Crafting with Data

Reality, Illusions, Truth & the Future

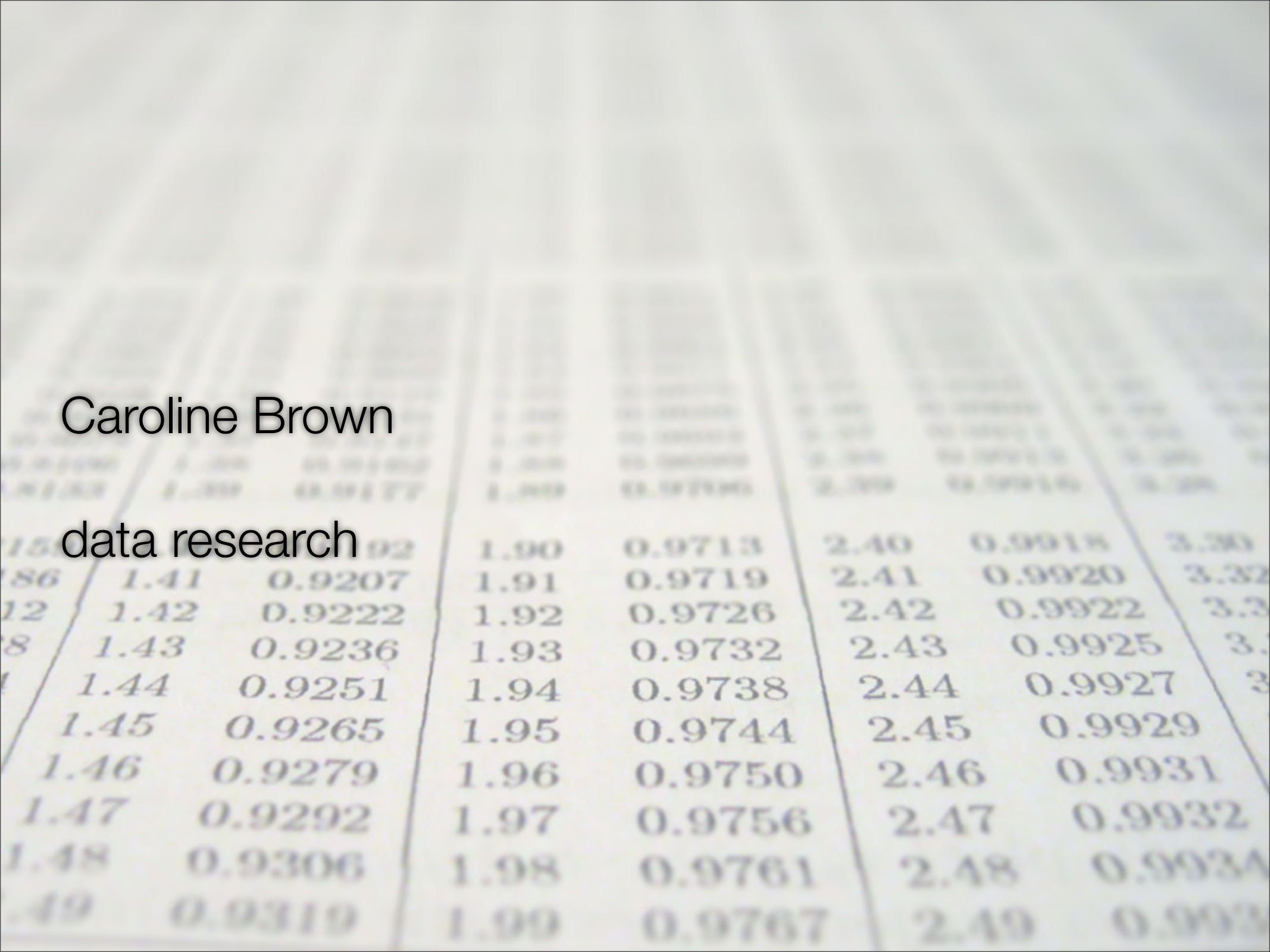
Instructor: Rob Faludi

Plan for Today

- Caroline Brown
- Exercise One: Families
- Exploratory Data Analysis
- Central Tendency & Measures of Spread
- Present New York Data project pilot results
- Exercise Two: Measuring (maybe)
- Readings & Assignments

Caroline Brown

data research



Exercise One: Families

- How many people in your family?
- Average in US is about 2.0
- How do we compare?

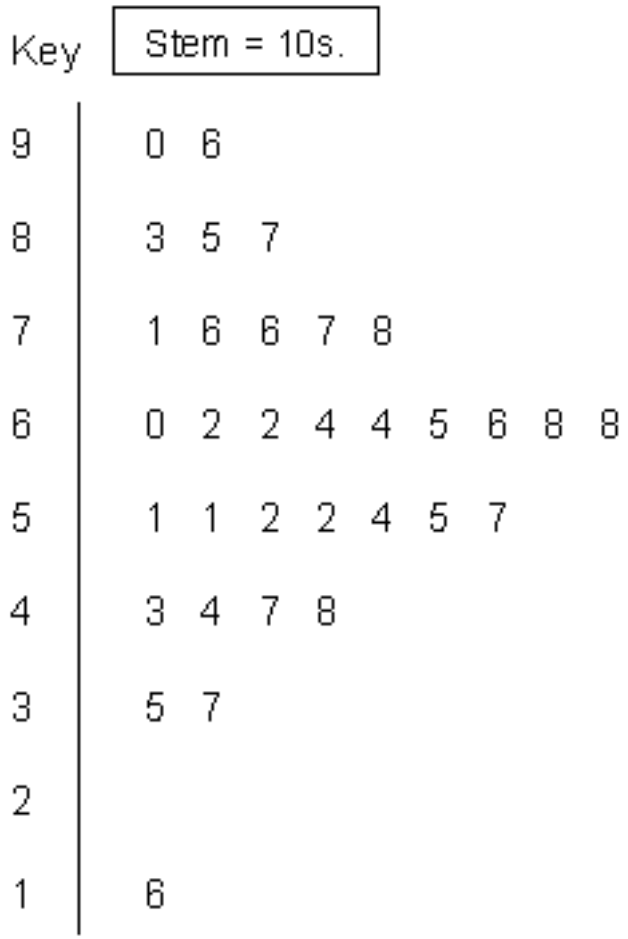
Exploratory Data Analysis

- many wells, many routes
- state of mind
- data as a source of new ideas
- skepticism for summaries
- openness to new patterns and views
- not by statistical summary alone, that’s not open

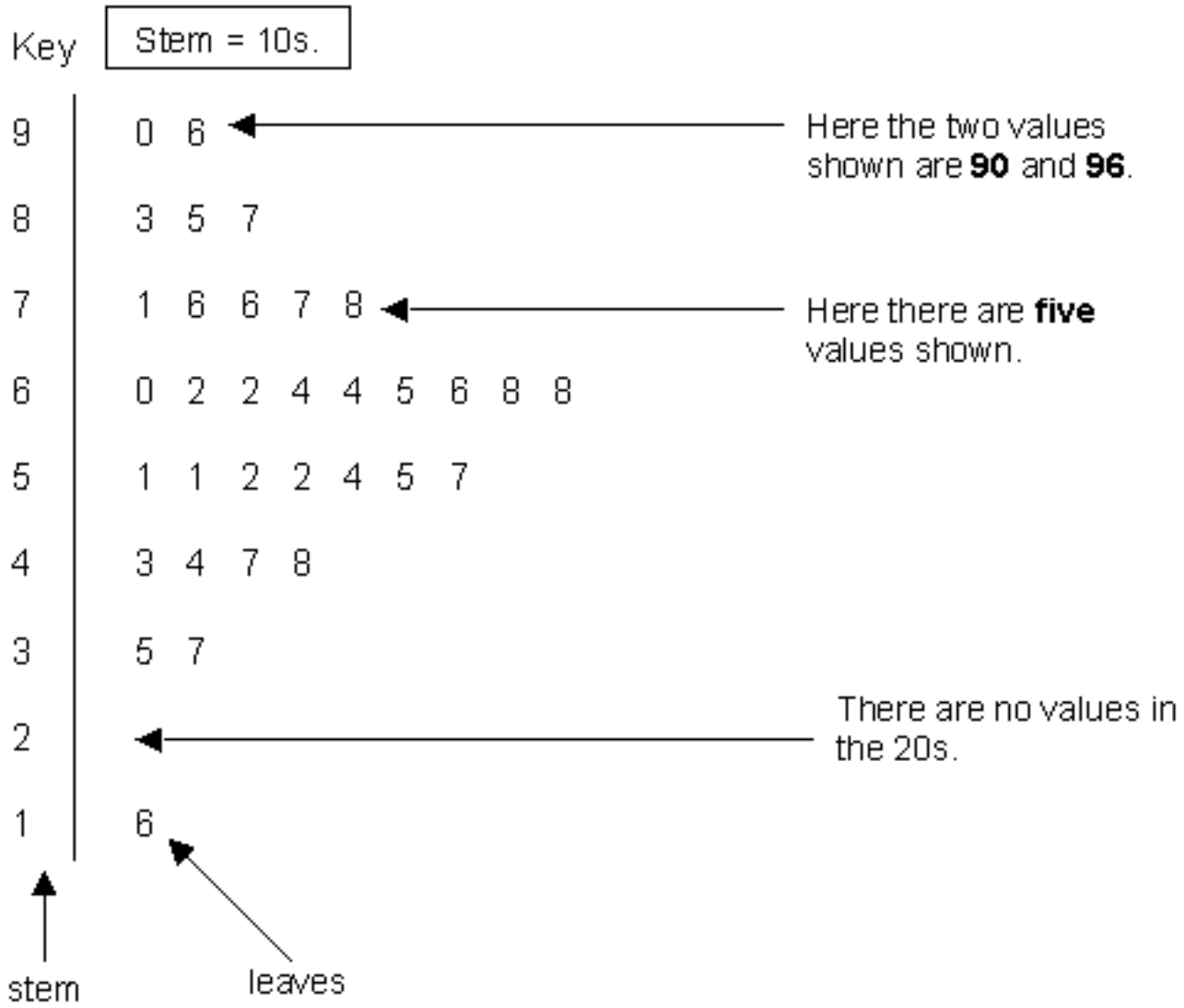
EDA

- Smooth vs. Rough
- inductive methods
- re-expression, scales
- visualization
- resistant or robust statistics
- outliers, include or exclude?

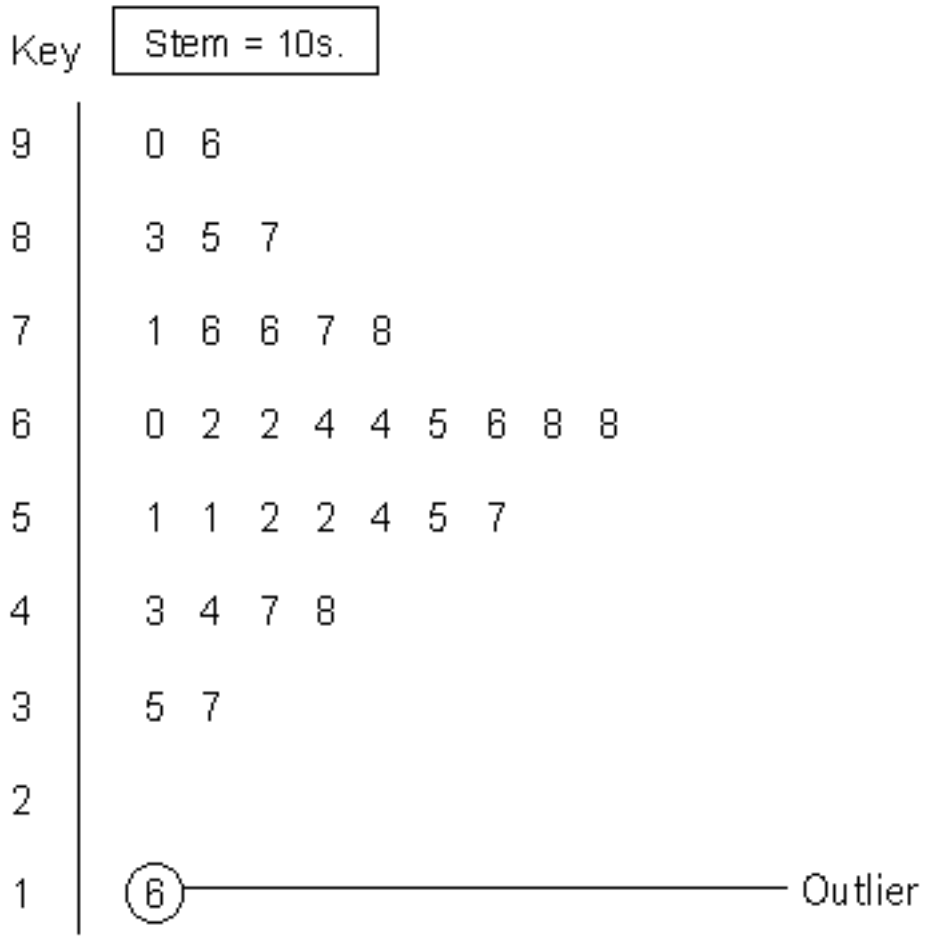
Visualization: Stem & Leaf



Visualization: Stem & Leaf

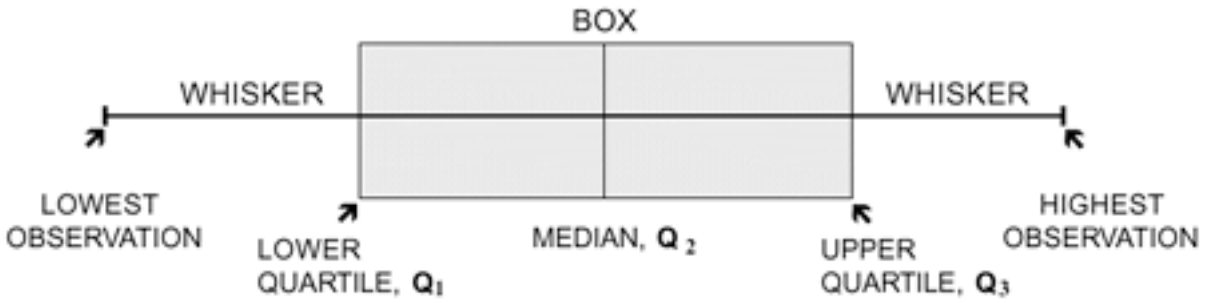


Visualization: Stem & Leaf

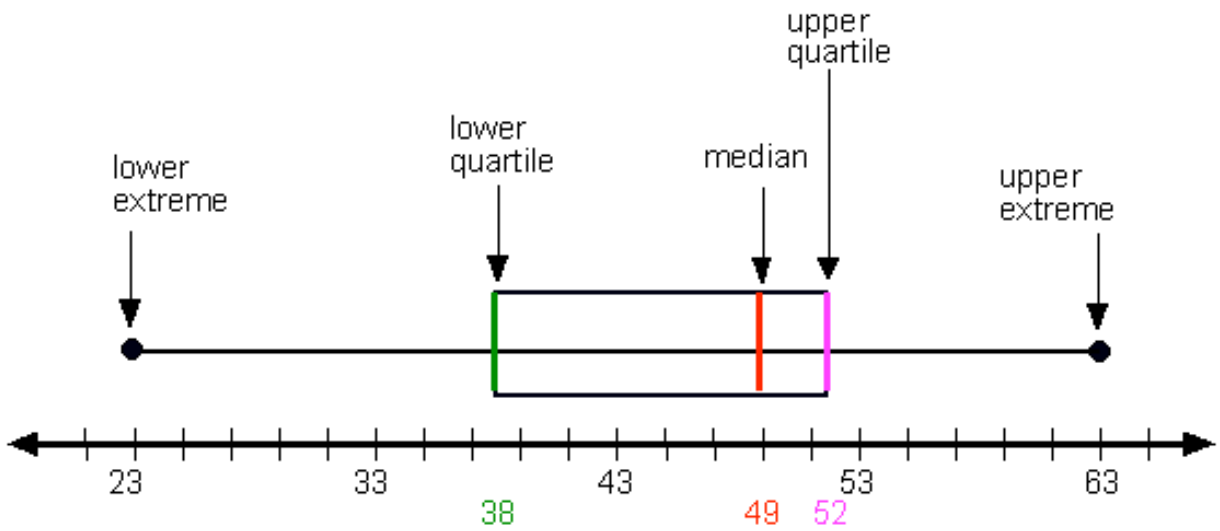


Total number of student marks = 33

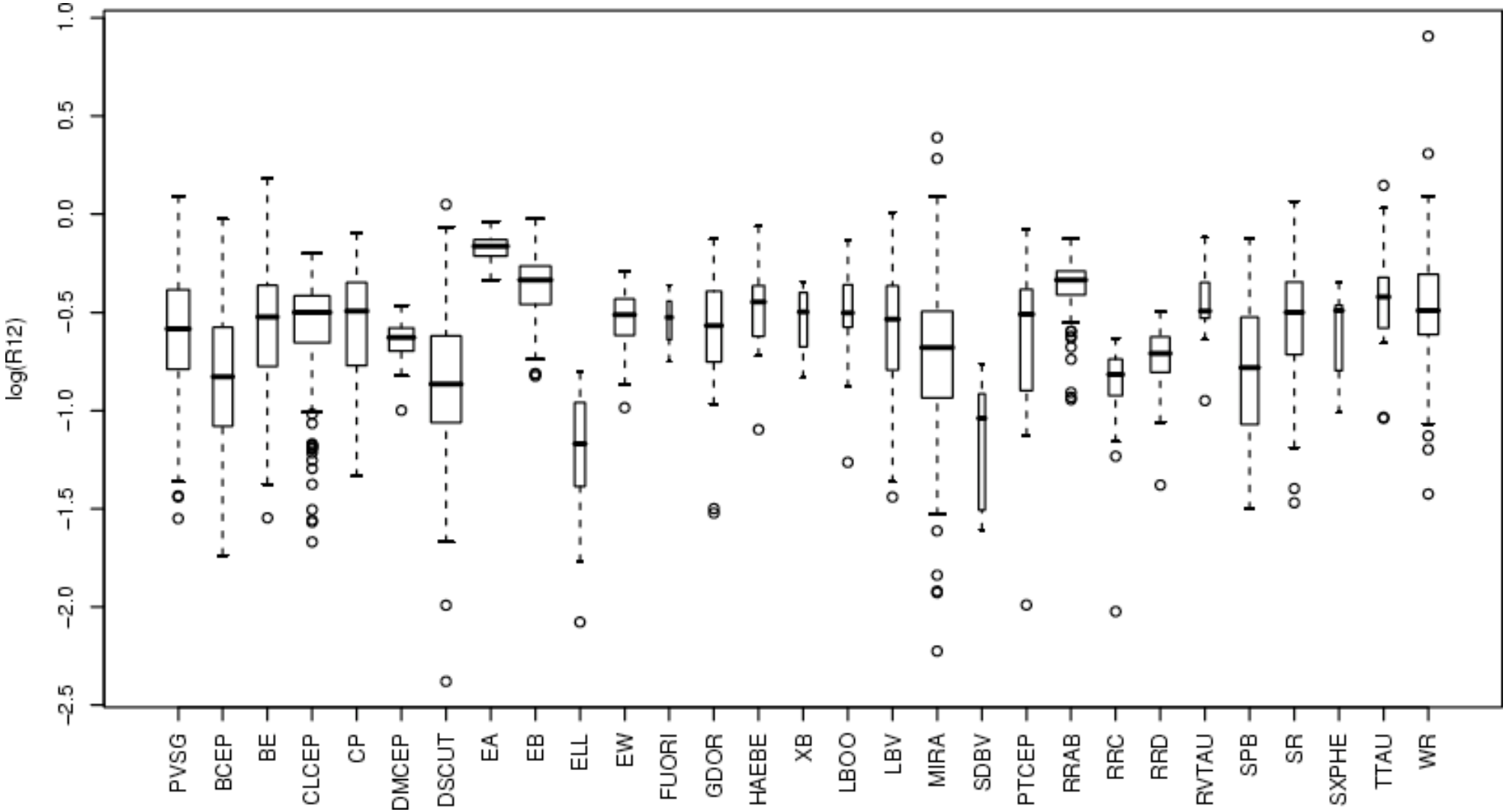
Visualization: Box & Whiskers



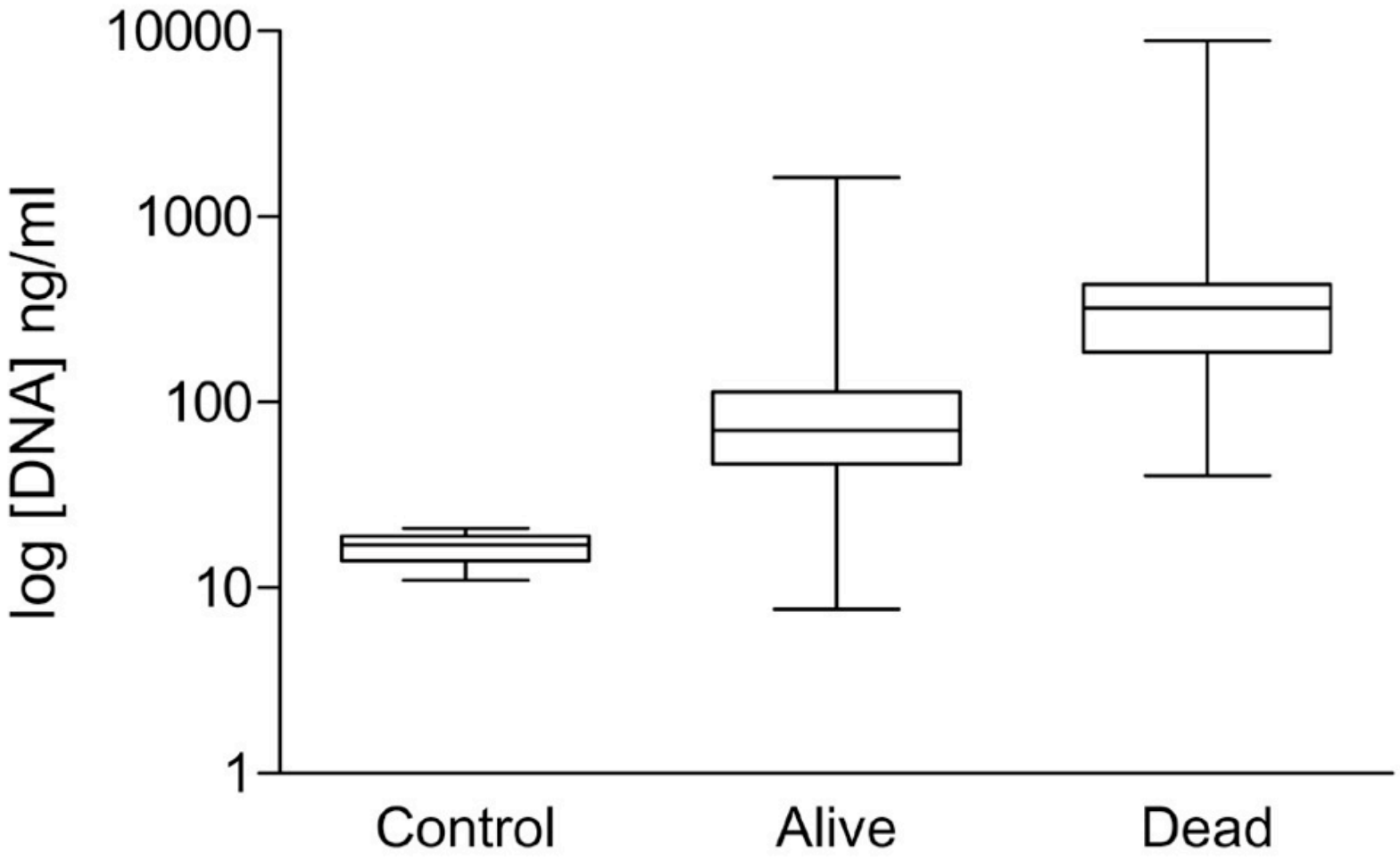
Visualization: Box & Whiskers



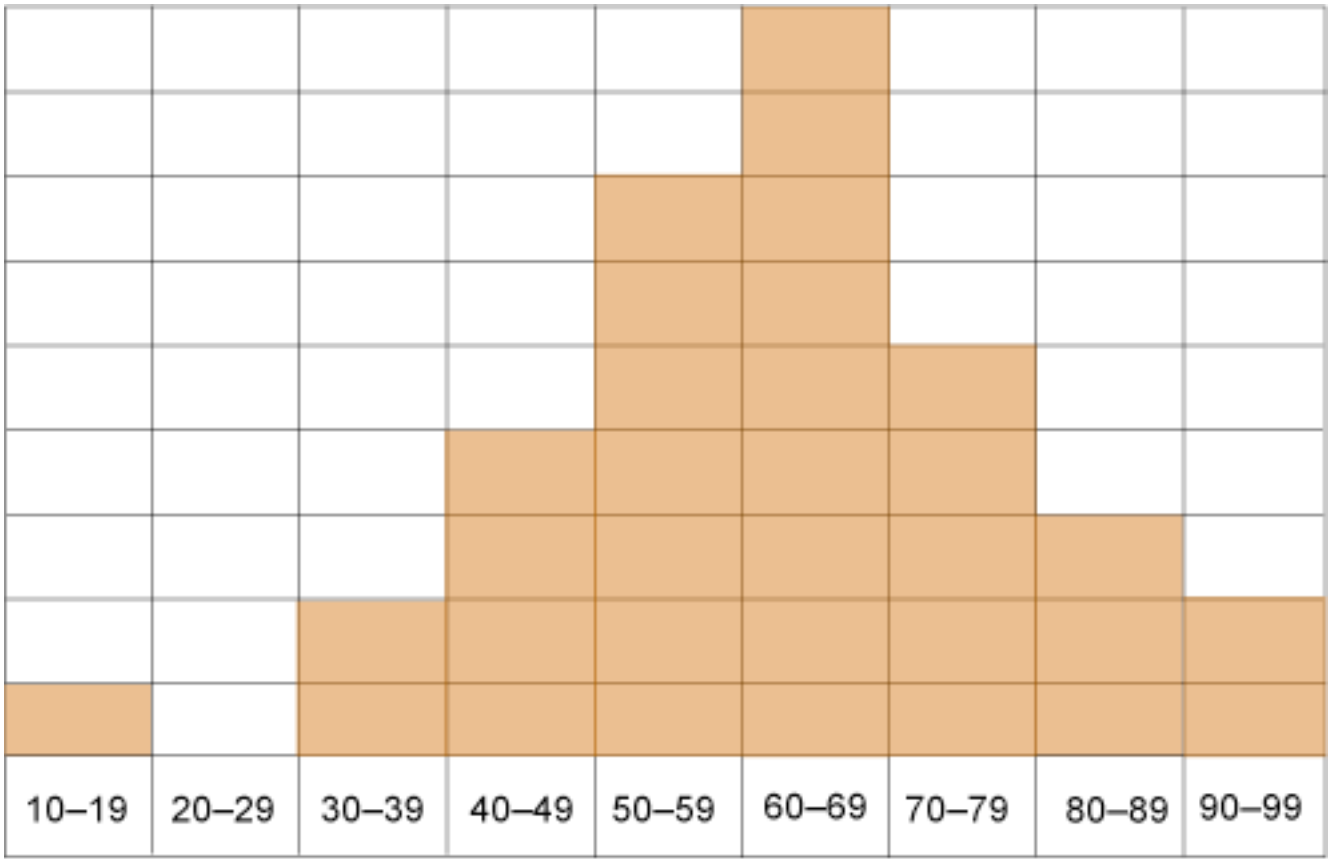
Visualization: Box & Whiskers



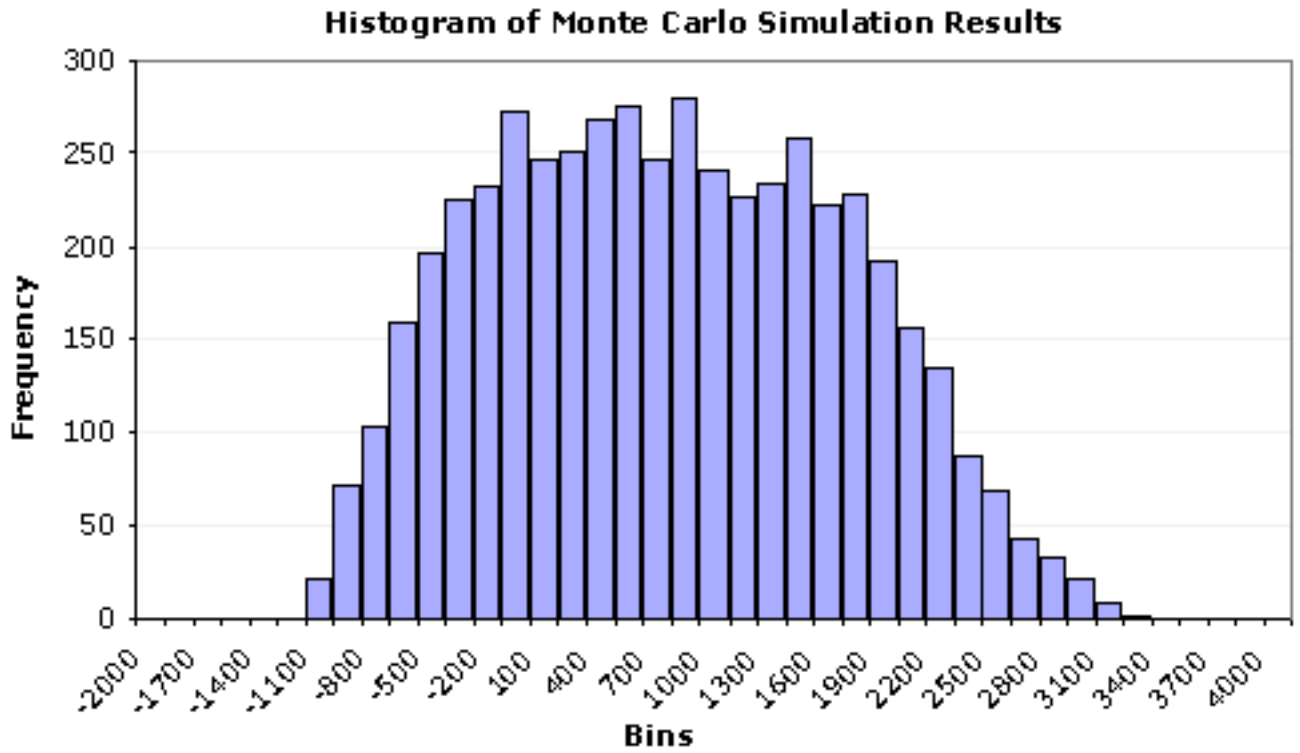
Visualization: Box & Whiskers



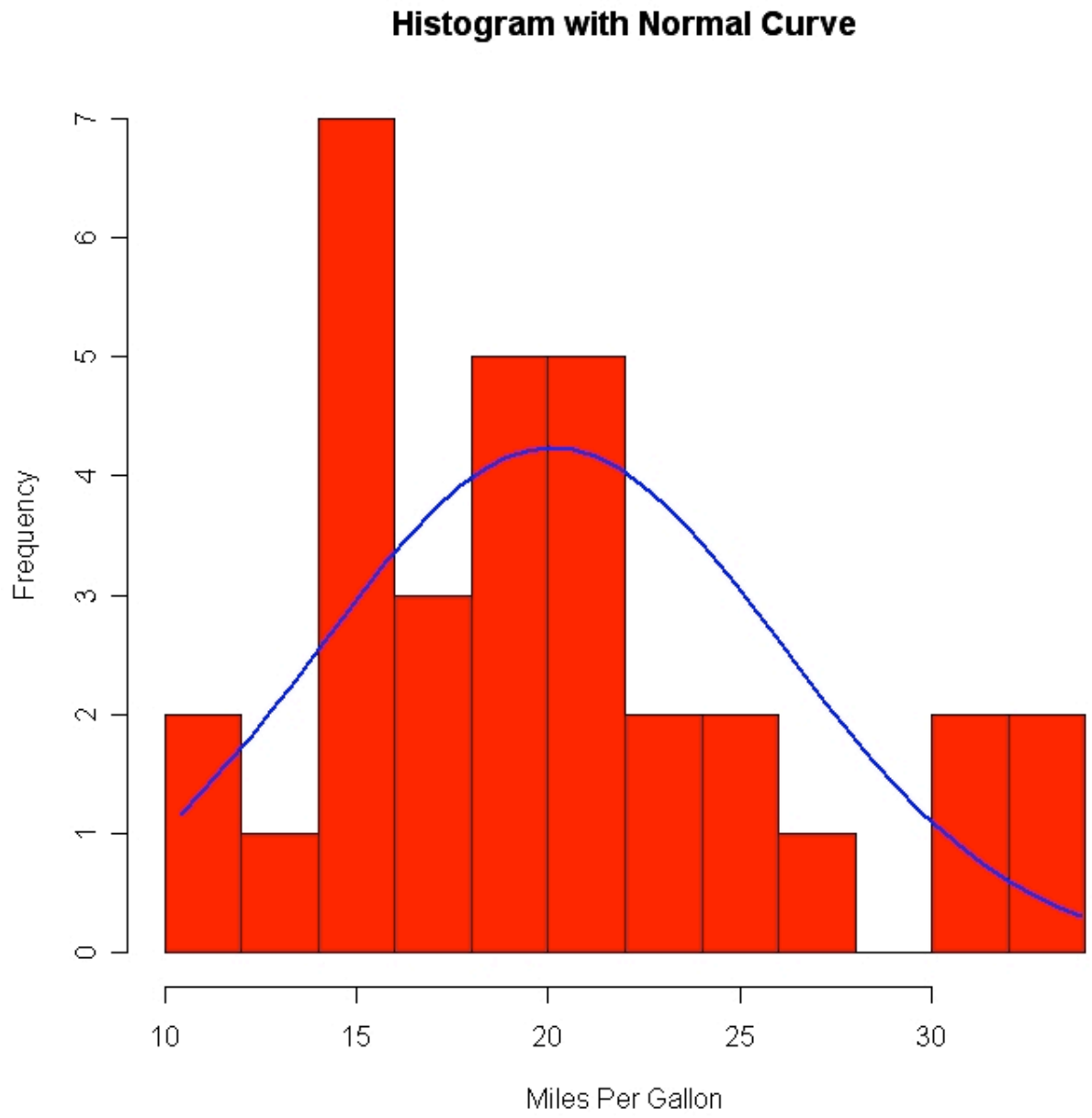
Visualization: Histograms



Visualization: Histograms



Visualization: Histograms



Central Tendency

Central Tendency

Central Tendency

- Mean $\overline{X} = \frac{\sum X_i}{n}$
- Median position (n + 1) / 2 and if no number then the mean of the nearest
- Mode most frequent number, but if none repeat there is no mode

Variation and Measures of Spread

Variation and Measures of Spread

Variation and Measures of Spread

Variation and Measures of Spread

- range high - low

- variance $\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = 0.667$

$$S^2 = \frac{\Sigma (X - \bar{X})^2}{n-1}$$

- standard deviation

$$\sigma = \sqrt{\frac{\Sigma\ (x - \bar{x})^2}{n - 1}}$$

σ = lower case sigma
Σ = capital sigma
$\bar{x}$ = x bar

New York Data: Proposals

- what you wanted to study
- what you expected to find
- how you collected the data

Readings and Assignments

- Readings
  - The Canon, by Natalie Angrier — Chapter 2: *Probabilities*
- Assignments
  - Question about New York assignment: revise method and gather final data. You'll be presenting your work at the next class, taking a first pass at analysis and visualization. After your presentation, you'll revise and re-present a final view.
  - Toss a coin 100 times and write down the results OR imagine tossing a coin 100 times and write down the results. You'll be assigned to one task.