IST687

NFL Punting and Concussions

Jon Kaplan and Jeff Bertucci

Dataset Info

- **666** games observed from 2016-2017 NFL seasons
- **6,681** punt plays analyzed
- **146,529** observations of player roles during punt plays
- **66,492, 490** observations of players locations and orientation during punt return play

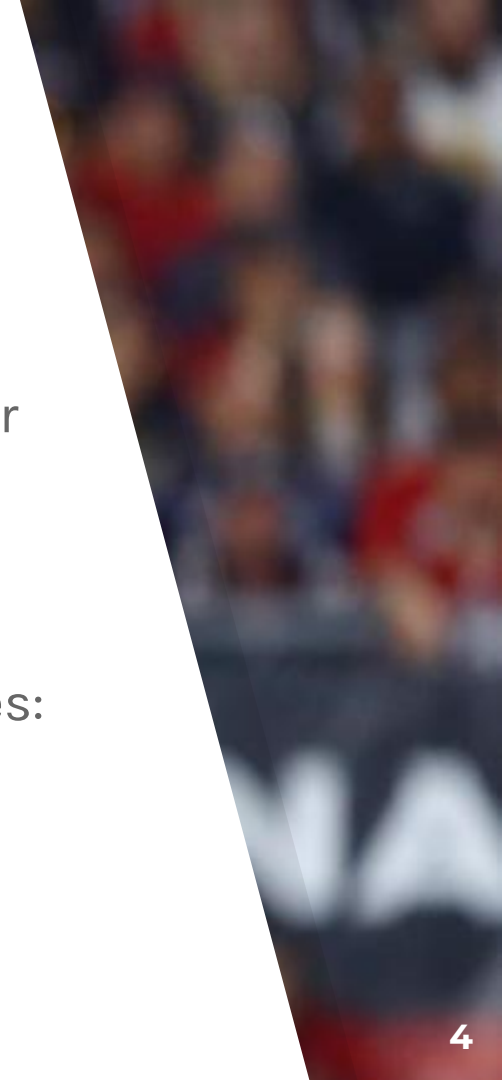
Data Questions

- What are the most common impact types resulting in concussions during punt plays?
- Is the distance from punter to punt returner a factor in more concussions?
- Does time in the game or Quarter play a factor in more concussions?
- Does certain weather or certain temperature range increase or decrease occurrences of concussions?
- Do more concussions happen on synthetic turf or natural? Outdoor or indoor stadium?
- Does game start/time of day play a factor in more concussions? Day of week?
- The more a team punts, the more concussions?
- What is the relationship of speed and direction for likelier occurrence of concussion?

Descriptive Info

	Count
Helmet-to-body	17
Helmet-to-ground	2
Helmet-to-helmet	17
Unclear	1

- **37** concussions from 2016-2017 NFL seasons :
  - **12** preseason, **25** regular season games
- Average temperature for concussion-involving games: **60.73° F**



Stadium

All 666 games:

StadiumType2 <chr>	n <int>
Indoor	126
Outdoor	463
Unknown	77

Concussion Games:

StadiumType <chr>	n <int>
NA	1
Dome	1
Indoors	2
Open	1
Outdoor	20
Outdoors	7
Outside	1
Retr. Roof - Closed	2
Retractable Roof	2

9 rows

Turf

Turf <chr>	n <int>
A-Turf Titan	1
Artificial	4
Field Turf	2
FieldTurf	4
FieldTurf 360	1
Grass	14
Natural grass	1
Natural Grass	8
UBU Speed Series-S5-M	2

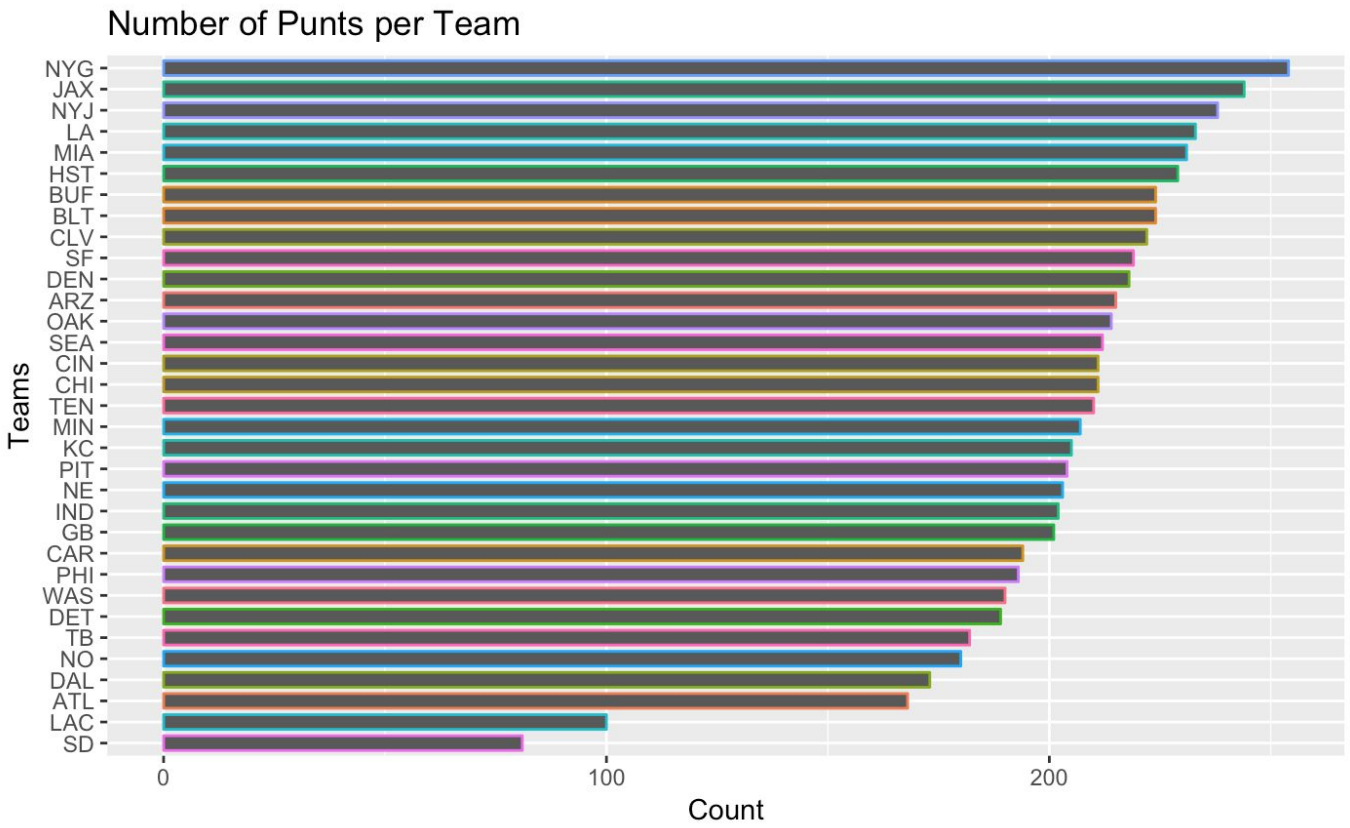
9 rows

23 concussions occured on natural grass

13 concussions occured on synthetic grass

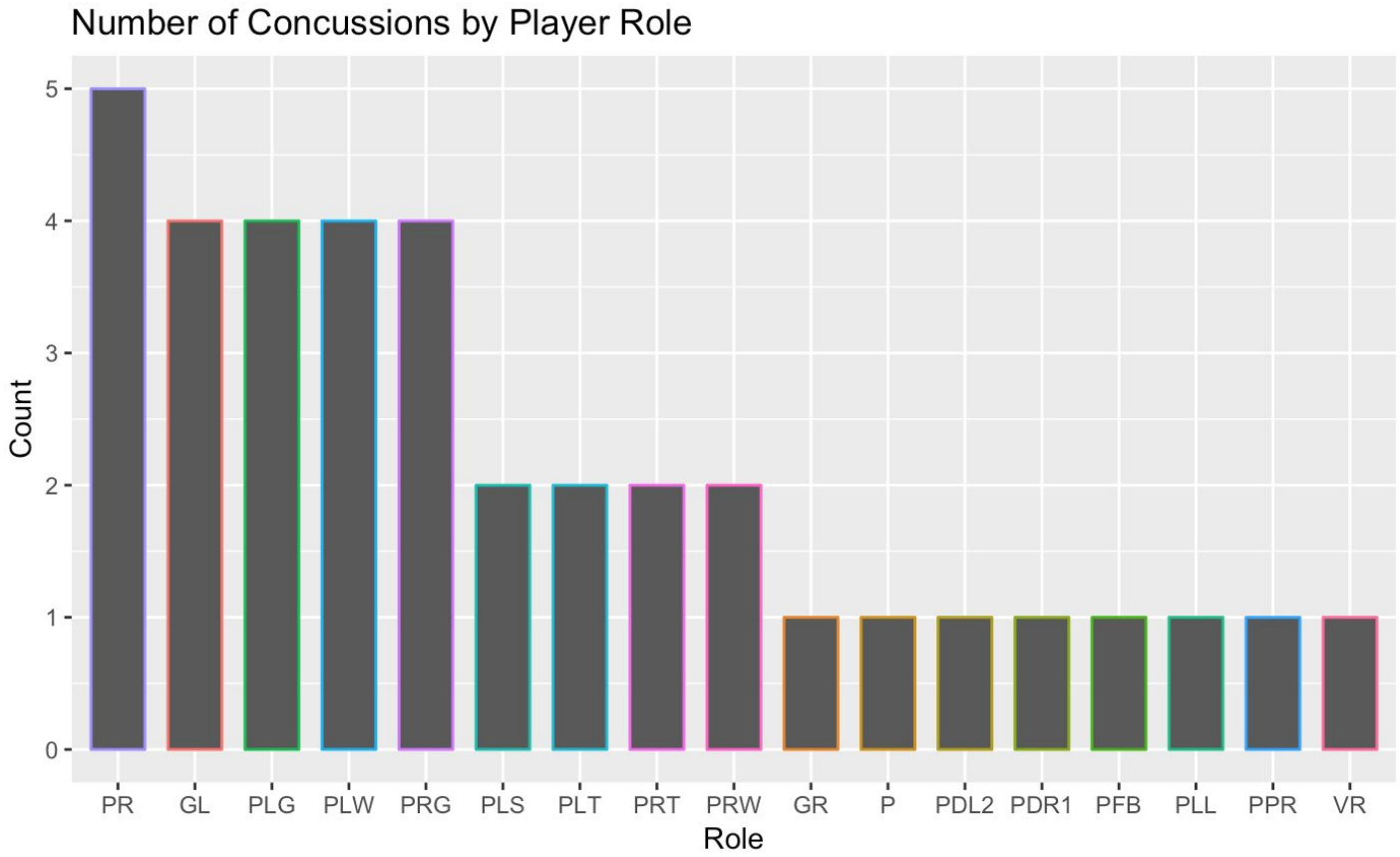
Types of Plays for Impacts

Number of Punts per Team



- Chance of concussion on punt play: **0.55% (37/6681)**

Player Roles



Regression

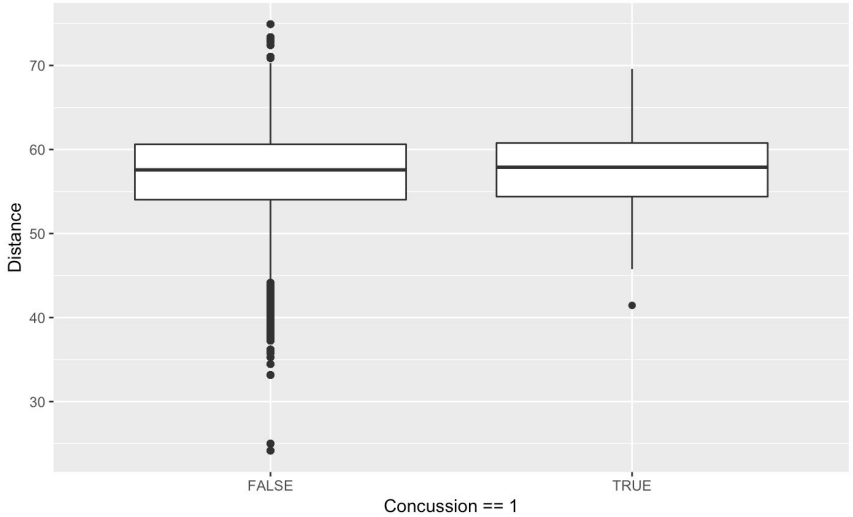
- Ran regressions on few columns in relation to # of concussions: **Game_Day, Turf, Temperature, and Quarter**
- For model 5, our Quarter regression, **Quarter 3** turned out to be significant with a value of **0.0428**

```
Call:
glm(formula = Concussion ~ Quarter, family = binomial(), data = puntmaster)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.0284  -0.0235  -0.0235  -0.0199   4.2129

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.8740     0.4472 -19.842   <2e-16 ***
Quarter2      0.6788     0.5394   1.258   0.2082
Quarter3      1.0554     0.5210   2.026   0.0428 *
Quarter4      0.3430     0.5856   0.586   0.5580
Quarter5    -10.6920   432.9405  -0.025   0.9803
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 687.01  on 146528  degrees of freedom
Residual deviance: 681.40  on 146524  degrees of freedom
AIC: 691.4

Number of Fisher Scoring iterations: 18
```

Distance between P and PR

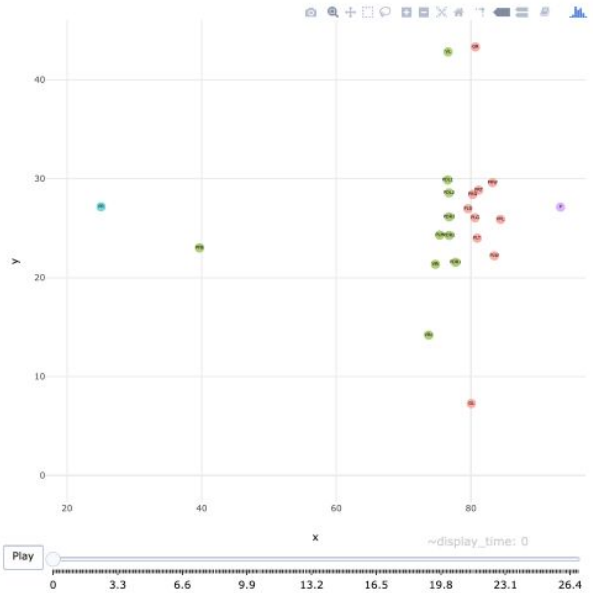
Looked at **distance** between **Punter** and **Punt Returner** on punt plays at time of punt to see if relationship for concussions

- Average distance: **57.51 yards**
- Concussion average distance: **56.62 yards**
- **Range (non-conc)**: **24.16 - 74.92 yards**
- **Range (conc)**: **41.45 - 69.59 yards**



Animated Concussion Plot

- Created animated plot to show specific concussion play based off PlayID
- Gives birds-eye view of positions and players involved



Machine Learning-RandomForest

- We ran a random forest analysis of all the punt plays from 2016-2017.
- “Unfortunately” there were only 36 concussions in both years combined
- Of those, we could only assume 6 were derivative of the punt play
- Those would be concussions involving the punter or the punting reciever

Machine Learning-RandomForest 2

- Nonetheless, we ran a random forest to see if we could glean ANYTHING

- We take into consideration: Turf, Weather, Quarter, Season Week, and the score differentials

- Results: As expected, not stellar, but it did score a 99.99% on its test

	Predicted 0	Predicted 1
Actual 0	13278	0
Actual 1	6	0

	Importance
Season_Type	0.02738847
Week	1.00023094
Game_Day	0.45209800
GameWeather	0.03522921
Temperature	1.61750888
StadiumType2	0.09042840
Quarter	0.72063474
Role	0.64749300
Score_difference	1.94196796
NaturalTurf	0.21137473

```
Call:
 randomForest(formula = Concussion ~ .,
data = puntmaster4)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 0.05%
Confusion matrix:
      0     1  class.error
0 13278     0            0
1     6     0            1
```

Improving our RandomForest

- More data, specifically more actual concussion data.
- If we were to run again maybe consider concussion per game rather than per play
- That may be more interesting since we are considering global effects more so than exact moment to moment effects.

Machine Learning-KSVM

- Sparing most of the similar details of RF
- KSVM was surprisingly worse.
- The goal: rout out the 1's in the concussion category
- The result: KSVM Model guessed the opposite of Random forest and got 0.001% correct, but it definitely did get all the 1's

```
> ML2
Support Vector Machine object of class
"ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 5

Polynomial kernel function.
 Hyperparameters : degree =  1  scale =  1
offset =  1

Number of Support Vectors : 8

Objective Function Value : -40
Training error : 0.000452
Cross validation error : 0.000452
Probability model included.
```

```
> summary(as.factor(comptable$prediction))
   1
4428
```

Lesson Learned:

Computers do exactly what you tell them to do

Resources

Animated Plot

- http://jimkloet.com/animated_nfl_play.html
- https://www.kaggle.com/c/NFL-Punt-Analytics-Competition