Statistics is about variation. Recognize, quantify and try to explain variation. Variation in contents of cola cans can be explained, in part, by the type of cola in the cans.

Response variable, y
– The variable of primary interest.
Explanatory variable, x
– A variable used to try to explain variation in the response.

When both the response and the explanatory variables are quantitative, display them both in a scatter plot. Look for a general pattern of association.
Scatter plots and Association

- Example: Tar (mg) and nicotine (mg) in cigarettes.
  - $y$, Response: Nicotine (mg).
  - $x$, Explanatory: Tar (mg).
  - Cases: 25 brands of cigarettes.

Positive Association

- Positive Association
  - Above average values of Nicotine are associated with above average values of Tar.
  - Below average values of Nicotine are associated with below average values of Tar.
Negative Association

- Example: Outside temperature and amount of natural gas used.
  - Response: Natural gas used (1000 ft\(^3\)).
  - Explanatory: Outside temperature (°C).
  - Cases: 26 days.

Negative Association

- Above average values of gas are associated with below average temperatures.
- Below average values of gas are associated with above average temperatures.
Correlation

- Linear Association
  - How closely do the points on the scatter plot represent a straight line?
  - The correlation coefficient gives the direction of the linear association and quantifies the strength of the linear association between two quantitative variables.

Correlation

- Standardize $y$
  \[ z_y = \frac{y - \bar{y}}{s_y} \]

- Standardize $x$
  \[ z_x = \frac{x - \bar{x}}{s_x} \]