NFL Analysis

Miłosz Groch

18 04 2023

NFL Analysis Project

This is a data analysis project that uses NFL data from kaggle (https://www.kaggle.com/competitions/nfl-big-data-bowl-2023/data?select=players.csv) I have divided it into several parts:

1. Weight and height analysis for position

WEIGHT AND HEIGHT ANALYSIS FOR POSITION

I would like to see what the differences in height and weight look like between the positions and what can be deduced from them

Let's start by loading libraries and data

```
library(readxl)
```

```
Dane <- read_excel("C:/Users/groch/Desktop/Programowanie/R/nfl-big-data-bowl-2023/players2.xlsx")
```

Now I want to change the type of weight and height columns to numeric and convert height from inches to centimeters

```
Dane$weight <- as.numeric(Dane$weight)
```

```
Dane$height_inch <- as.numeric(gsub("-", ".", Dane$height)) * 12 + as.numeric(gsub(".*-", "", Dane$heigh
Dane$height_cm <- Dane$height_inch * 2.54
```

I want to create the list of positions to use it in the loop that creates separate datasets for each position

```
Positions <- unique(Dane$officialPosition)
Positions
```

```
##  [1] "QB"  "T"    "TE"  "WR"  "DE"  "SS"  "C"    "FS"  "NT"  "DT"  "CB"  "G"
## [13] "OLB" "RB"   "MLB" "ILB" "LB"   "FB"   "DB"
```

```
for(position in Positions){
  assign(paste0("Dane_", position), Dane[Dane$officialPosition == position,])
}
```

Next I used sapply() to create Dane_weight where we can store mean of weight for each position

```
Dane_weight <- sapply(Positions, function(x) mean(get(paste0("Dane_", x))$weight))
Dane_weight
```

```
##       QB       T       TE       WR       DE       SS       C       FS
## 220.7333 315.5410 250.0000 199.3667 272.6620 204.2319 307.6909 202.3611
##       NT       DT       CB       G       OLB       RB       MLB       ILB
## 321.6176 306.8087 193.1094 317.0707 248.4412 214.1905 233.1154 235.5679
```

```
##      LB      FB      DB
## 247.5000 245.7333 198.0000
```

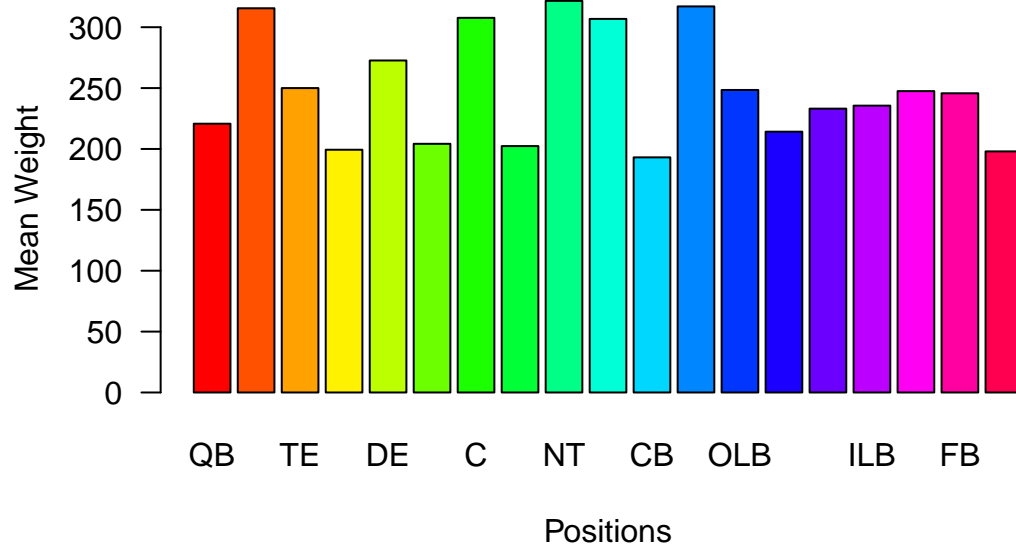We are ready to see the first barplot

Figure 1: A caption

For any NFL fan, these results should not come as a surprise, but we can see how relatively easy it is to analyze values of this type with this code. From my perspective, it's interesting to see the clear weight difference between QB and positions like WR or CB. The reason may be the likely average difference in height in favor of QB, which will be the subject of further analysis.

We use sapply() again this time to create Dane_height for mean of weight for each position

```
Dane_height <- sapply(Positions, function(x) mean(get(paste0("Dane_", x))$height_cm))
Dane_height
```

```
##       QB        T       TE       WR       DE       SS        C       FS
## 199.4510 215.0339 209.1575 192.7067 205.5468 186.8144 203.9112 187.7935
##       NT       DT       CB        G      OLB       RB      MLB      ILB
## 199.5036 200.4215 187.2948 207.7720 198.8282 187.5028 191.9068 191.8785
##       LB       FB       DB
## 196.8500 190.7574 181.8640
```

Let's display two barplots horizontal and next to each other so we can analyze them easier.

```
par(mfrow=c(1,2), mar=c(5, 7, 4, 2) + 0.1, las=1, oma=c(0, 0, 3, 0))
barplot(Dane_weight, names.arg=Positions, xlab="Positions", ylab="Mean Weight", horiz = TRUE,col = rain
barplot(Dane_height, names.arg=Positions, xlab="Positions", ylab="Mean height", horiz = TRUE,col = rain
```
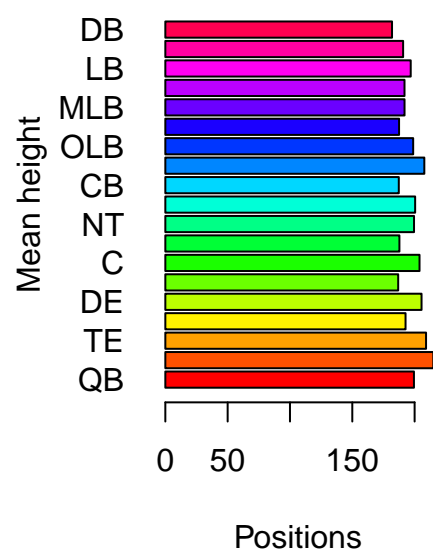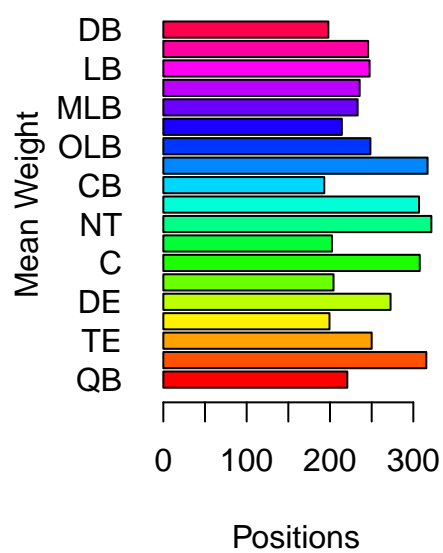
Figure 2: A caption

My earlier hypothesis that the higher weight of the quarterbacks than WR and CB may be due to greater height was supported by the second barplot. Another interesting thing that I noticed is a clear difference in height in favor of Tackle compared to the other positions of the Offensive Line, i.e. Guard and Center, despite the fact that the average weight is similar. For those more interested in the NFL, however, it is clear that the best players playing as Tackle are more athletic, and their height does not obscure the quaterback's field of view, as they are further to his flanks than the other linemen a second after the snap. Examples of great Tackle include: Trent Williams i Tyron Smith