Section 2.5

Two Quantitative Variables: Scatterplot and Correlation

Outline

- Two quantitative variables
  - Visualization: scatterplot
  - Summary statistic: correlation

Direction of Association

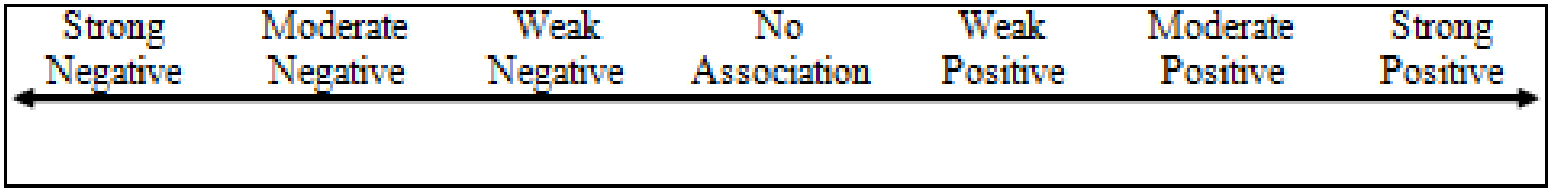
- A ***positive association*** means that values of one variable tend to be higher when values of the other variable are higher
- A ***negative association*** means that values of one variable tend to be lower when values of the other variable are higher
- Two variables are ***not associated*** if knowing the value of one variable does not give you any information about the value of the other variable

Cars Data (Variables and Relationships)

- Quantitative Variables:
  - Weight (pounds)
  - City MPG
  - Fuel capacity (gallons)
  - Page number (in Consumer Reports)
  - Time to go ¼ mile (in seconds)
  - Acceleration time from 0 to 60 mph
- Relationships
  - Weight vs. CityMPG
  - Weight vs. FuelCapacity
  - PageNum vs. Fuel Capacity
  - Weight vs. QtrMile
  - Acc060 vs. QtrMile
  - CityMPG vs. QtrMile

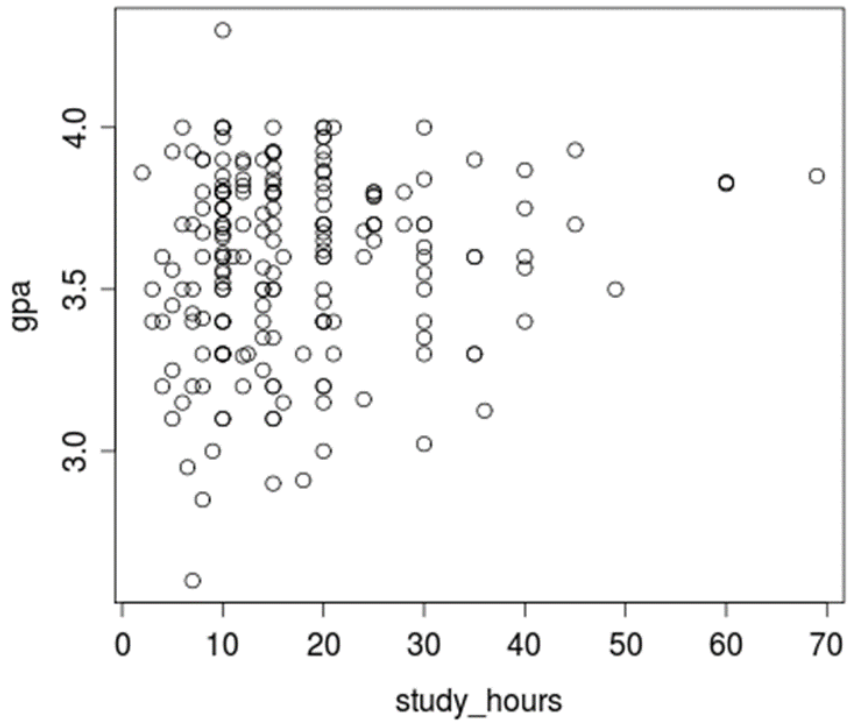
Cars Data (Strength and Direction)

- Make initial guesses for the strength and direction of association for each of the following:
  1) Weight vs. CityMPG
  2) Weight vs. FuelCapacity
  3) PageNum vs. Fuel Capacity
  4) Weight vs. QtrMile
  5) Acc060 vs. QtrMile
  6) CityMPG vs. QtrMile

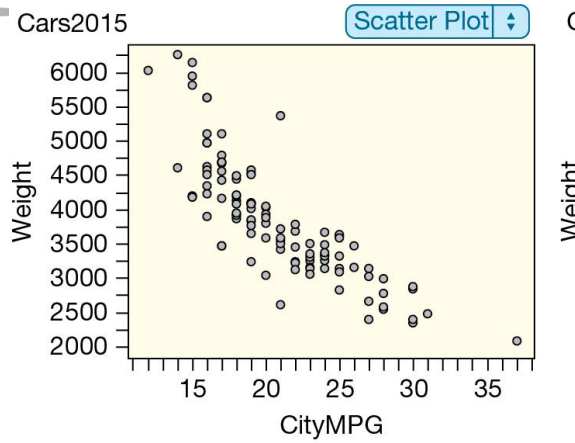


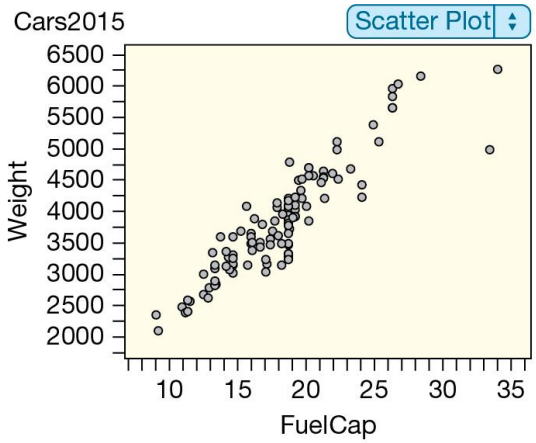
Scatterplot

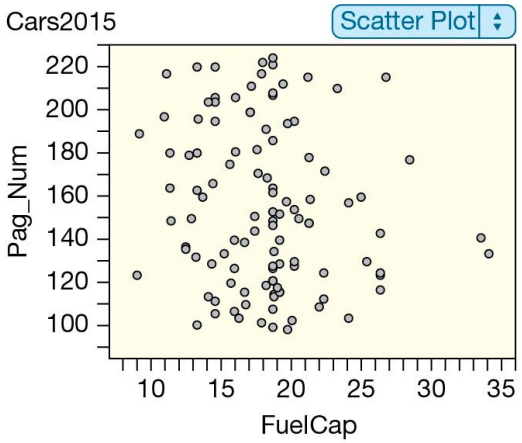
A ***scatterplot*** is the graph of the relationship between two quantitative variables.

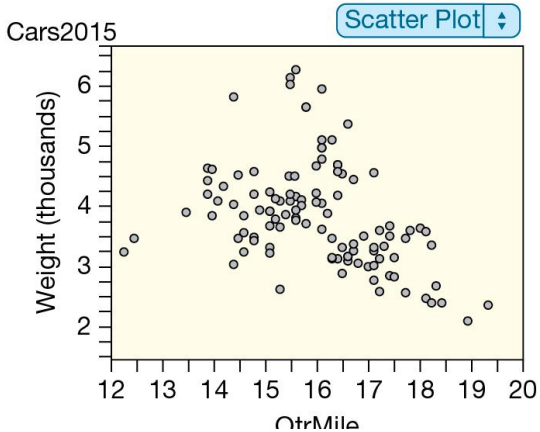


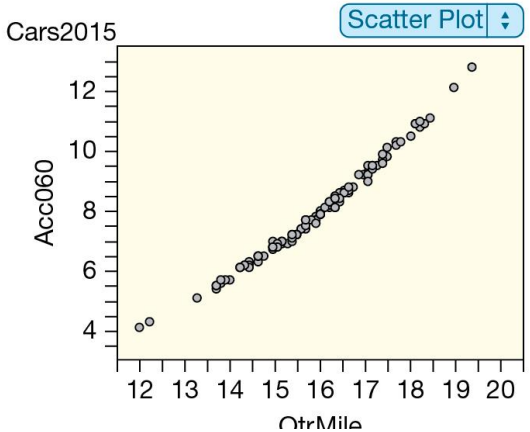
Car Associations

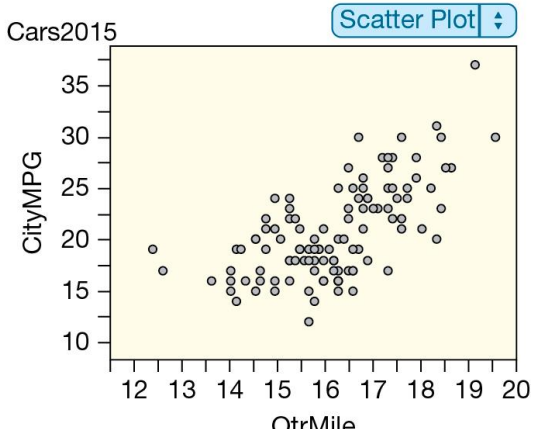


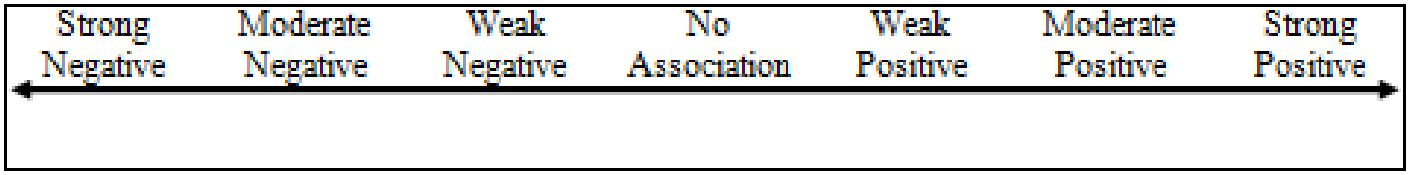










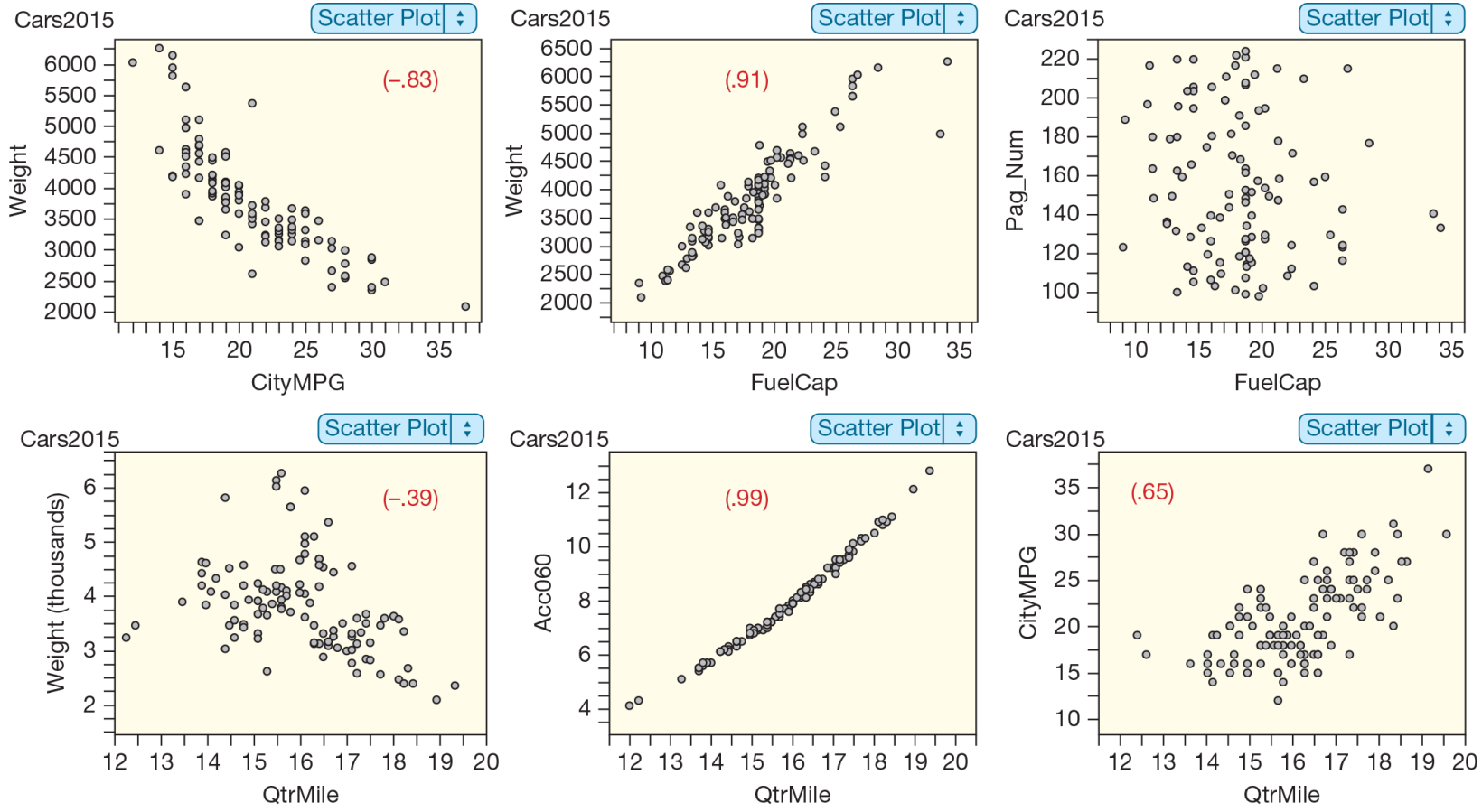


Correlation (Definition)

The ***correlation*** is a measure of the strength and direction of linear association between two quantitative variables

- Sample correlation: r
- Population correlation: ρ (“rho”)

Car Correlations



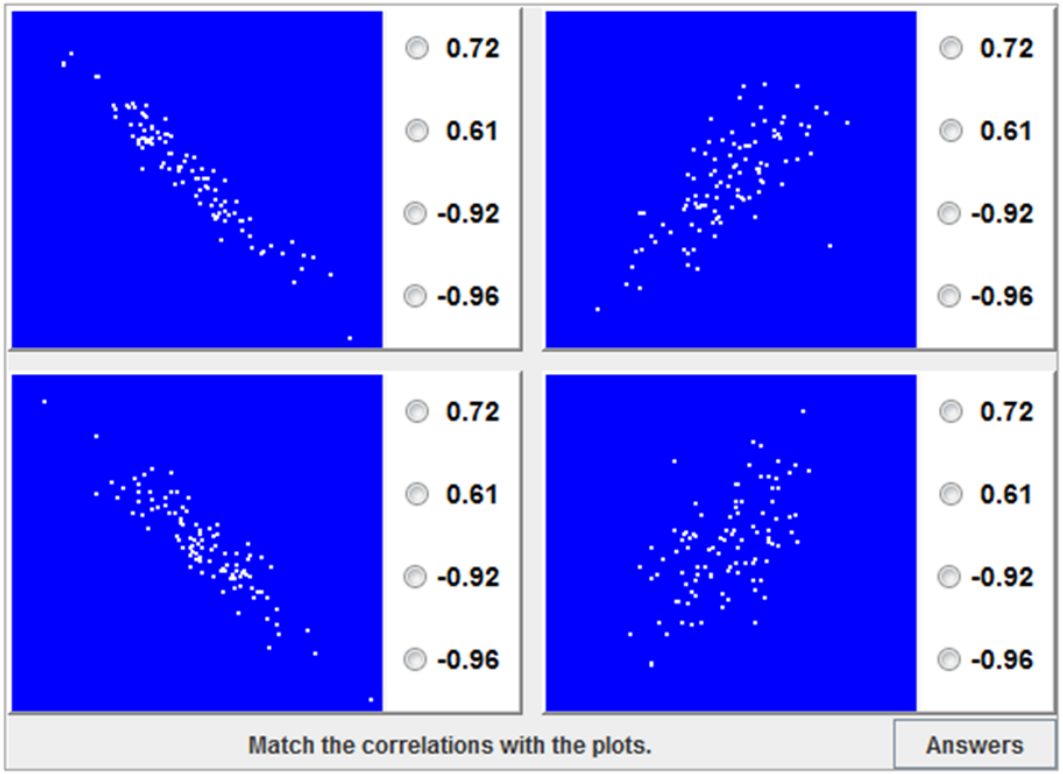
What are the properties of correlation?

Correlation (What Does It Tell Us?)

1. $-1 \leq r \leq 1$
2. The sign indicates the direction of association
    1. positive association: $r > 0$
    2. negative association: $r < 0$
    3. no linear association: $r \approx 0$
3. The closer r is to ± 1, the stronger the linear association
4. r has no units and does not depend on the units of measurement
5. The correlation between X and Y is the same as the correlation between Y and X

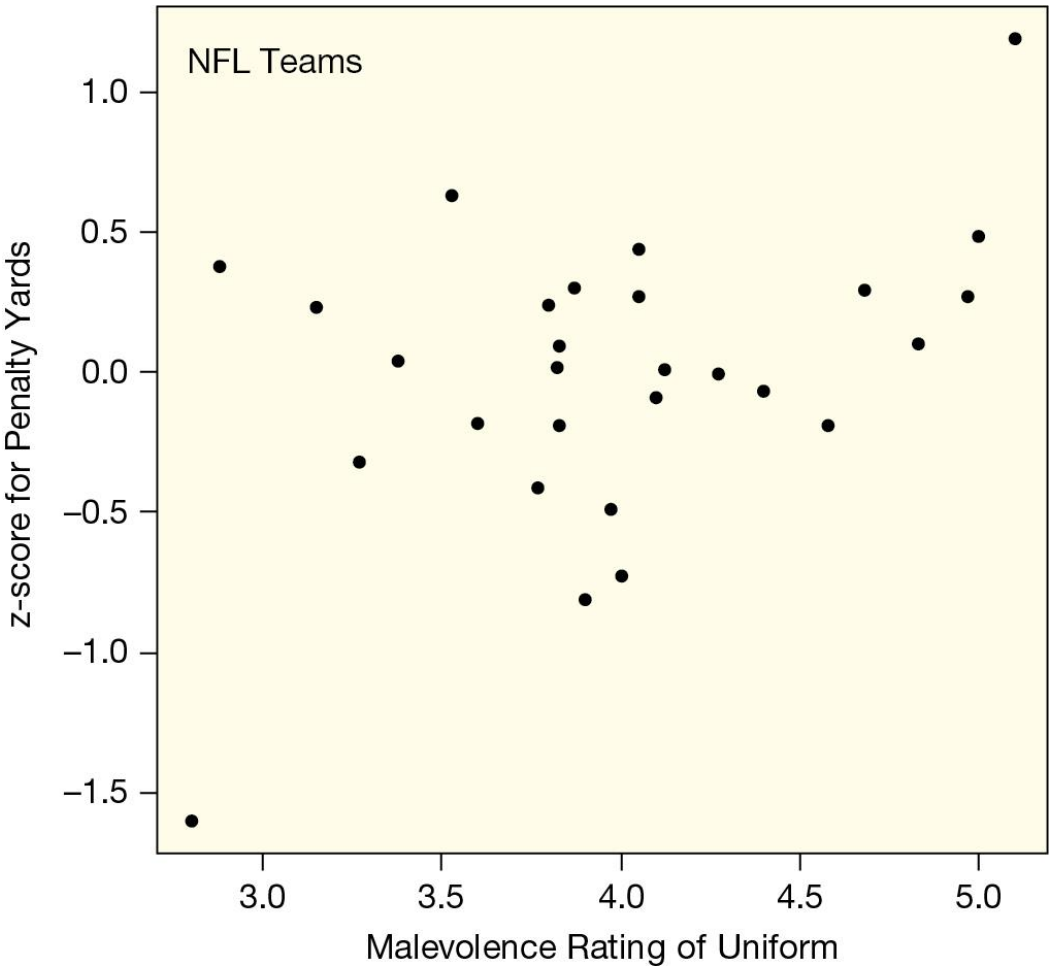
Correlation Guessing Game

http://istics.net/stat/correlations/

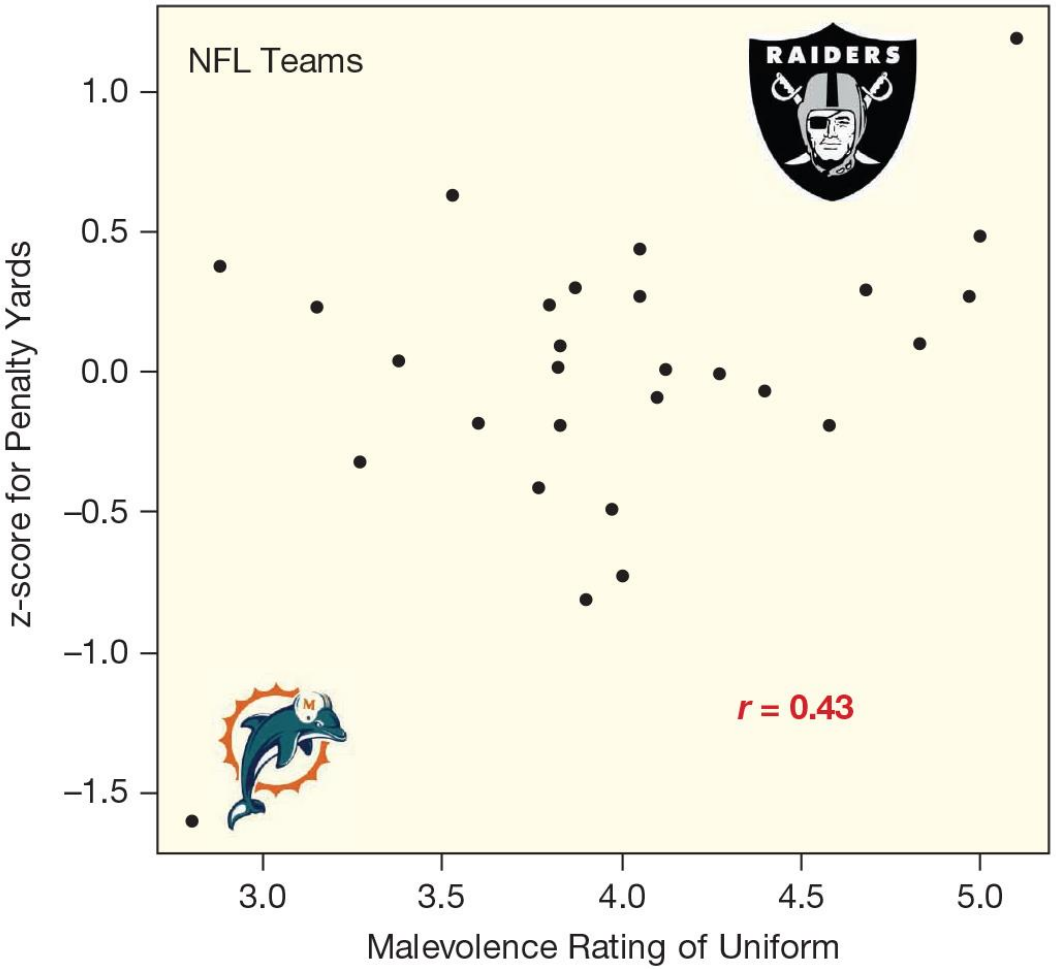


Highest scorer in the class gets an extra point on the first exam!

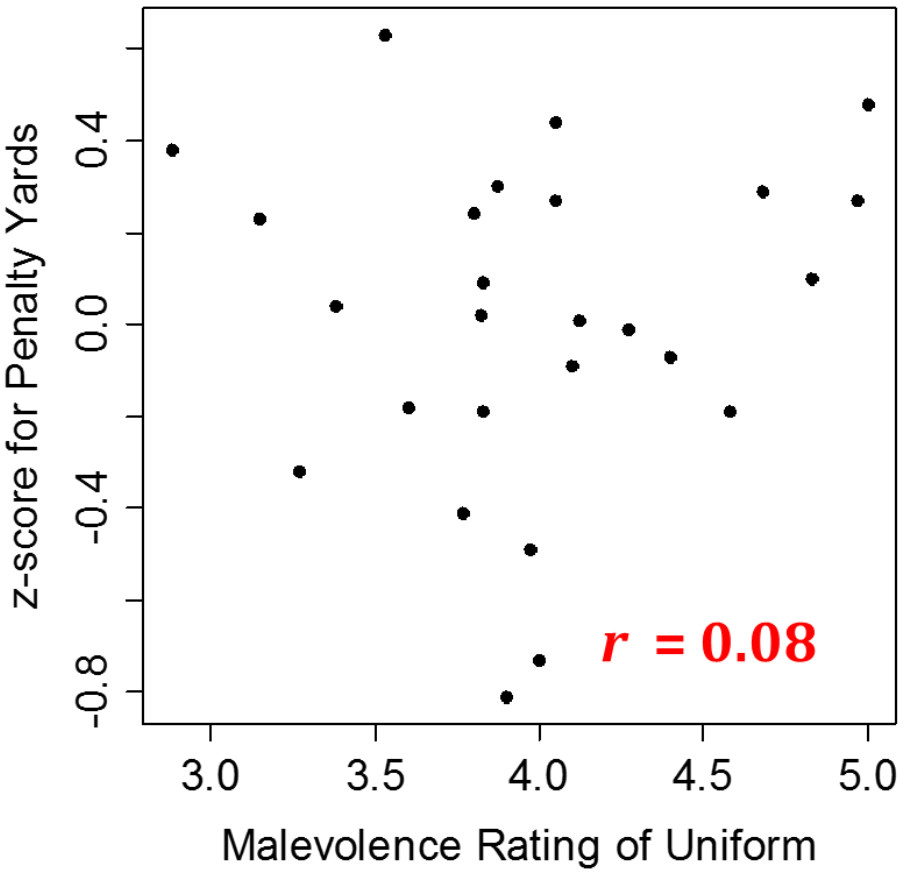
Correlation (NFL Teams)



Correlation (Who are the Outliers?)



Correlation (Outliers Removed)

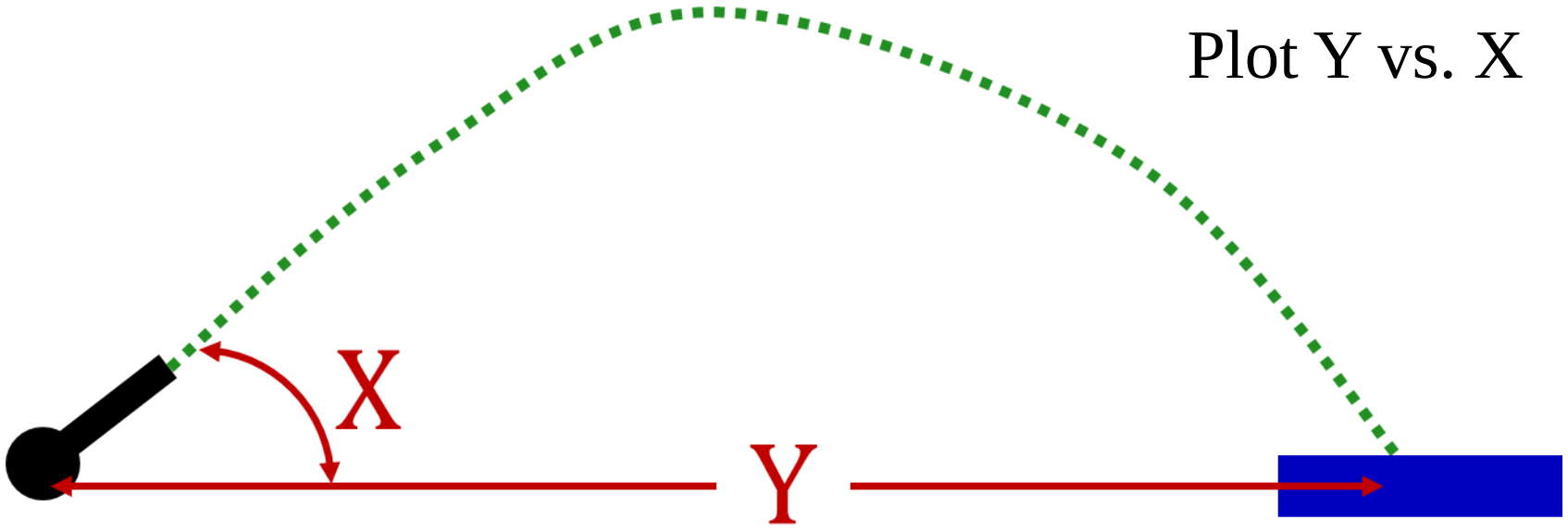


Same plot, but with Dolphins and Raiders (outliers) removed

Correlation Cautions (Outliers)

1. Correlation can be heavily affected by outliers. Always plot your data!

Human Cannonball



What is the correlation between X and Y?

$r \approx 0$

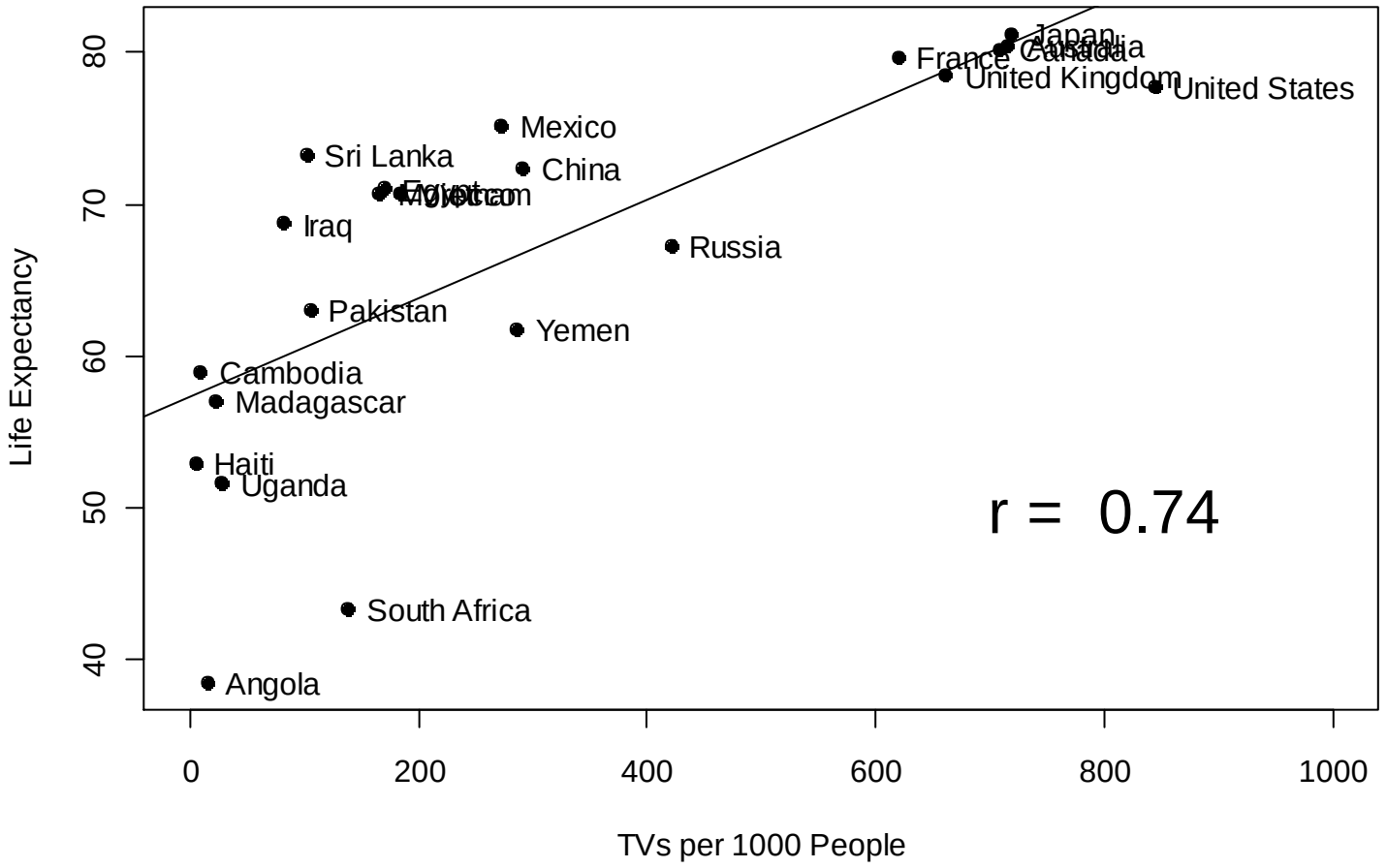
Are X and Y associated?

Yes!

Correlation Cautions ($r = 0$)

1. Correlation can be heavily affected by outliers. Always plot your data!
2. $r = 0$ means no *linear* association. The variables could still be otherwise associated. Always plot your data!

TVs and Life Expectancy



Correlation Cautions (and Causation)

1. Correlation can be heavily affected by outliers. Always plot your data!
2. $r = 0$ means no *linear* association. The variables could still be otherwise associated. Always plot your data!
3. Correlation does not imply causation!