---

title: "NBA Data"

author: "Oscar Sucre"

output: html_document

````{r setup, include=FALSE}

library(httr)

library(readxl)

library(tidyverse)

library(rvest)

library(plotly)

library(here)

```

## R Markdown

I scraped this data from data world from this link. https://data.world/gmoney/nba-players-birthplaces

````{r data,echo = FALSE, include=FALSE}

GET("https://query.data.world/s/tl43nqmrafxmxi34n6okb74lrzjvfp", write_disk(tf <- tempfile(fileext = ".xlsx")))

NBA <- read_excel(tf)

NBA

```

First off I needed to clean the data. The first thing I noticed where that there were mulitple rows that had no data just the column names so I filtered them out using a not equal operation.Then I had to simutaneously rename the coulmn names to be understandable and had to fix them from chars to doubles using as numeric, all by using mutate. Then for the Column that could stay as Char I just used

the rename function. Lastly there are many columns in the data set so I just chose the ones I believed to be the most important/useful.

```r
```{r cleaning data, echo = FALSE}
NBA_Clean <- NBA %>%
  filter(Player != "Player")%>%
  mutate(G = as.numeric(G), `FG%` =as.numeric(`FG%`), PTS...23 = as.numeric(PTS...23) )%>%
  rename(Games = G, PPG = PTS...23)%>%
  select(Player,Games,PPG,`FG%`,City,State)

NBA_Clean
```
```

This clean data set would be useful for whoever would want to do further research on NBA players from this era, being all American players who played before 2017.

```r
```{r csv file, echo =FALSE}
NBA_Clean%>%
  write_csv(here(paste0("NBAPlayerData", lubridate::today(), ".csv")))
```
```

This is where I manipulated the raw data to help me make some use of it and achieve my goal of finding the cities that produce quality NBA Players. There are many parts of a basketball player that can determine if they are skilled or not, but I figured that If I were to pick from only a few of the tracked statistics I would just make sure to disqualify the players who generally are useless on offense. Firstly the player must've played at least one full season in the NBA, which are 82 games long. Secondly the players need to average at least 10 points per game, which isn't very high for a player with decent NBA minutes, and while averaging .3 field goal percentage. .40 FG% is relatively low because 47.6% is league average so it if a player is only averaging 10 PPG they should at least be somewhat efficient to be considered a quality player.

Now that I had narrowed down the players my focus was to organize the data to a form that was organized by quality player per city. I organized it by City and had originally strugled to also include State until I remembered I could group by two columns. Using summarise function I made the most important

column of Players Per City to track the most successful cities that had born good NBA players. Then To be able to distinct from largest to smallest I arranged it descending.

```{r manipulating data,echo = FALSE}
NBA_Good_Players<- NBA_Clean%>%
  filter(Games > 82)%>%
  filter(PPG > 10 ,`FG%` > .40 )


NBA_Good_Players



NBA_City_Hotspots<-NBA_Good_Players%>%
  arrange(City)%>%
  group_by(City, State)%>%
  summarise(Players_Per_City = n())%>%
  filter(Players_Per_City >= 5)%>%
  arrange(desc(Players_Per_City))


NBA_City_Hotspots
```

I used ggplot to make a good visualization of the Players Per City variable. I also made the colors be for the state the city was in, if potentially some states had multiple city hotspots. Theme was used to make the data appealing and easy to understand.

```{r plotting results,echo=FALSE}
NBA_plot<-NBA_City_Hotspots%>%
  ggplot(aes(x = reorder(City, -Players_Per_City),Players_Per_City , fill= State))+
  geom_col(alpha =.8)+
  labs(title = "NBA Player City Hotspots",x = "Cities",y = "Number of Players")+
```

```
  theme(axis.text.x = element_text(size = 12, angle = 45, hjust = 1),

     axis.text.y = element_text(size = 15, hjust = 1),

     legend.position = "none")


NBA_plot
```

On my honor, I have neither received nor given any unauthorized assistance on this assignment.

Oscar Sucre