

Correlation vs. Causation

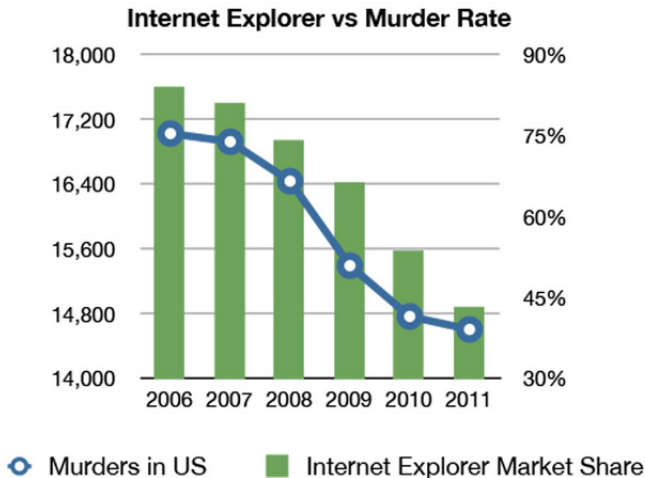
Kyle Burris

Duke University

February 9, 2016

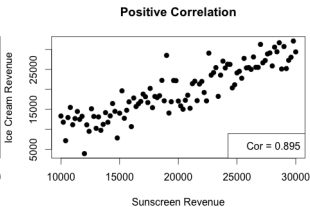
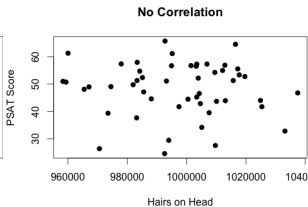
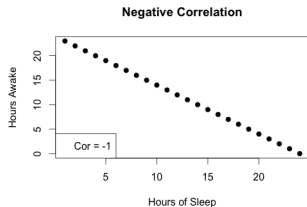
Motivating Example

- Does using Internet Explorer make you want to kill someone?
- Studies show that Internet Explorer causes murder!



What is correlation?

- Correlation is a measure of statistical association between two variables
- Ranges from -1 (perfect negative relationship) to +1 (perfect positive relationship)



- The covariance between two variables X and Y can be found by the following formula:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

- We can put covariance on the same scale by specifying

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

What can correlation mean?

- One of the following:
 - Nothing: relationship is spurious!
 - X causes Y
 - Y causes X
 - Z causes both X and Y
- Ultimately correlation cannot determine causality
- Is a good measure of the amount of variation of Y that can be explained by X

Review Questions

- What would you expect the correlation to be between the number of sunbathers at a lake and the number of fisherman at the same lake?
- Studies show that the more emergency vehicles that show up at an accident, the more fatalities there are. Should we stop sending in emergency vehicles to reduce deaths?
- Can you imagine a situation in which correlation would imply causality? If so, what is it?