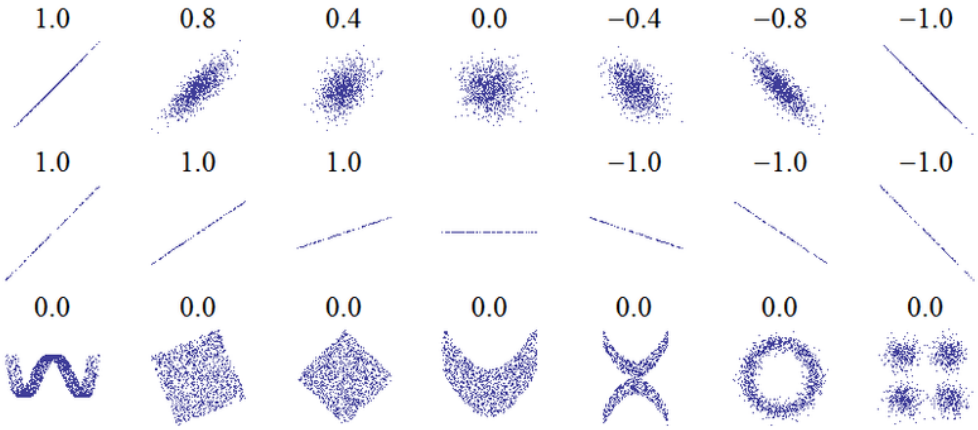
Lesson 12. Regression and Correlation

1 Overview

- **Correlation** quantifies the strength of the linear relationship between X and Y
- **Population correlation**:
- **Sample correlation**:
- Some examples that illustrate different correlation values:



https://commons.wikimedia.org/wiki/File:Correlation_examples.png

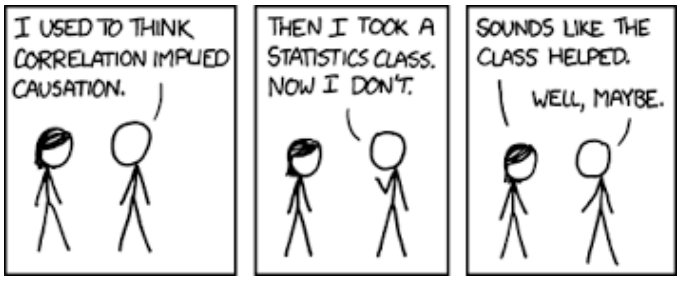
2 Properties

1. Possible values are from to
2. A larger magnitude means a linear relationship
3. $\rho > 0$ means larger values of Y are associated with values of X
4. $\rho < 0$ means larger values of Y are associated with values of X
5. Relation to slope:

⇒ In simple linear regression, testing whether $\beta_1 = 0$ (versus $\beta_1 \neq 0$) is equivalent to testing whether $\rho = 0$ (versus $\rho \neq 0$)

3 Correlation does not imply causation

- Example:
  - X = number of firefighters
  - Y = damage in dollars
  - X and Y probably have a strong correlation
  - Do more firefighters present cause more damage?
  - Size of fire is responsible for both
- A significant correlation only means the variables are associated, not that one causes the other



4 Coefficient of determination (r^2)

- The **coefficient of determination** r^2 tells us how much of the variability in the response variable is explained by the regression model

Example 1. Consider once again our regression model with the `PorschePrice` data. Look at the R output in Lessons 10 and 11.

a. Using the ANOVA table output by R, calculate the coefficient of determination (r^2). Interpret it.
b. Look at the summary output by R, where do you see the value you calculated in part a?