

---

# ISyE 2028 Basic Statistical Methods - Fall 2015

## Analysis of NFL Team Performance Against the Spread Betting Line

---

Author: Jonathan Pang

### Introduction

Sports betting, although controversial, is an extremely popular form of gambling in the United States. Although many different types of bets can be placed on popular websites or in casinos, the most popular type of bet that is placed on the results of major American sports is the spread bet.

A spread bet is a wager that the difference between the scores of two teams will be less than or greater than a certain value, which is specified by the bookmaker. For example, if the spread betting line for the Detroit Lions at Kansas City Chiefs game in Week 8 is +3.5 for the Detroit Lions, a spread bet placed on the Detroit Lion would be won if the score for the Detroit Lions ( $D$ ) + the spread betting line (+3.5) is greater than the score for the Kansas City Chiefs ( $K$ ).

$$D + 3.5 > K$$

If the score for the Detroit Lions ( $D$ ) + the spread betting line (+3.5) is less than the score for the Kansas City Chiefs, the wager will be lost.

$$D + 3.5 < K$$

In other words, the bettors are wagering on whether or not the Kansas City Chiefs will outscore the Detroit Lions by +3.5 points. It is also noted that most sports scores are only represented by integers. Thus, oftentimes the bookmakers will set the betting line at half a point intervals, making it impossible for neither side to win the bet. Otherwise, if the Kansas City Chiefs outscored the Detroit Lions by exactly the spread betting line, a push would occur and neither side would win (or lose).

In order to maximize the odds of winning a spread bet, historical data can be utilized to determine patterns in the final spread and the actual results in games. In this study, trends will be analyzed to determine whether or not any teams in the NFL have historically consistently “beat the spread”. This will be answered by hypothesis testing, comparing the spread betting lines and score differentials for each game a team plays. If the difference is significant, it is considered a safer bet to wager money on the team. However, it is expected that there is no significant difference between historical spread betting lines and score differentials for each game a team played. In addition, it is expected that historically no team has outperformed or underperformed the spread betting line on a consistent basis.

An aggregate of historical and current sports betting information was found at the Odds Sharks website, on which the final point spreads and over under betting lines are recorded for professional sports such as the NFL (National Football League), MLB (Major League Baseball), NBA (National Basketball Association), and NHL (National Hockey League) as well as college sports such as NCAAF (National Collegiate Athletic Association Football) and NCAAB (National Collegiate Athletic Association Basketball). A web scraper was utilized in order to extract past sports betting lines as well as actual scores of games.

### Data Collection

Data was collected from the Odds Sharks website using a CasperJS web scraper and processed by NodeJS. Relevant information was then inserted into a CSV document. Source code is attached in Appendix B. Data was collected for all regular season games played in the NFL since the 1995-1996 season. This resulted in approximately 20 years of data (including the present 2015-2016 season which has not yet been completed).

This data was then processed using R to determine the average amount by which each team scored above the betting line as well as how often each team was able to beat the betting line.

## Data Processing

The data collected from the Odds Shark website was first analyzed to determine the amount on average that a team overperformed the spread betting line in the past 20 seasons. This information is summarized in Appendix B in Table 1 and shown in graphical form in Figure 1.

As can be seen in Table 1, it is evident that the team that most outperformed the spread betting line on average was the New England Patriots, which scored on average, 2.798761610 points above the spread. Other teams such as the Green Bay Packers and the Pittsburgh Steelers also performed well against the spread betting line. On the other hand, it is evident that the team that on average most underperformed the spread betting line was the Arizona Cardinals, which scored on average,  $-1.743710692$ , points below the spread. Likewise, the Cleveland Browns and St. Louis Rams also underperformed against the spread betting line.

A two-tailed t-test with an alpha value of 0.05 was then constructed to determine if any teams in the past 20 years significantly outperformed or underperformed the spread betting lines. The t-tests were performed using R. It was determined that the following teams have historically performed significantly different than the spread betting lines: the Arizona Cardinals, the Cleveland Browns, the Green Bay Packers, and the New England Patriots. Both the Arizona Cardinals and the Cleveland Browns historically underperformed the spread betting lines and the Green Bay Packers and the New England Patriots historically outperformed the spread betting lines. The results of the t-tests are printed below.

ARI

One Sample t-test

```
data: team_games$pointsATS
t = -2.3237, df = 317, p-value = 0.02077
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-3.2200848 -0.2673366
sample estimates:
mean of x
-1.743711
```

CLE

One Sample t-test

```
data: team_games$pointsATS
t = -2.154, df = 277, p-value = 0.0321
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-2.9534503 -0.1328806
sample estimates:
mean of x
-1.543165
```

GB

One Sample t-test

```
data: team_games$pointsATS
```

```

t = 2.4456, df = 330, p-value = 0.01498
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
0.3528634 3.2543874
sample estimates:
mean of x
1.803625

```

NE

One Sample t-test

```

data: team_games$pointsATS
t = 3.4761, df = 322, p-value = 0.0005785
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
1.214757 4.382766
sample estimates:
mean of x
2.798762

```

Thus, the data support that the Arizona Cardinals and Cleveland Browns were historically underperformed the spread betting line and the Green Bay Packers and New England Patriots historically outperformed the spread betting line.

Though the previous observations can be interesting to bettors, a more relevant observation would not be how much a team scores above the spread betting line but rather how often a team scores above the spread betting line. Using R, the proportion of games in which a team was able to outperform the spread betting line was calculated and summarized in Table 1 and shown in graphical form in Figure 2.

This time, it is noted that the team that most consistently outperformed the spread betting line was the Carolina Panthers while the New England Patriots again fared well as the second most consistent with the Philadelphia Eagles and Green Bay Packers close behind. On the other end, we find that the St. Louis Rams was the most consistent underperformer while the San Francisco 49ers and the Washington Redskins were marginally better.

Furthermore, in order to determine if any team consistently outperformed or underperformed the spread betting line in the past 20 years, a z-test for proportions was performed on each of the sets of historical data. The two-tailed z-tests were constructed to compare the percentage of times that a team “beat the spread” with the null hypothesis that each team “beat the spread” 50% of the time. This was equivalent to stating that the probability of “beating the spread” was random. An alpha value of 0.05 was utilized and the tests were processed using R. It was determined that no team performed significantly different from the null hypothesis percentage of 50%. Thus, the hypothesis that no team has historically consistently outperformed or underperformed against the spread is supported.

## Conclusion

The purpose of this study was to determine if any team significantly and consistently beat the spread betting lines placed on games in casinos and on popular betting websites. The spread betting lines are essentially a prediction of how much one team will beat another team and are moved throughout the week in order to maintain a 50% balance on both sides of the spread (50% betting under the spread and 50% betting over the spread). Since a team has no incentive to “beat the spread” on their own as players and team personnel are prohibited from betting on games, a team that is able to consistently “beat the spread” can be interpreted as a team that is consistently underrated by the bookmakers and by the general public. Likewise, a team that consistently underperforms against the spread betting line can be interpreted as a team that is consistently overrated by the bookmakers and by the general public. From this study, it was determined that four teams on average perform significantly different from the spread betting lines. These

teams are the Arizona Cardinals and the Cleveland Browns, both of which on average underperform against the spread betting line and the Green Bay Packers and the New England Patriots both of which on average outperform against the spread betting line. However, of these four teams, none consistently outperform or underperform against the spread. This would mean that on the occasions that the teams do outperform or underperform the spread, the teams greatly outperform or underperform. However, these teams do not outperform or underperform the spread on a more consistent basis than do any other teams. In fact, the percentage of beating the spread is, for all these teams, very close to 50% which corresponds to random chance. In conclusion, no team was found to have historically outperformed or underperformed significantly against the spread and the proportion of beating the spread for all teams was found to be 50%.

The data collected is not consistent with the expectation that there is no difference between score differentials and the spread betting line. The data collected, however, is consistent with the expectation that no team consistently outperformed or underperformed the spread. The sample size utilized was certainly large enough with 325 game results and spread betting lines collected for each team. The sample is representative of the true population and can be utilized by bettors in order to determine whether to bet above or below the spread when looking at certain teams when placing a spread bet.

## Appendix A

Table 1: Points Scored Above the Spread for All Teams

Team	Mean	Percent
ARI	-1.743710692	0.4779874
CLE	-1.543165468	0.4580645
STL	-1.312703583	0.4462541
OAK	-1.117283951	0.4444444
DET	-0.938906752	0.4533762
DAL	-0.923870968	0.4580645
TEN	-0.848056537	0.4558304
SF	-0.838762215	0.4495114
WAS	-0.656146179	0.4518272
CIN	-0.492354740	0.4709480
NO	-0.400974026	0.4837662
BUF	-0.392857143	0.5031056
TB	-0.374137931	0.4827586
JAC	-0.360856269	0.4831804
ATL	-0.335889571	0.4938650
HOU	-0.239669421	0.5082645
NYG	-0.112804878	0.4939024
MIN	-0.022865854	0.5030488
MIA	0.007692308	0.4676923
CHI	0.253105590	0.4689441
SD	0.285493827	0.4845679
NYJ	0.328703704	0.4938272
DEN	0.387537994	0.4893617
PHI	0.419558360	0.5236593
IND	0.435975610	0.5030488
SEA	0.641033435	0.4924012
CAR	0.847692308	0.5292308
BAL	1.183121019	0.4904459
KC	1.290123457	0.5123457
PIT	1.307692308	0.5032051
GB	1.803625378	0.5226586
NE	2.798761610	0.5263158

Figure 1: Mean Points Above Spread for Teams

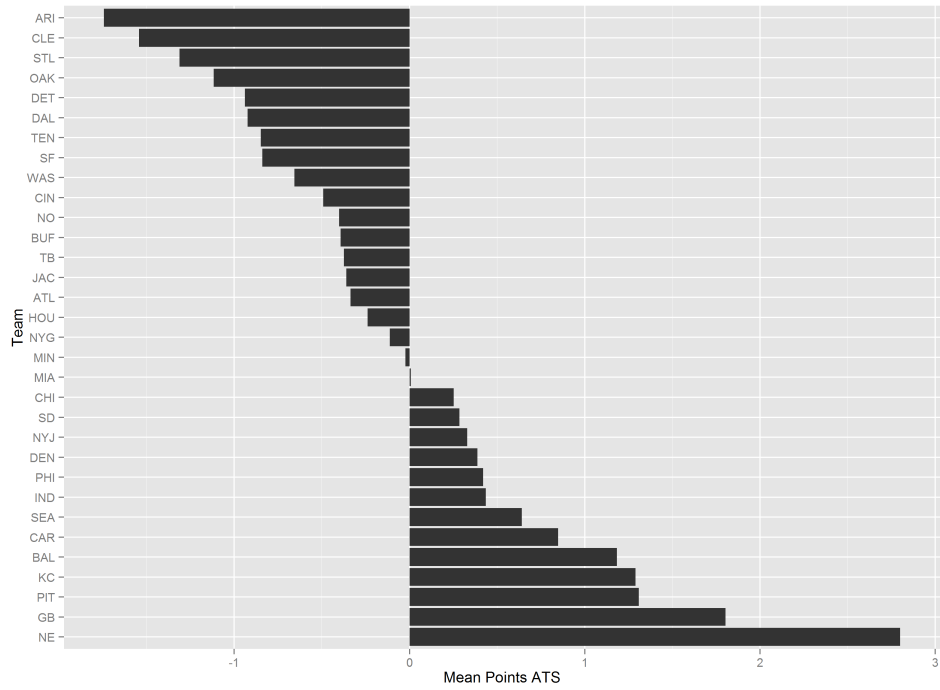
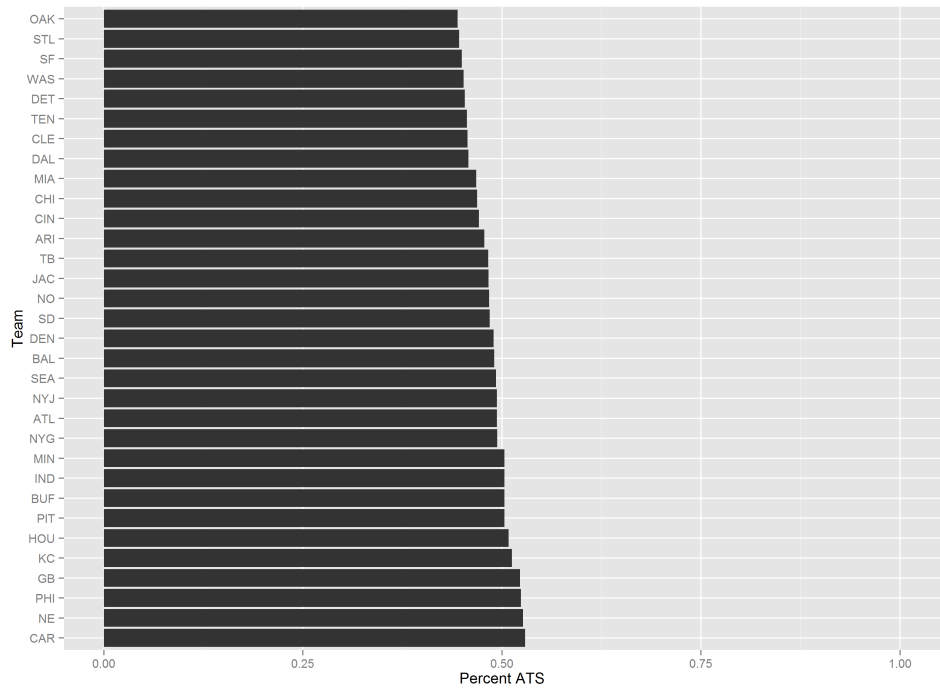


Figure 2: Mean Points Above Spread for Teams



## Appendix B

### main.js

```
var path = require('path');
var childProcess = require('child_process');
var cheerio = require('cheerio');
var fs = require('fs');
var csvWriter = require('csv-write-stream');

var separator = ',';
var newline = '\n';

fs.writeFile('output.csv', "date" + separator +
    "away" + separator +
    "away_score" + separator +
    "home" + separator +
    "home_score" + separator +
    "result" + separator +
    "home_spread" + separator +
    "ats" + separator +
    "total" + separator +
    "ou" + newline, function(err, data) {
    }
);

var teams = ["Arizona",
    "Atlanta",
    "Baltimore",
    "Buffalo",
    "Carolina",
    "Chicago",
    "Cincinnati",
    "Cleveland",
    "Dallas",
    "Denver",
    "Detroit",
    "Green",
    "Houston",
    "Indianapolis",
    "Jacksonville",
    "Kansas",
    "Miami",
    "Minnesota",
    "Giants",
    "Jets",
    "England",
    "Orleans",
    "Oakland",
    "Philadelphia",
    "Pittsburgh",
    "Diego",
    "Francisco",
    "Seattle",
```

```

    "Louis",
    "Tampa",
    "Tennessee",
    "Washington"];

team_index = 1;

while(team_index <= 32) {
    for(i = 1; i < teams.length; i++) {
        console.log(i)
        console.log(teams[0]);
        console.log(teams[i]);
        var result = childProcess.execSync('casperjs test.js ' + teams
            [0] + " " + teams[i]);

        var $ = cheerio.load(result);

        var date, away, away_score, home, home_score, result,
            home_spread, ats, total, ou;

        var json = {date: "", away: "", away_score: "", home: "",
            home_score: "", result: "", home_spread: "", ats: "",
            total: "", ou: ""};

        var no_entries = $("#block-system-main table:nth-of-type(2)
            tbody").find('tr').length;
        var no_columns = $("#block-system-main table:nth-of-type(2)
            tbody tr").eq(1).find('td').length;

        for(j = 0; j < no_entries; j++){
            date = $("#block-system-main table:nth-of-type(2)
                tbody tr").eq(j).children('td').eq(0).text();
            away = $("#block-system-main table:nth-of-type(2)
                tbody tr").eq(j).children('td').eq(1).text();
            away_score = $("#block-system-main table:nth-of-type
                (2) tbody tr").eq(j).children('td').eq(2).text();
            home = $("#block-system-main table:nth-of-type(2)
                tbody tr").eq(j).children('td').eq(3).text();
            home_score = $("#block-system-main table:nth-of-type
                (2) tbody tr").eq(j).children('td').eq(4).text();
            result = $("#block-system-main table:nth-of-type(2)
                tbody tr").eq(j).children('td').eq(5).text();
            home_spread = $("#block-system-main table:nth-of-type
                (2) tbody tr").eq(j).children('td').eq(6).text();
            ats = $("#block-system-main table:nth-of-type(2) tbody
                tr").eq(j).children('td').eq(7).text();
            total = $("#block-system-main table:nth-of-type(2)
                tbody tr").eq(j).children('td').eq(8).text();
            ou = $("#block-system-main table:nth-of-type(2) tbody
                tr").eq(j).children('td').eq(9).text();

            fs.appendFile('output.csv', "" + date + "" +
                separator +
                away + separator +

```

```
        away_score + separator +
        home + separator +
        home_score + separator +
        result + separator +
        home_spread + separator +
        ats + separator +
        total + separator +
        ou + newline, function(err){
            console.log('File successfully written
                !')
        });
    });
};

var first = teams.shift();
teams.push(first);
team_index = team_index + 1;
};
```

**test.js**

```

var casper = require('casper').create({
    verbose: true,
    logLevel: "debug"
});
casper.start("http://www.oddsshark.com/nfl/database");
casper.thenEvaluate(function() {
    // Number of Games
    document.getElementById('games-30-h2h').checked = 'checked';
    document.getElementById('location-home-h2h').checked = 'checked';

    // Month
    document.getElementById('chalk-select-month-h2h').getElementsByTagName('option')[0].selected = 'selected';

    // Regular Season or Postseason
    document.getElementById('chalk-select-game-type-h2h').getElementsByTagName('option')[1].selected = 'selected';

    // Favorite or Underdog
    document.getElementById('chalk-select-odds-h2h').getElementsByTagName('option')[4].selected = 'selected';

    // Line Range
    document.getElementById('txt-min-h2h').value = '-100';
    document.getElementById('txt-max-h2h').value = '100';
});
casper.then(function() {
    // Team Name
    this.sendKeys('input#team-search-h2h', casper.cli.get(0), {keepFocus: true});
    casper.page.sendEvent("keypress", casper.page.event.key.Down);

    this.click('ul.ui-autocomplete li.ui-menu-item:first-child a');

    // Opponent Name
    this.sendKeys('input#opponent-search-h2h', casper.cli.get(1), {keepFocus: true});
    casper.page.sendEvent("keypress", casper.page.event.key.Down);
    casper.page.sendEvent("keypress", casper.page.event.key.Down);
    casper.page.sendEvent("keypress", casper.page.event.key.Enter);
});
casper.then(function() {
    // Click Submit
    this.click('#submit-h2h');
});
casper.then(function() {
    console.log(casper.getPageContent());
});
casper.run();

```

**plots.R**

```

library(ggplot2)

#### Read CSV
nfl_data <- read.csv(file = '../data_collection/output.csv', sep = ',', header
  = TRUE)

#### Set Date as Date Type
nfl_data$date <- as.Date(nfl_data$date, "%b %d, %Y")
#### Take 20 Year Subset of Data
nfl_data <- subset(nfl_data, subset = nfl_data$date >= '1995-07-01')

#### Process Home Data Frame to Include Only Relevant Fields ####
nfl_home_df <- data.frame(home = nfl_data$home, pointsATS =
  nfl_data$home_score + nfl_data$home_spread - nfl_data$away_score, ats =
  nfl_data$ats)
#### Set ATS as Binary
nfl_home_df$ats <- ifelse(nfl_home_df$ats == "W", 1, 0)

#### Aggregate Data Frame According to Means
nfl_home_mean_df <- setNames(aggregate(nfl_home_df$pointsATS, by = list(
  nfl_home_df$home), FUN = mean), c("team", "mean"))

#### Make Bar Graph of Means
ggplot(data = nfl_home_mean_df, aes(x = reorder(team, -mean), y = mean)) +
  geom_bar(stat = "identity") +
  xlab("Home Team") +
  ylab("Mean Points ATS") +
  #ggtitle("Mean Points Above Spread for Home Teams") +
  coord_flip()
ggsave('images/home_team_mean.png')

#### Aggregate Data Frame According to Percents
nfl_home_percent_df <- setNames(aggregate(nfl_home_df$ats, by = list(
  nfl_home_df$home), FUN = mean), c("team", "percent"))

#### Make Bar Graph of Percents
ggplot(data = nfl_home_percent_df, aes(x = reorder(team, -percent), y =
  percent)) +
  geom_bar(stat = "identity") +
  xlab("Home Team") +
  ylab("Percent ATS") +
  ylim(0, 1) +
  #ggtitle("Percent Above Spread for Home Teams") +
  coord_flip()
ggsave('images/home_team_percent.png')

#### Process Away Data Frame to Include Only Relevant Fields ####
nfl_away_df <- data.frame(away = nfl_data$away, pointsATS =
  nfl_data$away_score - nfl_data$home_spread - nfl_data$home_score, ats =
  nfl_data$ats)
#### Set ATS as Binary

```

```

nfl_away_df$sats <- ifelse(nfl_away_df$sats == "L", 1, 0)

#### Aggregate Data Frame According to Means
nfl_away_mean_df <- setNames(aggregate(nfl_away_df$pointsATS, by = list(
  nfl_away_df$away), FUN = mean), c("team", "mean"))

#### Make Bar Graph of Means
ggplot(data = nfl_away_mean_df, aes(x = reorder(team, -mean), y = mean)) +
  geom_bar(stat = "identity") +
  xlab("Away Team") +
  ylab("Mean Points ATS") +
  #ggtitle("Mean Points Above Spread for Away Teams") +
  coord_flip()
ggsave('images/away_team_mean.png')

#### Aggregate Data Frame According to Percents
nfl_away_percent_df <- setNames(aggregate(nfl_away_df$sats, by = list(
  nfl_away_df$away), FUN = mean), c("team", "percent"))

#### Make Bar Graph of Percents
ggplot(data = nfl_away_percent_df, aes(x = reorder(team, -percent), y =
  percent)) +
  geom_bar(stat = "identity") +
  xlab("Away Team") +
  ylab("Percent ATS") +
  ylim(0, 1) +
  #ggtitle("Percent Above Spread for Away Teams") +
  coord_flip()
ggsave('images/away_team_percent.png')

#### Process Home and Away Data Frames for Means
nfl_home_sum_mean_df <- setNames(aggregate(nfl_home_df$pointsATS, by = list(
  nfl_home_df$home), FUN = sum), c("team", "sum"))
nfl_away_sum_mean_df <- setNames(aggregate(nfl_away_df$pointsATS, by = list(
  nfl_away_df$away), FUN = sum), c("team", "sum"))

nfl_home_length_df <- setNames(aggregate(nfl_home_df$pointsATS, by = list(
  nfl_home_df$home), FUN = length), c("team", "length"))
nfl_away_length_df <- setNames(aggregate(nfl_away_df$pointsATS, by = list(
  nfl_away_df$away), FUN = length), c("team", "length"))

nfl_total_mean_df <- data.frame(team = nfl_home_sum_mean_df$team, mean = (
  nfl_home_sum_mean_df$sum + nfl_away_sum_mean_df$sum)/(
  nfl_home_length_df$length + nfl_away_length_df$length))

nfl_total_mean_df <- nfl_total_mean_df[order(nfl_total_mean_df$mean),]

#### Make Bar Graph of Means
ggplot(data = nfl_total_mean_df, aes(x = reorder(team, -mean), y = mean)) +
  geom_bar(stat = "identity") +
  xlab("Team") +
  ylab("Mean Points ATS") +
  #ggtitle("Mean Points Above Spread for Teams") +
  coord_flip()

```

```
ggsave('images/team_mean.png')

#### Process Home and Away Data Frames for Percents
nfl_home_sum_percent_df <- setNames(aggregate(nfl_home_df$ats, by = list(
  nfl_home_df$home), FUN = sum), c("team", "sum"))
nfl_away_sum_percent_df <- setNames(aggregate(nfl_away_df$ats, by = list(
  nfl_away_df$away), FUN = sum), c("team", "sum"))

nfl_total_percent_df <- data.frame(team = nfl_home_sum_percent_df$team,
  percent = (nfl_home_sum_percent_df$sum + nfl_away_sum_percent_df$sum)/(
  nfl_home_length_df$length + nfl_away_length_df$length))

nfl_total_percent_df <- nfl_total_percent_df[order(
  nfl_total_percent_df$percent),]

#### Make Bar Graph of Percents
ggplot(data = nfl_total_percent_df, aes(x = reorder(team, -percent), y =
  percent)) +
  geom_bar(stat = "identity") +
  xlab("Team") +
  ylab("Percent ATS") +
  ylim(0, 1) +
  #ggtitle("Percent Above Spread for Teams") +
  coord_flip()
ggsave('images/team_percent.png')
```

**t-test.R**

```

#### Read CSV
nfl_data <- read.csv(file = '../data_collection/output.csv', sep = ',', header
  = TRUE)

#### Set Date as Date Type
nfl_data$date <- as.Date(nfl_data$date, "%b %d, %Y")
#### Take 20 Year Subset of Data
nfl_data <- subset(nfl_data, subset = nfl_data$date >= '1995-07-01')

#### Process Home Data Frame to Include Only Relevant Fields ####
nfl_home_df <- data.frame(team = nfl_data$home, pointsATS =
  nfl_data$home_score + nfl_data$home_spread - nfl_data$away_score, ats =
  nfl_data$ats)
#### Set ATS as Binary
nfl_home_df$ats <- ifelse(nfl_home_df$ats == "W", 1, 0)

#### Process Away Data Frame to Include Only Relevant Fields ####
nfl_away_df <- data.frame(team = nfl_data$away, pointsATS =
  nfl_data$away_score - nfl_data$home_spread - nfl_data$home_score, ats =
  nfl_data$ats)
#### Set ATS as Binary
nfl_away_df$ats <- ifelse(nfl_away_df$ats == "L", 1, 0)

#### List of Teams
teams <- data.frame(unique(nfl_data$home));

#### T-tests for Mean Points Above the Spread
for(i in seq(1, nrow(teams), by = 1)){
  home_subset <- subset(nfl_home_df, subset = nfl_home_df$team == teams[i,1])
  away_subset <- subset(nfl_away_df, subset = nfl_away_df$team == teams[i,1])

  team_games <- rbind(home_subset, away_subset)
  test <- t.test(x = team_games$pointsATS, mu = 0, alternative = c("two.sided
    "), conf.level = 0.95)

  if(test$p.value < 0.05) {
    print(teams[i,1])
    print(test)
  }
}

#### Z Test for Proportion of Time Beating the Spread
nfl_home_length_df <- setNames(aggregate(nfl_home_df$pointsATS, by = list(
  nfl_home_df$home), FUN = length), c("team", "length"))
nfl_away_length_df <- setNames(aggregate(nfl_away_df$pointsATS, by = list(
  nfl_away_df$away), FUN = length), c("team", "length"))

nfl_home_sum_percent_df <- setNames(aggregate(nfl_home_df$ats, by = list(
  nfl_home_df$home), FUN = sum), c("team", "sum"))
nfl_away_sum_percent_df <- setNames(aggregate(nfl_away_df$ats, by = list(
  nfl_away_df$away), FUN = sum), c("team", "sum"))

```

```
for(i in seq(1, nrow(teams), by = 1)){
  test <- prop.test(x = nfl_home_sum_percent_df$sum[i] +
    nfl_away_sum_percent_df$sum[i], n = nfl_home_length_df$length[i] +
    nfl_away_length_df$length[i], p = rep(0.5, times = 32)[i], alternative =
    "two.sided", conf.level = 0.95)
  if(test$p.value > 0.05) {
    print(teams[i,1])
    print(test)
  }
}
```