EDA_A1_singhvis

Vishal Singh

September 1, 2019

1.

Why I want to learn EDA

I aspire to be a data scientist and good visualization skills are integral to being a good data scientist. I believe EDA STAT S-670 is a course that will help me in honing visualization and data analyzing skills. Exploratory data analysis is a good way to learn more about data and make preliminary conclusions about the trend, exploring data also helps us in developing new features(variable) for performing various analysis and predictive modeling. Throughout my work experience as a data analyst/data scientist I've experienced that having good visualization skills help communicate technical aspect of the study in layman terms to business professionals. It is essential in informing business decision from the results obtained through technical analysis.

2.

```r
setwd("C:\\Users\\singh\\Documents\\Acads\\EDA\\A1")
nba.df = read.csv("nba2017-18.csv")
pacers.df = read.csv("pacers2017-18.csv")
```

```r
library(tidyverse)
```

Data Transformation

player heights

```r
pacers.heights = strsplit(as.character(pacers.df$Ht), split = "-")
pacers.heights = matrix(unlist(pacers.heights), byrow = T, ncol = 2)
feet = as.numeric(pacers.heights[, 1])
inches = as.numeric(pacers.heights[, 2])
Height = 12 * feet + inches
```
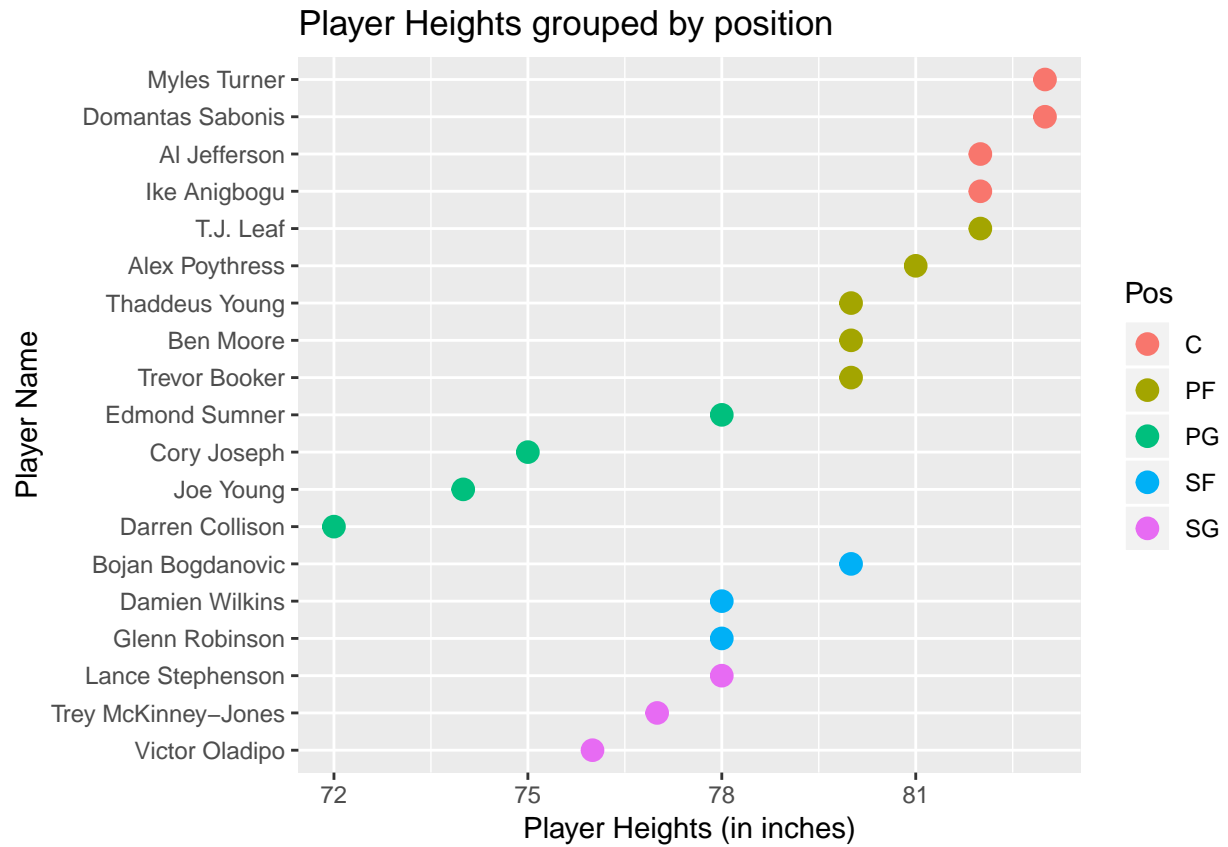
```r
pacers.df$Height = Height
```

player names

```r
pacers.names = strsplit(as.character(pacers.df$Player), "\\\\")
pacers.names = matrix(unlist(pacers.names), byrow = T, ncol = 2)
Name = pacers.names[, 1]
PlayerID = pacers.names[, 2]
pacers.df$Name = Name
pacers.df$PlayerID = PlayerID
```

```r
pacers.df$Name = factor(pacers.df$Name, levels = pacers.df$Name[order(pacers.df$Pos, -pacers.df$Height,
pacers.df = pacers.df[order(pacers.df$Pos),]
```

Plotting data

```r
ggplot(pacers.df, aes(Height, Name)) + geom_point(aes(color=Pos, group=Pos), stroke = 2) + ggtitle("Play
```

Player Heights grouped by position

The plot shows the player heights for different Indiana Pacers players, players of same position are shown with same color. The players of same position are together on the y-axis and are ordered by decreasing order of heights.