3.2 Correlation

Dr. Lauren Perry

Goals

1. Interpret a correlation coefficient.
   - Describe the strength of a linear relationship.
   - Relate correlation coefficients to scatter plots.

Correlation

Goal: formalize the concept of the *strength* of a linear relationship.

The **correlation** (or **correlation coefficient**) R between two variables describes the strength of their linear relationship.

$$R = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \times \frac{y_i - \bar{y}}{s_y} \right)$$

- s_x and s_y are the respective standard deviations for x and y
- The sample size n is the total number of (x, y) pairs.
- R always takes values between -1 and 1.

This is a pretty involved formula! We'll let a computer handle this one.

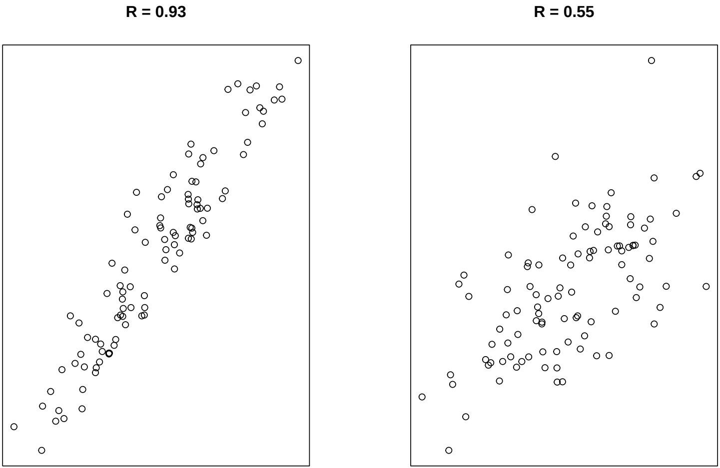
Correlations

- close to -1 suggest strong, negative linear relationships.
- close to $+1$ suggest strong, positive linear relationships.
- close to 0 have little-to-no linear relationship.

Note: the sign of the correlation will match the sign of the slope!

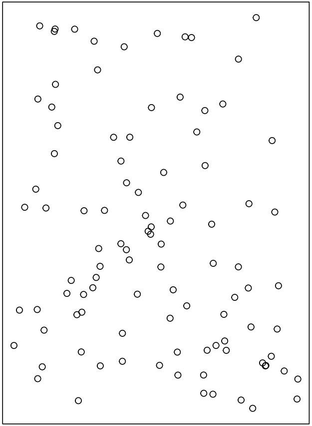
- If $R < 0$, there is a downward trend and $b_1 < 0$.
- If $R > 0$, there is an upward trend and $b_1 > 0$.
- If $R \approx 0$, there is no relationship and $b_1 \approx 0$.

Examples

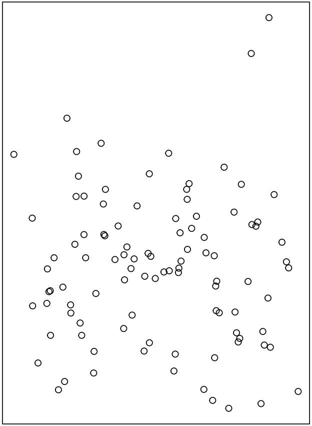


More Examples

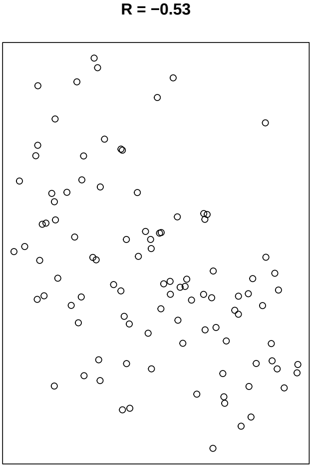
R = 0.07

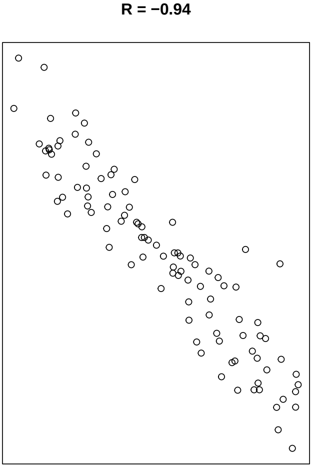


R = −0.05



Even More Examples





When two variables are highly correlated (R close to -1 or 1)

- we know there is a strong *relationship* between them
- BUT we do not know what *causes* that relationship

That is, **correlation does not imply causation**.

A final note:

- Obviously there's a strong relationship between x and y.
  - In fact, $y = x^2$.
- But the *correlation* between x and y is 0!