

ATS Record and ST DVOA

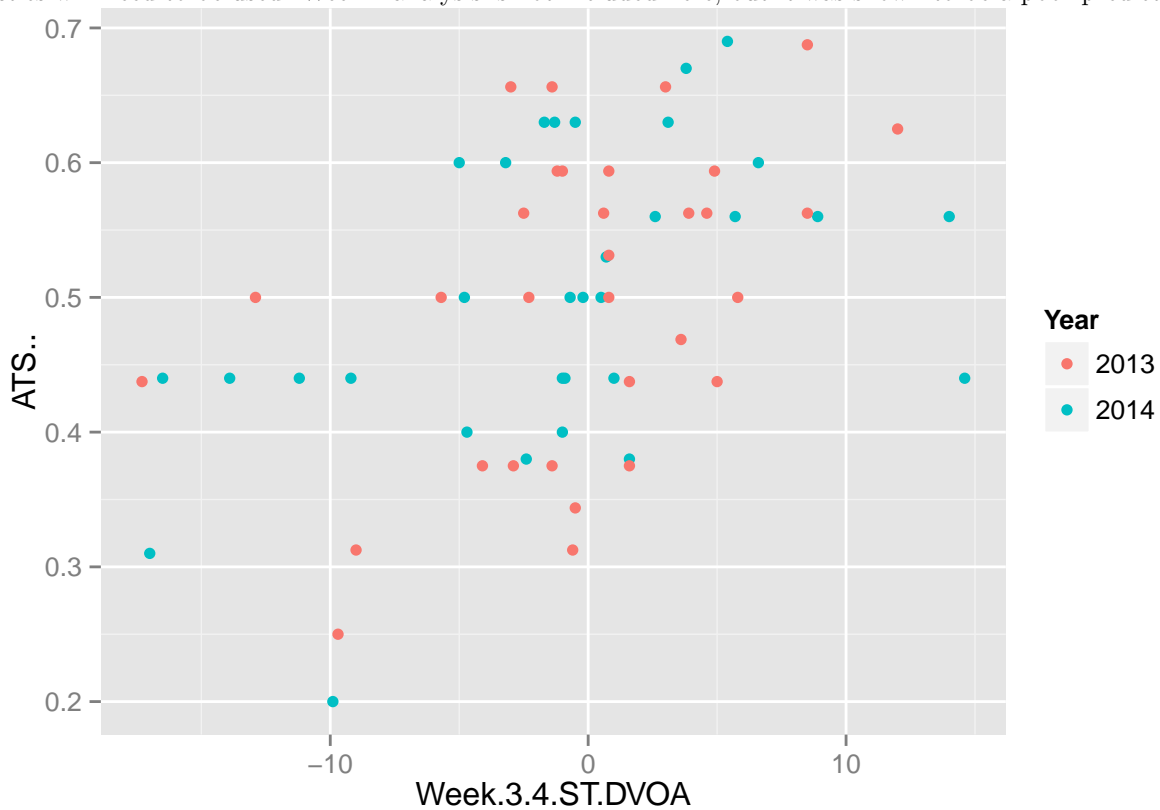
James Whedee

September 28, 2015

Exploratory Analysis

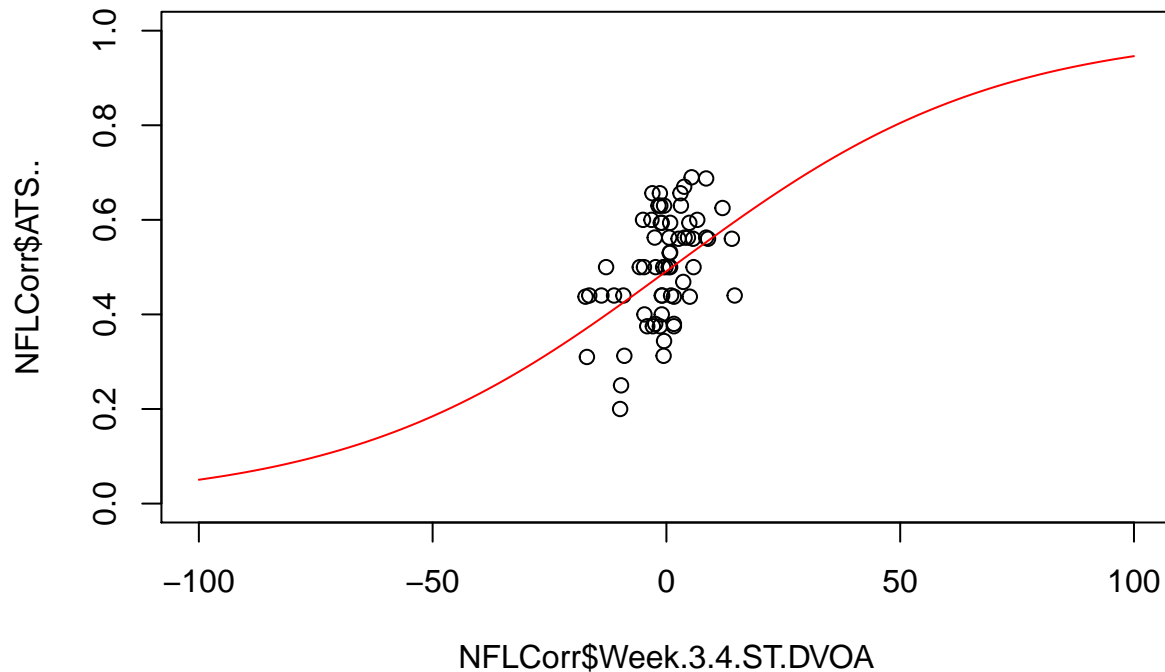
The following is exploratory analysis on using special teams DVOA as tracked by Football Outsiders as a predictive tool for against the spread winning percentage of NFL teams.

Here are the 2013 and 2014 ATS winning percentages of teams plotted as a function of their Week 3/4 ST DVOA, with line of best fit. Week 3/4 was chosen because for this to be of predictive value, early season statistics will need to be used. Week 1 analysis is not included here, but it was shown to be a poor predictive



tool.

Here is the same data with a logistic fit, zoomed out so that the shape of the curve is visible. A logistic fit is more intuitive for modeling winning percentage since its domain is (0,1)



Summary of significance of line of best fit:

```
##
## Call:
## lm(formula = "NFLCorr$ATS..~NFLCorr$Week.3.4.ST.DVOA")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23023 -0.07104  0.01499  0.07701  0.17311
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.506149   0.012594  40.190 < 2e-16 ***
## NFLCorr$Week.3.4.ST.DVOA 0.007668   0.001855   4.133 0.000109 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1002 on 62 degrees of freedom
## Multiple R-squared:  0.216, Adjusted R-squared:  0.2033
## F-statistic: 17.08 on 1 and 62 DF, p-value: 0.0001094
```

Results of correlation testing:

```
##
## Pearson's product-moment correlation
##
## data: NFLCorr$Week.3.4.ST.DVOA and NFLCorr$ATS..
## t = 4.1329, df = 62, p-value = 0.0001094
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2471740 0.6377049
## sample estimates:
```

```
##          cor
## 0.4647464
```

Summary of significance of logistic fit:

```
##
## Call:
## glm(formula = cbind(floor(16 * ATS..), ceiling(16 * (1 - ATS..))) ~
##      Week.3.4.ST.DVOA, family = binomial(logit), data = NFLCorr)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98315  -0.67475   0.09007   0.57808   1.27787
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.034562   0.063125  -0.548  0.58403
## Week.3.4.ST.DVOA  0.029011   0.009427   3.077  0.00209 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 49.072  on 63  degrees of freedom
## Residual deviance: 39.438  on 62  degrees of freedom
## AIC: 248.63
##
## Number of Fisher Scoring iterations: 3
```

Exploratory analysis finds a significant correlation, suggesting week 3/4 ST DVOA can be used to predict ATS winning percentage.

It should be noted that there is no significant correlation between week 3/4 ST DVOA and ATS in the 2012 season. This could be a result of changes in Vegas spreads, the calculation of ST DVOA, or conditions and rules within the NFL. Still, it is unlikely the 2013 and 2014 seasons can be passed off as coincidental.

Concrete Betting Strategy

The optimal betting strategy would likely be a probit/logit regression taking into account ST DVOA and any other identified non-trivial regressors. This section will cover another possible strategy. The idea will be to find the optimal threshold for ST DVOA. It will propose to choose the top X ST DVOA teams after week 3/4 and bet on them to beat the spread for the remaining weeks of the season.

There will be two factors that will affect the choice of X. First is the expected ATS record of teams above that threshold. Second is the number of teams above that threshold. X will be determined by maximizing the expected return in a single season given these competing factors.

```
expectedValue <- function(team,method)
  #method 1 maximizes expected returns
  #method 2 attempts to protect against variance by multiplying returns by a factor of games to bet on
  {
    STDVOA <- team[2]
    aboveST <- NFLCorr[NFLCorr$Week.3.4.ST.DVOA>=STDVOA,]
```

```

ATS <- mean(aboveST$ATS..)
returns <- (21*ATS-11)/11 #returns based on Vegas -110 vig
numRows <- nrow(NFLCorr[NFLCorr$Week.3.4.ST.DVOA>=STDVOA,])
if(method==1) factor <- 1
else factor <- numRows
c(numRows,returns*factor)
}
maxVector <- apply(select(NFLCorr,c(ATS..,Week.3.4.ST.DVOA)),1,expectedValue,method=1)
factorVector <- apply(select(NFLCorr,c(ATS..,Week.3.4.ST.DVOA)),1,expectedValue,method=2)
maxTeamsIndex <- which.max(maxVector[2,])
factorTeamsIndex <- which.max(factorVector[2,])
maxNumTeams <- maxVector[1,maxTeamsIndex]/2 #divide by two because 2 seasons of data used
factorNumTeams <- factorVector[1,factorTeamsIndex]/2

```

The results indicate that 5 teams per season maximizes returns and that 9.5 maximizes returns multiplied by the number of available games to bet on.

Since we must settle on an integer, we choose 9 because the expected value for 9 is higher than for 10. By examining the full vectors, we can also see that the max return for 9 teams is hardly different than for 5 teams, yet we have almost double the protection from variance.

In conclusion, this model proposes to choose the top 9 ST DVOA teams after week 3/4 and bet on them to beat the spread for the remaining weeks of the season.

Reversing the analysis for bottom ST DVOA teams pleasantly shows that the bottom 9 ST DVOA teams are the optimal choice for betting against to beat the spread for the remaining weeks of the season.

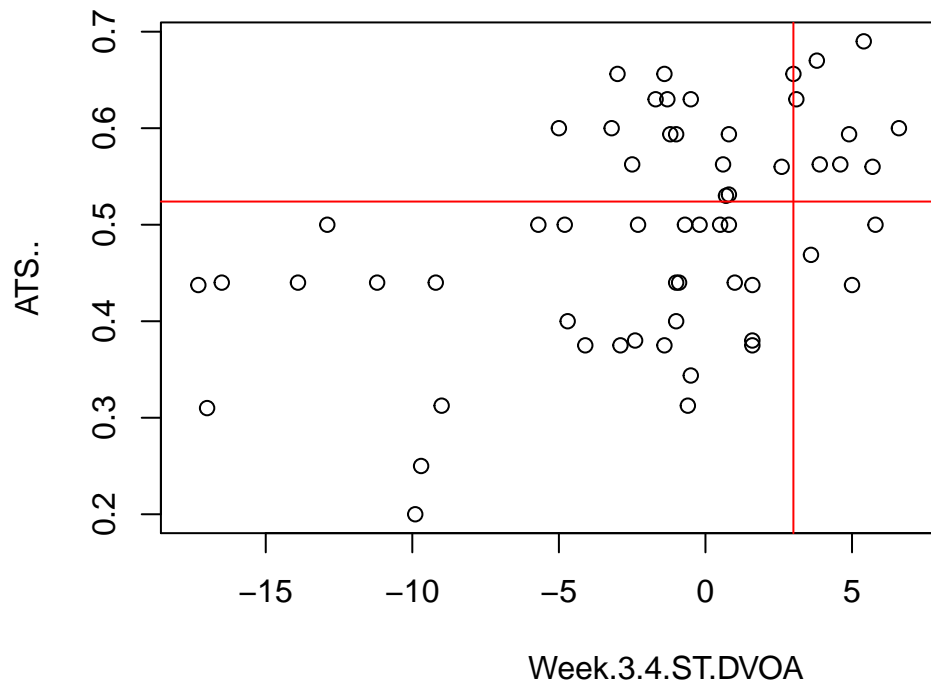
In concrete terms:

The mean ATS in 2013 and 2014 for the top 9 ST DVOA teams was .556 and .586 respectively. Assuming a vig of -110, these equate to returns of 6.06% and 11.79%.

The mean ATS in 2013 and 2014 for the bottom 9 ST DVOA teams was .441 and .419 respectively. Assuming a vig of -110, these equate to returns of 6.72% and 10.94% when betting against.

Over those two years, a proportion of .570 of bets made would be won, meaning a return of 8.88%.

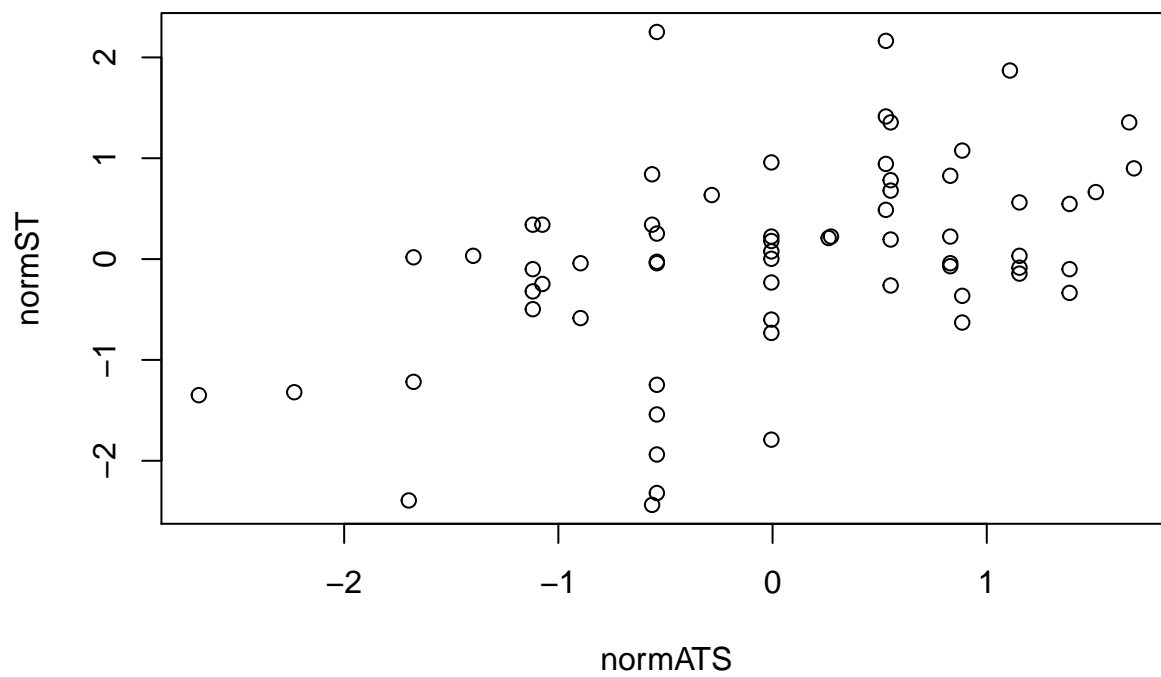
Finally, here is the original plot, with a vertical line showing the threshold of 9 teams per season and a horizontal line showing the break-even winning percentage given a -110 vig. The plot made for the bottom teams looks



much the same, but is omitted here for clarity.

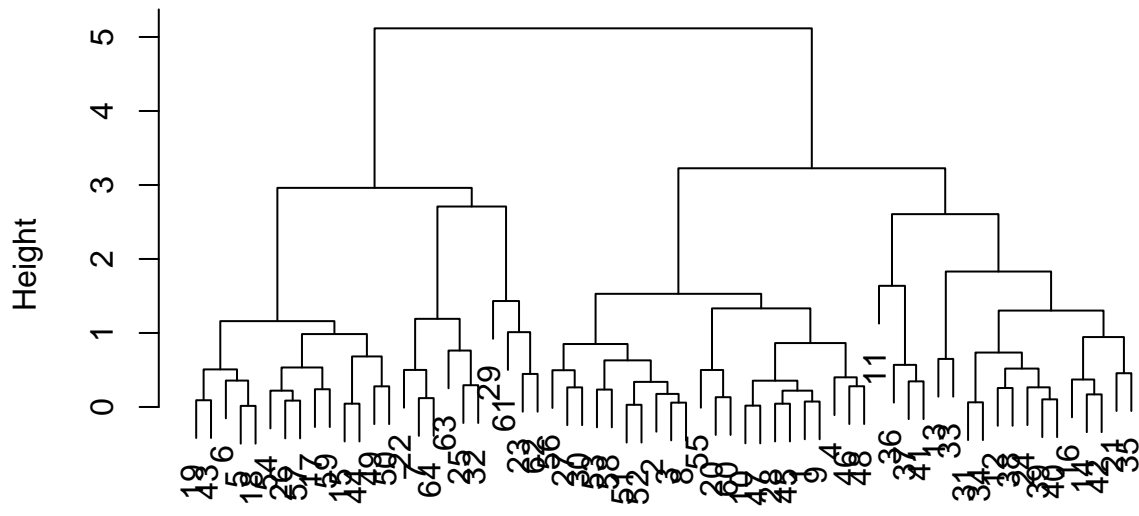
Exploring Clustering

As an afterthought, let's look at a hierarchical clustering of our plot. Since we'll be using Euclidean distance in this investigation, we will first norm our data so as to not give undue preference to ST DVOA (which has much greater variance).



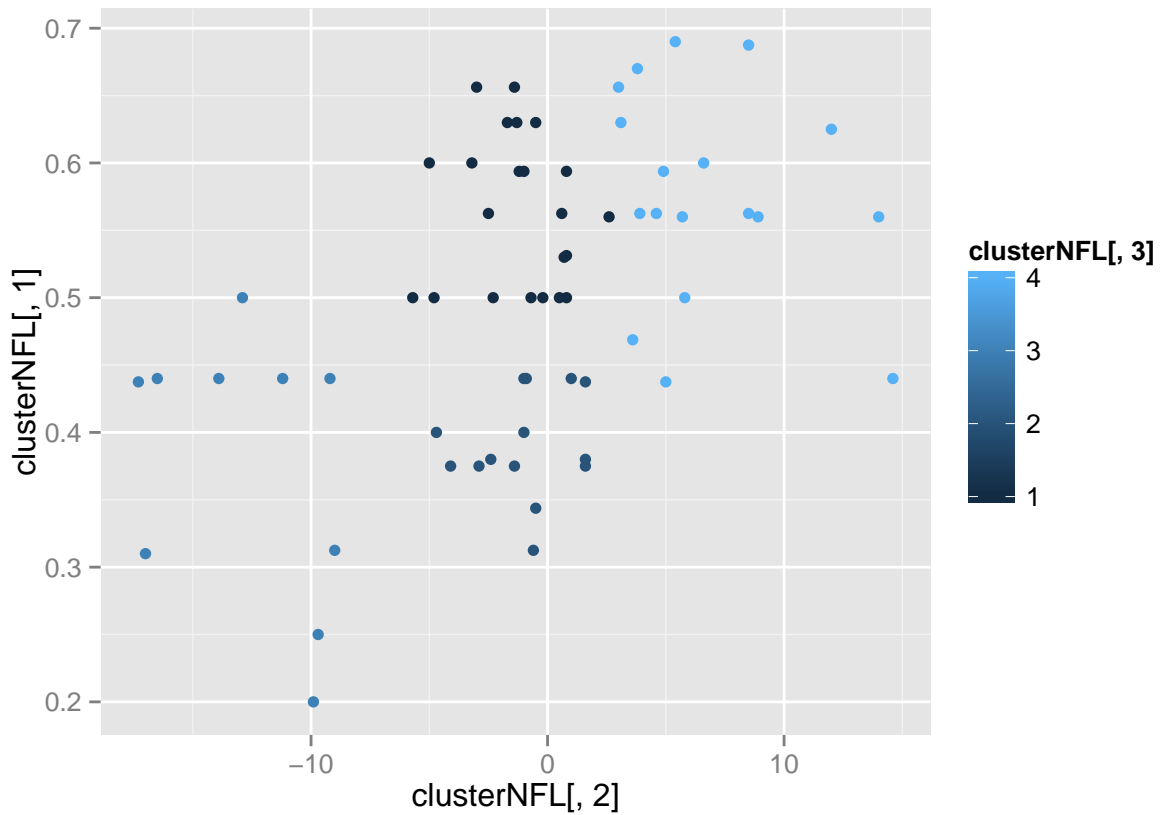
Now, we will cluster our normed data.

Cluster Dendrogram



distNormATSandST
hclust (*, "complete")

There appear to be 4 roughly even clusters, so we will make our cuts there.



This actually isolates the same 18 teams (average 9 per season) selected to maximize ATS in the earlier

analysis. However, it selects 10 teams (average 5 per season) on the lower end.

Sources

ATS data for 2013 pulled from <http://www.sportingcharts.com/articles/nfl/nfl-against-the-spread-and-total-standings-2013-14.aspx>

ATS data for 2014 pulled from <http://www.predictem.com/nfl/2014-ats-records.php>

ST DVOA pulled from various weeks by selecting the desired weeks from https://web.archive.org/web/*/http://www.footballoutsiders.com/stats/teameff