Problem Set 1

Raj Thakkar

15 January 2019

Using the files nba2017-18.csv and pacers2017-18.csv, draw a dot plot of the heights (in inches) and positions of all players who appeared for the Indiana Pacers during the 2017-18 season that improves on the plot below that we drew on Tuesday (for a different team.) "Improve" might mean grouping, better color choices, or some other improvement of your choice. All data cleaning must be done through R.

Note: I worked with Abhilash Kuhikar in the class

First we will load the dataset in R.

```
nba <- read.csv("nba2017-18.csv", header = TRUE)
pacers <- read.csv("pacers2017-18.csv", header = TRUE)
```

We will make the graph better by ordering on position and then also by heights. The paired palette had been used so that it's easy to distinguish between the colors used for various positions.

We can clearly see that players playing at the "Center" position are the tallest while the players playing at the "Point Guard" position are the shortest in Pacers basketball team.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

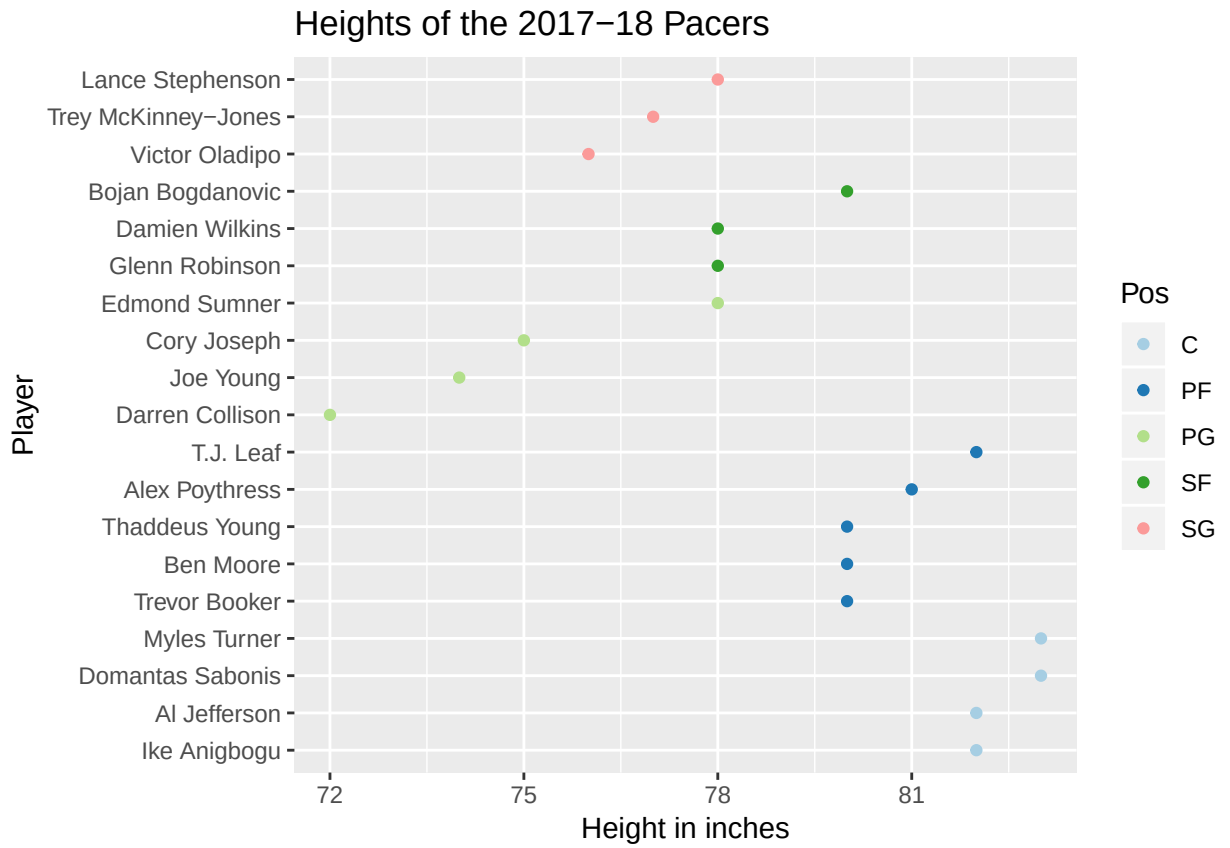
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(tidyr)
```

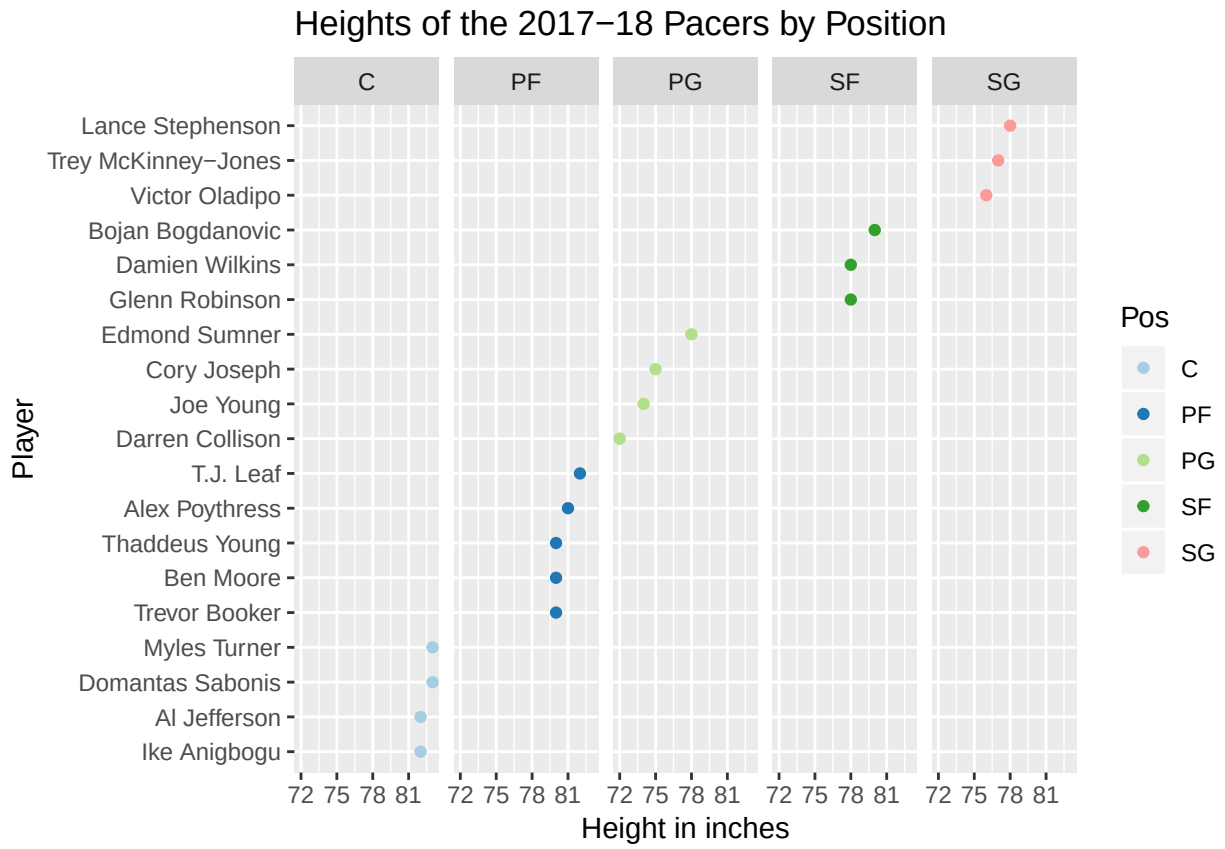
```
## Warning: package 'tidyr' was built under R version 3.5.2
```

```
# Converting the given height in inches

pacers_updated <- pacers %>% separate(Ht, c("Ht_feet", "Ht_inches"), sep = "-")
pacers_updated <- pacers_updated %>% separate(Player, c("Player", "player_id"),
    sep = "\\\\")
pacers_updated$Height <- as.numeric(pacers_updated$Ht_feet) * 12 + as.numeric(pacers_updated$Ht_inches)


o = order(pacers_updated$Pos, pacers_updated$Height)
pacers_updated$Player = factor(pacers_updated$Player, levels = pacers_updated$Player[o])
gg = ggplot(pacers_updated, aes(x = Height, y = Player, color = Pos)) + geom_point()

gg + scale_color_brewer(palette = "Paired") + xlab("Height in inches") + ggtitle("Heights of the 2017-18
```



I tried to imporve the graph further by adding facet grid based on player position. Seperating out the positions using facets leads to information loss as relationship between the position and height of basketball player is not as clear as it is in the graph above.

```
o = order(pacers_updated$Pos, pacers_updated$Height)
pacers_updated$Player = factor(pacers_updated$Player, levels = pacers_updated$Player[o])
gg = ggplot(pacers_updated, aes(x = Height, y = Player, color = Pos)) + geom_point()

gg + facet_grid(~Pos) + scale_color_brewer(palette = "Paired") + xlab("Height in inches") +
    ggtitle("Heights of the 2017-18 Pacers by Position")
```


Heights of the 2017−18 Pacers by Position
C
PF
PG
SF
SG
Lance Stephenson
Trey McKinney−Jones
Victor Oladipo
Bojan Bogdanovic
Damien Wilkins
Glenn Robinson
Edmond Sumner
Cory Joseph
Joe Young
Darren Collison
T.J. Leaf
Alex Poythress
Thaddeus Young
Ben Moore
Trevor Booker
Myles Turner
Domantas Sabonis
Al Jefferson
Ike Anigbogu
Player
72 75 78 81
Height in inches
Pos
C
PF
PG
SF
SG