

Finding Betting Inefficiencies in the NFL

By: Gavin Wood and Samuel Ouellette



Project Overview

- The prompt for this project was to build a model that identifies inefficiencies in the NFL betting market given data from every NFL game played since 1980
- The project was completed in Python, but the original codebase was lost due to a hardware failure and is no longer recoverable

Betting on Indoor Games

- Favored team wins 63.2% of the time in indoor games
- Given that the favored team wins an indoor game, they cover the spread about 69% of the time
- If you're going to take the favorite team's money line, chances are, they'll cover the spread
 - The odds of covering the spread are about 2.24 times the odds of not covering the spread
- Does the degree of the point spread in an indoor game affect whether or not you should pick the favorite team to cover the spread?

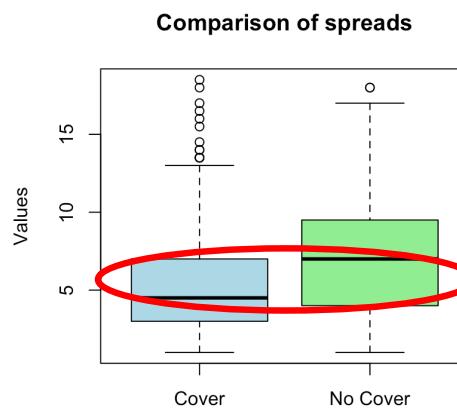
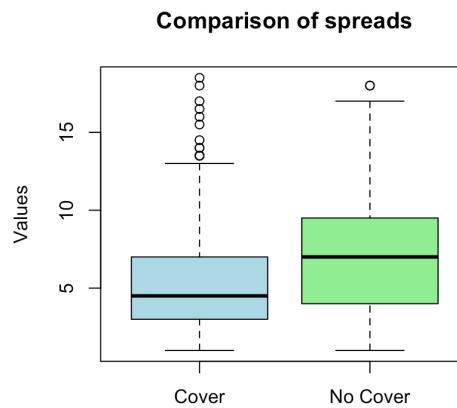
***All data is based on the given the favorite team wins an indoor game

If the spread is more than 7, its two times as likely that the favored winning team won't cover compared to if they cover

If the spread is less than 4, its about two times as likely that they cover compared to if they don't cover (46% of data lies below 4 for cover compared to 20.2% of data below 4 for no cover)

Average spread when covered	5.28
Average spread when NOT covered	7.19

A two-sample Z-test in R concluded that these averages are significantly different.
(p-value=.01338)



Median of not covering = 7 = Q3 of covering

Median of covering = 4.5
Q1 of not covering = 4

If the favorite team wins, there is about an equally likely chance that that they cover or don't cover if the spread is between 4 and 7.

- Probability of covering between 4.5 and 7 point spread is .25 (q3-median)
- Probability of not covering between 4.5 and 7 point spread is .243 (median-25.7th percentile of not covering)
- Running a two sample proportion test in R, these are not significantly different
- Conclusion from this?
 - When the spread is between 4.5 and 7, the chances of covering are about equal to the chances of not covering
 - If you want to beat the breakeven point of 52.4%, do not bet on a spread between 4.5 and 7 given that you think the favorite team will win in an indoor game.

Point Spread	Pick to Cover? (F=Favorite, U=Underdog)
X<4	F (probability is twice as likely compared to not covering)
4<x<7	NULL (probabilities are equally likely)
X>7	U (prob of F not covering is twice probability of covering)

75th of not covering is 9.5

9.5 is 88th percentile of covering

Favorite team Twice as likely to not cover compared to cover at 9.5 or above (25% of data>9.5 compared to 12%)

25th percentile of cover is 3

3 is 13th percentile of not covering

About half as likely to cover when spread is 3 or less (25% of data below 3 compared to 13%)

Using Logistic Regression to Model Probabilities

- We wanted to investigate the relationship between the spread in an indoor game, and the proportion of the time that the favored team covers the spread given they won
- Using the property of logistic regression, the probability of covering the spread given the favored team has already one, for all levels of historical spread=

$$\text{historical spread} = \frac{e^{1.81335 - 0.16393x}}{1 + e^{1.81335 - 0.16393x}}$$

Call:

```
glm(formula = y.cover/n ~ x, family = binomial(link = logit),  
  data = log.data, weights = n)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.81335	0.11702	15.50	<2e-16 ***
x	-0.16393	0.01618	-10.13	<2e-16 ***

The odds of the favorite team covering given they won multiply by about 84% ($e^{-0.16}$) for each 1 increase in the point spread

We did the same process for the probability of not covering given the favorite team had won. The results and interaction are modeled: ---→

Likelihood Ratio Test on the Model

Analysis of Deviance Table (Type II tests)

Response: y.cover/n

	LR	Chisq	Df	Pr(>Chisq)
x	109.27	1	< 2.2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

This allows us to conclude that point spread is significant in determining whether or not a team covers the spread given the favorite wins.

x	prob to cover (given win)
1	0.838812646
1.5	0.827420496
2	0.815400381
2.5	0.802742657
3	0.789441133
3.5	0.77549351
4	0.760901807
4.5	0.745672749
5	0.729818122
5.5	0.713355066
6	0.696306295
6.5	0.678700232
7	0.660571041
7.5	0.641958556
8	0.622908083
8.5	0.603470089
9	0.583699757
9.5	0.563656439
10	0.543402987
10.5	0.523004996
11	0.502529978
11.5	0.482046472
12	0.461623129
12.5	0.441327797
13	0.421226626
13.5	0.401383219
14	0.381857857
14.5	0.362706809
15	0.343981751
15.5	0.325729299
16	0.307990659
16.5	0.290801408
17	0.274191388
17.5	0.258184713
18	0.242799882
18.5	0.228049979

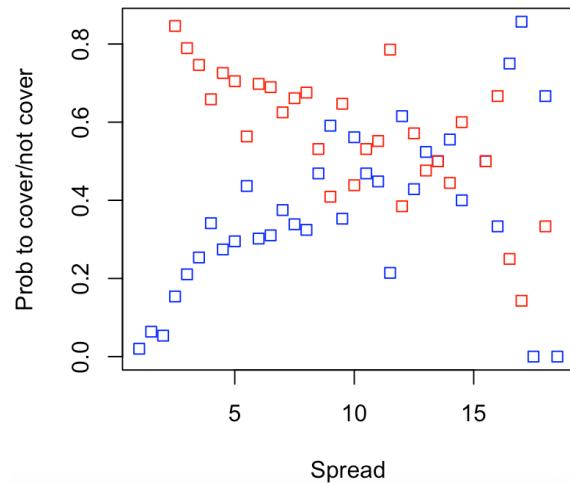
Given the favorite team wins, they cover the spread more than 52.4% (breakeven) of the time when the spread is less.

If the spread is 10.5-11.5 there is an equally likely chance that the favorite team could cover or not cover given they have one. This is a bet you would not want to make. (too much uncertainty)

If the spread is more than 11.5, you can bet the underdog to cover the spread even if the favorite team wins and succeed more than 52.4% of the time.

What if you only want to bet the spread and think the favorite team will win?

Interaction Effect



The most interaction takes place at about a 50% probability and 11 spread

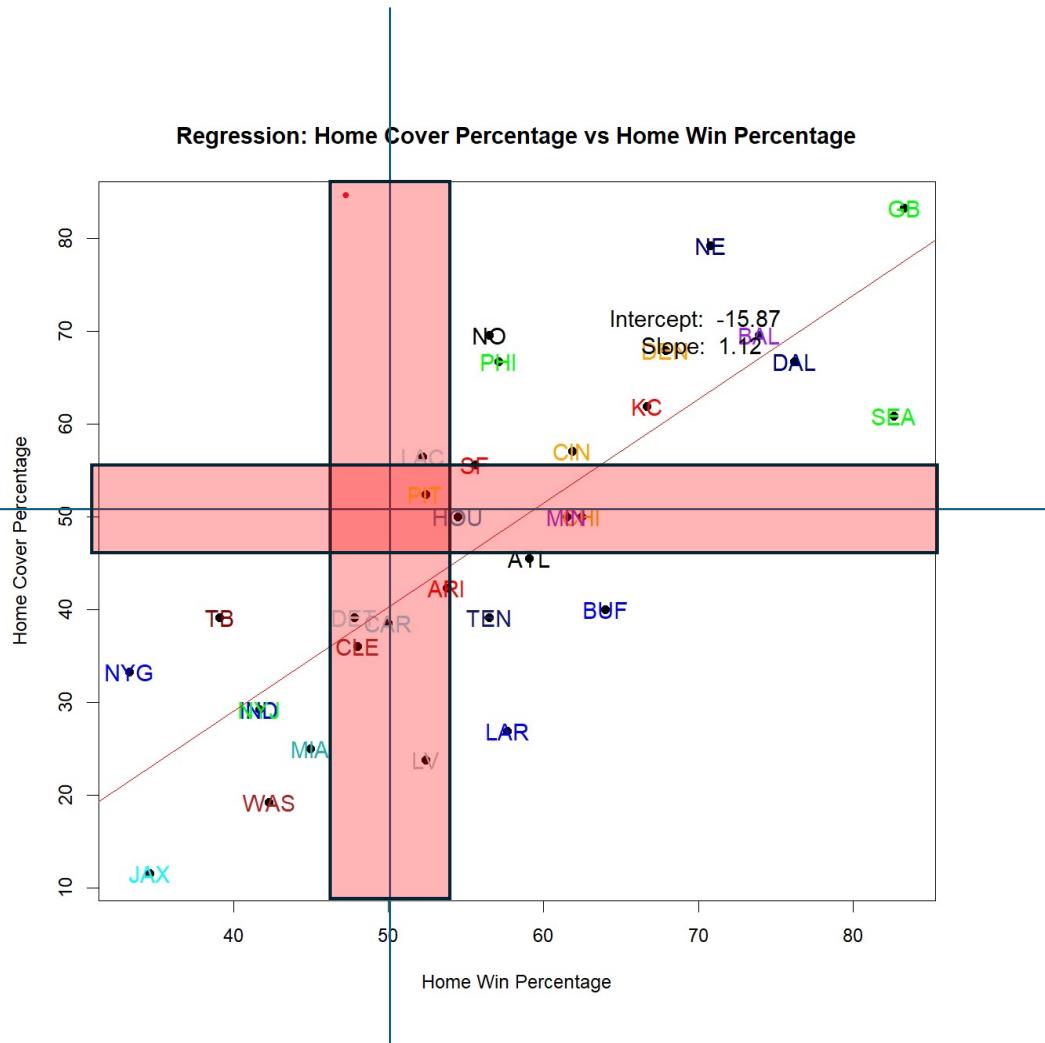
Comparing: **Home Covering % vs Home Win %** since 2010

What are we looking for?

- Teams that win a lot at home but don't cover often.
- Teams that cover despite losing often.
- Teams that cover and win extremely often(best of the best).

What do we want to stay away from?

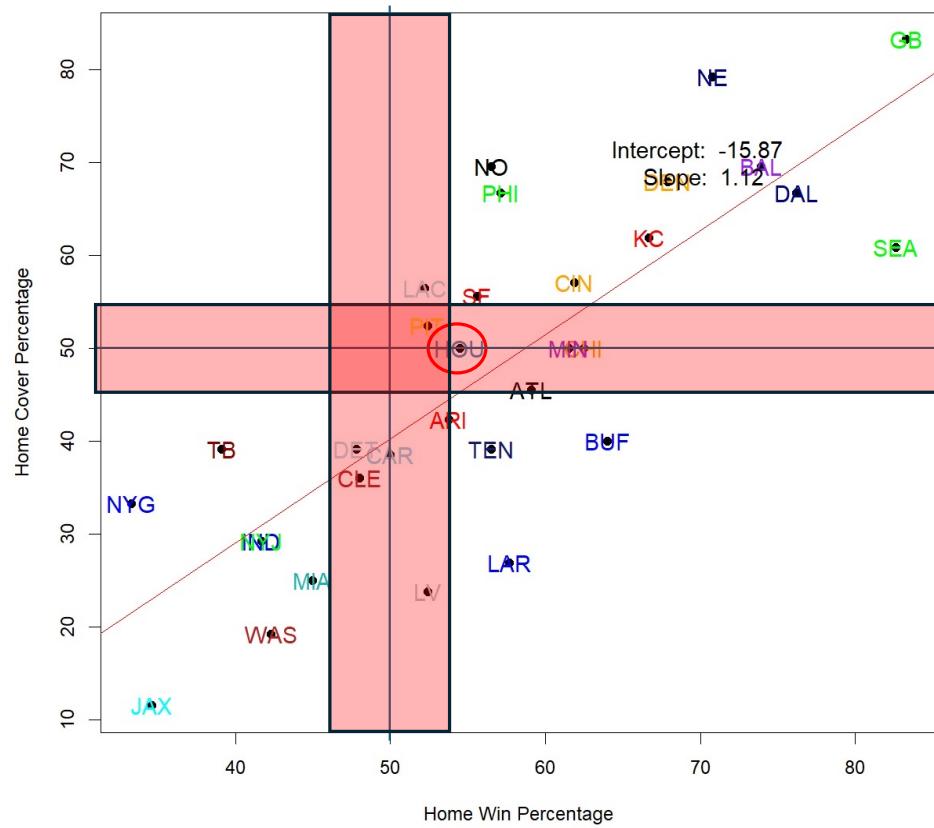
- Teams that perform along the trend.
 - What does that mean?
 - Teams that have an equal chance in both home cover % and home win %. This is significant because if you are making a bet, these teams are basically a coin flip.



- Teams above the horizontal line you want to bet to cover at home
- Teams below the horizontal line you want to away team to cover
- Team right of vertical line you want to bet to win at home
- Team left of vertical line you want to bet to lose at home
- Teams in the redzone have too much uncertainty to bet in one direction
- If in horizontal but not vertical zone, bet to win or lose but not to cover
- Vice versa if in vertical zone but not horizontal
- If in both zones, DO NOT BET ON THIS TEAM!!!!

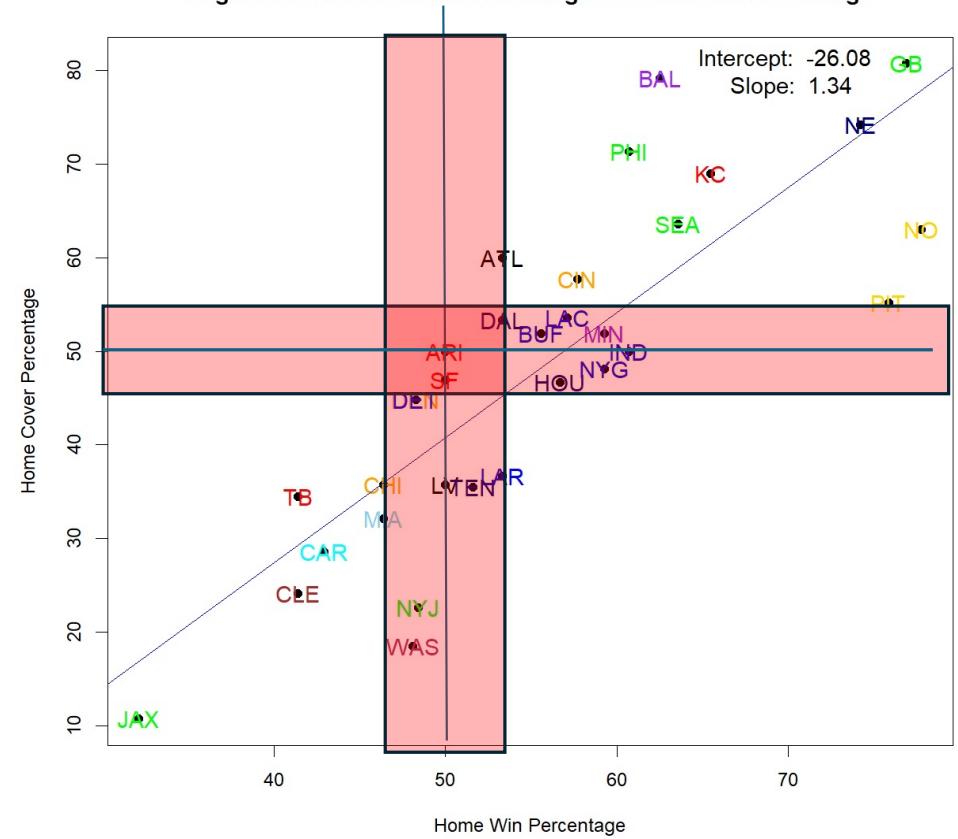
SEPTEMBER

Regression: Home Cover Percentage vs Home Win Percentage



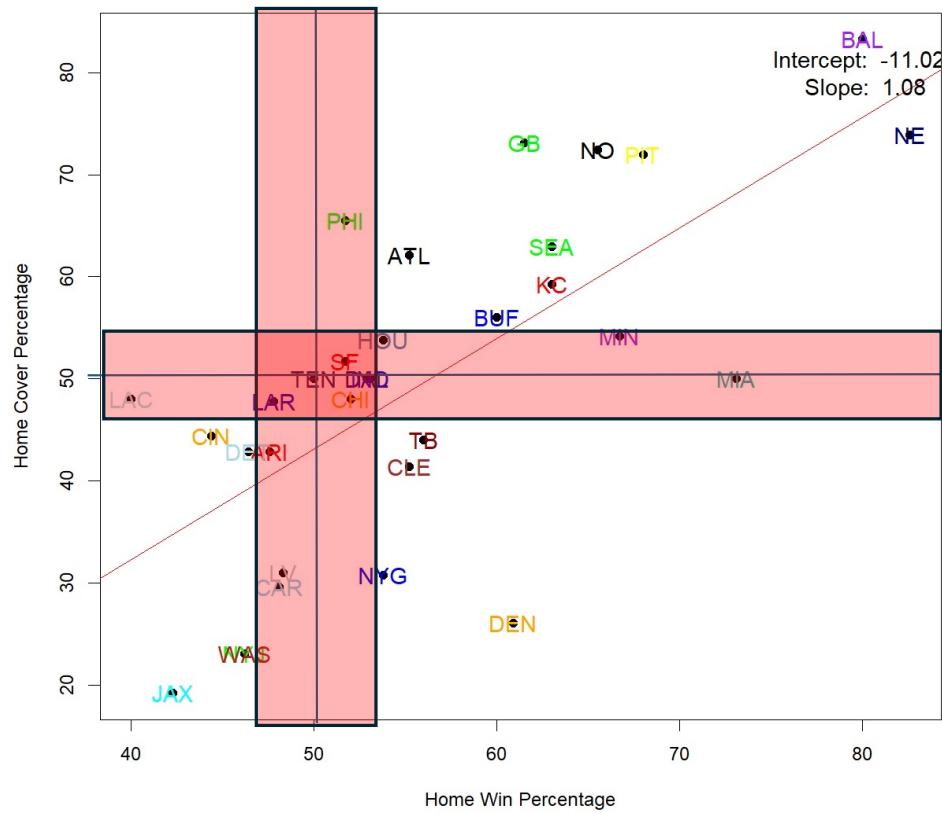
OCTOBER

Regression: Home Cover Percentage vs Home Win Percentage



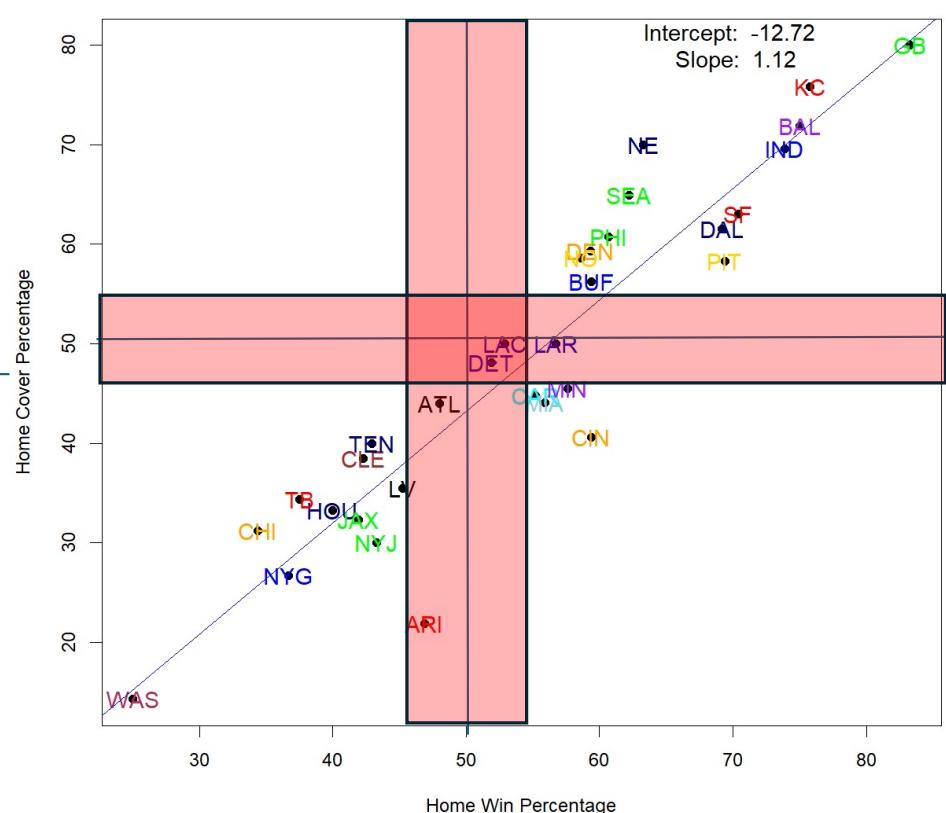
NOVEMBER

Regression: Home Cover Percentage vs Home Win Percentage



DECEMBER

Regression: Home Cover Percentage vs Home Win Percentage



What teams do you want to bet on/against in each month?

- **September:**
 - New Orleans Saints
 - LA Rams
- **October:**
 - Washington Commanders
 - New York Jets
- **November:**
 - Denver Broncos
 - Pittsburgh Steelers
- **December:**
 - Green Bay Packers
 - Kansas Chiefs

What teams do you want to stay away from each month?

- **September:**
 - Pittsburgh Steelers
 - Houston Texans
- **October:**
 - Arizona Cardinals.
 - San Francisco 49ers
- **November:**
 - Tennessee Titans
 - San Francisco 49ers
- **December:**
 - LA Chargers
 - Detroit Lions



Let the gambling
begin!