

Stat 130 - Day 22
Chapter 14: Describing Relationships: Scatterplots and Correlation

Many statistical studies involve relationships between two quantitative variables measured on the same individuals. As before, we'll use both graphical displays and numerical summaries, watching for both general patterns and deviations from them.

Example 1: Scrabble Names

Is there a relationship between the number of letters in a name and the number of “Scrabble points” in the name? The number of Scrabble points that each letter is worth follows:

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|----|
| 1 | 3 | 3 | 2 | 1 | 4 | 2 | 4 | 1 | 8 | 5 | 1 | 3 | 1 | 1 | 3 | 10 | 1 | 1 | 1 | 1 | 4 | 4 | 8 | 4 | 10 |

(a) Determine how many letters are in your name (as you would sign it on a check or credit card receipt) and how many Scrabble points those letters are worth.

Letters:

Points:

(b) Identify the individuals and the two variables in this study. What type of variables (categorical or numerical) are they?

Individuals:

Variable 1:

Variable 2:

- A **scatterplot** displays the relationship between two quantitative variables. The explanatory variable appears on the horizontal axis and the response on the vertical. Each individual is represented by one point fixed by the values of both variables for that individual.

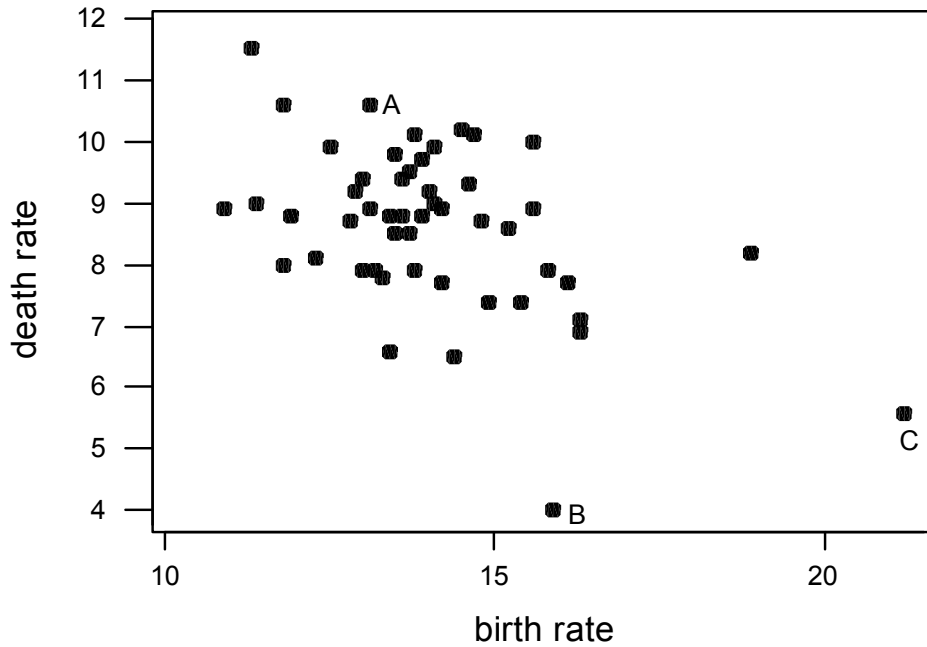
(c) Draw a scatterplot to reveal the relationship between Scrabble points and number of letters.

(d) Does the scatterplot reveal a tendency for people with more letters to have more Scrabble points in their name? Explain.

- (e) Is it the case that *every* student with more letters than another has more Scrabble points as well? If not, identify a pair such that the student with more letters has fewer points.

Example 2: Birth and Death Rates

The following scatterplot displays the relationship between the death rate and the birth rate of the 50 states, both measured per 1000 residents, as of 1997:

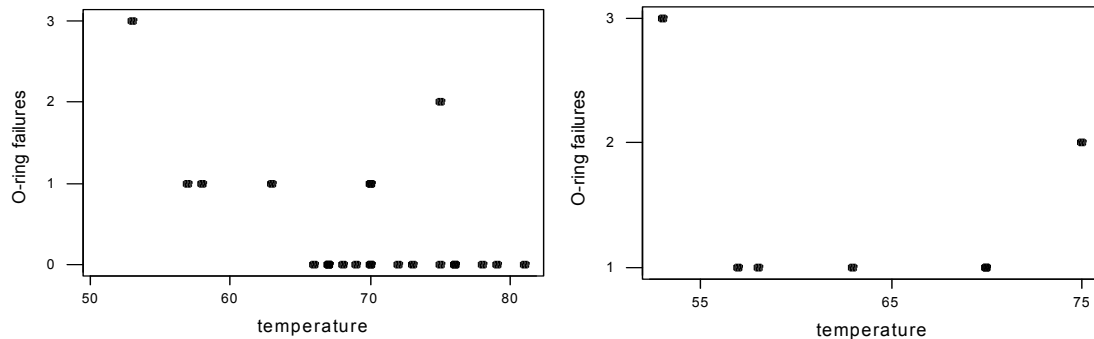


- (a) Describe the overall pattern in this scatterplot. Specifically, do states with higher birth rates tend to have lower, higher, or the same death rates as states with lower birth rates?
- (b) For each of the lettered states, describe how its birth and death rates compare to the others.
- (c) These three states are Alaska, Florida, and Utah. Guess which state goes with which letter.

- The overall pattern of a scatterplot includes the **form**, **direction**, and **strength** of the relationship.
 - A common form is for the relationship to be **linear** (follow a straight line), although curved forms are also common, and another form is for the data to fall in clusters.
 - Two variables are **positively associated** if above-average values of one tend to accompany above-average values of the other and below-average values of one tend to accompany below-average values of the other (scatterplot slopes upward from the left).
 - Two variables are **negatively associated** if above-average values of one tend to accompany below-average values of the other, and vice versa (scatterplot slopes downward).
 - The strength of the relationship concerns how closely the points follow a clear form.

Example 3: Space Shuttle Challenger Disaster

The following scatterplots display the number of O-ring seals showing evidence of thermal distress vs. the air temperature at launch for the 23 space shuttle missions preceding the fatal launch of Challenger in January 1986:



(a) Explain the difference in how the two graphs were constructed.

(b) Would you say that the graph on the left reveals an association between number of O-ring failures and launch temperature? If so, describe its form, direction, and strength.

(c) Repeat (b) for the graph on the right.

