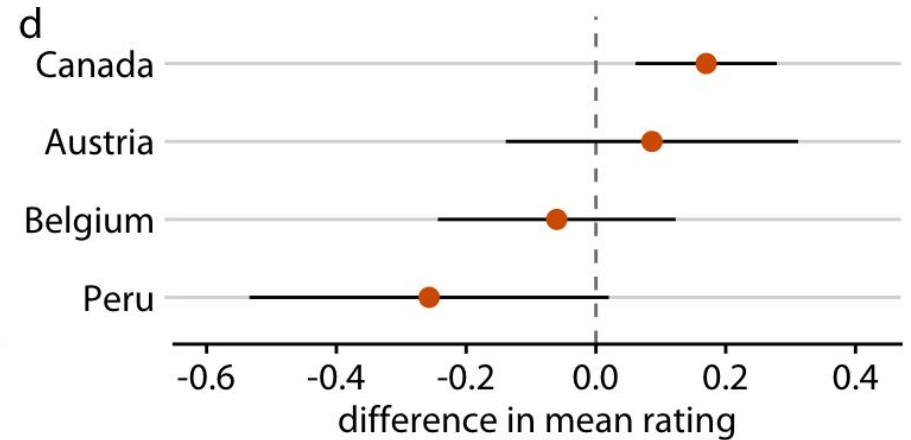
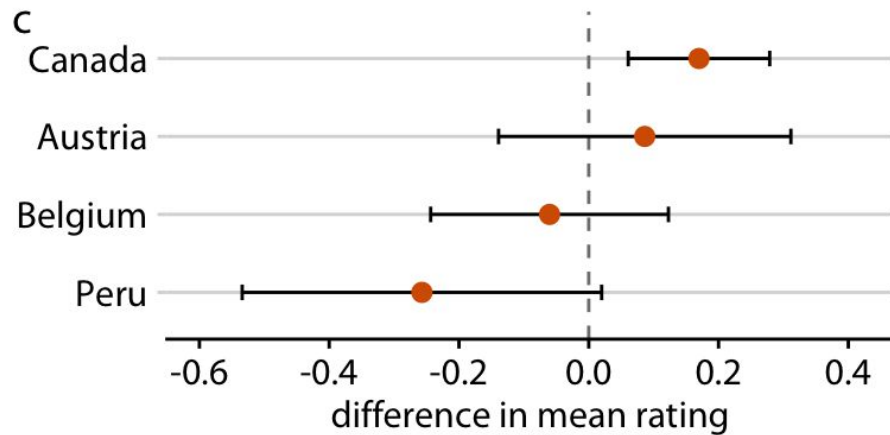
?



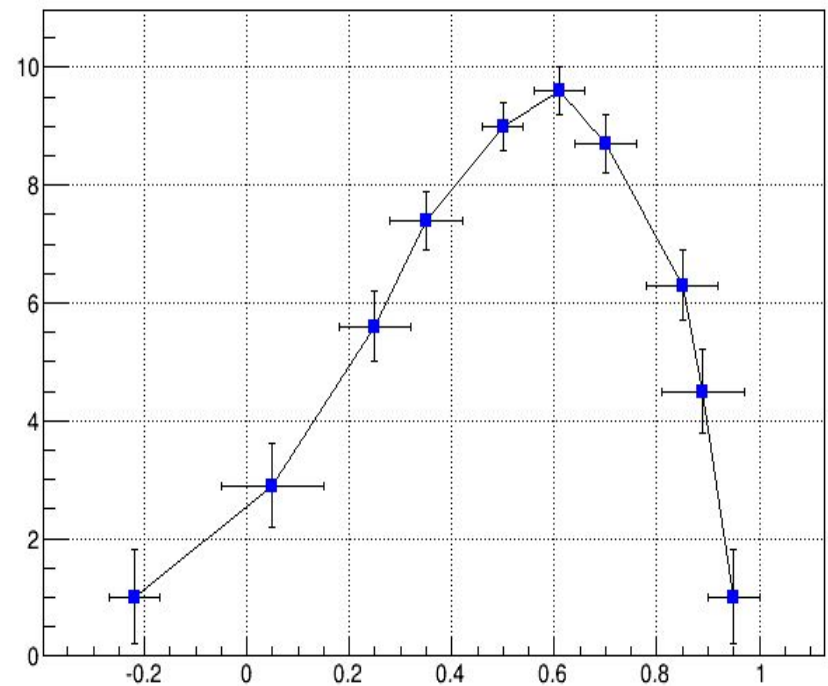
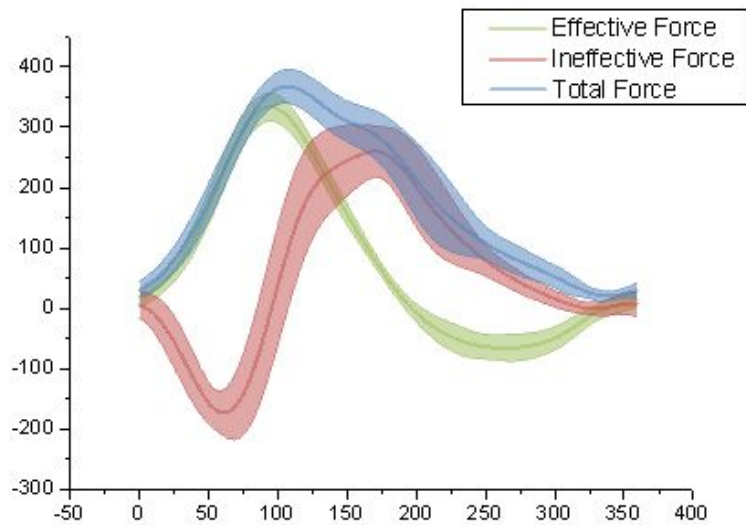
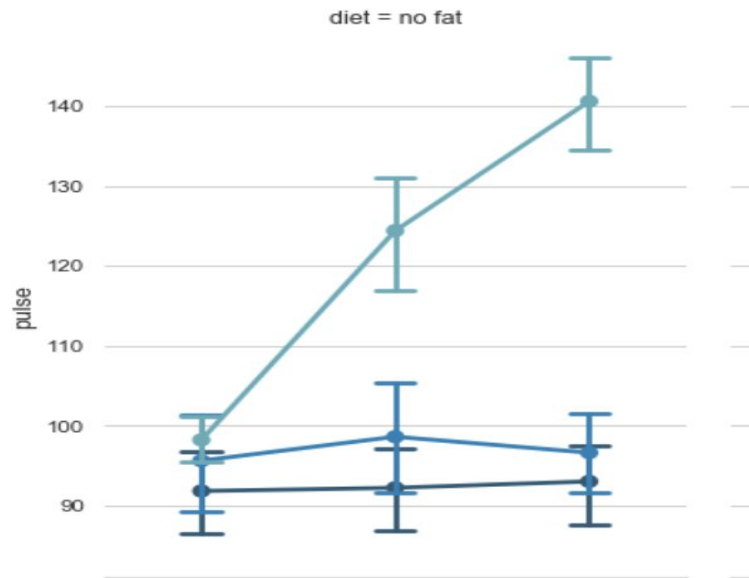
No significant difference

Significance difference

Uncertainty



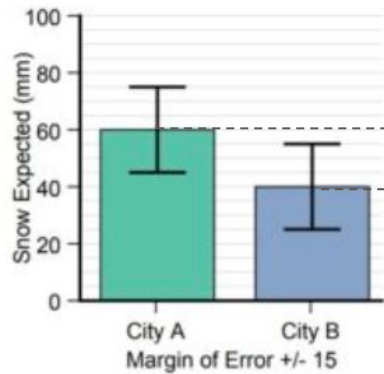
Uncertainty



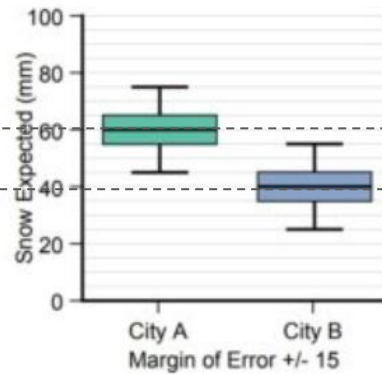
Comparing plots

!

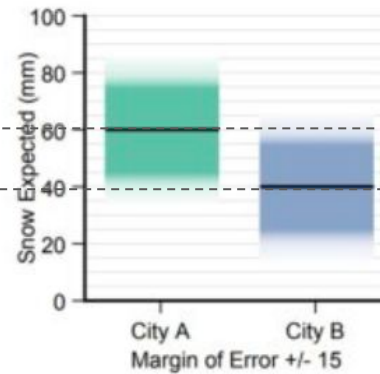
!



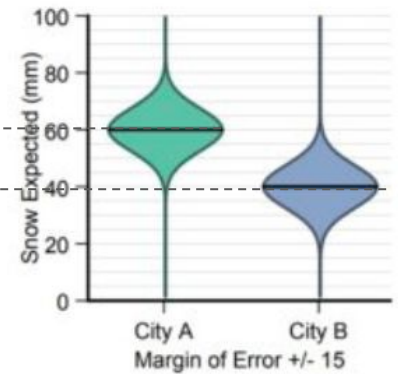
Bar plot



Box plot



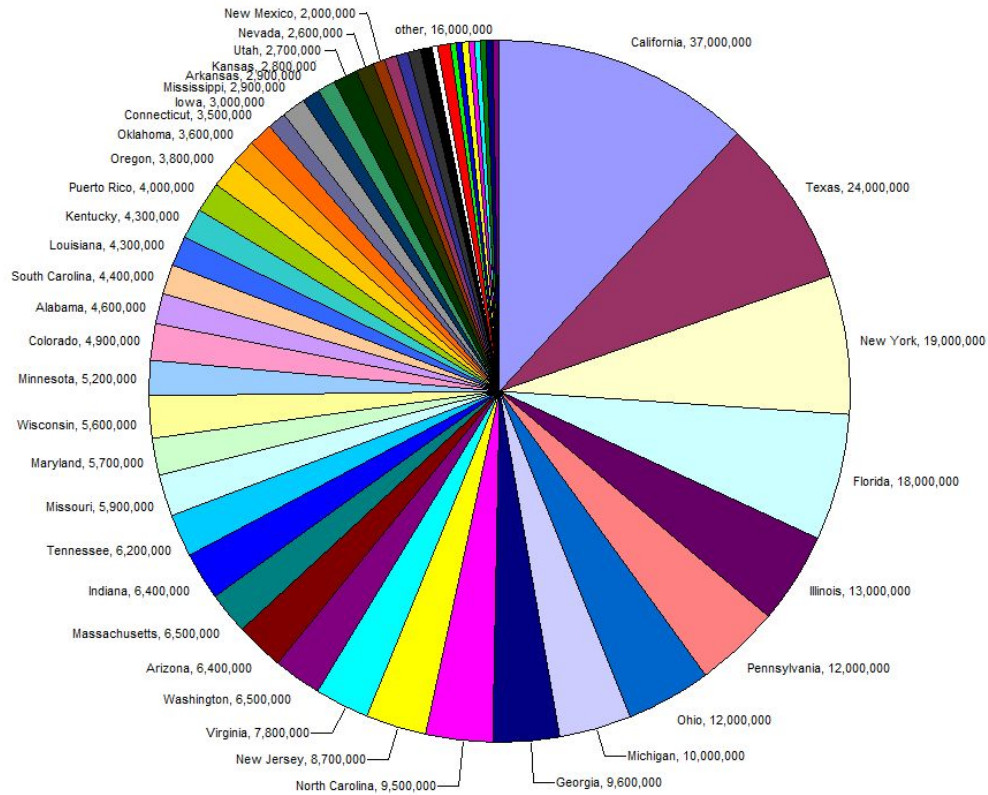
Gradient plot



Violin plot

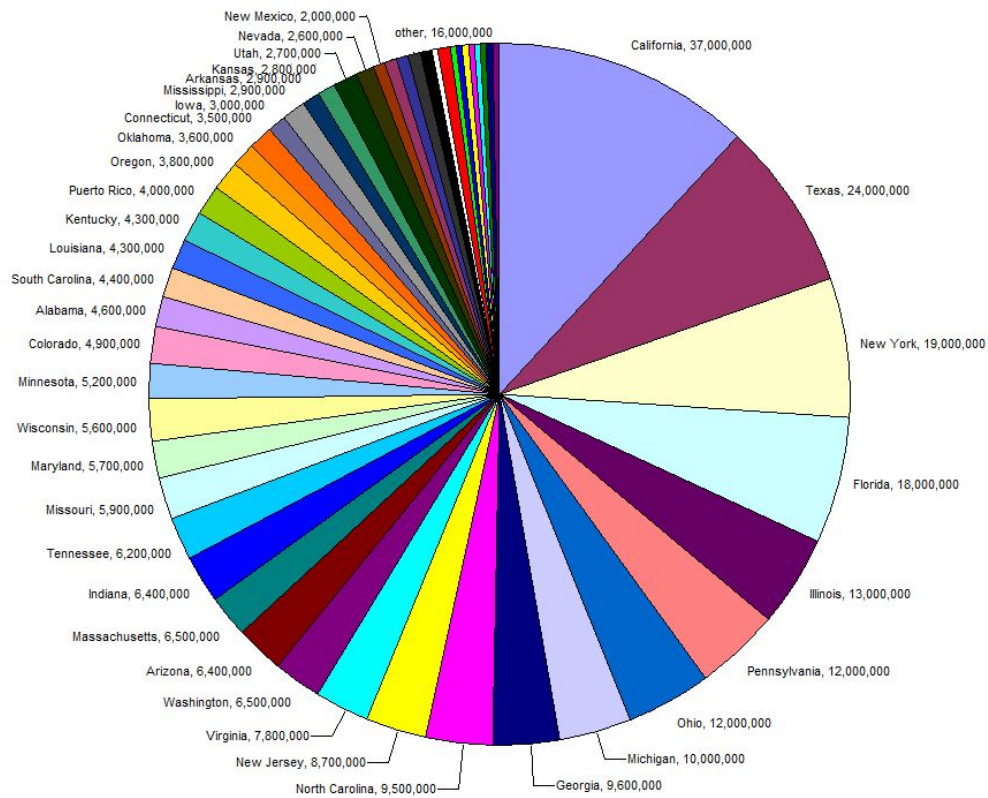
A word on pie charts...

Bad Pie Charts

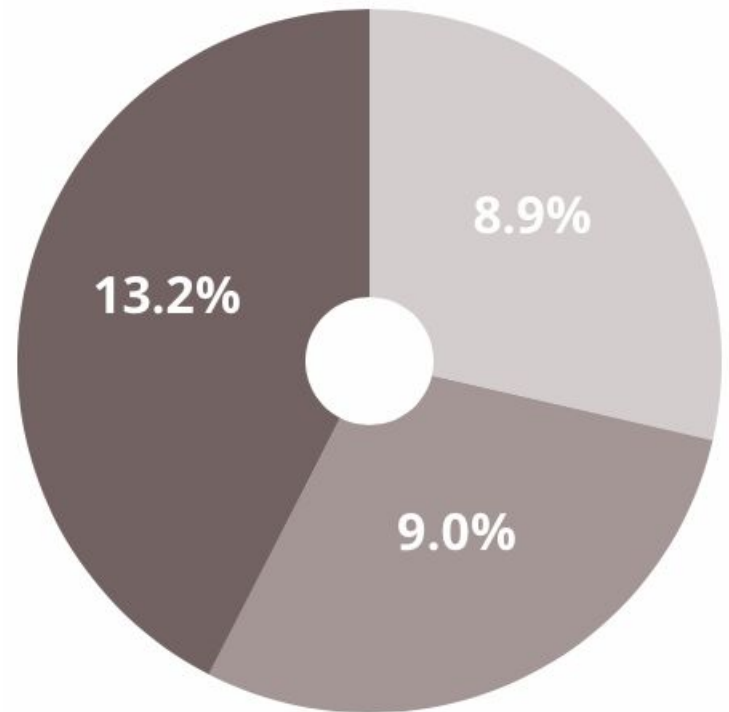


- Too many values
- Differences hard to understand
- Distribution?
- (Confusing colors)

Bad Pie Charts



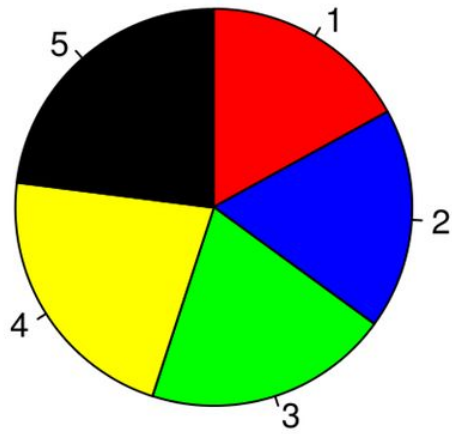
- Too many values
- Differences hard to understand
- Distribution?
- (Confusing colors)



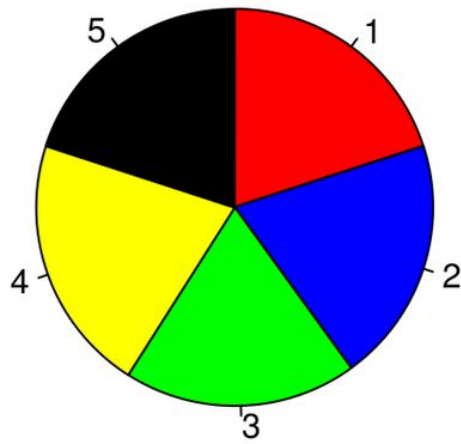
- Not adding up to 100%

Pie Charts

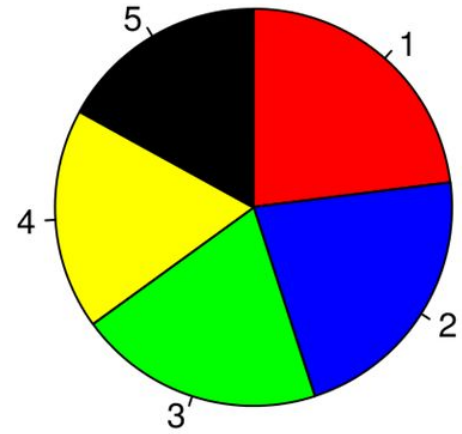
A



B

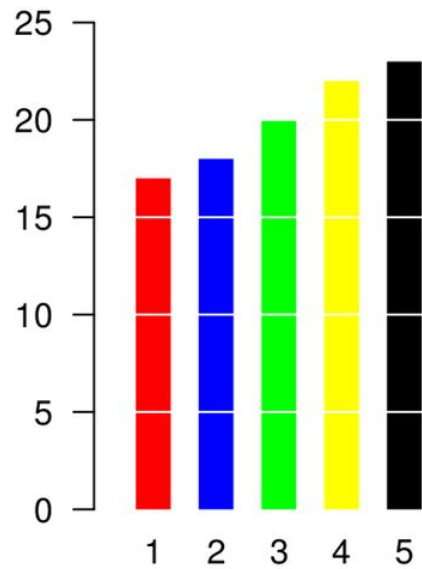
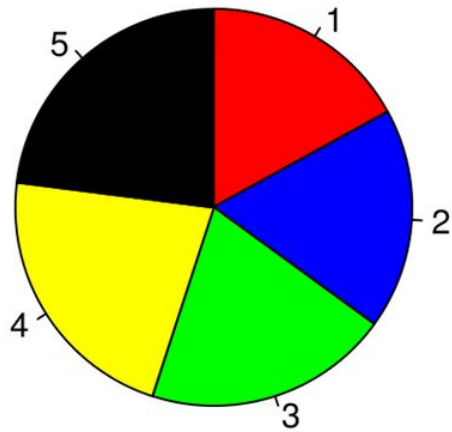


C

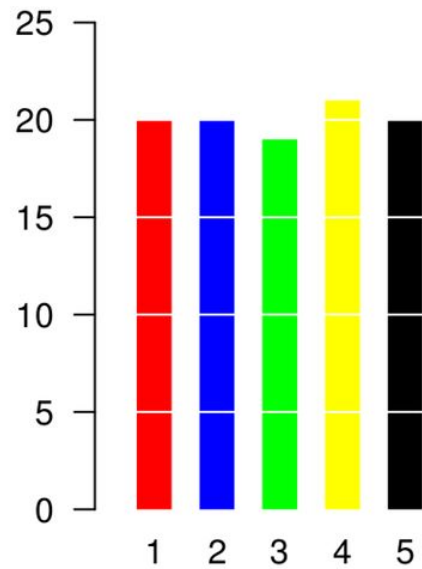
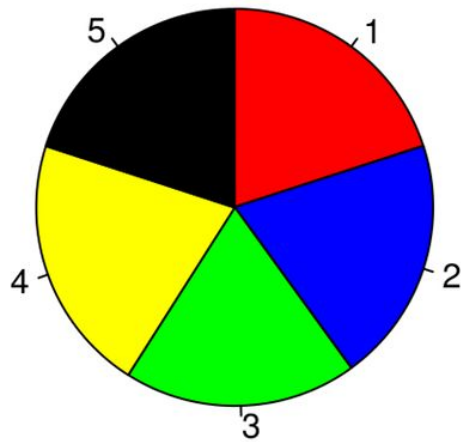


Pie Charts

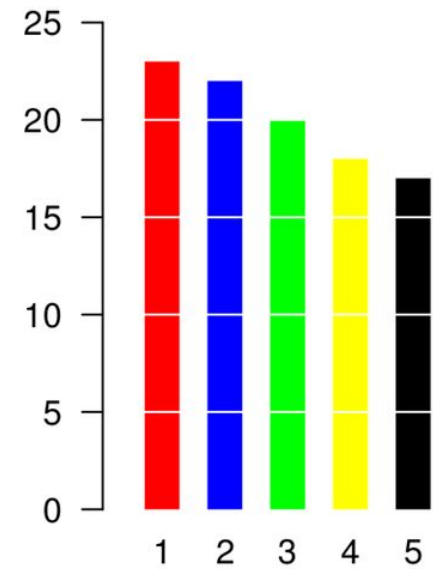
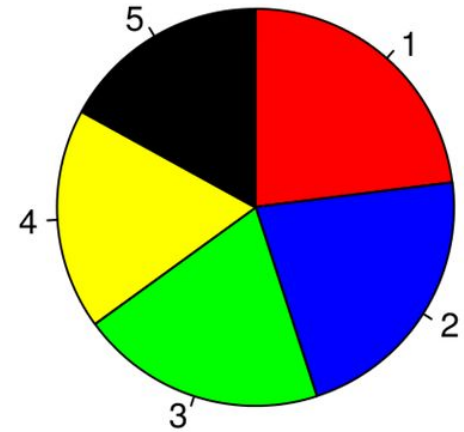
A



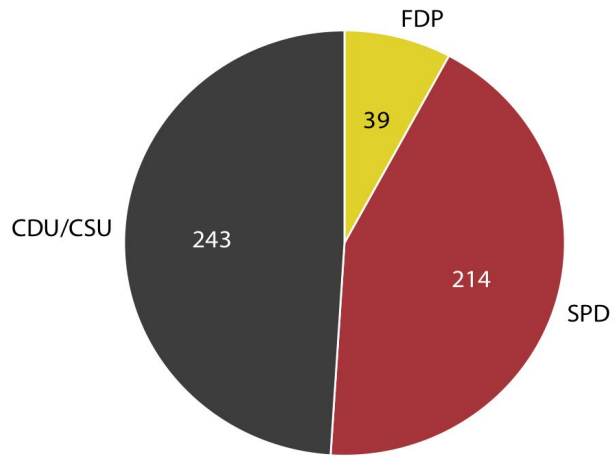
B



C

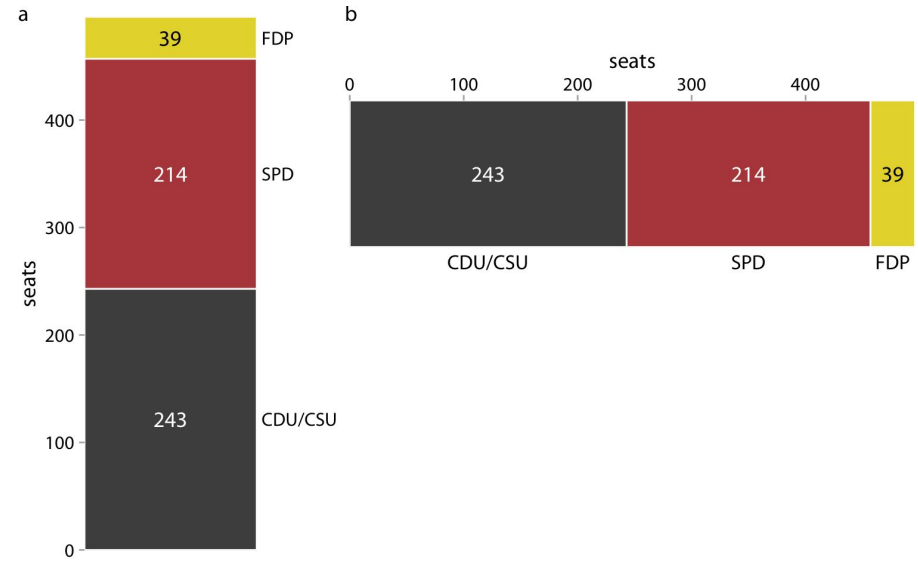


Alternatives?

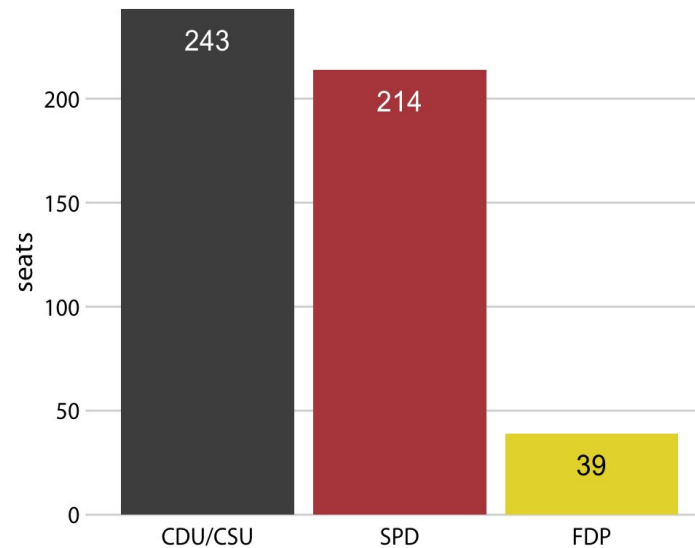


Pie chart

Stacked bars



Bar chart



Pie Chart Alternatives

Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗

Pie Chart Alternatives

Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗
Allows easy visual comparison of the relative proportions	✗	✗	✓

Pie Chart Alternatives

Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗
Allows easy visual comparison of the relative proportions	✗	✗	✓
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✓	✗	✗

Pie Chart Alternatives

Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗
Allows easy visual comparison of the relative proportions	✗	✗	✓
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✓	✗	✗
Looks visually appealing even for very small datasets	✓	✗	✓

Wilke, Claus O. *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media, 2019.

Pie Chart Alternatives

Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗
Allows easy visual comparison of the relative proportions	✗	✗	✓
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✓	✗	✗
Looks visually appealing even for very small datasets	✓	✗	✓
Works well when the whole is broken into many pieces	✗	✗	✓

Wilke, Claus O. *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media, 2019.

Pie Chart Alternatives

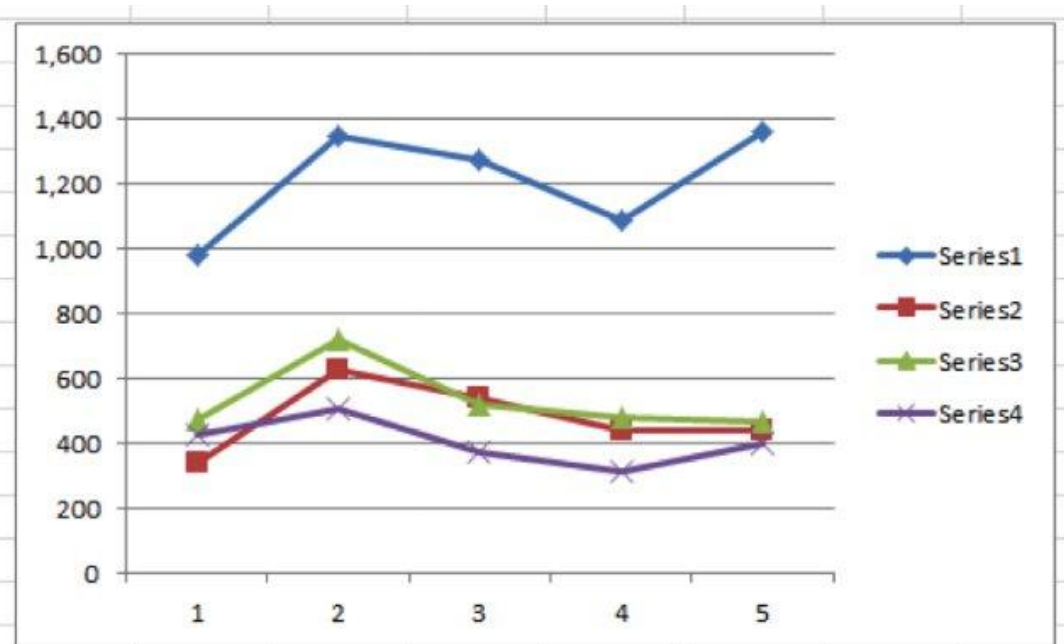
Table 10.1: Pros and cons of common approaches to visualizing proportions: pie charts, stacked bars, and side-by-side bars.

	Pie chart	Stacked bars	Side-by-side bars
Clearly visualizes the data as proportions of a whole	✓	✓	✗
Allows easy visual comparison of the relative proportions	✗	✗	✓
Visually emphasizes simple fractions, such as 1/2, 1/3, 1/4	✓	✗	✗
Looks visually appealing even for very small datasets	✓	✗	✓
Works well when the whole is broken into many pieces	✗	✗	✓
Works well for the visualization of many sets of proportions or time series of proportions	✗	✓	✗

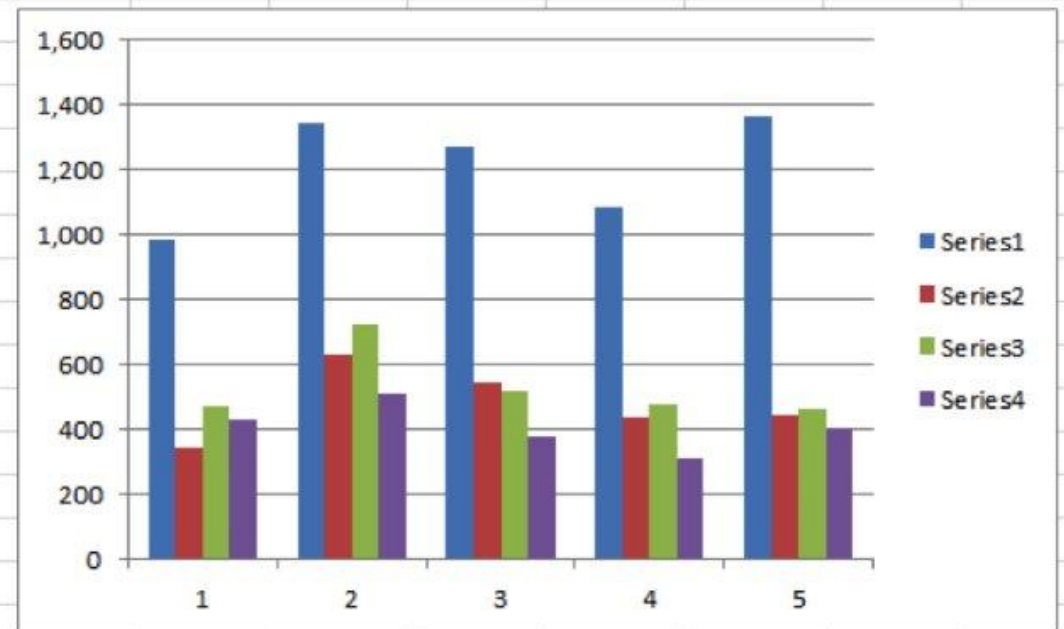
Line Charts vs. Bar Charts

Line Charts vs. Bar Charts?

Continuous values
e.g., time series



Discrete values
e.g., countries



Wrap up

- Choose appropriate bin-sizes
- violin plots for compare distributions
- Label error bars
- Don't confound box plots and error bars
- Avoid pie charts
- Line charts for continuous data

Further Readings

- Wilke, Claus O. **Fundamentals of data visualization: a primer on making informative and compelling figures**. O'Reilly Media, 2019.
 - <https://serialmentor.com/dataviz/visualizing-proportions.htm>
 - <https://serialmentor.com/dataviz/histograms-density-plots.html>
 - <https://serialmentor.com/dataviz/boxplots-violins.html>
- Hullman, Jessica, Paul Resnick, and Eytan Adar. "**Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering**." *PloS one* 10.11 (2015).
- Skau, Drew, and Robert Kosara. "**Arcs, angles, or areas: Individual data encodings in pie and donut charts**." *Computer Graphics Forum*. Vol. 35. No. 3. 2016.
- Cairo, Alberto. **The truthful art: data, charts, and maps for communication**. New Riders, 2016.
 - Chapter 6: Exploring Data with Simple Charts
 - Chapter 7: Visualizing Distributions