Section 3.1 (Scatterplots)

Explanatory variable is \( x \) and the response variable is \( y \).

When looking for an overall pattern of a scatterplot, look for form (linear, curved, etc.), direction (positive or negative), strength (how close do the points lie to simple form), and outliers.

Section 3.2 (Correlation)

Correlation \( r \) measures the strength and direction of a linear association. The value of \( r \) is always between -1 and 1 inclusively and is not affected by changes in the unit measurement of either variable. Correlation is not resistant, so outliers can greatly change the value of \( r \).

Section 3.3 (Least Squares Regression)

A regression line is a straight line that describes how a response variable \( y \) changes as an explanatory variable \( x \) changes.

The least-squares regression line is used to predict values of \( y \) for certain values of \( x \). However, you must be cautious when predicting for values outside the given data (extrapolation). The least-squares regression line is \( y = a + bx \), where \( a = \bar{y} + b\bar{x} \)

\[ b = r \frac{s_y}{s_x} \]

It always passes through the point \((\bar{x}, \bar{y})\).

The square of the correlation \( r^2 \) is the fraction of the variance of one variable that is explained by least-squares regression on the other variable.

Residuals are the differences between an observed value of the response variable and the value predicted by the regression line (observed – expected).

Things to look for in a residual plot:

1. A curved pattern shows that the relationship is not linear.
2. Increasing or decreasing spread about the line as \( x \) increases indicates that the prediction of \( y \) will be less accurate for larger \( x \).
3. Individual points with large residuals are outliers in the vertical direction because they lie far from the line that describes the overall pattern.
4. Individual points that are extreme in the \( x \) direction may not have large residuals, but they can be influential points (if removing the point would markedly change the result of the least-squares regression line).

An outlier is an observation that lies outside the overall pattern of the other observations. Influential points are often outliers in the \( x \) direction, but may not have large residuals.