NFL 2015: A Statistical Analysis

Dave Dyer[1], Jostein Barry-Straume[1], Robert Flamenbaum[1], & Bryan Cikatz[1]

[1] Southern Methodist University

Abstract

This study investigates how average and variance of yards gained impacts a given team's ability to get touchdowns in the NFL. The theory is that teams with high average yards gained, but with low variance, would get more touchdowns than those with higher average yards gained, but higher variance. In other words, consistency is more important than big plays.

The Data

The data are the 2015 play-by-play records available at Kaggle (https://www.kaggle.com/maxhorowitz/nflplaybyplay2015). These data are comprised of 46,129 rows describing every NFL play run over the 2015 season, by 32 teams. There are 65 columns that include multiple values, including text description of the play / penalty, names of players , touchdown boolean flags, down markers, timestamps, game ids, and yardages.

The Tools

For this analysis, we use MongoDB for data mining and basic statistics, and R for data visualization and Markdown. We originally had planned on using MySQL on bluemix, but abandoned it in favor of Mongo after some serious issues whist loading the data.

EDA

We counted the number of plays per team. Here's that breakdown:

Table 1
W/L Data

teams	wins	losses
broncos	10	4
chargers	12	2
oilers	5	9

Results

Discussion

We used R (3.4.2, R Core Team, 2017) and the R-packages *knitr* (1.17, Xie, 2015), and *papaja* (0.1.0.9492, Aust & Barth, 2017) for all our analyses.

Conclusion

Appendices

References

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/