STAT-670-Assignment 1

Prashanth Thirukkurungudi Sekar (UID: 2000376633)

January 16, 2019

Worked with Pravin Sundar and Barathwaaj Parthasarathy

Loading the libraries

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.4.3
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 3.4.4
```

Reading the data

```
nba2017_18 <- read_csv("nba2017-18.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Player = col_character(),
##   Season = col_character(),
##   Tm = col_character(),
##   Lg = col_character(),
##   Ht = col_character(),
##   `FG%` = col_double(),
##   `2P%` = col_double(),
##   `3P%` = col_double(),
##   `eFG%` = col_double(),
##   `FT%` = col_double(),
##   `TS%` = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
pacers2017_18 = read_csv("pacers2017-18.csv")

## Warning: Missing column names filled in: 'X7' [7]

## Parsed with column specification:
## cols(
##   No. = col_integer(),
##   Player = col_character(),
##   Pos = col_character(),
##   Ht = col_character(),
##   Wt = col_integer(),
##   `Birth Date` = col_character(),
##   X7 = col_character(),
##   Exp = col_character(),
##   College = col_character()
## )
```

Cleaning the name variable

```
pacers2017_18 = data.frame(pacers2017_18)
pacers2017_18$Player = as.character(pacers2017_18$Player)
pacers2017_18$Player = str_split_fixed(pacers2017_18$Player, "\\\\", 2)[,1]
```

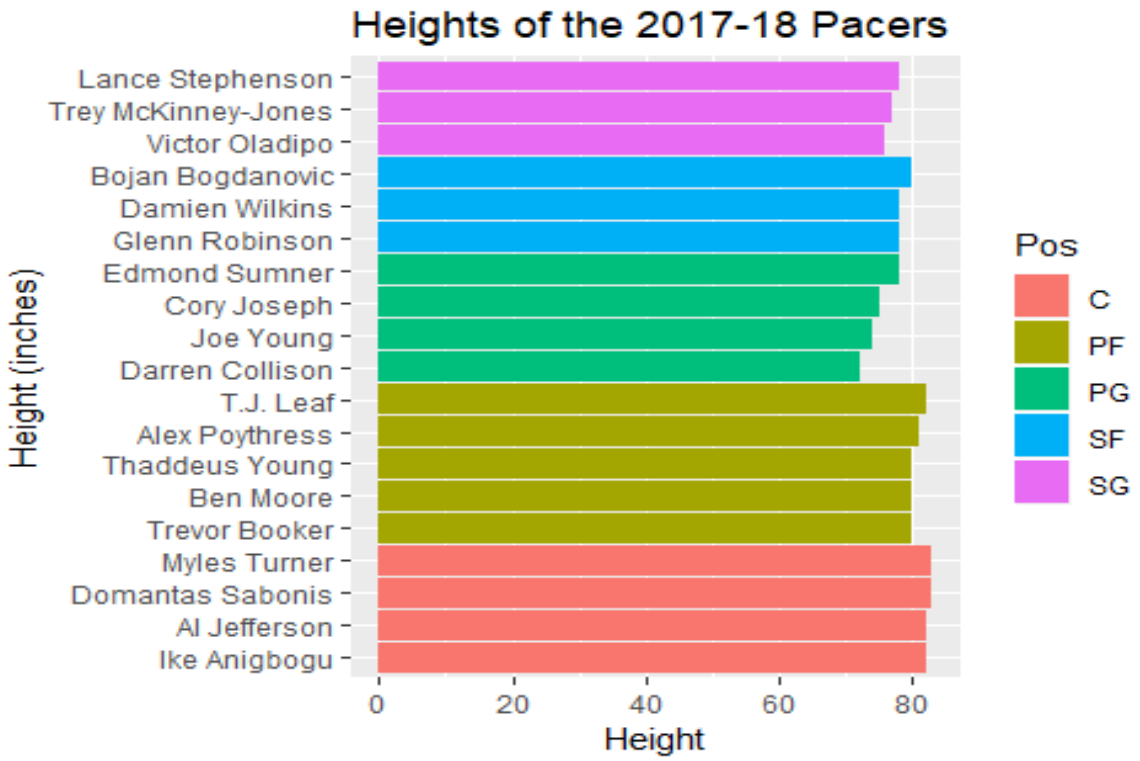
Creating height variable with inches

```
pacers2017_18$Height = as.numeric(substring(pacers2017_18$Ht,1,1)) * 12 + as.
numeric(substring(pacers2017_18$Ht,3,nchar(pacers2017_18$Ht)))
```

In order to improve scatter plot provided, a horizontal histogram visualization is used with the names in the y-axis and the heights (in inches) in the x-axis.

Improvisation of the plot

```
o = order(pacers2017_18$Height)
pacers2017_18$Player = factor(pacers2017_18$Player, levels = pacers2017_18$Pl
ayer[order(pacers2017_18$Pos,pacers2017_18$Height)])
gg = ggplot(pacers2017_18, aes(x = Player, y = Height, fill = Pos)) + geom_ba
r(stat="identity") + coord_flip()
gg + xlab("Height (inches)") + ggtitle("Heights of the 2017-18 Pacers")
```



The names are ordered by position first and then according to the heights within each position. Using this visualization, it is easier to observe that “Centers” are the tallest in the team and the “Guards” being the shortest in the team.