

# NFL TOUCHDOWNS!

MSDS 7330 – 404

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

# INTRODUCTION

# ABSTRACT

This study investigates how the 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

- 2015 play-by-play records available at Kaggle
  - 14.7 MB
  - 46,129 rows
  - 65 columns

# THE TOOLS

- MongoDB
  - Data mining and basic statistics
- R
  - Data visualization and Markdown
  - Mongolite package

# PREVIOUS WORK

- Underrated NFL Stats
  - Based on 2010 season
  - Looks at the relationship between big-plays allowed and touchdowns allowed
  - Big-play defined as 20 or more yards
- How to Quantify the NFL
  - Based on 2015 season
  - Looks at five different statistics – we focus on big-play percentage
  - Big-play defined as runs of more than 10 yards or passes of more than 25 yards

# "UNDERRATED NFL STATS"

## BEST TEAMS AT PREVENTING BIG-PLAYS

- 1. Steelers** (36, 5 TDs)
- 2. Chargers** (43, 5 TDs)
- 3. Falcons** (48, 9 TDs)
- 4. Vikings** (49, 9 TDs)
- 5. Buccaneers** (51, 9 TDs)

## WORST TEAMS AT PREVENTING BIG-PLAYS

- 32. Broncos** (84, 20 TDs)
- 31. Seahawks** (76, 14 TDs)
- 30. Cardinals** (74, 14 TDs)
- 29. Jaguars** (72, 16 TDs)
- 28. Redskins** (70, 10 TDs)

# "UNDERRATED NFL STATS" (CONT.)

## TAKEAWAY POINTS

- "Only six of [the top ten teams] finished .500 or better last season, but those six teams fall in the top seven on the list."
- "Of [the bottom ten teams], only Seattle, with a sub-.500 record, and Philadelphia made the playoffs."



# "HOW TO QUANTIFY THE NFL"

## TAKEAWAY POINTS

- Teams with the best big-play percentage were the Bills, Vikings, Seahawks, Panthers, Chiefs, and Steelers (roughly 9% of all plays were big-plays).
- Five of these six teams made the playoffs.

# EXPANDING UPON PREVIOUS WORK

Ultimately, we are looking to expand upon the role of big-plays in the NFL. Previous work looked more at the defensive side (only touching lightly on offense) whereas we are focusing on the offensive side of the equation.

Furthermore, we seek to discover the importance of consistency in relation to big-plays. We want to know if smaller but more consistent gains are more effective than larger but more irregular gains.

# DATA & ANALYSIS

# INITIAL QUERIES IN MONGODB

```
1 db.NFLPlaybyPlay2015.aggregate([
2   { "$group": {
3     "_id": { "posteam": "$posteam"},
4     "Touchdown": { "$sum": "$Touchdown" },
5     "Penalty_Yards": { "$sum": "$Penalty.Yards" }
6   }
7 })
```

Shell Output

```
1 { "_id" : { "posteam" : "MIN" }, "Touchdown" : 35, "Penalty_Yards" : 730 }
2 { "_id" : { "posteam" : "DAL" }, "Touchdown" : 31, "Penalty_Yards" : 822 }
3 { "_id" : { "posteam" : "DEN" }, "Touchdown" : 38, "Penalty_Yards" : 698 }
4 { "_id" : { "posteam" : "BAL" }, "Touchdown" : 35, "Penalty_Yards" : 809 }
5 { "_id" : { "posteam" : "SF" }, "Touchdown" : 26, "Penalty_Yards" : 853 }
6 { "_id" : { "posteam" : "CIN" }, "Touchdown" : 55, "Penalty_Yards" : 973 }
7 { "_id" : { "posteam" : "OAK" }, "Touchdown" : 45, "Penalty_Yards" : 1012 }
8 { "_id" : { "posteam" : "TB" }, "Touchdown" : 38, "Penalty_Yards" : 1152 }
9 { "_id" : { "posteam" : "IND" }, "Touchdown" : 36, "Penalty_Yards" : 1067 }
10 { "_id" : { "posteam" : "NYJ" }, "Touchdown" : 49, "Penalty_Yards" : 692 }
11 { "_id" : { "posteam" : "GB" }, "Touchdown" : 45, "Penalty_Yards" : 1304 }
12 { "_id" : { "posteam" : "CHI" }, "Touchdown" : 40, "Penalty_Yards" : 971 }
13 { "_id" : { "posteam" : null }, "Touchdown" : 0, "Penalty_Yards" : 0 }
14 { "_id" : { "posteam" : "WAS" }, "Touchdown" : 46, "Penalty_Yards" : 899 }
15 { "_id" : { "posteam" : "NE" }, "Touchdown" : 55, "Penalty_Yards" : 1069 }
16 { "_id" : { "posteam" : "JAC" }, "Touchdown" : 48, "Penalty_Yards" : 813 }
17 { "_id" : { "posteam" : "PIT" }, "Touchdown" : 43, "Penalty_Yards" : 1041 }
18 { "_id" : { "posteam" : "MIA" }, "Touchdown" : 37, "Penalty_Yards" : 920 }
19 { "_id" : { "posteam" : "NYG" }, "Touchdown" : 46, "Penalty_Yards" : 1048 }
```

- We ran queries on:
  - Total TD's for the season by team
  - Total TD's for each quarter by team
  - Average yards gained per play by team
  - Average TD's earned per play by team
  - Total TD's with penalty yards for the team

# LIST OF R PACKAGES

- `library("mongolite")`
- `library("dplyr")`
- `library("knitr")`
- `library("tibble")`
- `library("kableExtra")`
- `library("ggplot2")`

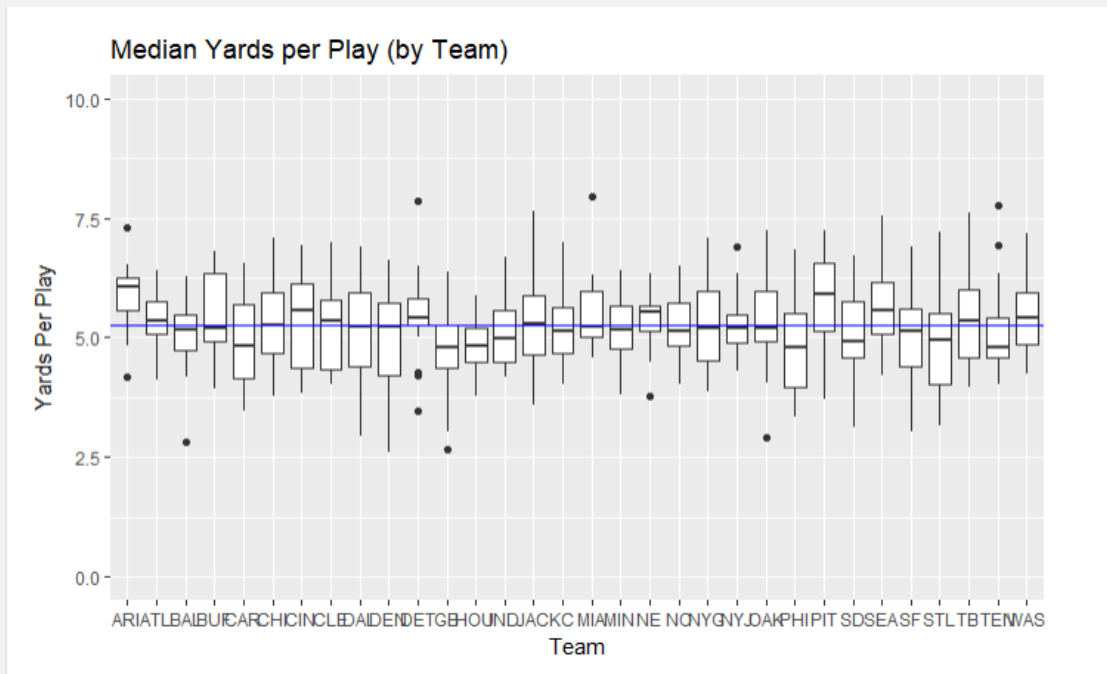
# MONGOLITE

# Returns sum of touchdowns and sum of penalty yards per team

```
sumTdsPenYds <- m$aggregate ( ' [  
  { "$group": {  
    "_id": { "posteam": "$posteam"},  
    "Touchdown": { "$sum": "$Touchdown" },  
    "Penalty_Yards": { "$sum": "$Penalty.Yards" }  
  }  
}] ' )
```

# MEDIAN YARDS PER PLAY

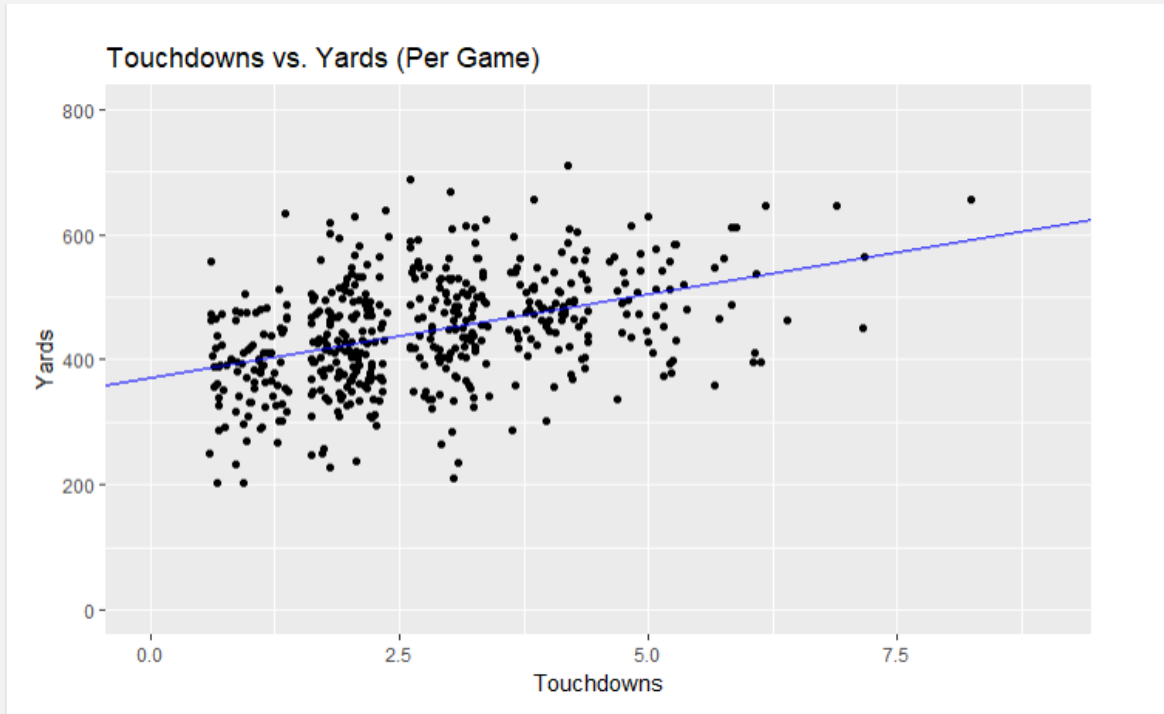
## INSIGHTS



- Not a large amount of variance between medians
- What we expected – NFL is highly competitive
- Large difference in outer quartiles and number of outliers
- Supports our investigation into whether or not a large variance changes team performance

# TD'S VS. YARDS

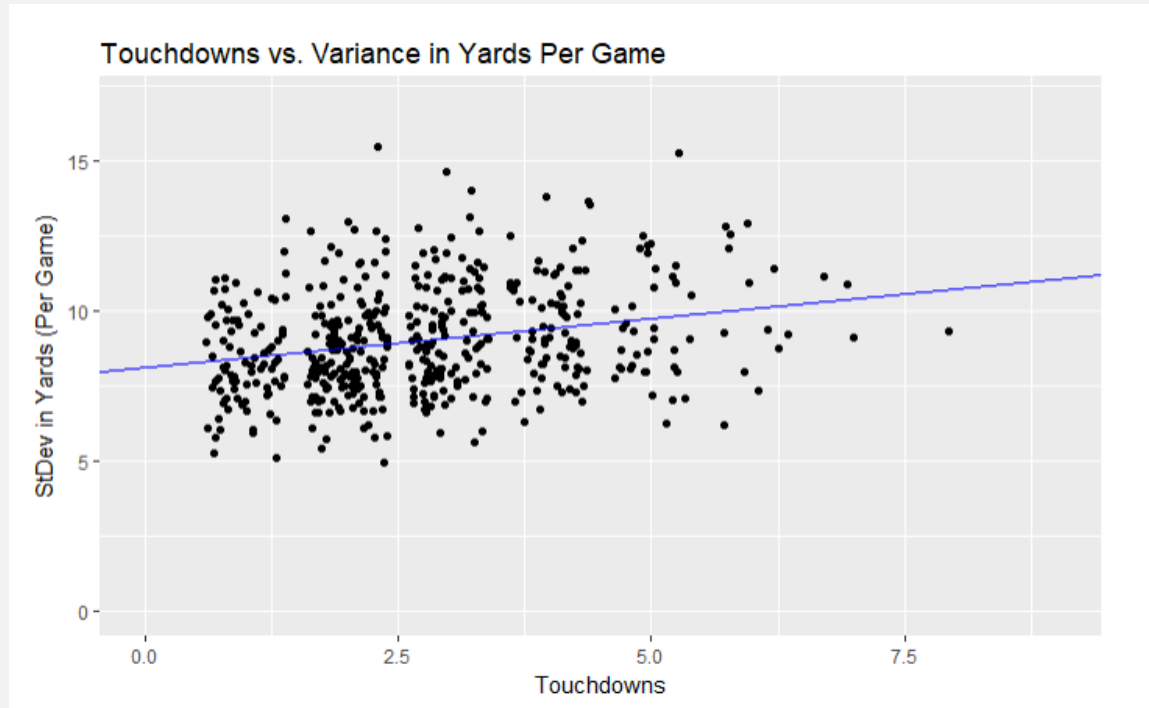
## INSIGHTS



- Upward trend
  - As yards increase, so do the number of touchdowns
- However, yards per play is not normalized



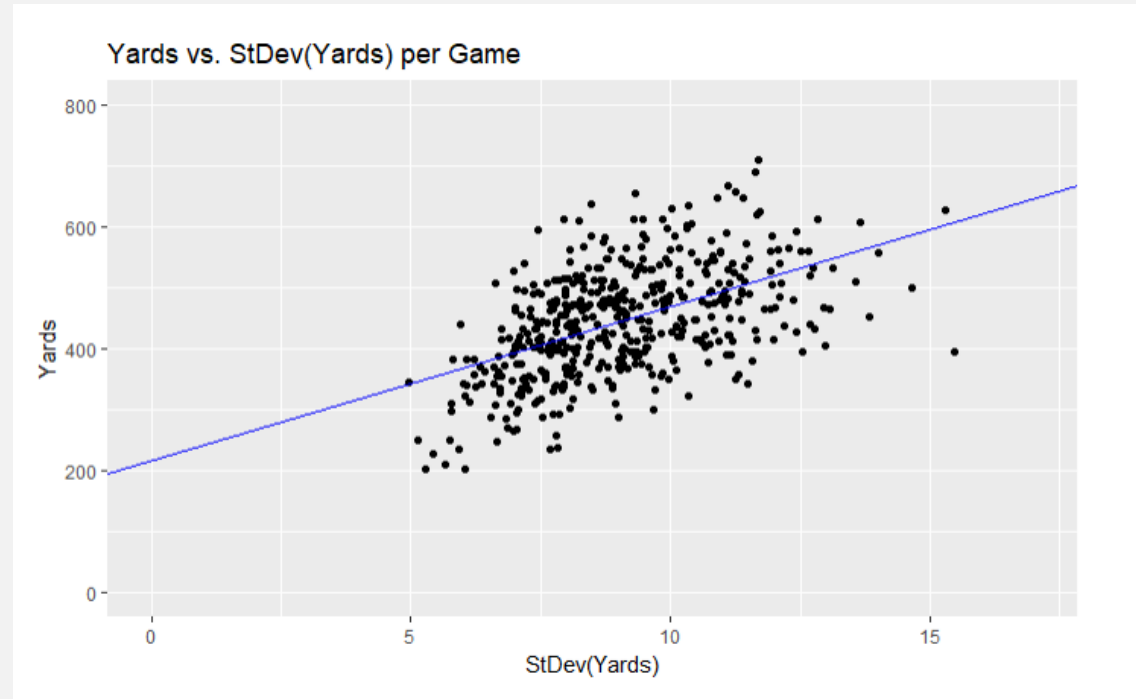
# TD'S VS. STDDEV IN YARDS



## INSIGHTS

- Relationship is relatively static, with a slight rise as standard deviations go up.
- Again, data are not normalized

# YARDS VS. STDDEV OF YARDS



## INSIGHTS

- Spread increases as yards and standard deviation increase
- You can get more yards with higher a standard deviation but run risk of not getting as many yards

# CONCLUSION

- Higher, not lower, standard deviation results in more touchdowns based on unnormalized data
- Need to remove impact of yards per game on touchdowns per game but is not feasible
- Ultimately, unclear relationship between standard deviation of yards and touchdowns
- Warrants more analysis

# LESSONS LEARNED

- Anticipated using MySQL
- Great learning experience with MongoDB and R
  - Integration through mongolite
  - Combination of MongoDB and R will be a powerful tool in future projects

# SPECIFICATIONS & REFERENCES

# COMPUTER/PROGRAM SPECIFICATIONS

## MONGODB

Bluemix Storage: 1 GB

Data Size: 14.3 MB

Database Server: Compose for  
MongoDB-jj

Database Version: 3.2.11

Database Location: US South

Cloud Hosting Service: IBM Bluemix

1 x 2.0 GHz Cores

1 GB RAM

## OS X EL CAPITAN

VERSION 10.11.6

MacBook Pro (Retina 15-inch, Mid  
2015)

Processor: 2.2 GHz Intel Core i7

Memory: 16 GB 1600 MHz DDR3

Graphics: Intel Iris Pro 1536 MB

# REFERENCES

- 2015 NFL Data from Kaggle
  - <https://www.kaggle.com/maxhorowitz/nflplaybyplay2015>
- NFL Articles
  - <https://www.theringer.com/2016/8/4/16038580/five-better-nfl-stats-teddy-bridgewater-dwight-freeney-187cb19326f1>
  - <http://www.sportingnews.com/nfl/news/191984-underrated-nfl-stats-big-plays-allowed>
- GitHub Link For Our Project
  - <https://github.com/tigerninaproject1/nfl>

## REFERENCES (CONT.)

- R and R packages
  - <https://www.R-project.org/>
  - <https://yihui.name/knitr/>
  - <https://CRAN.R-project.org/package=dplyr>
  - <http://arxiv.org/abs/1403.2805> (mongolite)
  - <https://CRAN.R-project.org/package=bindrcpp>
  - <https://CRAN.R-project.org/package=tibble>
  - <http://ggplot2.org>



Q & A