

Assignment 1  
Due January 26

1) A 2012 Pew Research survey asked 2,373 randomly sampled registered voters their political affiliation (Republican, Democrat, or Independent) and whether or not they identify as swing voters. 35% of respondents identified as Independent, 23% identified as swing voters, and 11% identified as both.

- a) Are being Independent and being a swing voter disjoint, i.e. mutually exclusive?
- b) What percent of voters are Independent but not swing
- c) What percent of voters are Independent or swing voters?
- d) What percent of voters are neither Independent nor swing voters?
- e) Is the event that someone is a swing voter independent of the event that someone is a political Independent?

2) Load the Felix Hernandez dataset in R.

Date	Date of game	IP	Innings pitched
Month	Month of game	H	Hits
OPP	Opponent	R	Runs
away	Playing away from home?	HR	Home runs
W	Win	BB	Base on balls (walk)
L	Loss	SO	Strike outs
ERA	Earned run average		

- a) How many wins does Felix have this year?
- b) What is the mean, median, and mode number of strikeouts Felix threw over the 2015 season? Use this function to calculate the mode:  

```
Mode <- function(x) {  
  ux <- unique(x)  
  ux[which.max(tabulate(match(x, ux)))]  
}
```
- c) Plot the relationship between innings pitched and strikeouts and between innings pitched and walks (base on balls). Describe the patterns you see (decreasing relationship? No relationship?).
- d) Calculate the correlation coefficient between innings pitched and strikeouts and between innings pitched and walks. Do these align with what you saw in the plots?

e) Calculate the mean and variance of walks by month (hint: use the `by()` function like in lab). Do you see changing mean walks over time? What about the variability over time? What might the pattern mean?

f) Does Felix win more on the road or at home?

g) Load the other data set containing similar records for Randy Johnson in 1995. Does Randy Johnson outperform Felix in terms of strikeouts across the 1995 season?

3) Sophia who took the Graduate Record Examination (GRE) scored 156 on the Verbal Reasoning section and 157 on the Quantitative Reasoning section. The mean score for Verbal Reasoning section for all test takers was 151 with a standard deviation of 7, and the mean score for the Quantitative Reasoning was 153 with a standard deviation of 7.67. Suppose that both distributions are nearly normal.

a) What is Sophia's Z-score on the Verbal Reasoning section? On the Quantitative Reasoning section?

b) Draw a standard normal distribution curve and mark these two Z-scores.

c) Relative to others, which section did she do better on?

d) Find her percentile scores for the two exams.

e) What percent of the test takers did better than her on the Verbal Reasoning section? On the Quantitative Reasoning section?

f) Explain why simply comparing her raw scores from the two sections would lead to the incorrect conclusion that she did better on the Quantitative Reasoning section (2-3 sentences).