

MLB Analysis

15 October, 2020

Contents

Load MLB Data	1
Format data to find the averages for each MLB team	1
Create graph showing average payroll and wins	1
Display Graph	2
After looking at data it is clear that using averages there is a trend of higher payroll teams winning more	3

Load MLB Data

```
library(tidyverse)
library(mdsr)
MLB <- MLB_teams
```

Format data to find the averages for each MLB team

```
library(data.table)
avgWPct <- setDT(MLB)[, mean(WPct), by = teamID]
avgPayroll <- setDT(MLB)[, mean(payload), by = teamID]

averageMLB <- data.frame(teamID = avgPayroll$teamID,
  averagePayroll = avgPayroll$V1, averageWinPercent = avgWPct$V1)
options(scipen = 10000)
```

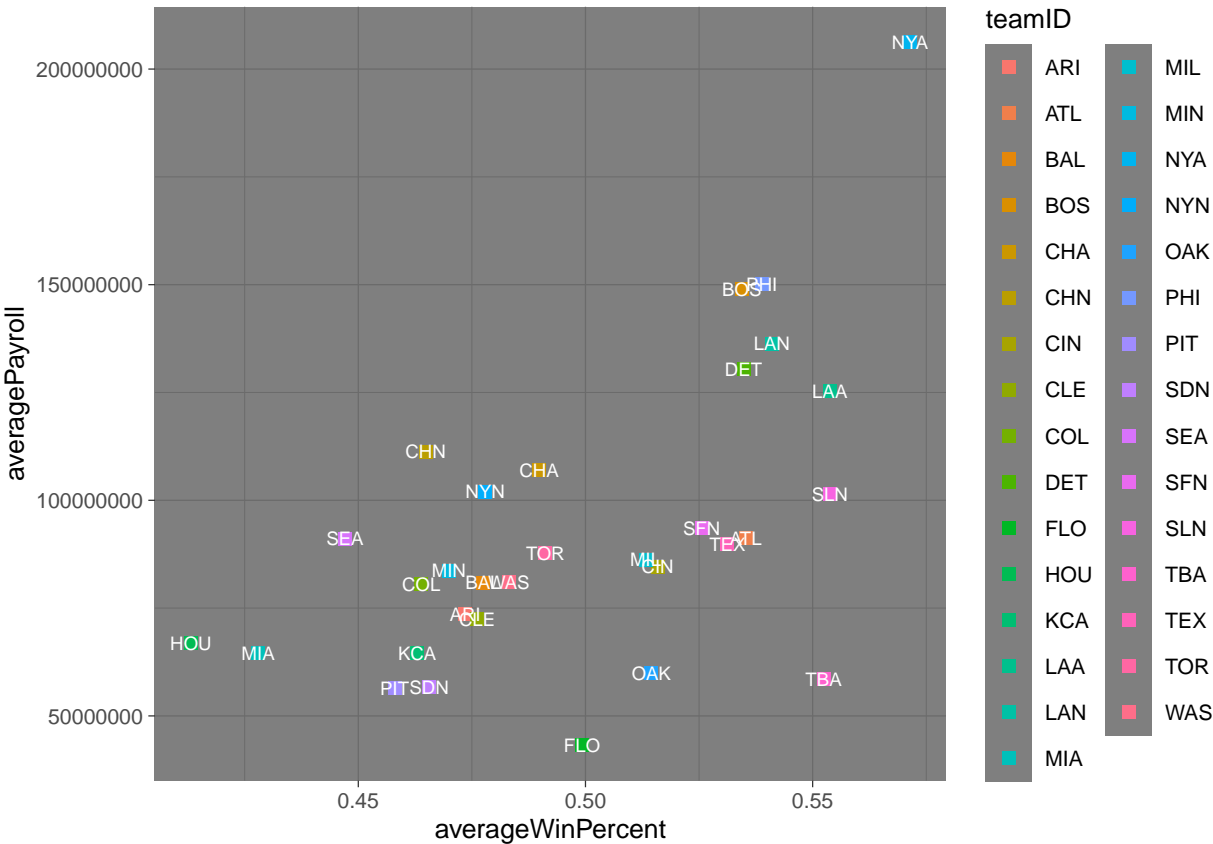
Create graph showing average payroll and wins

```
averageGraph <- ggplot(averageMLB, aes(x = averageWinPercent,
  y = averagePayroll, color = teamID)) + theme_dark(base_size = 10) +
  geom_point(size = 2, shape = 15) + geom_text(size = 2.5,
  label = averageMLB$teamID, color = "white")

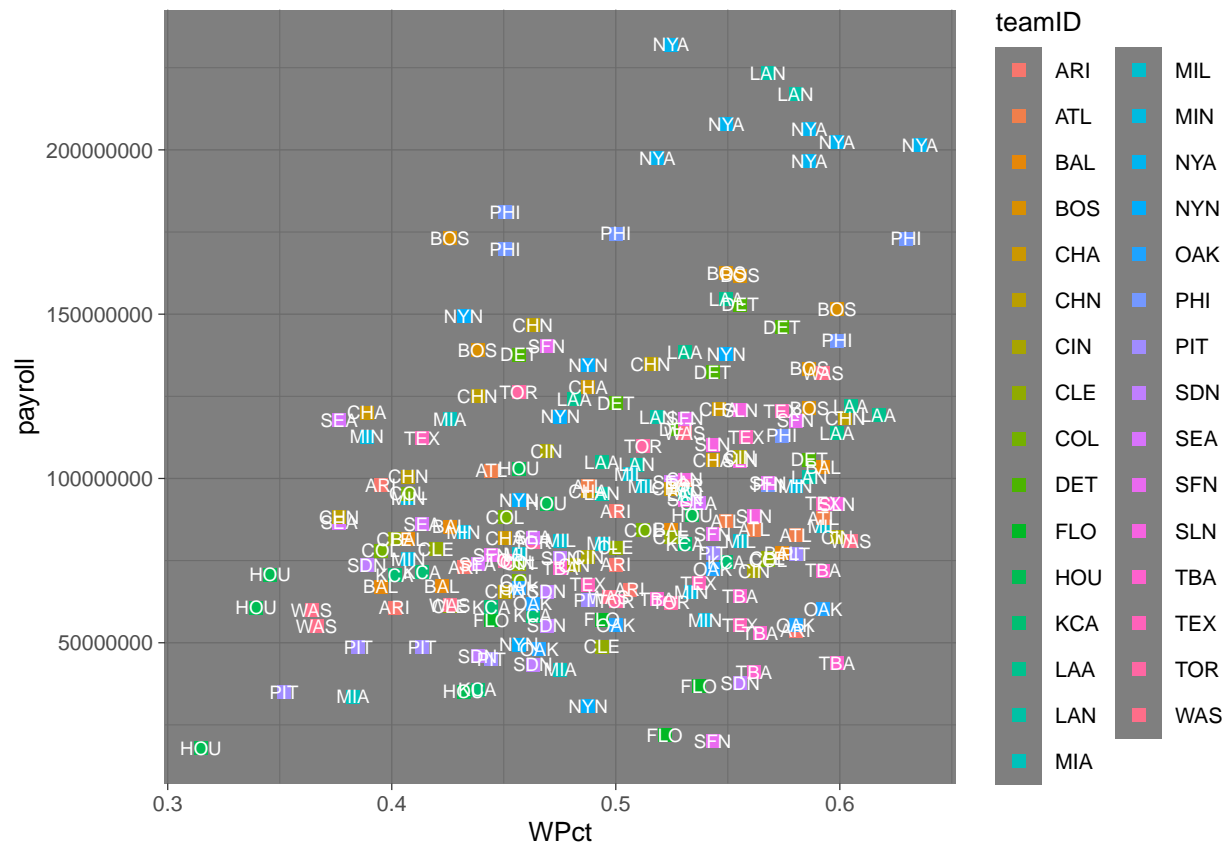
# create graph showing total payroll
totalGraph <- ggplot(MLB_teams, aes(x = WPct, y = payroll,
  color = teamID)) + theme_dark(base_size = 10) +
  geom_point(size = 2, shape = 15) + geom_text(size = 2.5,
  label = MLB_teams$teamID, color = "white")
```

Display Graph

averageGraph



totalGraph



After looking at data it is clear that using averages there is a trend of higher payroll teams winning more