Suggested reading: OpenIntro Statistics, 3rd edition, Chapter 1, Section 1.6

LO 1. Use scatterplots for describing the relationship between two numerical variables making sure to note the direction (positive or negative), form (linear or non-linear) and the strength of the relationship as well as any unusual observations that stand out.

LO 2. When describing the distribution of a numerical variable, mention its shape, center, and spread, as well as any unusual observations.

LO 3. Note that there are three commonly used measures of center and spread:

- center: mean (the arithmetic average), median (the midpoint), mode (the most frequent observation).
- spread: standard deviation (variability around the mean), range (max-min), interquartile range (middle 50% of the distribution).

LO 4. Identify the shape of a distribution as symmetric, right skewed, or left skewed, and unimodal, bimodal, multimodal, or uniform.

LO 5. Use histograms and box plots to visualize the shape, center, and spread of numerical distributions, and intensity maps for visualizing the spatial distribution of the data.

LO 6. Define a robust statistic (e.g. median, IQR) as a statistic that is not heavily affected by skewness and extreme outliers, and determine when such statistics are more appropriate measures of center and spread compared to other similar statistics.

LO 7. Recognize when transformations (e.g. log) can make the distribution of data more symmetric, and hence easier to model.

Test yourself:

1. *Describe what is meant by robust statistics and when they are used.*
2. *Describe when and why we might want to apply a log transformation to a variable.*

✓ Complete