

Data description and visualization

Moses Chan

April 4, 2024
DATA_ENG 200

Adapted from Health and Healthcare Data Visualization Module from Tableau

1

Quick recap

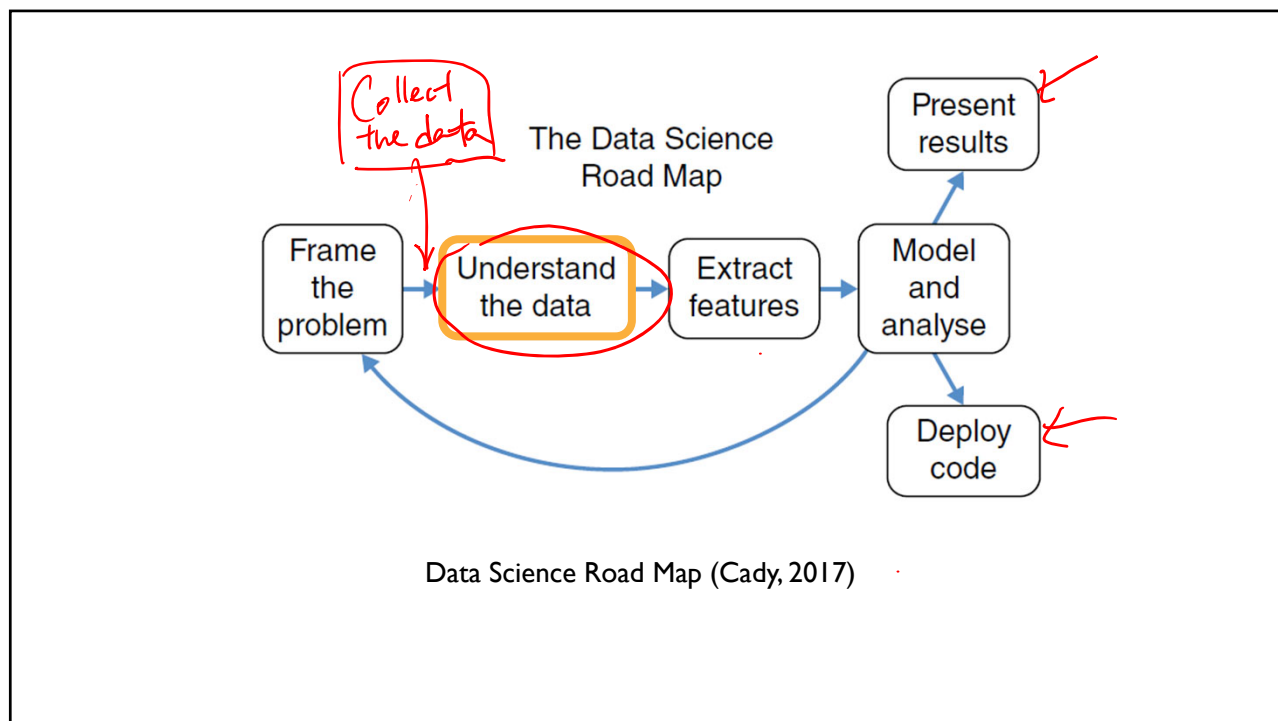
Data types

- Qualitative vs quantitative
- Categorical
 - Nominal, ordinal
- Numerical
 - Discrete, continuous

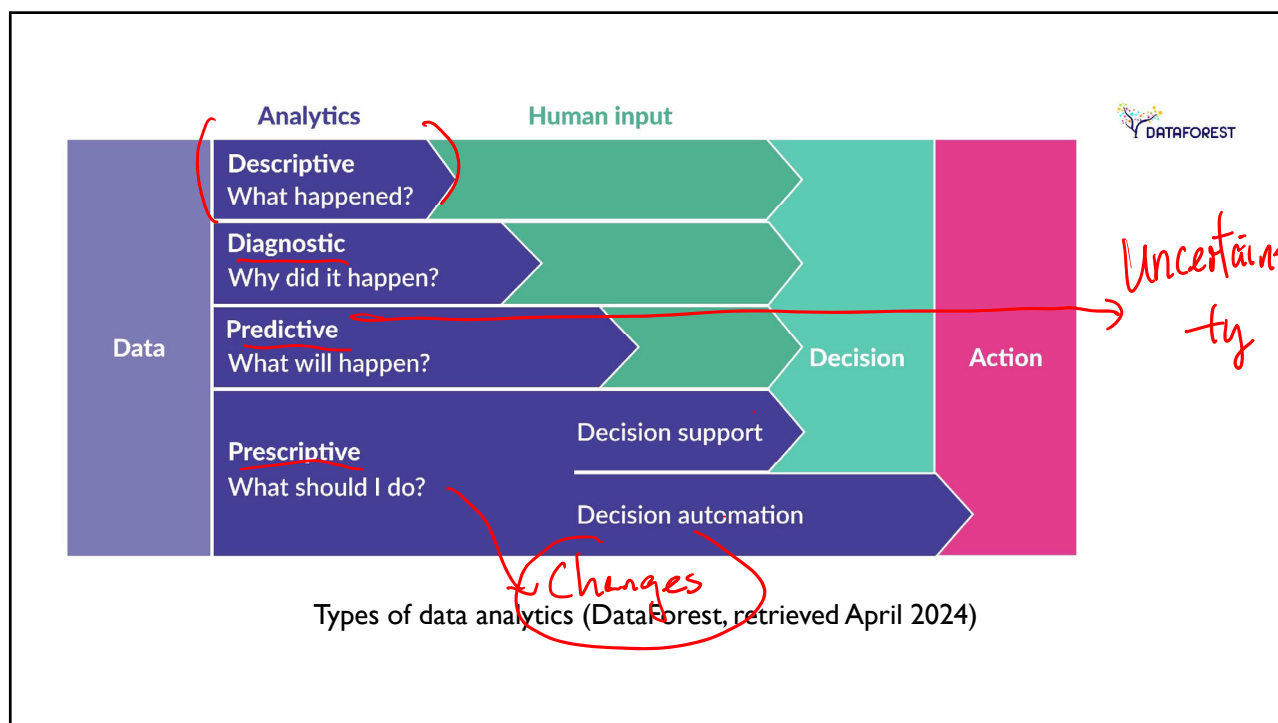
Tool

- Tableau demo

2



3



4

Descriptive analytics for single variable

7

Numerical
summary

or

Summary
statistics

Central tendency

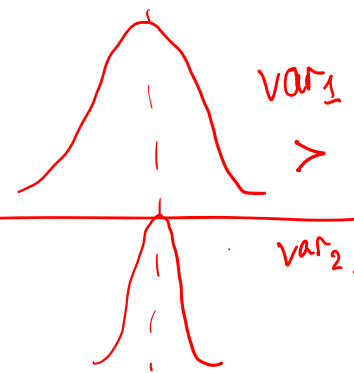
- Mean
 - Arithmetic average
- Median
 - The middle of the data points ($\frac{1}{2}$ above, $\frac{1}{2}$ below)
- Mode
 - The most frequent value

Variability

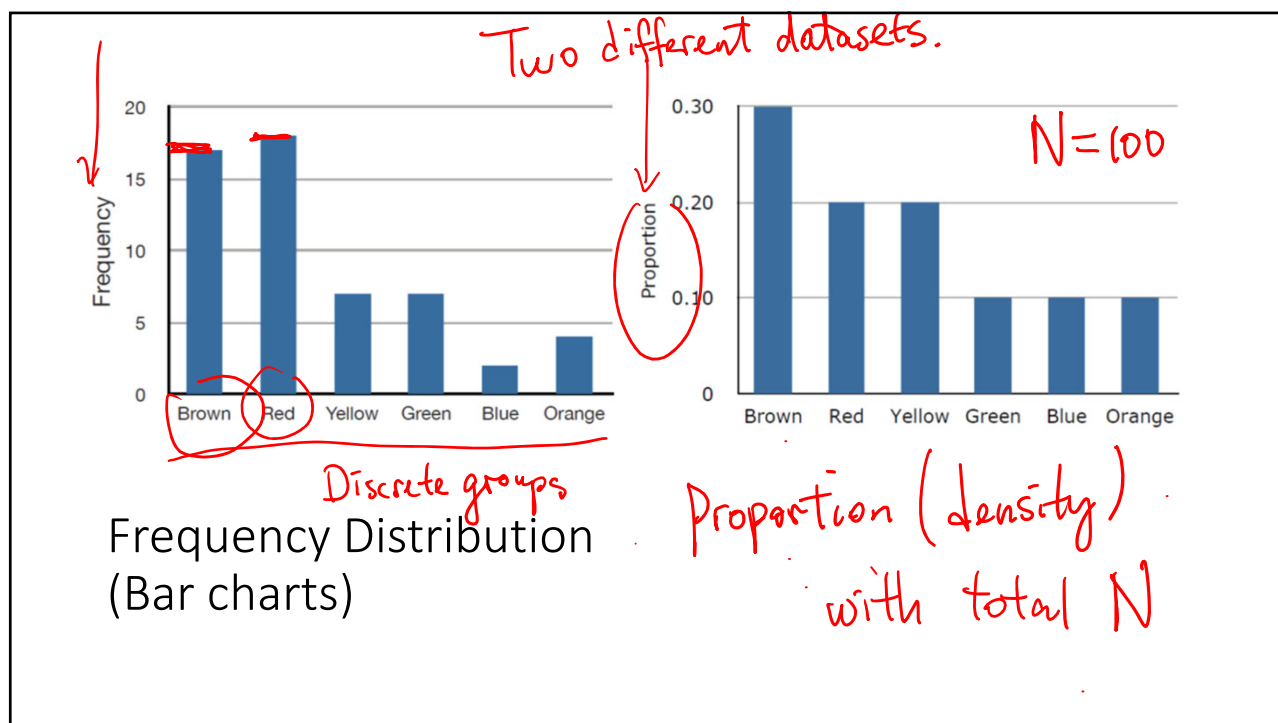
- Standard deviation
- Variance
- Range

Single numbers

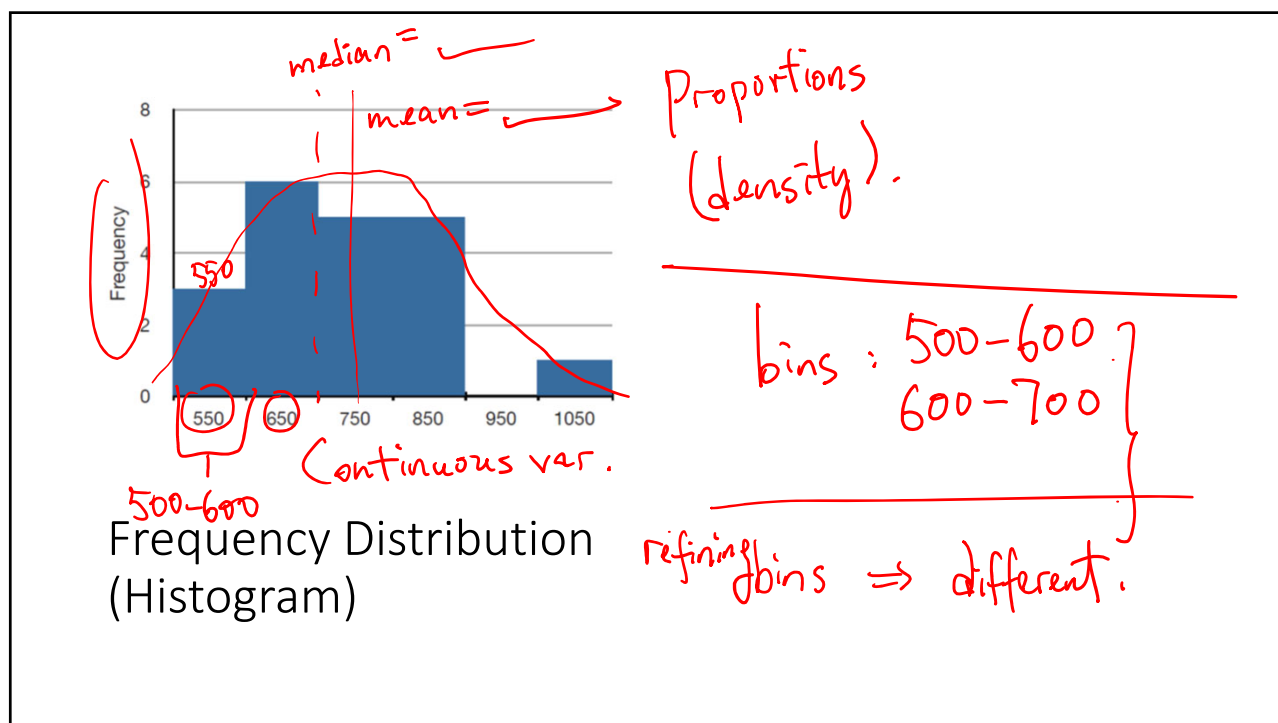
for single
variable



8



9



10

Density curve

- Useful to describe common distributions
- **"Our choice"** of distributions
 - e.g. Normal distribution is symmetric and bell-shaped.
- Informed by data?
 - *kernel density estimation*

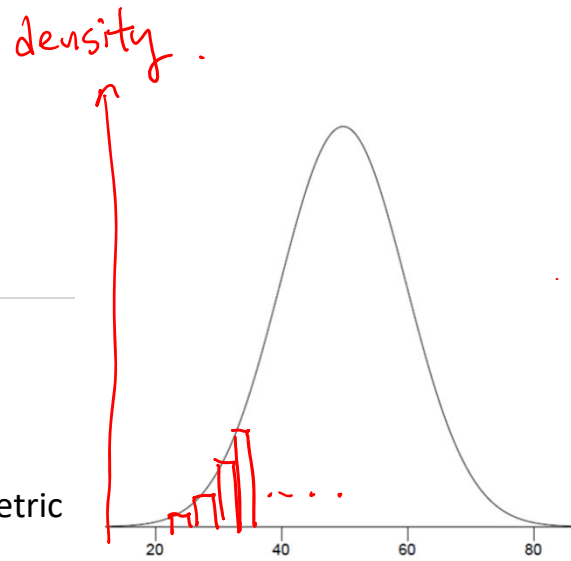
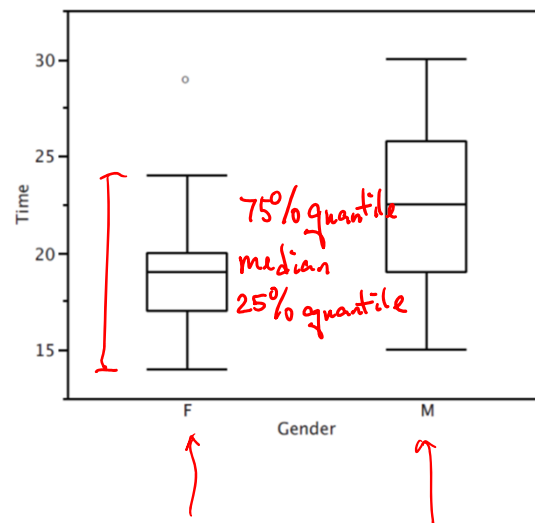


Figure 4. A normal distribution.

11

Boxplot

- Useful for identifying Outliers and comparing distributions
- Other name: Box and whiskers plot, 5-point summary



12

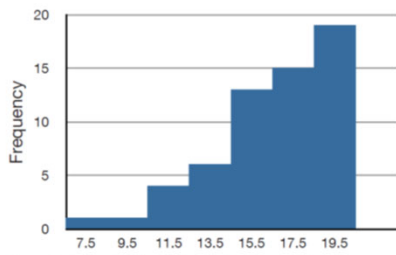


Figure 8. A distribution with negative skew. This histogram shows the frequencies of various scores on a 20-point question on a statistics test.

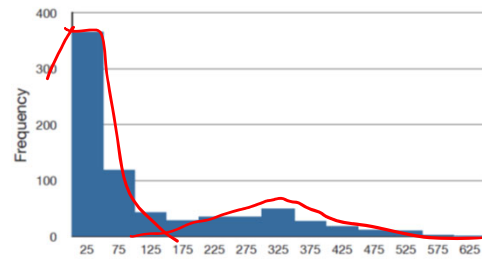
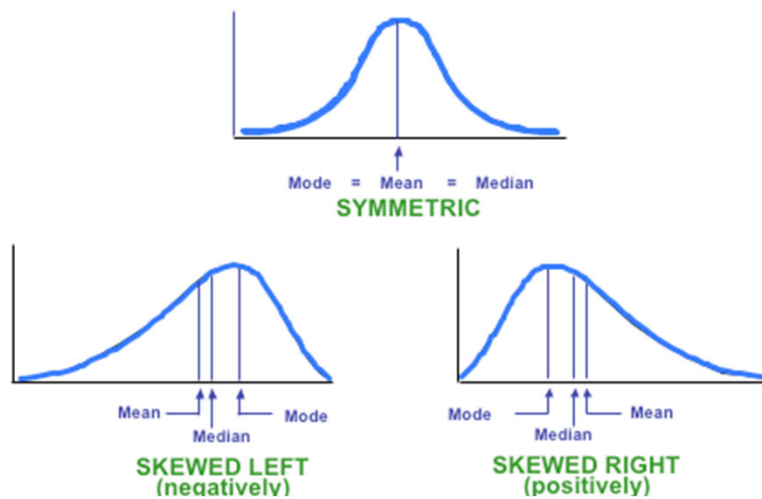


Figure 6. A distribution with a very large positive skew.

Distribution Skew

13

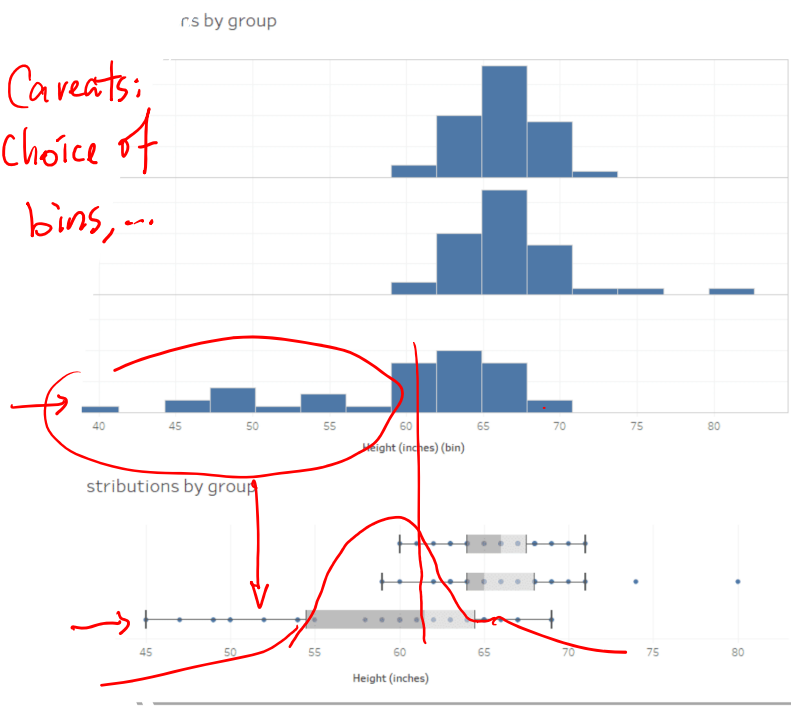
Means and Medians differ when distributions are skewed



14

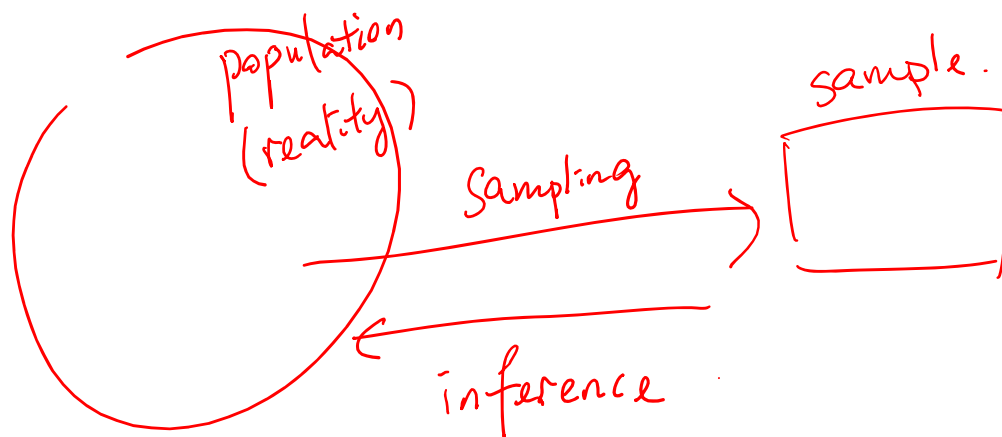
Boxplot vs Histogram

Caveats:
Choice of
bins, ...



15

All data are samples



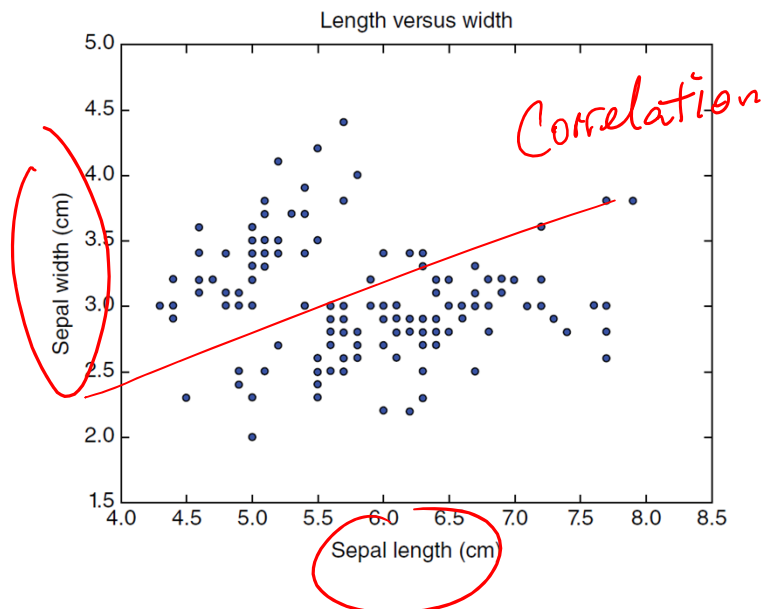
16

Descriptive analytics for multiple variables

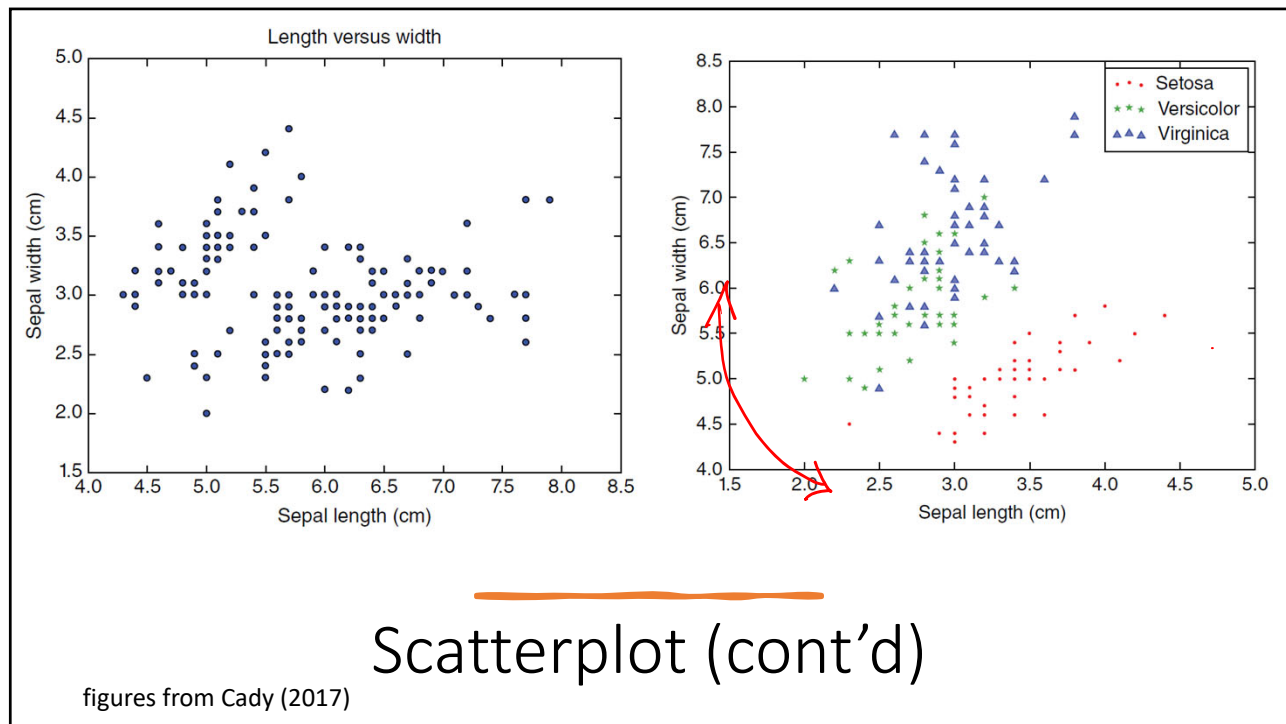
17

Scatterplot

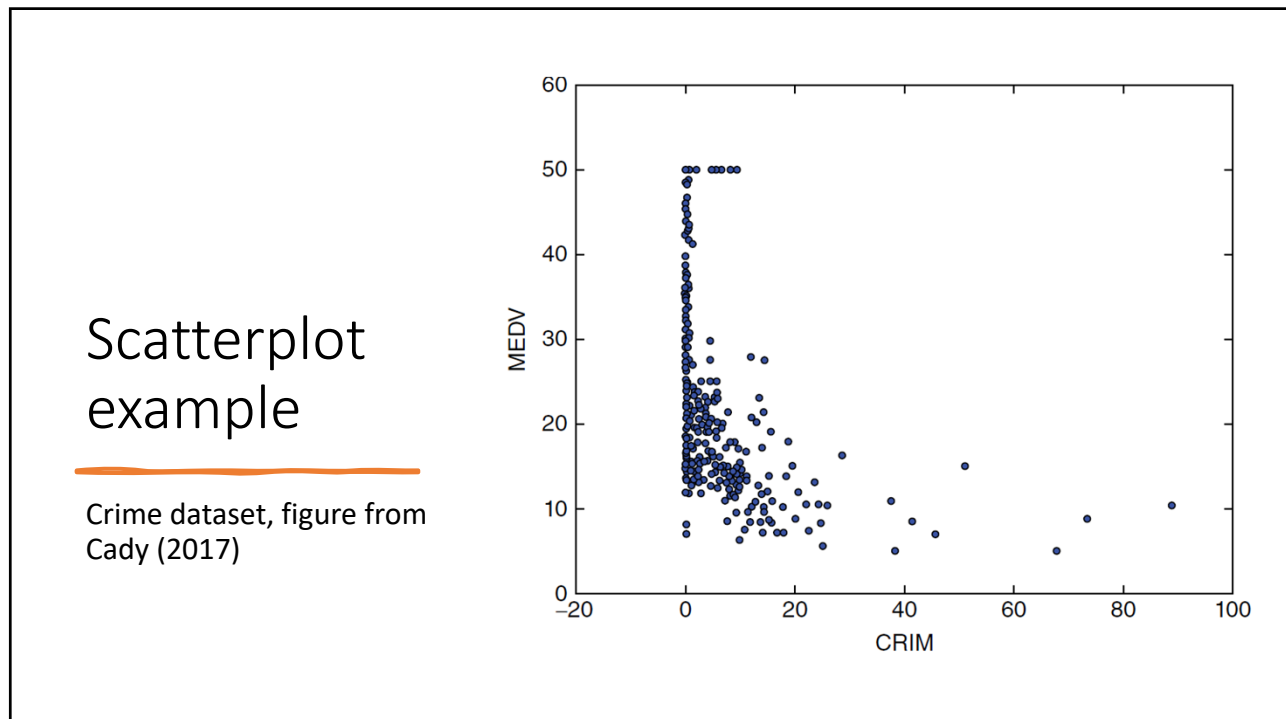
Iris dataset, figure from Cady, 2017.



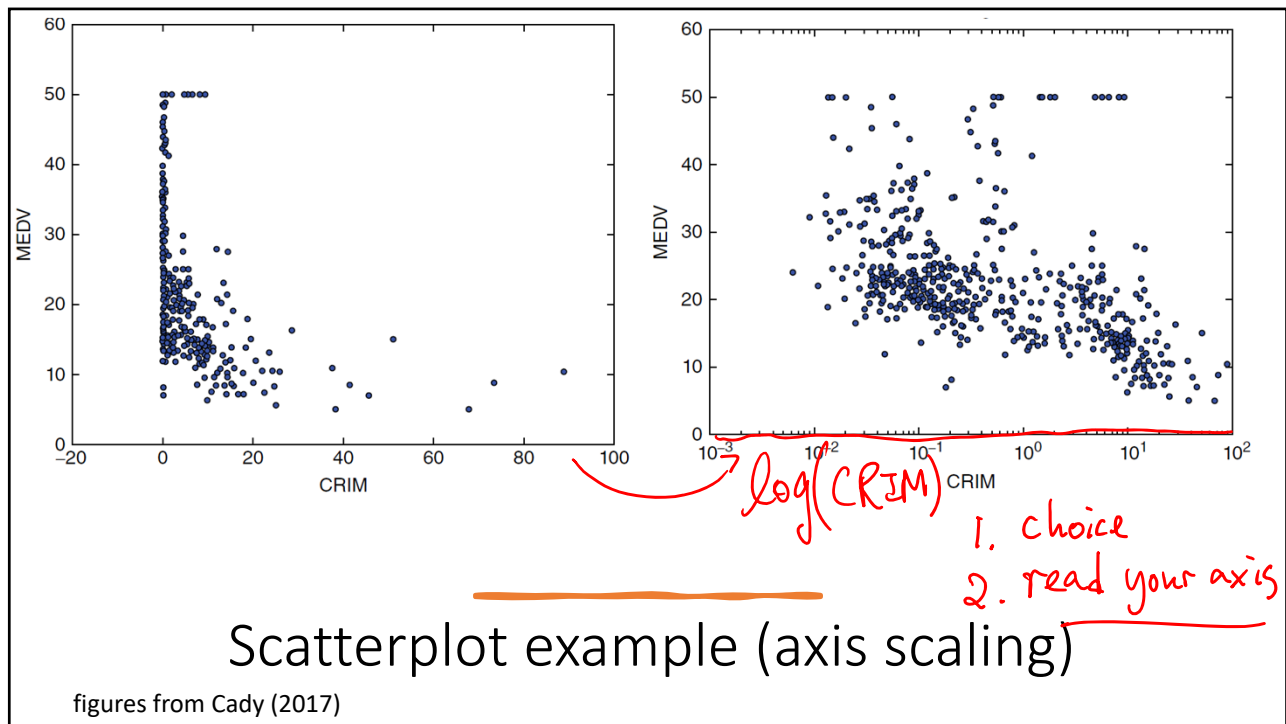
18



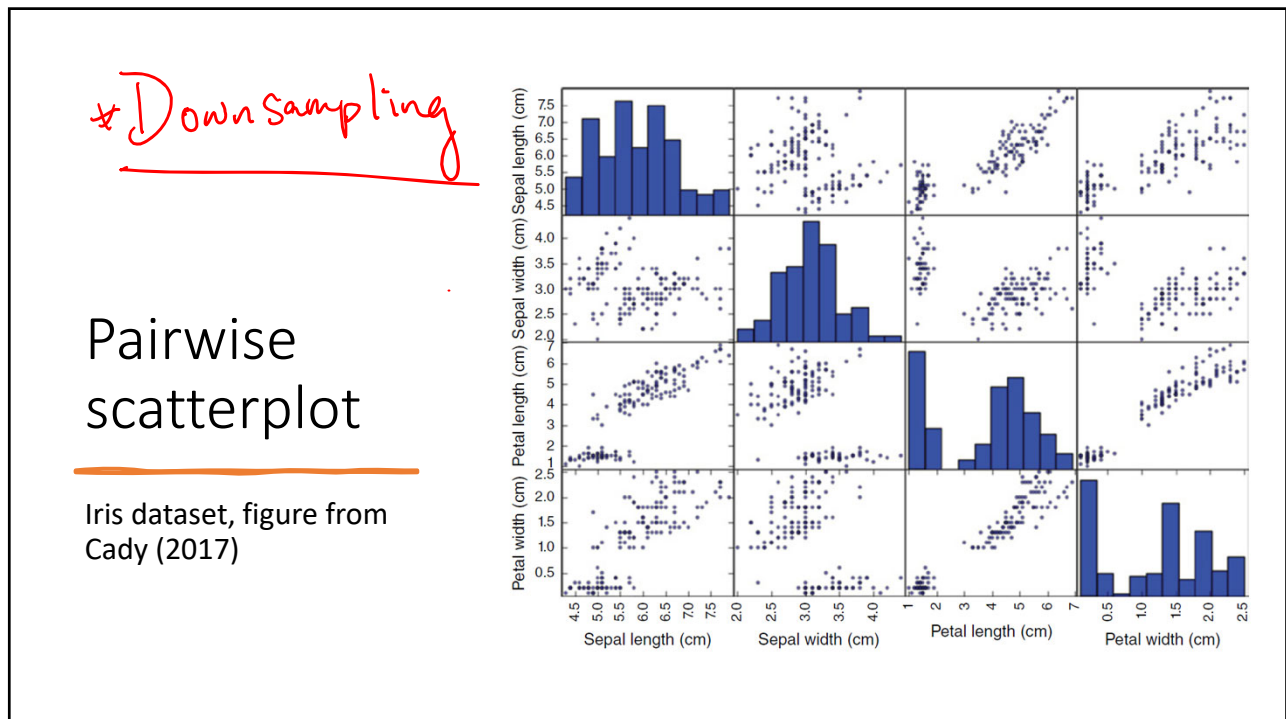
19



20



21

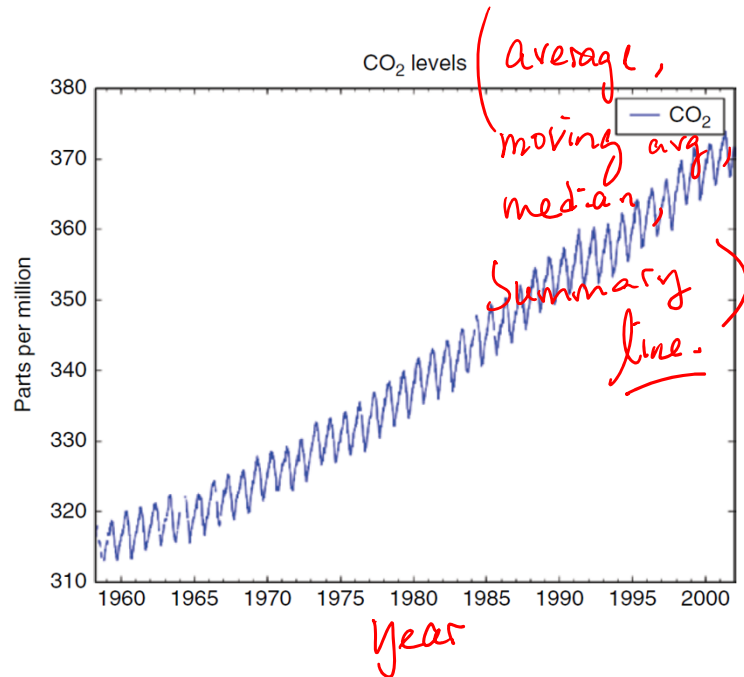


22

Time Series

CO2 level since 1960, figure from Cady (2017)

<https://gml.noaa.gov/ccgg/trends/global.html>



23

Descriptive analytics / visualizations are most useful for *understanding* and *building intuition*

24

Tableau Demo and Practice

Target

- Multiple datasets
- Calculated field
- Geographical *misc.*
- Dashboarding (if time allows)