While examining a set of data about the Super Bowl, I had a handful of questions I wanted to examine:

1) What are the best Super Bowl Squares to get?

- In my analysis, I have determined you should be looking for a 0, 1, 4, or 7. If you get a 2, 5, or 8, you may not have fun watching the game. I think if there was any “missing data”, it would be the scores at the end of the first quarter, second quarter, and third quarter as well. This way you could do an analysis on these squares as well, since there are typically payouts at these time points as well. It also may help to have some indication of AFC vs NFC team. Perhaps have columns afc\_team, nfc\_team, since this is the combination when you actually play the squares pool.

2) Can we use this data to bet the over/under?

- In my analysis, I think you can definitely use the history of the Super Bowl to assist your bet. Should it be the only thing you consider? Absolutely not. But can it be a piece of the puzzle you assemble when making a bet? Absolutely.

3) Has scoring increased in the more recent Super Bowls?

- In my analysis, this was a pretty hard “no”. Based on both PMF analysis and a difference in means test, it can be pretty safely concluded that there is no difference in scoring between the old Super Bowls (1-27) and the new ones (28-54). Perhaps the groups should not be an even 27 vs 27. Maybe the analysis would be better if I went to the extreme and used the first 5 and the last 5.

4) Do higher scoring games lead to higher ratings?

- This was a rather profound “no” as well. Based on a scatter plot and a linear regression analysis, it was rather clear that there was no relationship between these two variables. One struggle with this analysis is that I probably don’t have a full grasp on what TV ratings are. Maybe I should use number of viewers instead.

5) Can advertising costs be represented by a particular distribution?

- It seems, through analyses of scatter plots, CDFs, and CCDFs, that this could very reasonably be modeled using an exponential distribution.

One of the major challenges that I had was with the actual graphs that I produced. I could not figure out how to get the axis values to be what I wanted. For example, when I was producing histograms of whole numbers, the scale of the axis would have decimals and it is just not a 100% accurate way to represent the data. Other than that, I felt like Downey’s instruction/walkthrough of the code made it easy to understand and easy to perform our own analyses.