NBA

Narcel Reedus

October 11, 2017

## Load libraries into RStudio

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(magrittr)

## Warning: package 'magrittr' was built under R version 3.4.2

library(ggplot2)

## Read data

NBA\_train <- read.csv("C:/Users/narce/OneDrive/Documents/GitHub/NBA/NBA/NBA\_train.csv")  
  
NBA <-NBA\_train

## Examine the structure of the data (835 observations and 20 variables)

str(NBA)

## 'data.frame': 835 obs. of 20 variables:  
## $ SeasonEnd: int 1980 1980 1980 1980 1980 1980 1980 1980 1980 1980 ...  
## $ Team : Factor w/ 37 levels "Atlanta Hawks",..: 1 2 5 6 8 9 10 11 12 13 ...  
## $ Playoffs : int 1 1 0 0 0 0 0 1 0 1 ...  
## $ W : int 50 61 30 37 30 16 24 41 37 47 ...  
## $ PTS : int 8573 9303 8813 9360 8878 8933 8493 9084 9119 8860 ...  
## $ oppPTS : int 8334 8664 9035 9332 9240 9609 8853 9070 9176 8603 ...  
## $ FG : int 3261 3617 3362 3811 3462 3643 3527 3599 3639 3582 ...  
## $ FGA : int 7027 7387 6943 8041 7470 7596 7318 7496 7689 7489 ...  
## $ X2P : int 3248 3455 3292 3775 3379 3586 3500 3495 3551 3557 ...  
## $ X2PA : int 6952 6965 6668 7854 7215 7377 7197 7117 7375 7375 ...  
## $ X3P : int 13 162 70 36 83 57 27 104 88 25 ...  
## $ X3PA : int 75 422 275 187 255 219 121 379 314 114 ...  
## $ FT : int 2038 1907 2019 1702 1871 1590 1412 1782 1753 1671 ...  
## $ FTA : int 2645 2449 2592 2205 2539 2149 1914 2326 2333 2250 ...  
## $ ORB : int 1369 1227 1115 1307 1311 1226 1155 1394 1398 1187 ...  
## $ DRB : int 2406 2457 2465 2381 2524 2415 2437 2217 2326 2429 ...  
## $ AST : int 1913 2198 2152 2108 2079 1950 2028 2149 2148 2123 ...  
## $ STL : int 782 809 704 764 746 783 779 782 900 863 ...  
## $ BLK : int 539 308 392 342 404 562 339 373 530 356 ...  
## $ TOV : int 1495 1539 1684 1370 1533 1742 1492 1565 1517 1439 ...

## Variables definitions

## SeasonEnd: Year the season ended

## Team: Name of team

## Playoffs: Binary variable for playoff appearance

## W: Wins (regular season)

## PTS: Points scored (regular season)

## oppPTS:Oopponent points scored (regular season)

## FG, FGA: Field goals (including three pointers)

## X2P, X2PA: 2-pointers

## X3P, X3PA: 3-pointers

## FT, FTA: Free throws

## ORB, DRB: Offensive and defensive rebounds

## AST: Assists

## STL: Steals

## BLK: Blocks

## TOV: Turnovers

## Determine how many regular season wins needed to make playoffs

## by grouping the data by the number of wins and

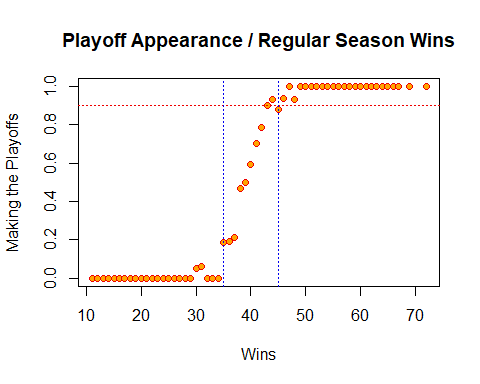
## creating three new features: number of playoff appearances,

## and the percentage of playoff appearances fracPO divided by total number of wins

tmp<- group\_by(NBA, W) %>% summarise(nTot = n(), nPO = sum(Playoffs), fracPO = nPO/nTot)  
  
View(tmp)

## Plot y=tmpfracPO

plot(tmp$W, tmp$fracPO, pch = 21, col = "red2", bg = "orange",   
 xlab = "Wins", ylab = "Making the Playoffs", main = "Playoff Appearance / Regular Season Wins")  
abline(h = 0.9, lty = 3, col = "red2")  
abline(v = 35, lty = 3, col = "blue2")  
abline(v = 45, lty = 3, col = "blue2")

 ## Results show that (45>) wins have a (90>) chance of making it to the playoffs ## Predicting wins by calculating the difference between points scored (PTS) and points allowed (oppPTS)

NBA$PTSdiff <- NBA$PTS - NBA$oppPTS

## Check the linear relationship between PTSdiff and Wins

plot(NBAW, pch = 21, col = "red2", bg = "orange", xlab = "PTS Difference", ylab = "Wins", main = "Wins / Points Scored")

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.