

Football Capstone

Vasu

Saturday, December 10, 2016

Football Statistics

Introduction

I will be analyzing past National Football League statistics to predict who will be the top players who play the running back position. The running back position gets the football handed to them and they run as far as they can behind their teammates that block the defense for them. You will see the term rush in this report, which means run. The number of yards rushing is how many yards the running back ran. This can be useful for people that take part in fantasy football leagues.

Load necessary packages

Load the data

```
setwd("C:/Users/Vasu/data")
data <- read.csv("pbp-2015.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE)
```

Mutate some of the data

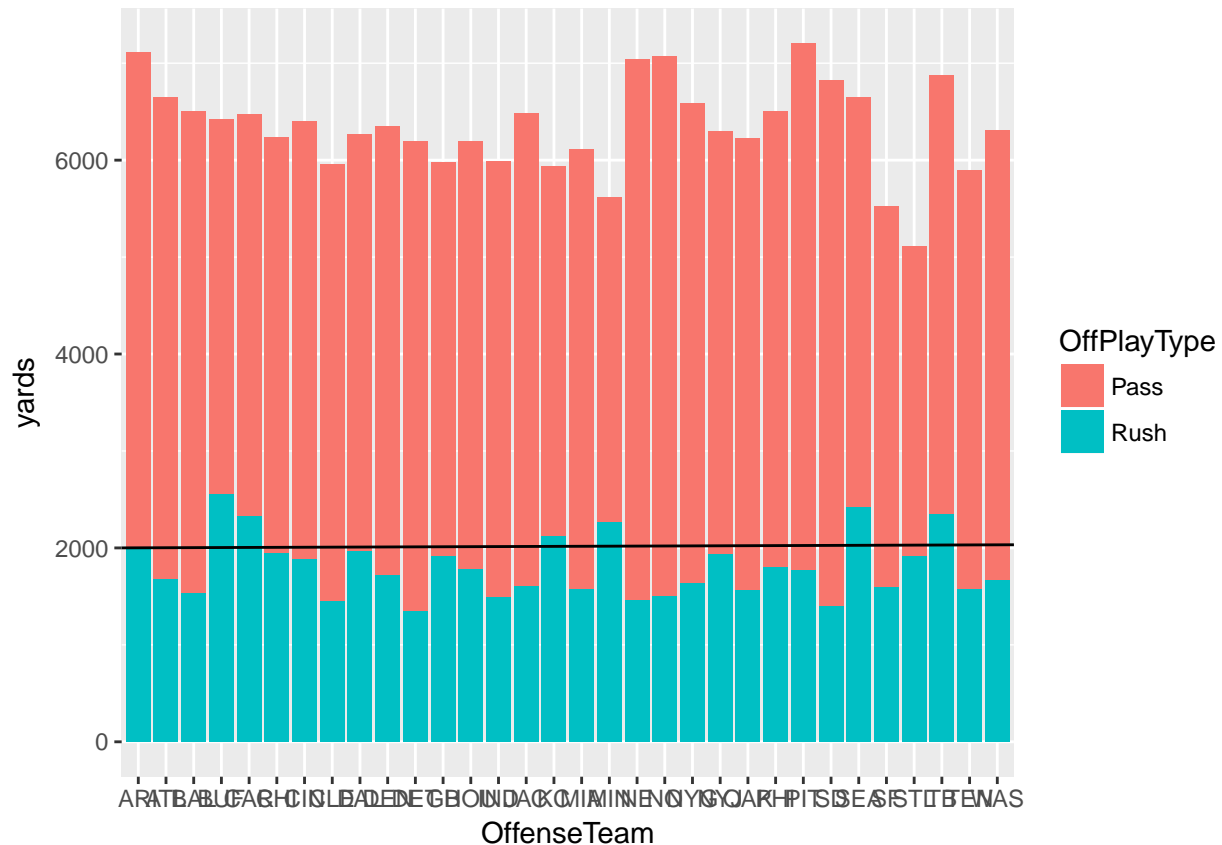
Create Date variable from string in data set. We will also only filter the plays that were rushing or passing plays.

```
data$date1 <- as.Date(data$GameDate)
dat2 <- data %>% filter(IsRush == 1 | IsPass == 1)
```

Total Yards Rushing vs Total Yards Passing

Add a new variable called OffPlayType to determine if its a running(Rush) or a passing play. I will group the total yards per team and OffPlayType. Then, create a stacked bar chart to show the yards per team and broken down by the OffPlayType. As you can see, there were 6 teams that had over 2000 yards rushing during the season. These 6 teams were Buffalo, Carolina, Kansas City, Minnesota, Seattle and Tampa Bay

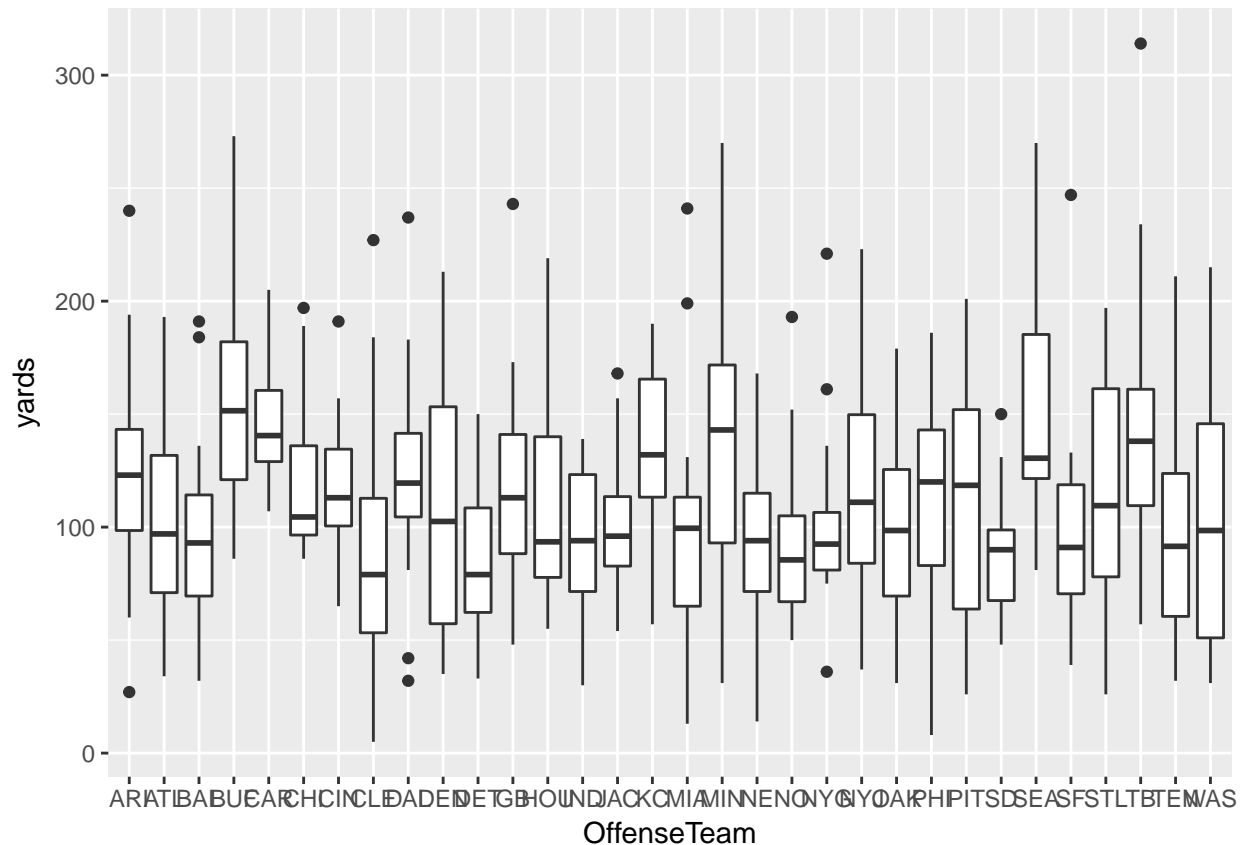
```
dat2 <- mutate(dat2, OffPlayType = ifelse(IsRush == 1, "Rush", ifelse(IsPass == 1, "Pass", "NA")))
dat2 <- dat2 %>% group_by(OffenseTeam, OffPlayType) %>% summarise(yards = sum(Yards))
ggplot(dat2, aes(x=OffenseTeam, y = yards, fill = OffPlayType)) + geom_bar(stat="identity") + geom_abline
```



Distribution of rushing yards

The following is a box plot of all of the teams and how many yards rushing occurred during the season. The plot shows that of the teams that had the most rushing yards, Buffalo and Seattle had more games above their median than below.

```
dat6 <- data %>% filter(IsRush == 1 & OffenseTeam != '')
dat6 <- dat6 %>% group_by(date1, OffenseTeam, DefenseTeam) %>% summarise(yards = sum(Yards))
ggplot(dat6, aes(x=OffenseTeam, y=yards)) + geom_boxplot()
```



Rushing Yards per quarter

The trend for most of the top rushing teams is that the rushing yards goes down from the 1st quarter to the 2nd quarter. The only team that this didn't happen for was Kansas City.

```
dat7 <- data %>% filter(IsRush == 1 & OffenseTeam != '' & Quarter != 5)
dat7 <- dat7 %>% group_by(OffenseTeam, Quarter) %>% summarise(yards = sum(Yards))
dat7 <- dat7 %>% filter(OffenseTeam == "BUF" | OffenseTeam == "CAR" | OffenseTeam == "KC" | OffenseTeam
ggplot(dat7, aes(x=Quarter, y=yards, color = OffenseTeam)) + geom_line()
```

