NFL Rest Days Effects on Injuries

Malcolm Gaynor & Will Gibbs

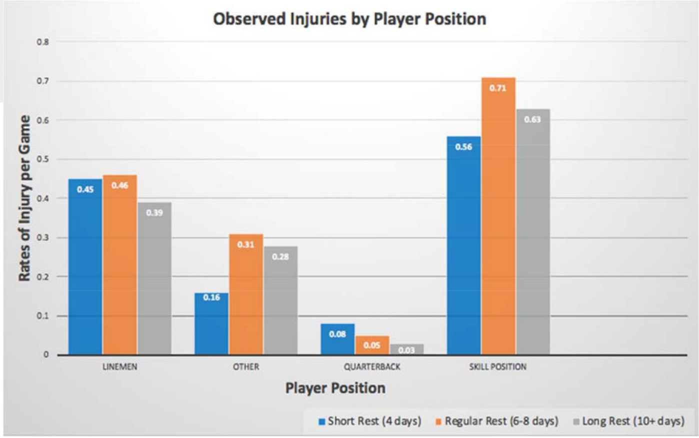
Motivation

- Patrick Mahomes, Christian McCaffrey, Devin McCourty, and Richard Sherman have all criticized Thursday Night Football Games

- There are only 4 days between Sunday and Thursday games

- However, the literature finds that Thursday Night Football games have LESS injuries:
  - National Institute of Health
    - https://pubmed.ncbi.nlm.nih.gov/30848976/#:~:text=Results%3A%20The%20all%2Dcause%20injury,6%2C072%20per%201%2C000%20athletic%20exposures.
  - Orthopaedic Journal of Sports Medicine
    - https://journals.sagepub.com/doi/abs/10.1177/2325967119S00341

Problems with the Literature



- Both research articles use a categorical independent variables
    - Days of rest: short (4 days) vs. regular (6-8 days) vs. long (10+ days)
    - Night of game: Thursday vs. Sunday vs. Monday
- However, days of rest can be a QUANTITATIVE variable
- Also, this research does not consider other independent variables that could impact the data

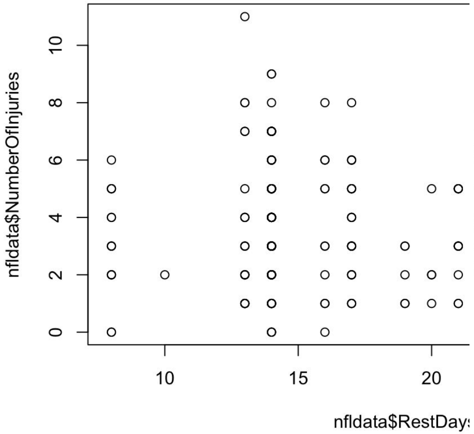
Our data

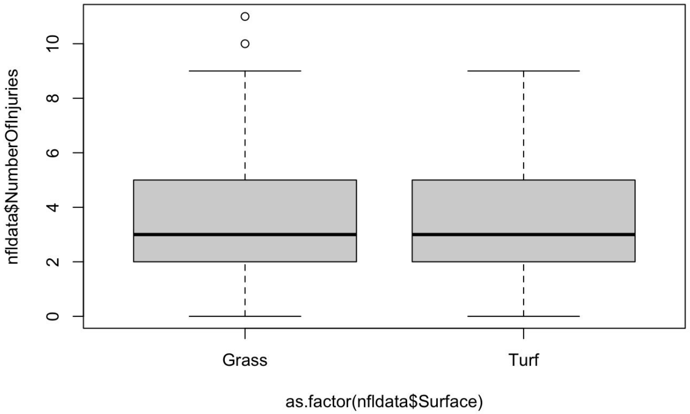
(5:43) (SHOTGUN) 16-T.LAWRENCE SACKED AT CIN 30 FOR -7 YARDS (91-T.HENDRICKSON). JAX-16-T.LAWRENCE WAS INJURED DURING THE PLAY. HIS RETURN IS QUESTIONABLE.

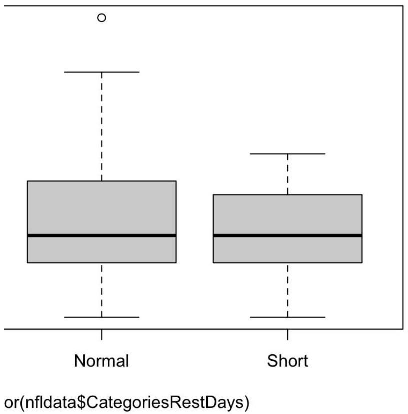
- NFL 2023 play-by-play data weeks 2-14
- Injuries counted when they occur in the play description
    - Flawed data…
- Other independent variables include Week, Rest days, Rush yards, Pass yards, Total yards, Points scored, percent of yards that are rush yards, surface (turf or grass)

Exploratory data analysis

- Initially, rest days (and all independent variables) do not appear significant







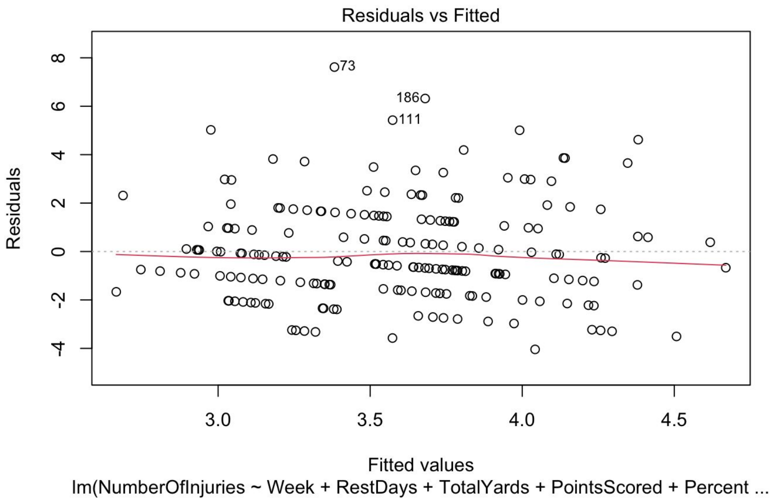
Linear Regression (full model)

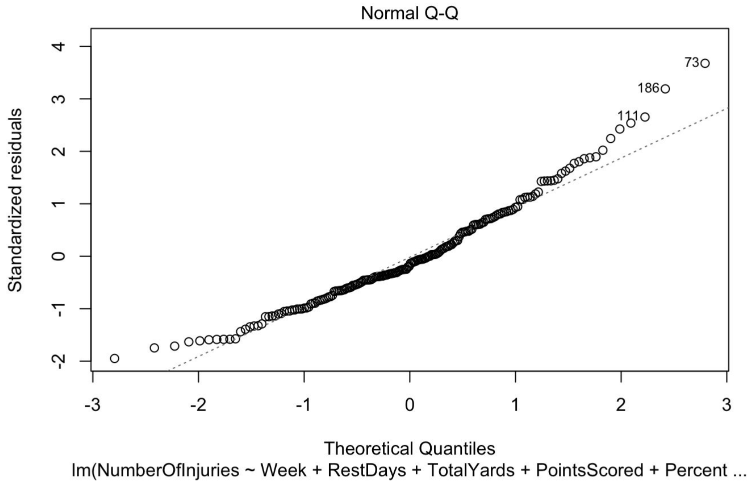
- One model treated rest days quantitatively, another treated rest days categorically.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.5402254  1.4847882   1.037    0.301
Week                    0.0672145  0.0407323   1.650    0.101
RestDays                0.0064948  0.0418274   0.155    0.877
TotalYards              0.0003422  0.0019518   0.175    0.861
PointsScored            0.0163917  0.0159339   1.029    0.305
PercentYardsRush        1.9044577  1.8899625   1.008    0.315
as.factor(Surface)Turf -0.3164677  0.3092772  -1.023    0.308

Residual standard error: 2.096 on 185 degrees of freedom
Multiple R-squared:  0.04026,	Adjusted R-squared:  0.009132
F-statistic: 1.293 on 6 and 185 DF,  p-value: 0.2622
```

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.3186133  1.4409895   0.915   0.3613
Week                      0.0804219  0.0409579   1.964   0.0511 .
CategoriesRestDaysNormal  0.3954929  0.3556429   1.112   0.2676
CategoriesRestDaysShort  -0.4273676  0.6193646  -0.690   0.4911
TotalYards                0.0003876  0.0019424   0.200   0.8421
PointsScored              0.0162101  0.0157549   1.029   0.3049
PercentYardsRush          1.8322066  1.8748151   0.977   0.3297
as.factor(Surface)Turf   -0.3314478  0.3084286  -1.075   0.2839
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.086 on 184 degrees of freedom
Multiple R-squared:  0.05454,	Adjusted R-squared:  0.01858
F-statistic: 1.516 on 7 and 184 DF,  p-value: 0.164
```

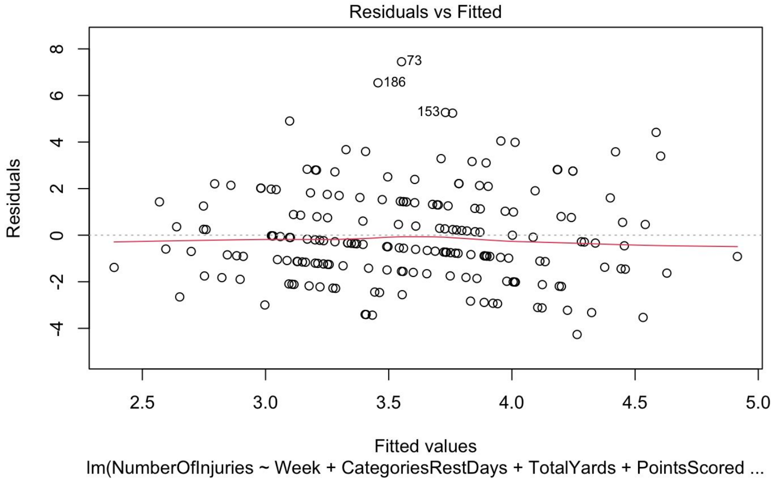
Conditions (continuous model):

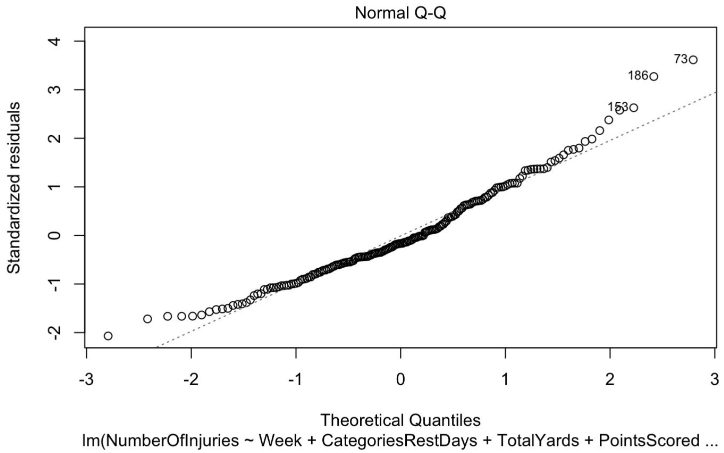





```
              Week
          1.037266
          RestDays
          1.073248
        TotalYards
          2.256925
      PointsScored
          2.116862
  PercentYardsRush
          1.173832
as.factor(Surface)
          1.031788
```

Conditions (categorical model):

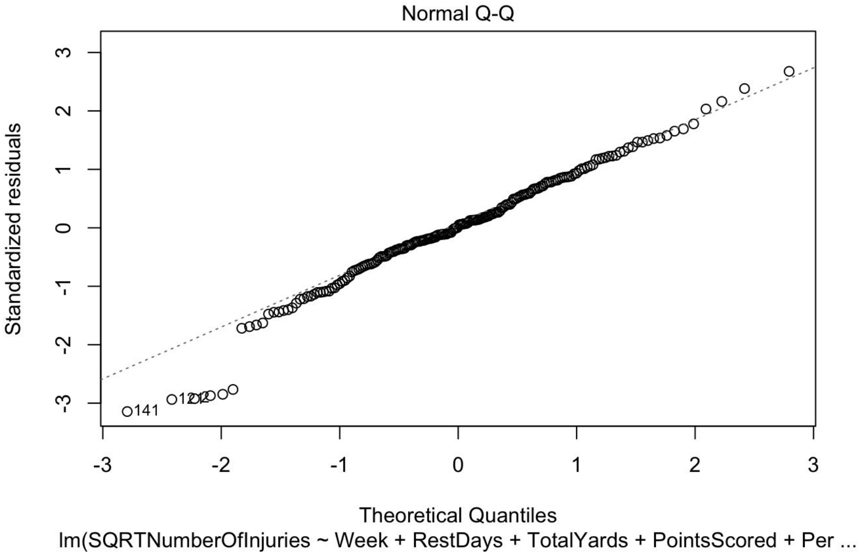


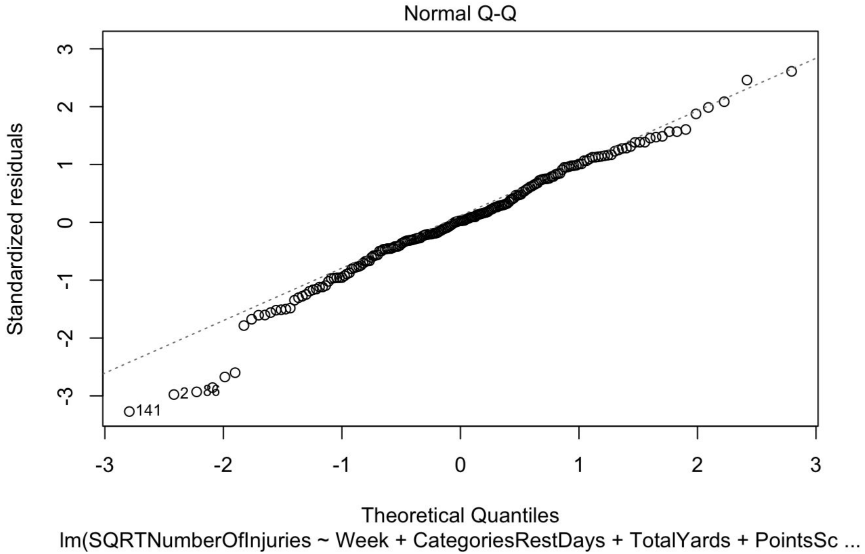


VIF:

1.058879
1.073520
2.256758
2.089491
1.166207
1.036008

Normal QQ-plots after square root transformation of injuries:





Best subsets model selection

- Quantitative days of rest for number of injuries and square root of number of injuries
- Days of rest not included in either model

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.80836    0.95991   1.884   0.0611 .
Week                    0.06822    0.03994   1.708   0.0893 .
PointsScored            0.01811    0.01118   1.620   0.1069
PercentYardsRush        1.82869    1.79138   1.021   0.3087
as.factor(Surface)Turf -0.31250    0.30669  -1.019   0.3096
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.085 on 187 degrees of freedom
Multiple R-squared:  0.03995,	Adjusted R-squared:  0.01941
F-statistic: 1.945 on 4 and 187 DF,  p-value: 0.1047
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.195231   0.263211   4.541 9.98e-06 ***
Week             0.017117   0.011568   1.480   0.1406
PointsScored     0.005879   0.003222   1.825   0.0696 .
PercentYardsRush 0.606221   0.514087   1.179   0.2398
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6039 on 188 degrees of freedom
Multiple R-squared:  0.03215,	Adjusted R-squared:  0.0167
F-statistic: 2.082 on 3 and 188 DF,  p-value: 0.1041
```

Best subsets model selection

- Categorical days of rest for number of injuries and square root of number of injuries
- Days of rest included in both models

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)   
(Intercept)                2.20839    0.71201   3.102  0.00222 **
Week                       0.08304    0.04074   2.038  0.04292 * 
CategoriesRestDaysNormal   0.39655    0.35391   1.120  0.26395   
CategoriesRestDaysShort   -0.44963    0.61603  -0.730  0.46639   
PointsScored               0.01622    0.01093   1.484  0.13960   
as.factor(Surface)Turf    -0.37243    0.30465  -1.222  0.22308   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.08 on 186 degrees of freedom
Multiple R-squared:  0.0496,	Adjusted R-squared:  0.02405 
F-statistic: 1.941 on 5 and 186 DF,  p-value: 0.08946
```

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               1.126972   0.275710   4.088 6.48e-05 ***
Week                      0.020152   0.011827   1.704   0.0901 .  
CategoriesRestDaysNormal  0.090017   0.102425   0.879   0.3806    
CategoriesRestDaysShort  -0.171928   0.177583  -0.968   0.3342    
PointsScored              0.006048   0.003221   1.878   0.0620 .  
PercentYardsRush          0.580820   0.513156   1.132   0.2592    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6025 on 186 degrees of freedom
Multiple R-squared:  0.0468,	Adjusted R-squared:  0.02118 
F-statistic: 1.827 on 5 and 186 DF,  p-value: 0.1096
```

K fold cross validation to compare models

- Method for splitting up data into test and training sets
- RMSE- error in the units of response variable (number of injuries). Objective is to minimize.

R code

- 10, 20, and 30 folds
- Key output: RMSE- error in the units of response variable (number of injuries). Objective is to minimize.

```
number = 10
#number = 20
#number = 30
kfoldedQuant <- train(NumberOfInjuries~Week+RestDays+TotalYards+PointsScored+PercentYardsRush+as.factor(Surface),
                 data=nfldata, trControl = trainControl(method = 'cv', number = number), model = 'lm')
print(kfoldedQuant)

kfoldedCat <- train(NumberOfInjuries~Week+CategoriesRestDays+TotalYards+PointsScored+PercentYardsRush+as.factor(Surface),
                  data=nfldata,trControl = trainControl(method = 'cv', number = number), model = 'lm')
print(kfoldedCat)
```

Output

- RMSE values minimized for categorical injuries, but still large contextually

- R-squared values still always tiny (well below 0.1)

RMSE Chart	10 folds	20 folds	30 folds
Quantitative Rest Days	2.135284	2.107371	2.106220
Categorical Rest Days	2.114944	2.095337	2.052833

Conclusions from linear analysis

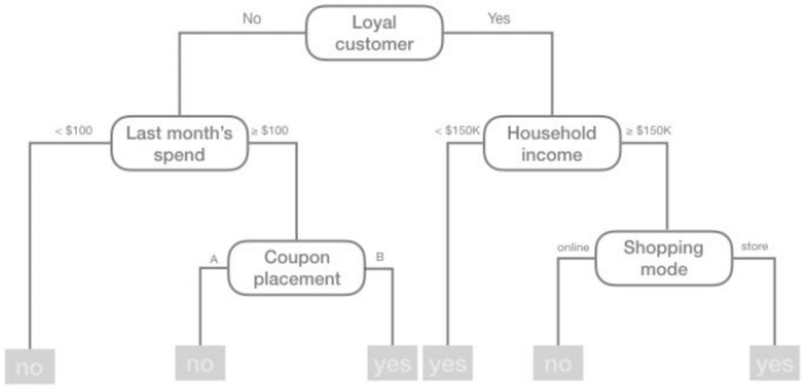
- We concur with the literature that days of rest is not a significant predictor of injuries
- We find no evidence that treating days of rest as a categorical variable was misleading
- Next, we will consider some non-linear analysis

Decision Trees

Use binary splitting rules

Such divide and conquer methods result in simple rules w
illustrated with tree diagrams

Can be viewed as a piecewise constant approximation



When to Stop Splitting a Tree?

If too many nodes the tree won't generalize as well

Less branches might not capture all interactions

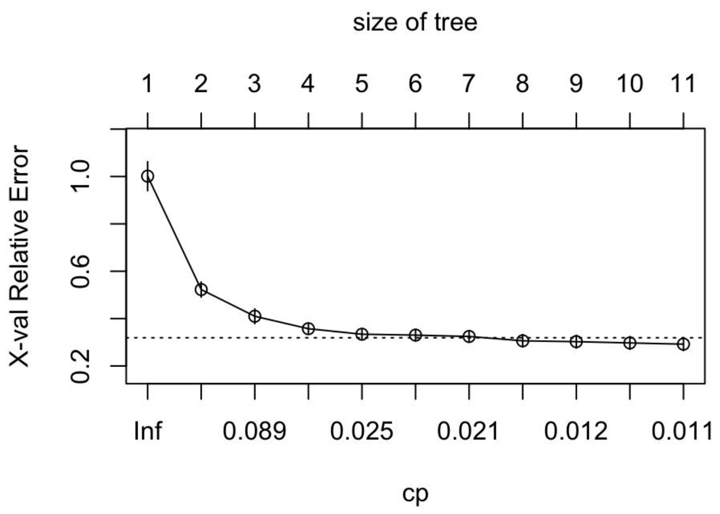
Less variance with more nodes

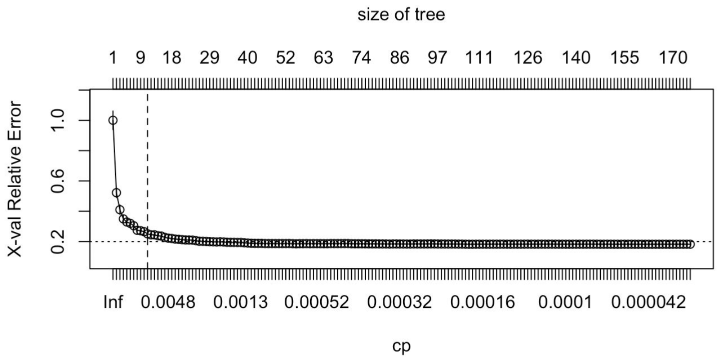
Pruning Complexity Parameter Plot

Prune - continue until diminishing returns

CP - probability to produce an output within certain boundaries

Dashed line represents smallest tree within 1 standard error of the minimum CV error

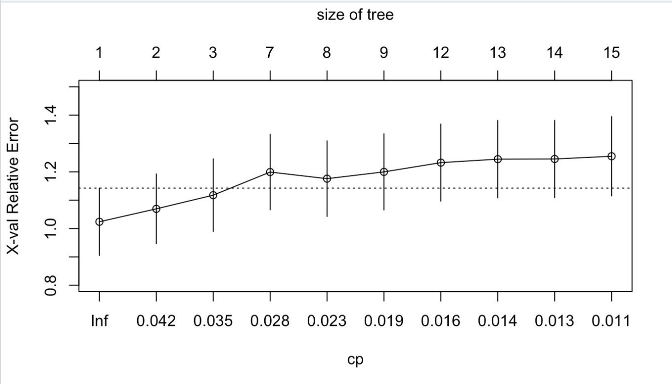


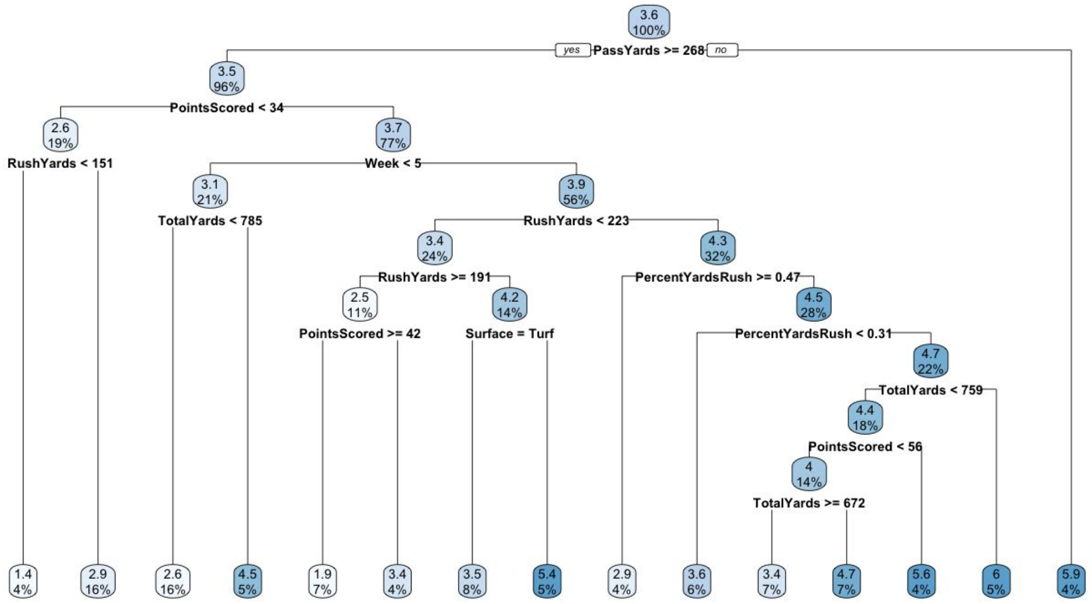


Our Decision Tree

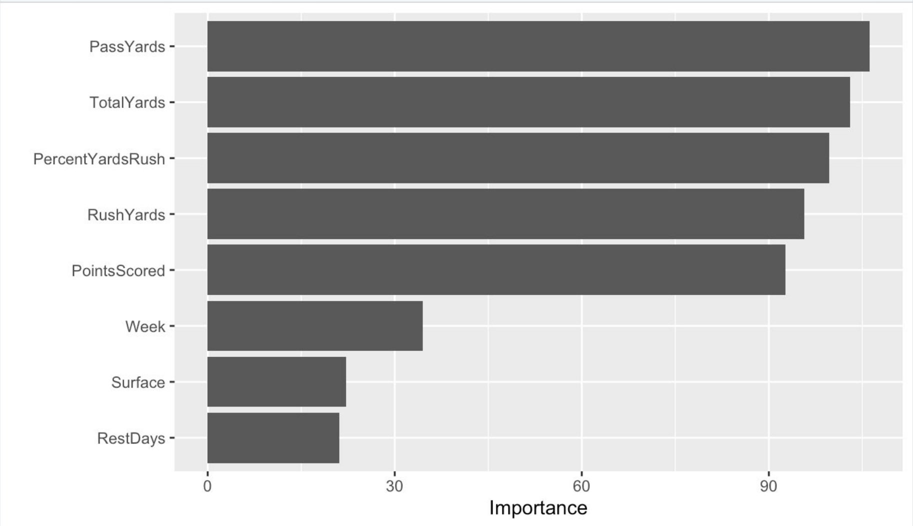
Same for categorical and quantitative rest day models

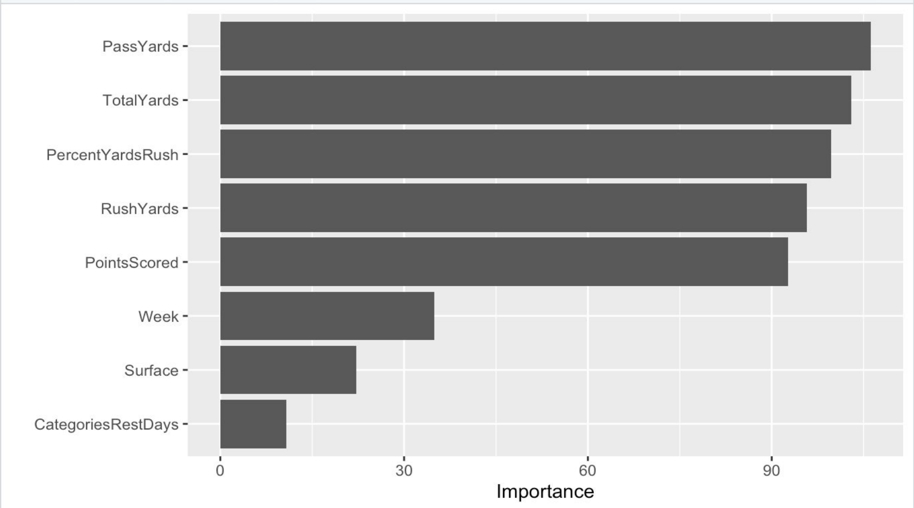
Generally the higher the node the more important it is





Importance of Predictors





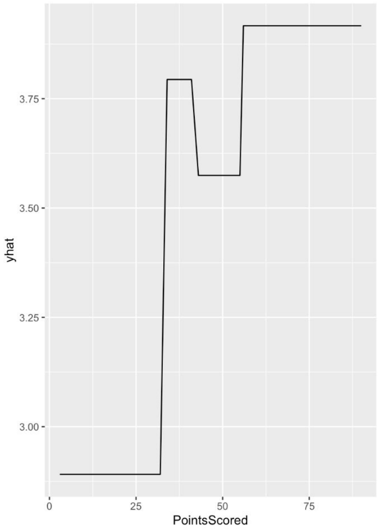
Problems with Decision Trees

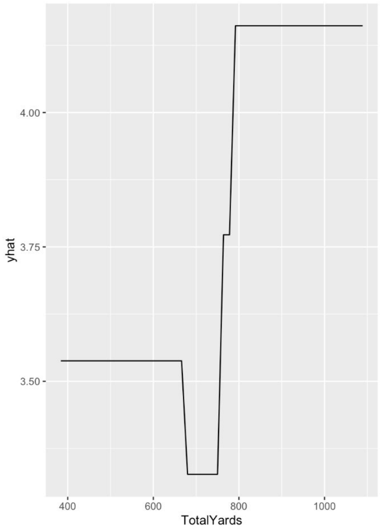
Lack in predictive performance relative to more complex algorithms such as neural networks and MARS

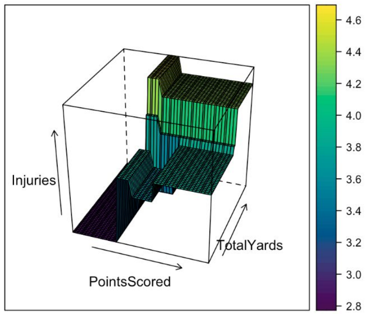
Rather staggered in approach

Greedy algorithm

More powerful ensemble algorithms include decision trees such as random forests and gradient boosting machines







If We Had More Time

- Expand on other linear models (such as rank-based regression)
- Expanded our data as not to “overfit” such a small sample, or just found a better dataset
- Explored potential correlations within our Decision Tree
- Tried a more advanced ensemble algorithms, such as random forests, gradient boosting machines, and bagging
- Specialize based on types of injuries or position groups

Sources

- https://sports.yahoo.com/chargers-te-donald-parham-unconscious-chiefs-nfl-014744061.html
- https://pubmed.ncbi.nlm.nih.gov/30848976/#:~:text=Results%3A%20The%20all%2Dcause%20injury,6%2C072%20per%201%2C000%20athletic%20exposures.
- https://journals.sagepub.com/doi/abs/10.1177/2325967119S00341
- https://www.si.com/nfl/2023/03/29/kansas-city-chiefs-patrick-mahomes-disapproval-thursday-night-football-change
- https://sotsports.com/2021/11/28/nfl-thursday-night-football-alternatives/
- https://nflsavant.com/about.php
- https://www.pro-football-reference.com/
- https://bradleyboehmke.github.io/HOML/

Questions?