Lecture 27
Correlation
Announcements

- Project 2, Part 2, due Friday 5:59PM
- Prelim 2, April 20, 8:30PM-10PM in Kennedy 116 (here) for Ithaca-resident students
  - Coverage from Lecture 12 - Lecture 26 (Monday)
  - Review session on Saturday 3:30PM-5:30PM, room TBA
  - Review sheet and sample exam posted on Canvas.
  - NB: The sample exam is not one I wrote, and is likely to be somewhat different than what I will do.
  - Table of functions included again, allowed a double-sided sheet of notes you make yourself
Prediction

• Guess outcomes in the future, based on available data
• Our simple goal: predict value of one variable based on another

(Demo)
Prediction

If we have a line describing the relation between two variables, we can make predictions.
Relation Between Two Variables

Visualize then quantify

- Any discernible pattern?
- Simplest kind of pattern: Linear? Non-linear?

(Demo)
The Correlation Coefficient $r$

- Developed by Karl Pearson (1857-1936) based on work of Francis Galton (1822-1911)
- Measures linear association
- $-1 \leq r \leq 1$
  - $r = 1$: scatter is perfect straight line sloping up
  - $r = -1$: scatter is perfect straight line sloping down
- $r = 0$: No linear association; *uncorrelated* (Demo)
Definition of $r$

**Correlation Coefficient** ($r$) =

| average of | (array) product of | x in standard units | and | y in standard units |

Measures how clustered the scatter is around a straight line