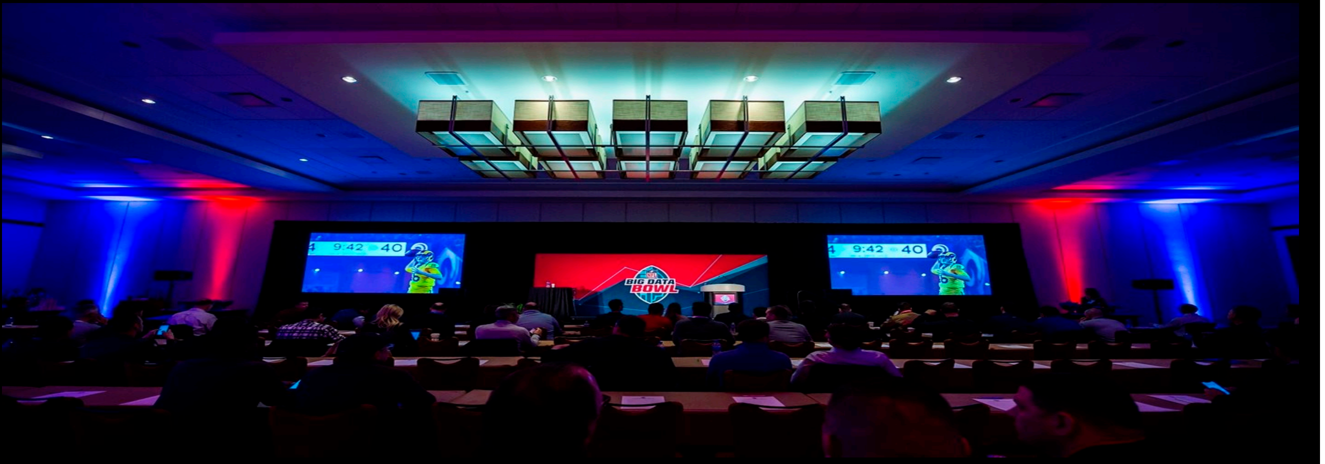
YAPR

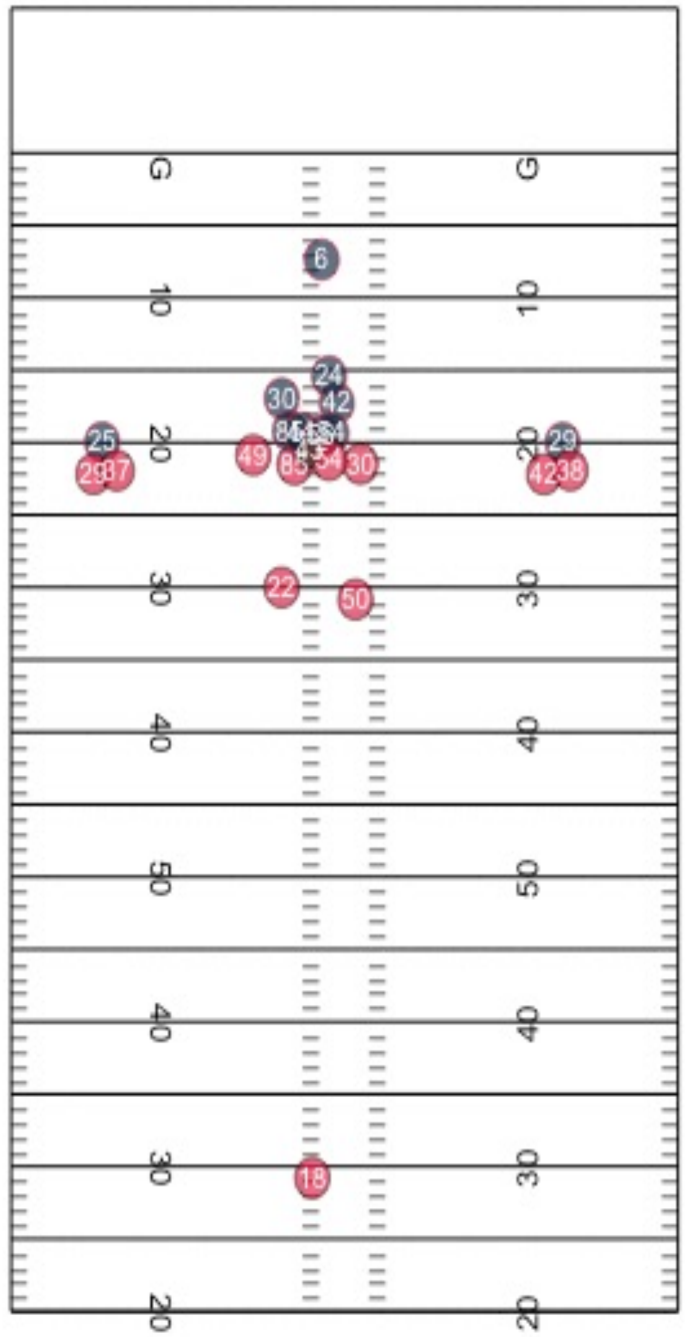
'Yards Above Predicted Return'

Nicholas Kondo, Max Batsch, Tino Diaz-Ordaz



2022 NFL Big Data Bowl

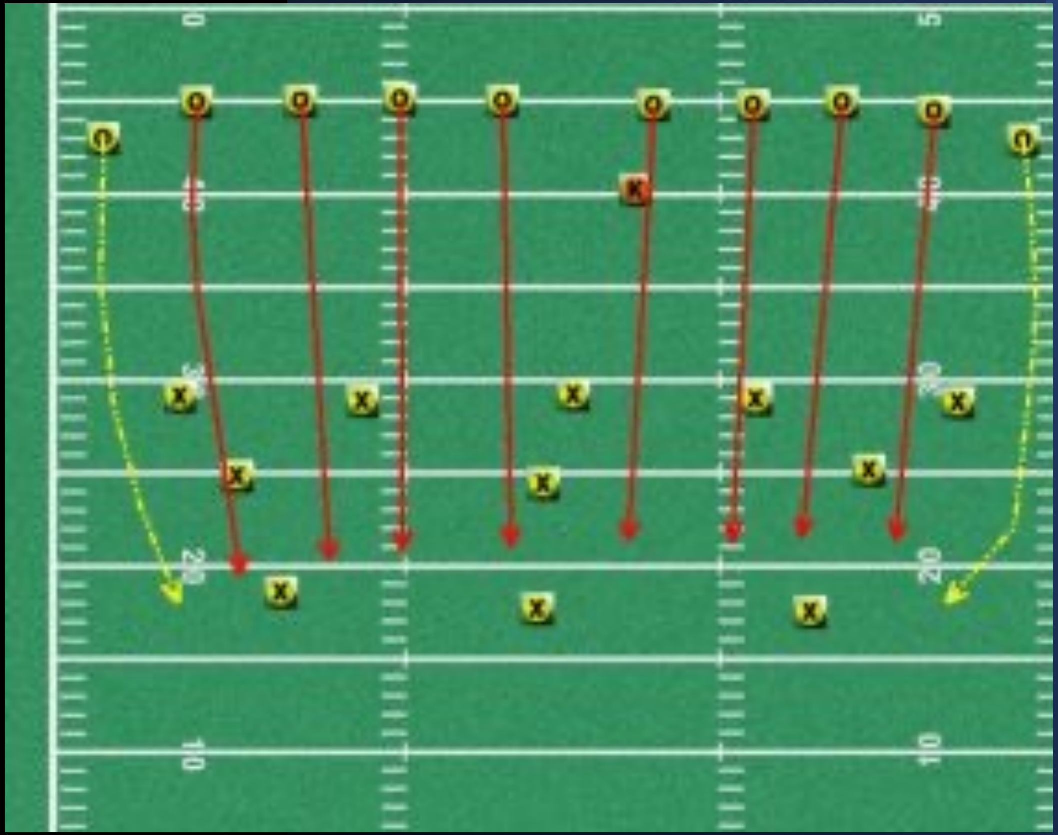
- Annual Big Data Bowl
- Innovation with player tracking data
- Hosted by Kaggle
- The goal:
    - 1. Create a new special teams metric
    - 2. Quantify special teams strategy
    - 3. Rank special teams players



Expected Return Yards

- How many yards will the returner gain at the moment they receive the ball?

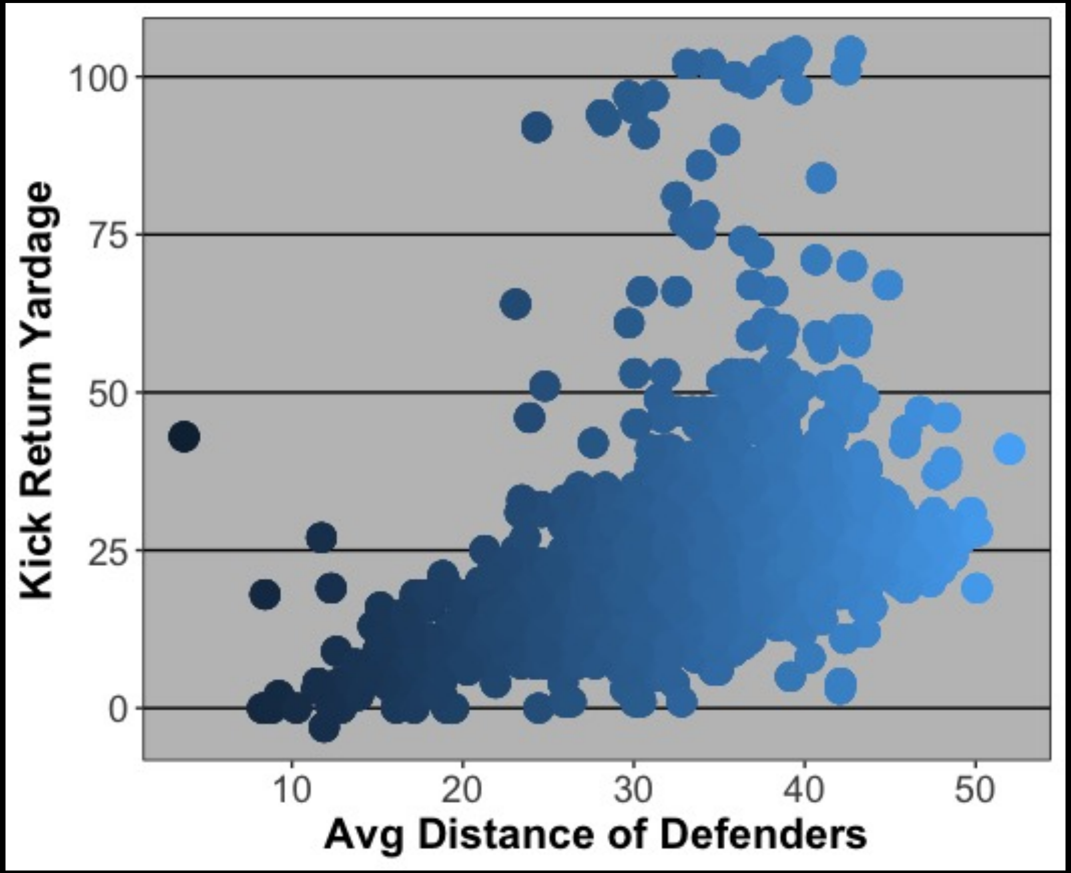
- Variables used to determine Expected Return Yards
    - The yard line the ball was kicked to
    - The location the returner received the ball
    - Distance between returner and defenders (kickoff team)
    - Speed of defenders
    - Kick Type (Flat, Deep, or Pooch)
    - Catch Type (Dropped or Caught)




YARD LINE RECEIVED: -2 yds
AVG DEFENDER DISTANCE: 33.7 yds
Y POSITION: Right
KICK TYPE: Deep
RESULT: Return
EXPECTED YARDS: 26 yds
RESULT: 24 yd line
LOADING...

Why Use Machine Learning for Player Tracking Data?

- Ranking
  - Developing an understanding of successful and unsuccessful attributes
- Payment
  - Creating a tangible metric that can be associated with an increase in pay
- Health
  - Understanding the likelihood of an injury or other statistics that can help make the game safer



Data Cleaning

- Removing NA values
  - Removed observations when football was tracked
- Filter
  - Kickoff plays only
  - Frame ID of 'Kick Received'
  - No squib or onside kicks
  - Returners only
    - Defenders become a separate feature

```{r}
sort(colSums(is.na(tracking2018)))
```

```
          time             x             y             s
             0             0             0             0
             a           dis         event   displayName
             0             0             0             0
          team       frameId        gameId        playId
             0             0             0             0
 playDirection             o           dir         nflId
             0        555537        555537        555537
  jerseyNumber      position
        555537        555537
```

```{r}
table(tracking2018[is.na(tracking2018$nflId), "displayName"])
```

```
football
  555537
```

Data Cleaning (Continued)

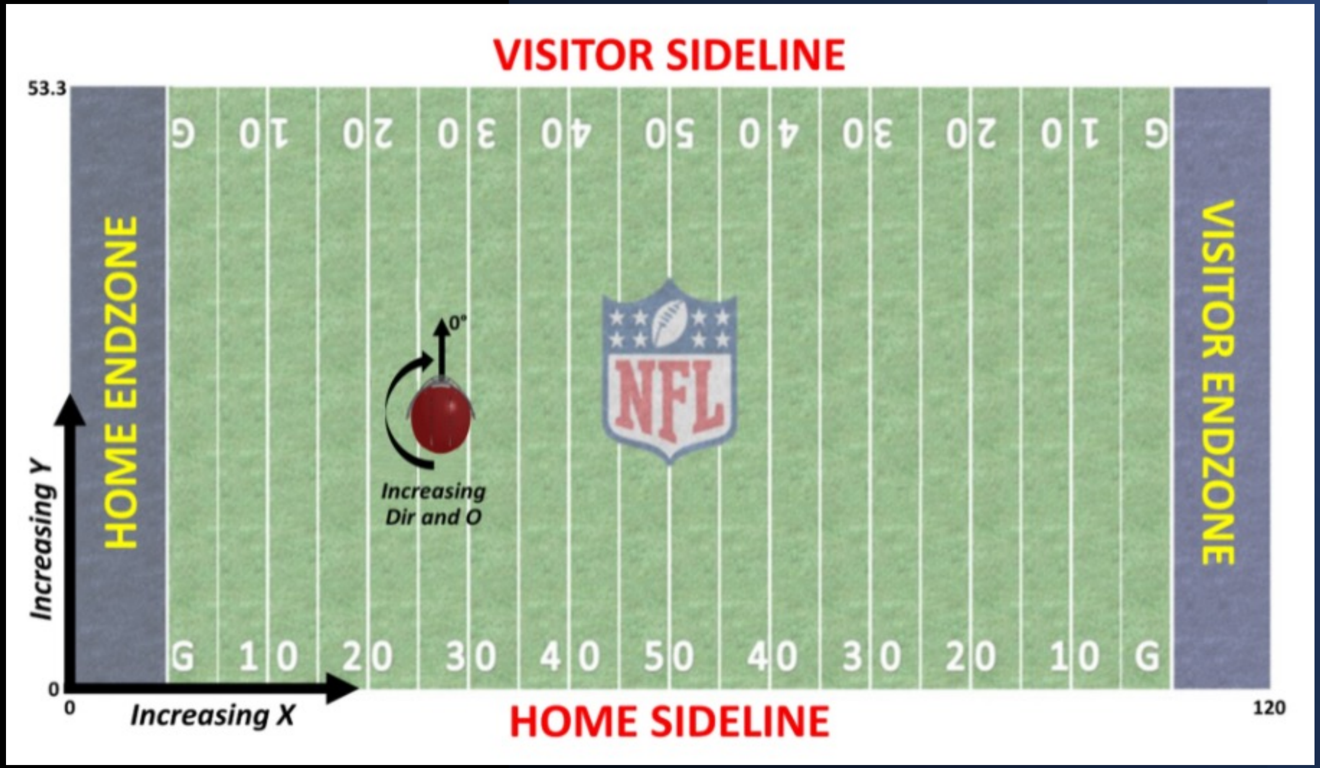
Display Name	X	Y	Team	Returner?
Player 1	39.08	24.57	Home	No
Player 2	41.11	32.25	Home	No
Player 3	43.08	38.77	Home	No
Player 4	2.5	30.5	Home	Yes
Player 5	44.11	24.92	Home	No
Player 6	43.9	21.75	Home	No
Player 7	39.07	41.27	Home	No
Player 8	45.07	49.79	Home	No
Player 9	32.17	38.71	Home	No
Player 10	50.5	46.24	Home	No
Player 11	21.67	43.81	Home	No
Defender 1	34.23	31.98	Away	No
Defender 2	50.44	17.86	Away	No
Defender 3	43.95	27.63	Away	No
Defender 4	39.83	44.49	Away	No
Defender 5	40.75	33.99	Away	No
Defender 6	11.53	42.45	Away	No
Defender 7	44.18	42.04	Away	No
Defender 8	37.43	29.16	Away	No
Defender 9	20.92	38.29	Away	No
Defender 10	33.04	34.62	Away	No
Defender 11	33.75	35.52	Away	No

Expanding the data so we go from 1 observation for each player, we have 11 observations for each player to include defenders.

Display Name	X	Y	Team	Returner?	Defender	X	Y
Player 4	2.5	30.5	Home	Yes	Defender 1	41.3	10.2
Player 4	2.5	30.5	Home	Yes	Defender 2	41.5	15.3
Player 4	2.5	30.5	Home	Yes	Defender 3	40.6	20.4
Player 4	2.5	30.5	Home	Yes	Defender 4	41.5	22.1
Player 4	2.5	30.5	Home	Yes	Defender 5	41.5	23.4
Player 4	2.5	30.5	Home	Yes	Defender 6	43.2	25.2
Player 4	2.5	30.5	Home	Yes	Defender 7	40.2	35.5
Player 4	2.5	30.5	Home	Yes	Defender 8	42.5	39.3
Player 4	2.5	30.5	Home	Yes	Defender 9	40.6	45.6
Player 4	2.5	30.5	Home	Yes	Defender 10	41.1	50.3
Player 4	2.5	30.5	Home	Yes	Defender 11	43.8	58.2

Feature Engineering

- Distance (Pythagorean Theorem Used)
- Yard Result
  - YardResult_Over25 (1 = yes)
- Yard line 'Kicked To'
  - 5 Yard Bin 'Kicked To'
- Y_Position (Left, Right, or Center)
- Defender Variables
  - Average Distance
  - Average Speed
  - Average Acceleration



Modeling Stages

- Linear Regression
- Logistic Regression for 25 Yard line
- ElasticNet
- Random Forest

Linear Regression

Adjusted R^2= .30

- As the yard line kicked to decreases by 1 yard, or the further the ball is kicked, the returner typically gains .28 more yards
- As the average distance between the defenders and the returners increases by 1, the returner typically gains an average of .43 more yards.

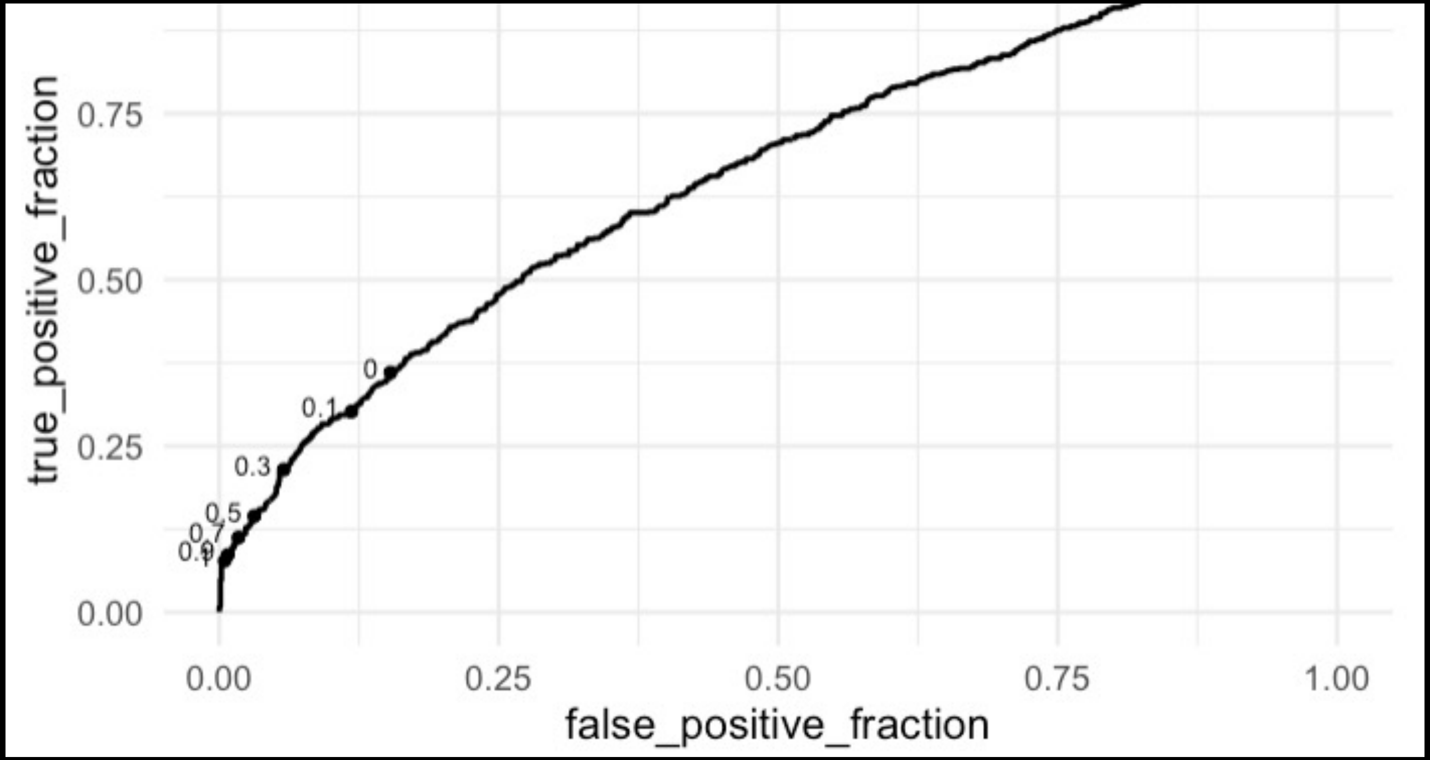
	kick Return Yardage		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-4.09	-8.53 – 0.35	0.071
kickedto yardline	-0.28	-0.36 – -0.19	**<0.001**
KoTeamAvgDist	0.43	0.35 – 0.52	**<0.001**
KoTeamAvgA	1.09	0.28 – 1.89	**0.008**
Y position [left]	-0.82	-1.53 – -0.12	**0.022**
Y position [right]	-0.98	-1.69 – -0.27	**0.007**
kickType [F]	1.00	-0.27 – 2.26	0.123
kickType [P]	-2.62	-3.76 – -1.49	**<0.001**
specialTeamsResult [Return]	10.15	8.22 – 12.07	**<0.001**
Observations	2576		
R^2 / R^2 adjusted	0.299 / 0.296		

Modeling Performance

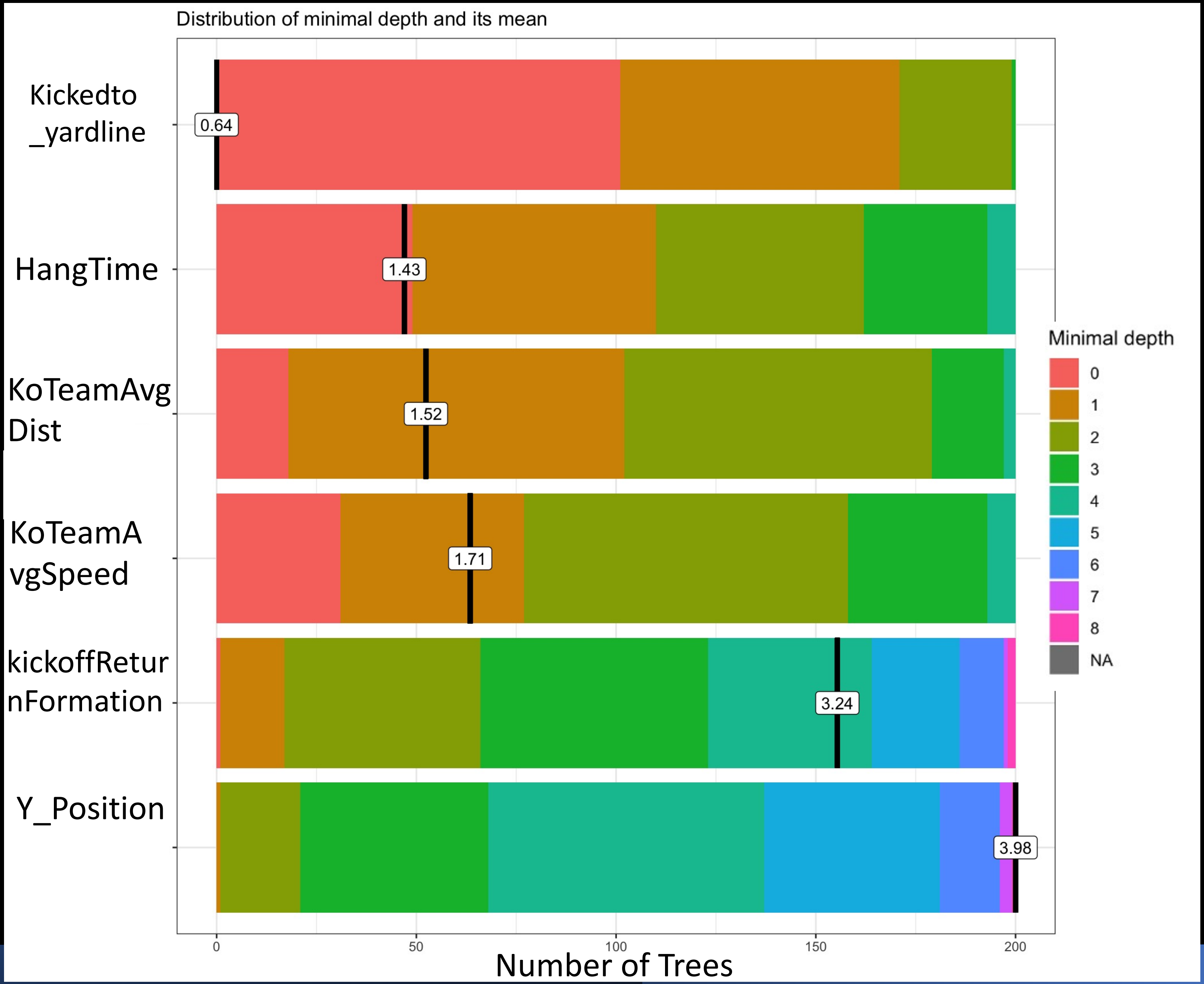
ROC Plot

Logistic Regression

- The AUC for the training and testing sets were .65 and .64 respectively

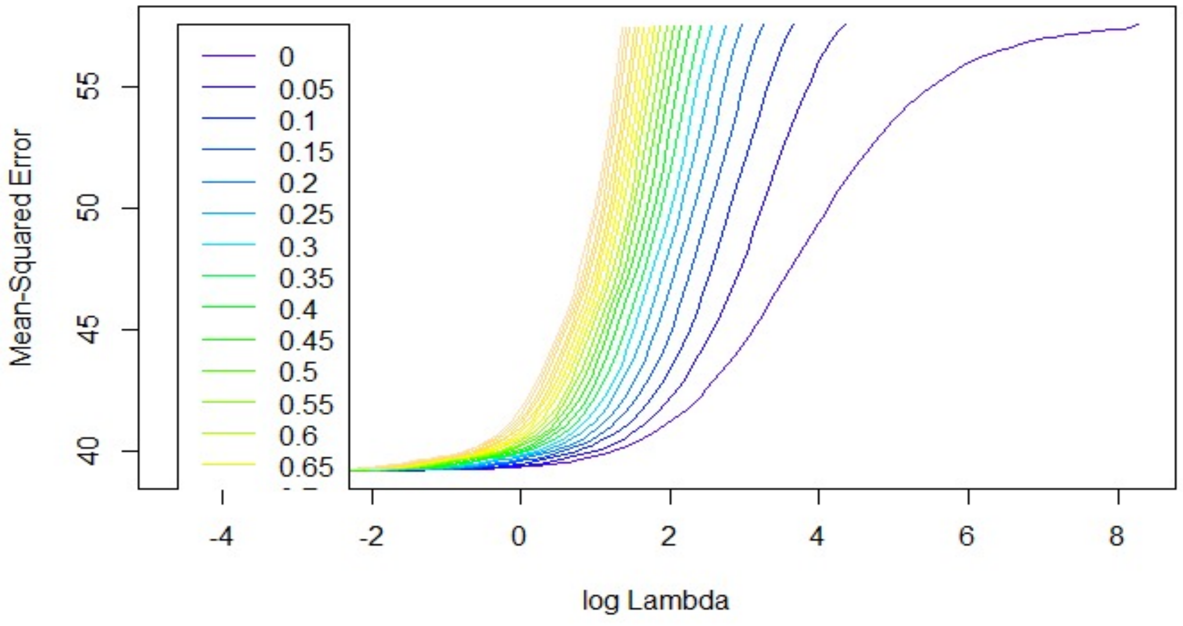


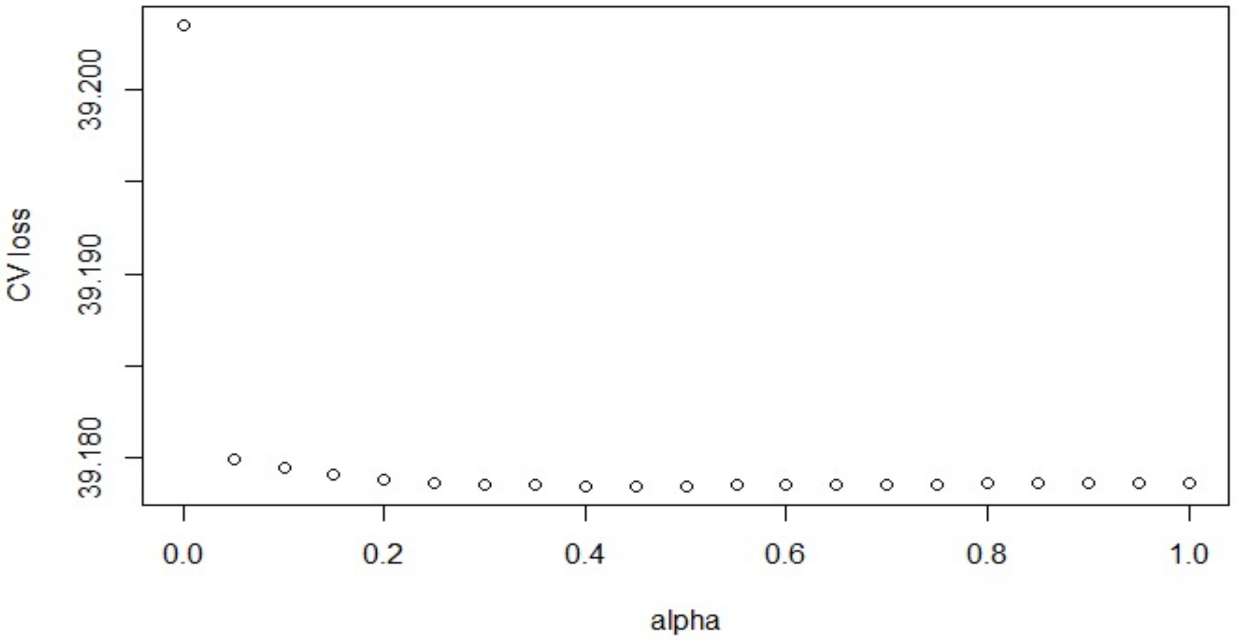
Random Forest



ElasticNet

Best alpha	Lambda.min	Lambda.1se	Error
0.45	0.039	1.10	39.18





How This Can Be Adopted?

1. Metrics from 'Expected Return Yards' Model

Expected Return Yards: Yards a returner is expected to gain given X's

Efficiency: (Actual Return Yards – Expected Return Yards)

25 Yard Line Probability: Likelihood of a returner reaching 25-yard line

2. Quantify Special Teams Strategy

- A player can have an idea of when they should or should not return the ball out of the endzone at the moment of catching the ball

3. Ranking Players

- Efficiency can be measured and used to rank special teams players

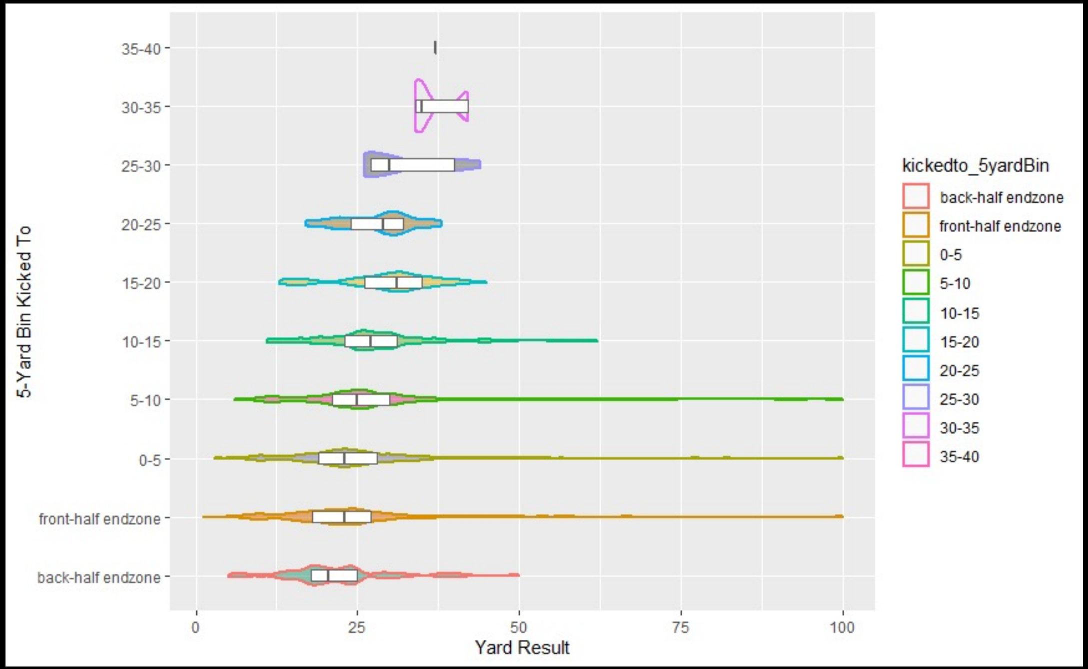
Thank You

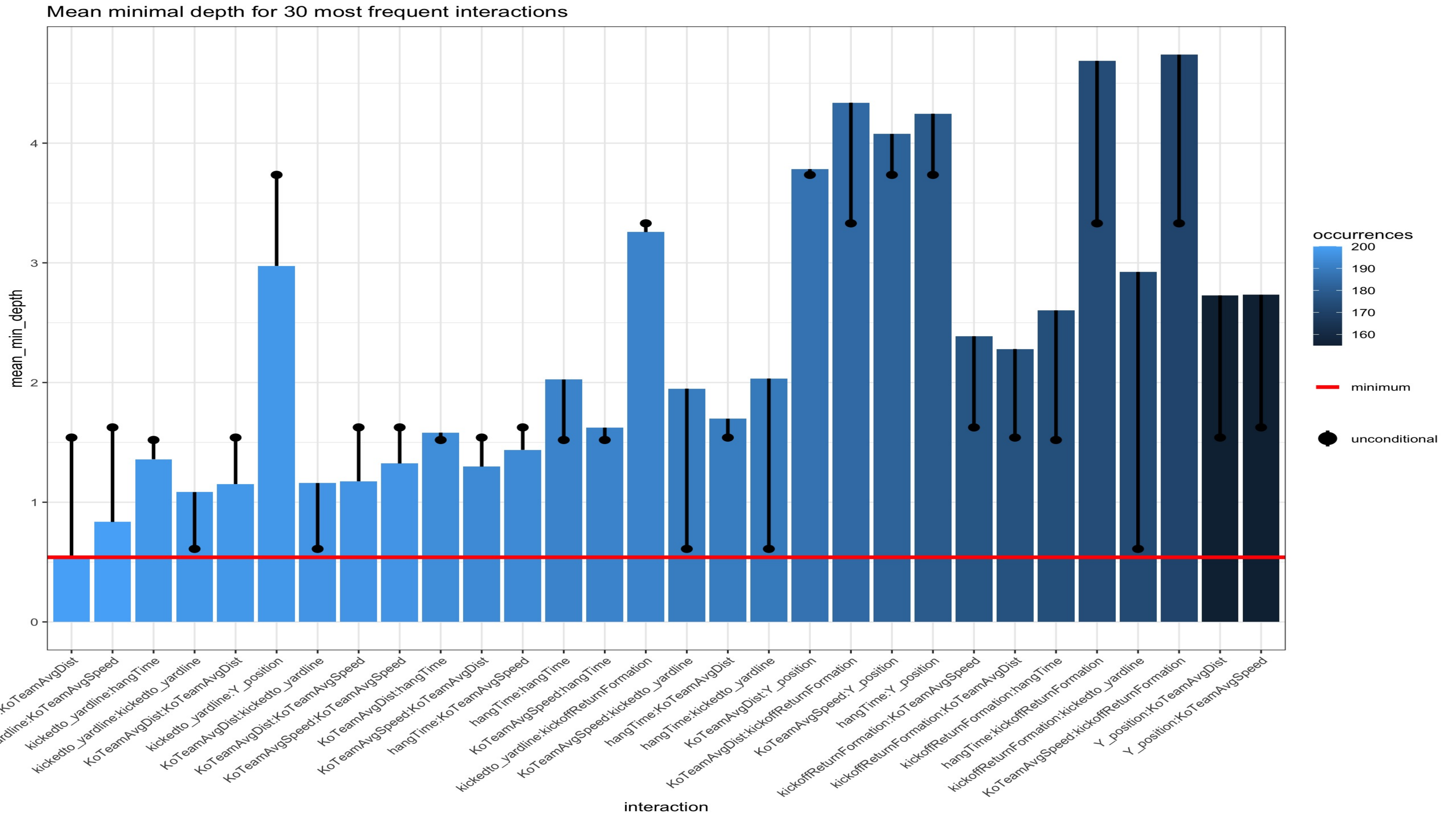
- Any Questions?

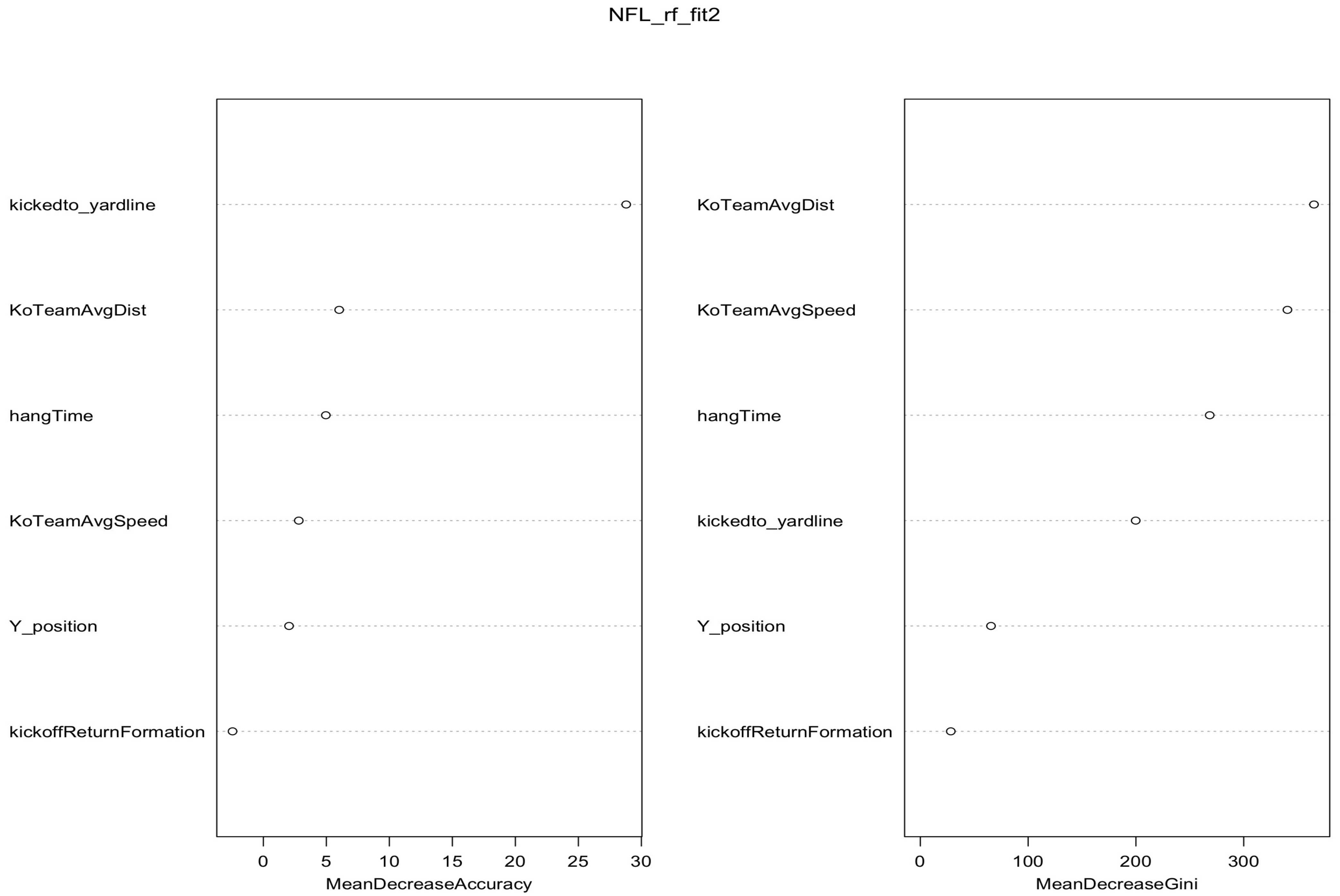
Appendix

Average Yard Result based on Kick_to bins

````
```{r}
#Here I am making a shortcut for the ggplot of the returns plot with Bin and yard result

g <- ggplot(returns, aes(x = kickedto_5yardBin, y = yard_result, color = kickedto_5yardBin)) +
     labs(x = '5-Yard Bin Kicked To', y = 'Yard Result')

#Since Boxplots are boring, I made a violin plot instead
g + geom_violin(aes(fill = kickedto_5yardBin), size = 1, alpha = .5) +
  geom_boxplot(outlier.alpha = 0, coef = 0,
  color = "gray40", width = .2) +
  scale_fill_brewer(palette = "Dark2", guide = "none") +
  coord_flip()
```
````

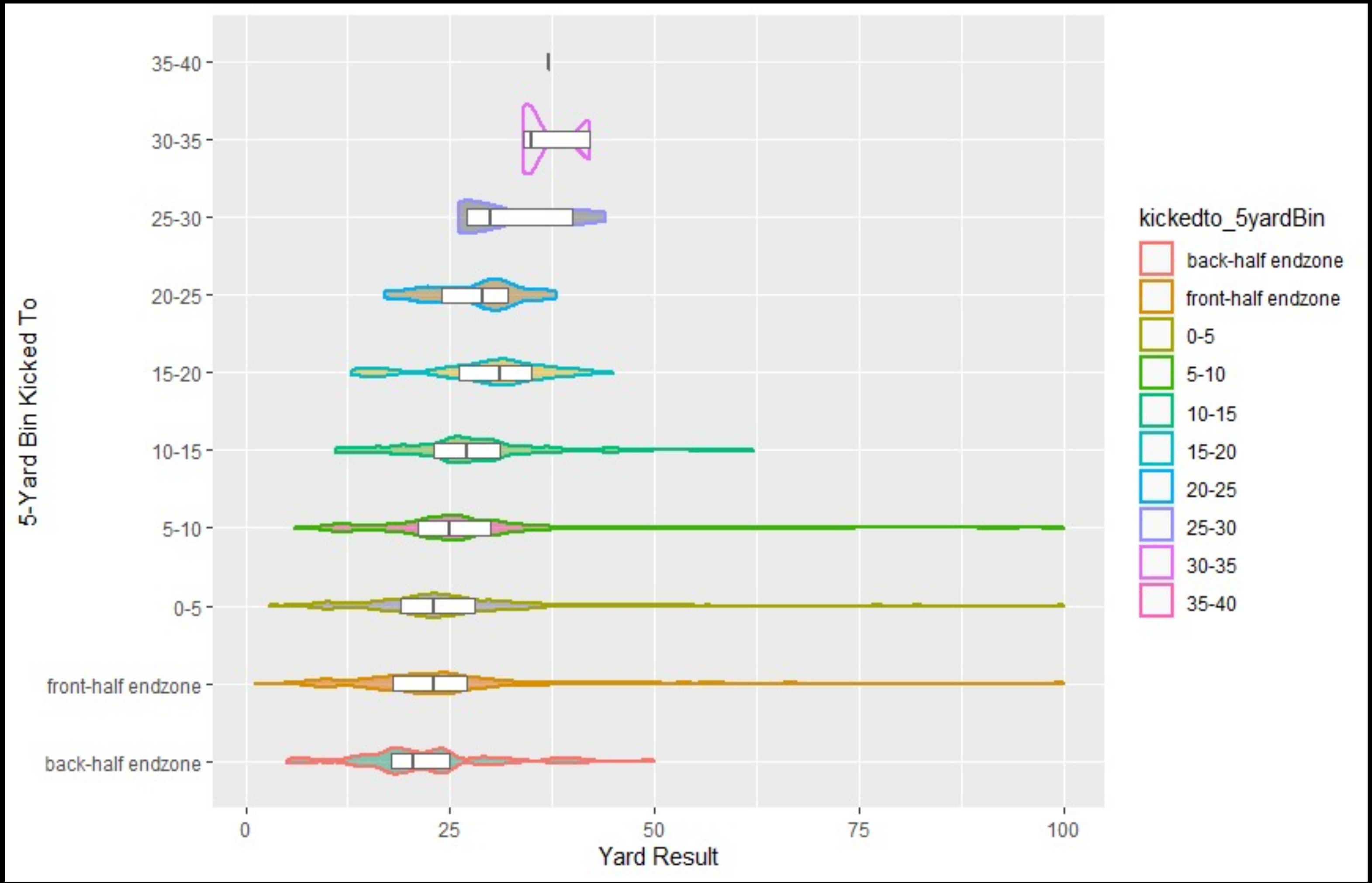


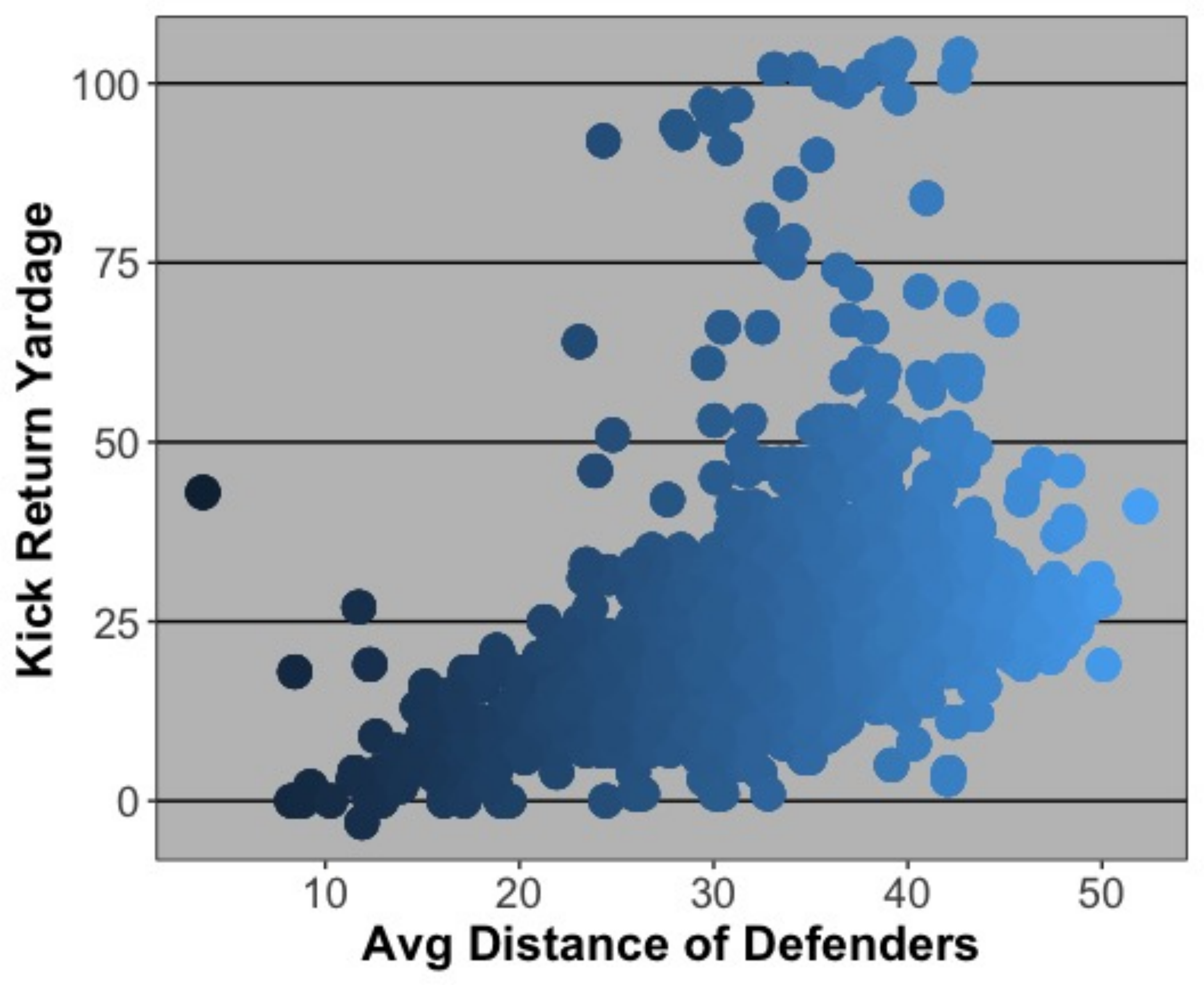

Mean minimal depth for 30 most frequent interactions
mean_min_depth
0
1
2
3
4
interaction
kickedto_yardline:hangTime
kickedto_yardline:kickedto_yardline
KoTeamAvgDist:KoTeamAvgDist
kickedto_yardline:Y_position
KoTeamAvgDist:kickedto_yardline
KoTeamAvgDist:KoTeamAvgSpeed
KoTeamAvgSpeed:KoTeamAvgSpeed
KoTeamAvgDist:hangTime
KoTeamAvgSpeed:KoTeamAvgDist
hangTime:KoTeamAvgSpeed
hangTime:hangTime
KoTeamAvgSpeed:hangTime
kickedto_yardline:kickoffReturnFormation
KoTeamAvgSpeed:kickedto_yardline
hangTime:KoTeamAvgDist
hangTime:kickedto_yardline
KoTeamAvgDist:Y_position
KoTeamAvgDist:kickoffReturnFormation
KoTeamAvgSpeed:Y_position
hangTime:Y_position
kickoffReturnFormation:KoTeamAvgSpeed
kickoffReturnFormation:KoTeamAvgDist
kickoffReturnFormation:hangTime
hangTime:kickoffReturnFormation
kickoffReturnFormation:kickedto_yardline
KoTeamAvgSpeed:kickoffReturnFormation
Y_position:KoTeamAvgDist
Y_position:KoTeamAvgSpeed
occurrences
200
190
180
170
160
minimum
unconditional


NFL_rf_fit2
kickedto_yardline
KoTeamAvgDist
hangTime
KoTeamAvgSpeed
Y_position
kickoffReturnFormation
0 5 10 15 20 25 30
MeanDecreaseAccuracy
KoTeamAvgDist
KoTeamAvgSpeed
hangTime
kickedto_yardline
Y_position
kickoffReturnFormation
0 100 200 300
MeanDecreaseGini

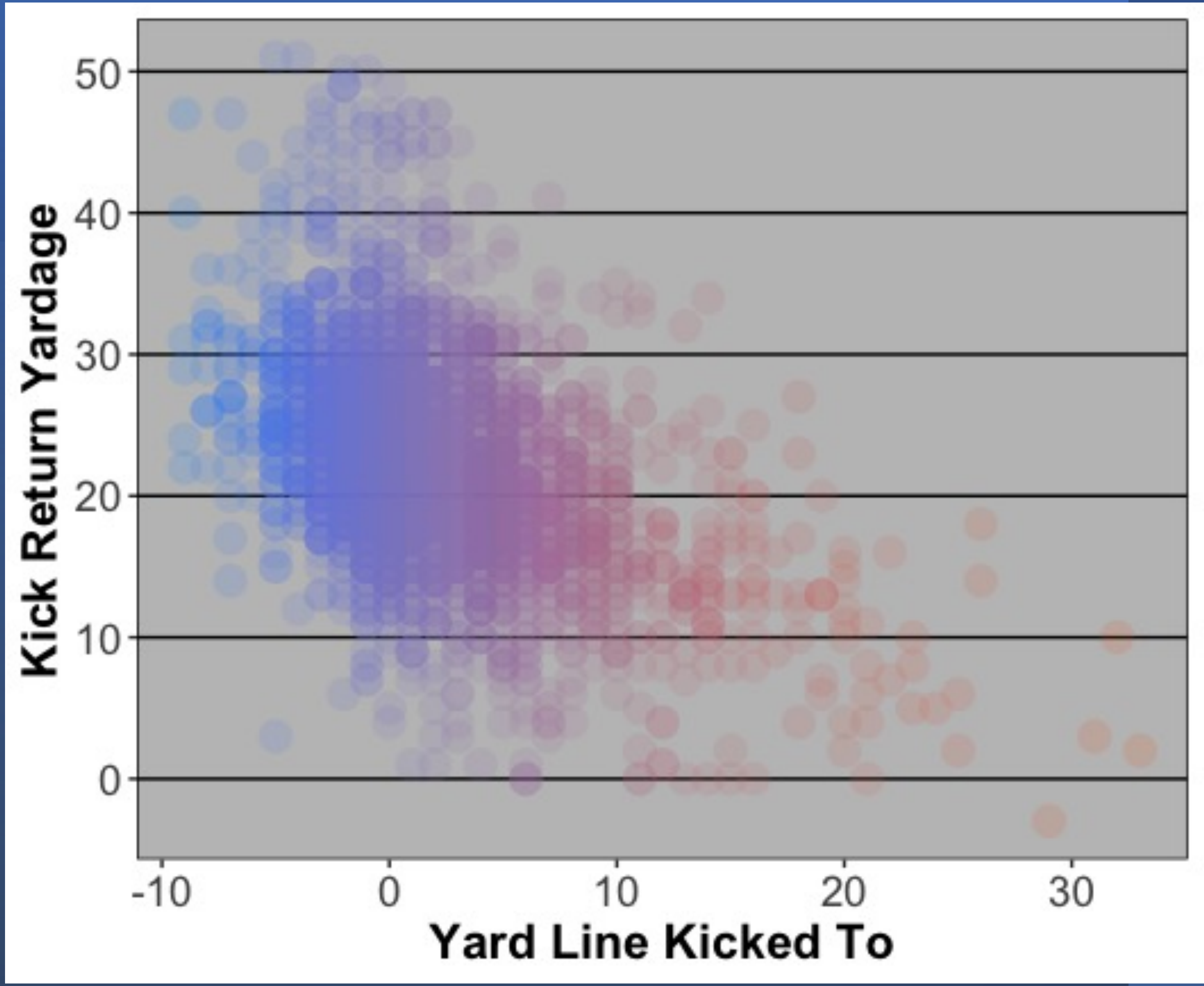
Average Yard Result based on Yard-line Kick to bins

````
```{r}
#Here I am making a shortcut for the ggplot of the returns plot with Bin and yard result

g <- ggplot(returns, aes(x = kickedto_5yardBin, y = yard_result, color = kickedto_5yardBin)) +
    labs(x = '5-Yard Bin Kicked To', y = 'Yard Result')

#Since Boxplots are boring, I made a violin plot instead
g + geom_violin(aes(fill = kickedto_5yardBin), size = 1, alpha = .5) +
  geom_boxplot(outlier.alpha = 0, coef = 0,
  color = "gray40", width = .2) +
  scale_fill_brewer(palette = "Dark2", guide = "none") +
  coord_flip()
```
````

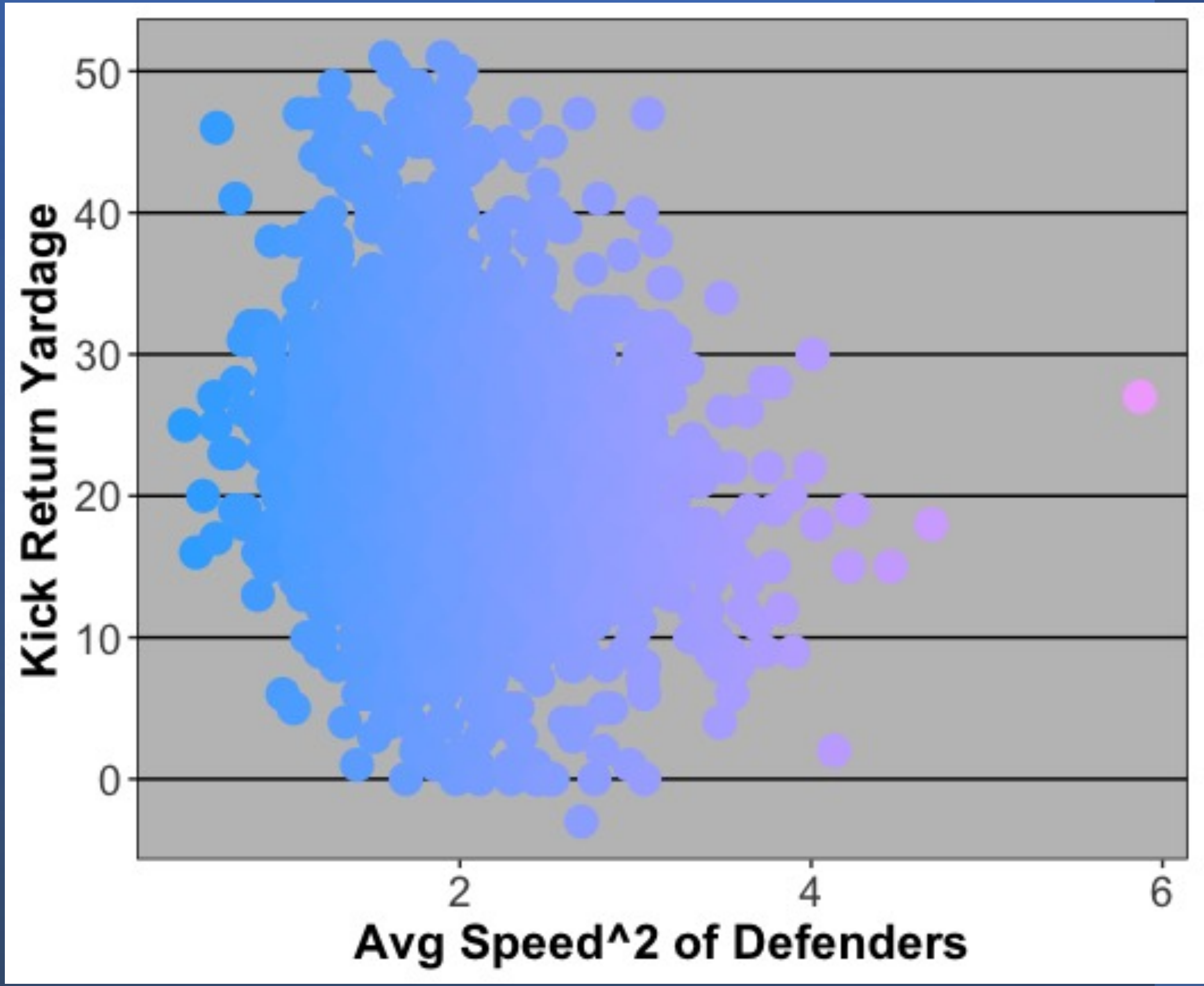




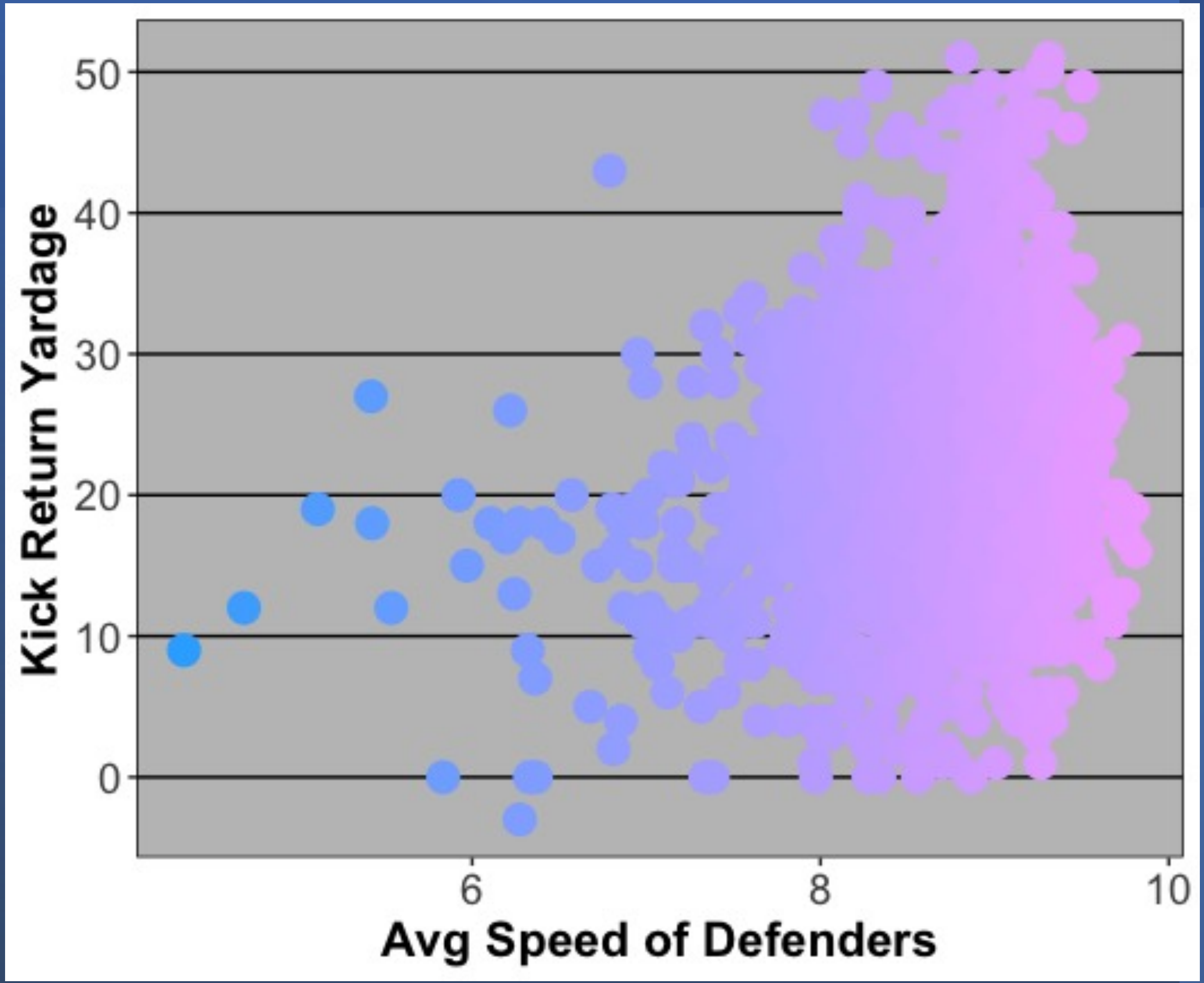
Kick Return Yardage
100
75
50
25
0
10
20
30
40
50
Avg Distance of Defenders



Kick Return Yardage
50
40
30
20
10
0
-10
0
10
20
30
Yard Line Kicked To

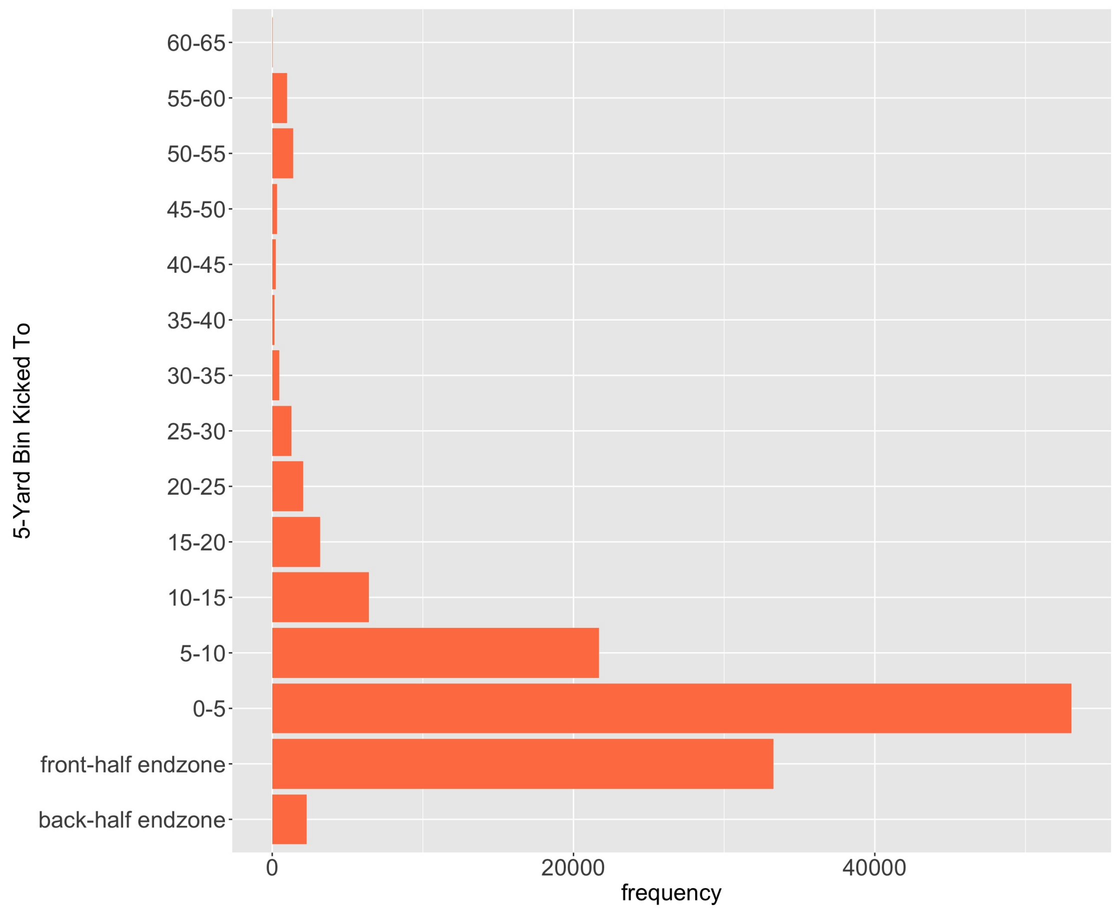


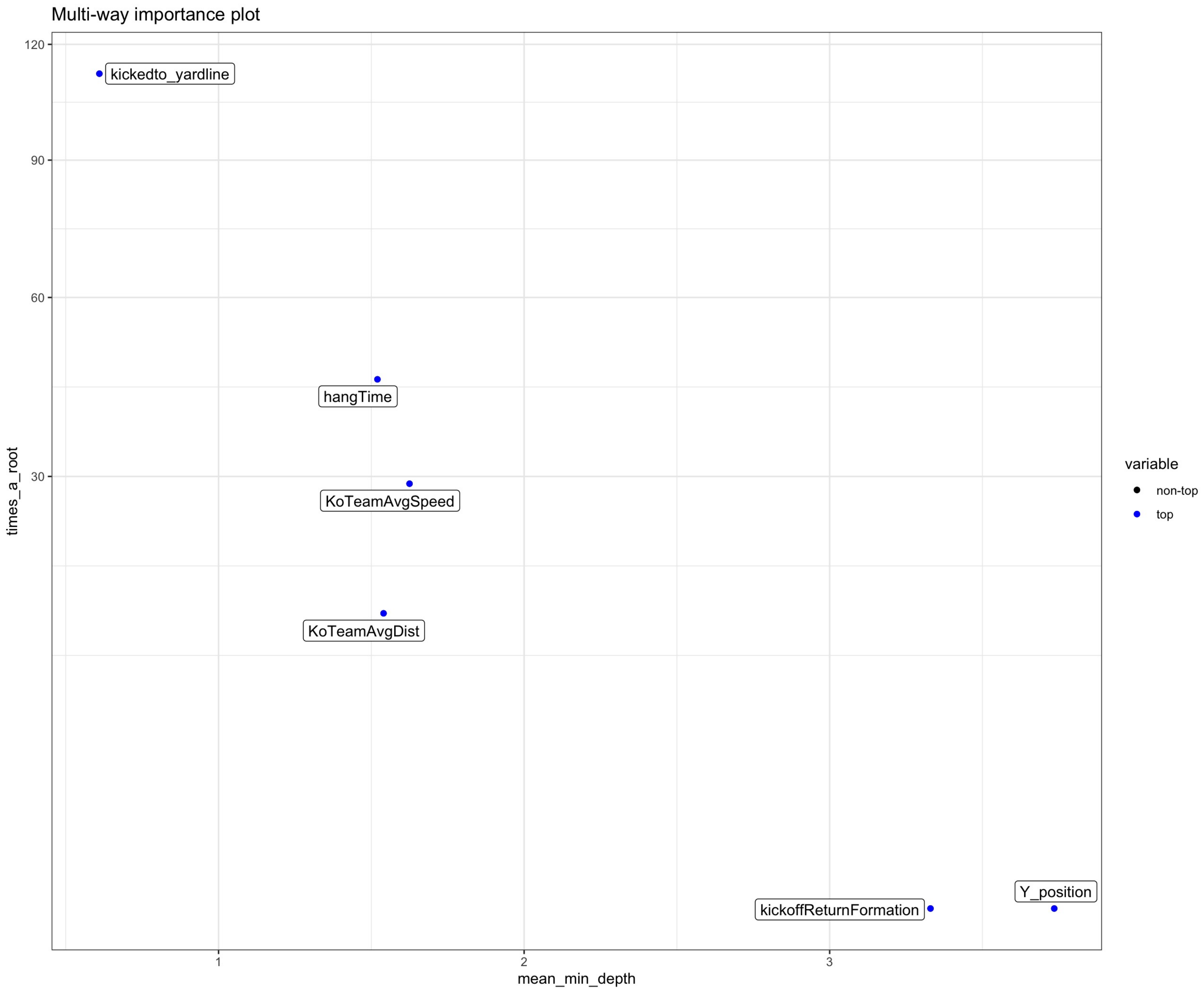
Kick Return Yardage
50
40
30
20
10
0
2
4
6
Avg Speed^2 of Defenders



Kick Return Yardage
50
40
30
20
10
0
6
8
10
Avg Speed of Defenders

```{r}

ggplot(kickoff_returns, aes(x = kickedto_5yardBin)) +
  geom_histogram(stat = "count") +
  coord_flip() +
  xlab('5-Yard Bin Kicked To') +
  ylab('frequency') +
  theme(axis.text=element_text(size=18),
        axis.title = element_text(size=18))

```




Multi-way importance plot
120
90
60
30
times_a_root
kickedto_yardline
hangTime
KoTeamAvgSpeed
KoTeamAvgDist
kickoffReturnFormation
Y_position
1
2
3
mean_min_depth
variable
non-top
top