Categorical Variables

Ryan Miller

Summarization

- Categorical variables tend to be much simpler to summarize (relative to quantitative variables)
- The two main summaries are *frequencies* and *proportions*
    - **Frequencies** are a tally of how many cases belong to a particular category
    - **Proportions** are the fraction of the total cases that belong to a particular category

Frequency Tables

Frequencies are usually displayed for *all categories* of a variable in a **frequency table**

Day	Frequency
Fri	19
Sat	87
Sun	76
Thur	62

Dividing each frequency by the total number of cases (244 in this dataset) yields proportions for each category

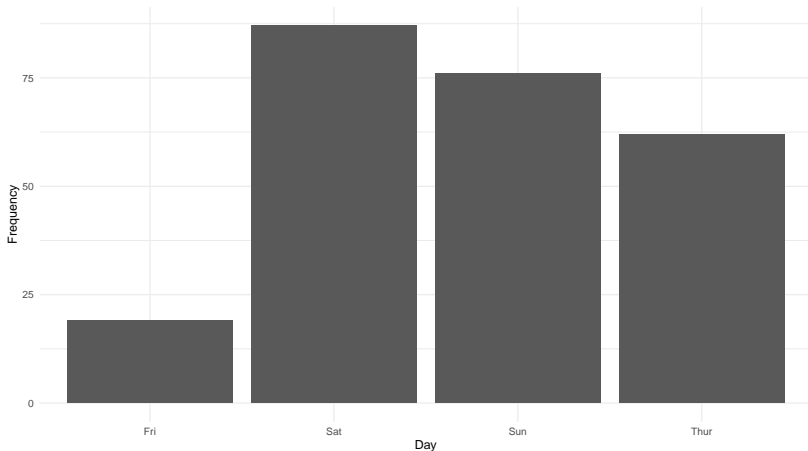| Day	Proportion
Fri	0.08
Sat	0.36
Sun	0.31
Thur	0.25

X

Proportions vs. Percentages

▶ Note that proportions are just percentages divided by 100
  ▶ Statisticians prefer proportions in most scenarios due to
    connections they share with probability (a topic for another
    time)

Day	Proportion	Percentage
Fri	0.08	8%
Sat	0.36	36%
Sun	0.31	31%
Thur	0.25	25%

Barcharts

- The most common way to visualize categorical variables is using a **bar chart**

Piecharts

- An alternative visualization is the **pie chart**, but research has shown that readers perceive information more accurately from bar charts