Lecture 1: 1/3/24

Key Terms

Population: The set of all subjects of interest
- could be tangible things, like people or manufactured goods
- could also be more abstract, like a physical process we are interested in

e.g., All college students

Sample: The subset of the population for which we have data
e.g., The college students in MATH 2310

Individual: one of the people/things in the sample being measured or described. Also called "observations" or "observational units."
e.g., one of the students in MATH 2310

Variable: Characteristic(s) that we measure for each individual
- Data can be univariate, bivariate, or multivariate (1, 2, or 3+ variables)
- Categorical Variable: non-numerical, or qualitative variable.
  e.g., glasses or no glasses, or letter grade?
- Numerical Variable: Quantitative Variable
  e.g., Numerical grade

Continuous Variable: Numerical variable that can take any value within some range

e.g., Time since birth

Discrete Variable: Numerical variable that can only take certain values (usually integers) within a range.

e.g., Number of pens in your backpack

Lecture 2: Visualizing Data

Visualizing Categorical Data

- 2 Primary Methods: Pie charts and bar charts

ex. Survey of tree species in a certain patch of forest:
105 Jeffrey Pines
50 Ponderosas
45 white firs

→ switch to Categorical_data.R and make Pie chart

- shows proportion of each category as the proportion of total circle area
- Helpful for comparing one category, or "slice" to the others, but difficult to compare between other categories

→ Now plot same data as bar plot.
- Now it's easier to see relative sizes of all categories, but harder to see that JP takes up more than half the data.

- overall, bar charts are preferred and have been shown to convey information more effectively
  - why? Rectangular sections easier to interpret than angular wedges