FIFA21 EDA

Lee Kiat Kai

4/5/2021

```
# First, read in the dataset
fifa = read.csv("C:/Users/Kiat Kai/Desktop/NOTES/Y4S2/ST4248/Group Project/Data/FIFA21_clean_2.csv")

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
require(scales)
```

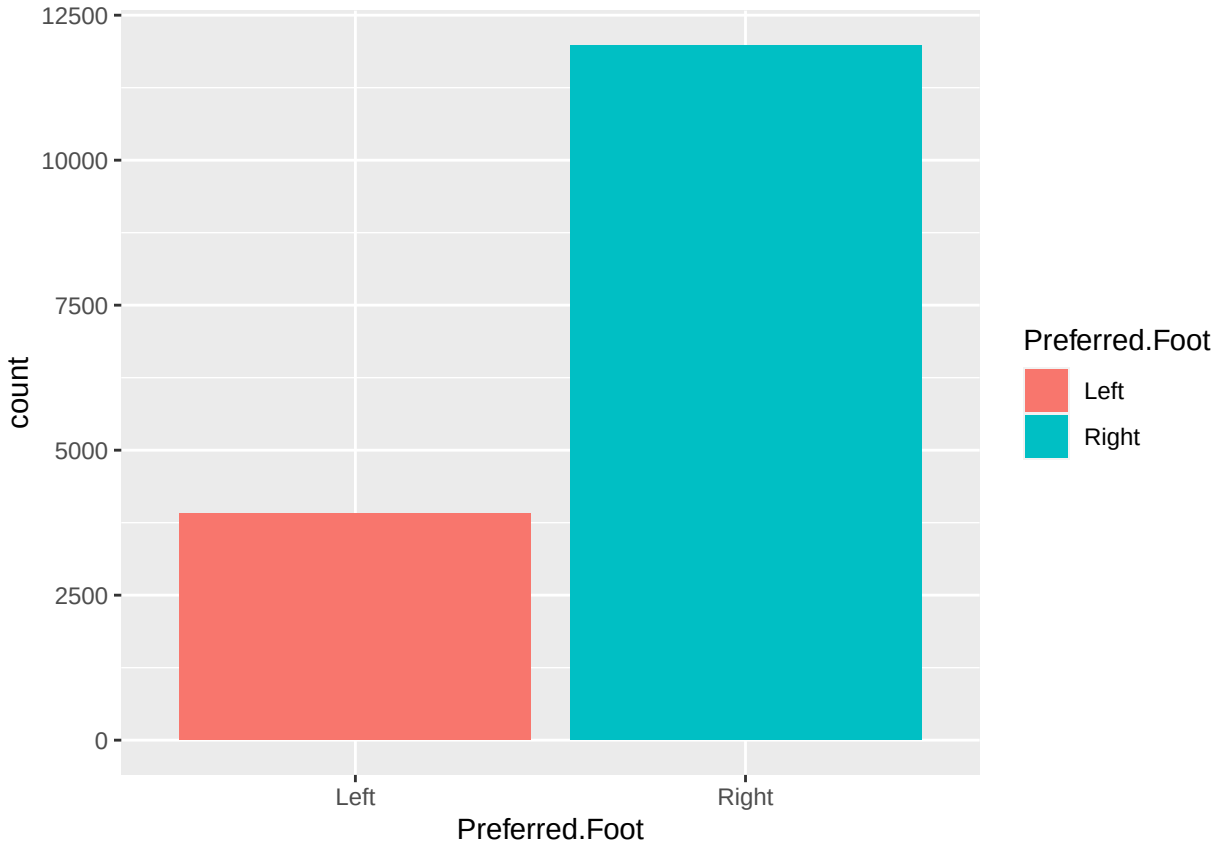
```
## Loading required package: scales

## Warning: package 'scales' was built under R version 3.6.3
```

```
library(fmsb)
```

Barplot of Preferred foot

```
# Barplot of preferred foot
ggplot(fifa, aes(Preferred.Foot, fill=Preferred.Foot)) + geom_bar()
```



Piechart of International Reputation

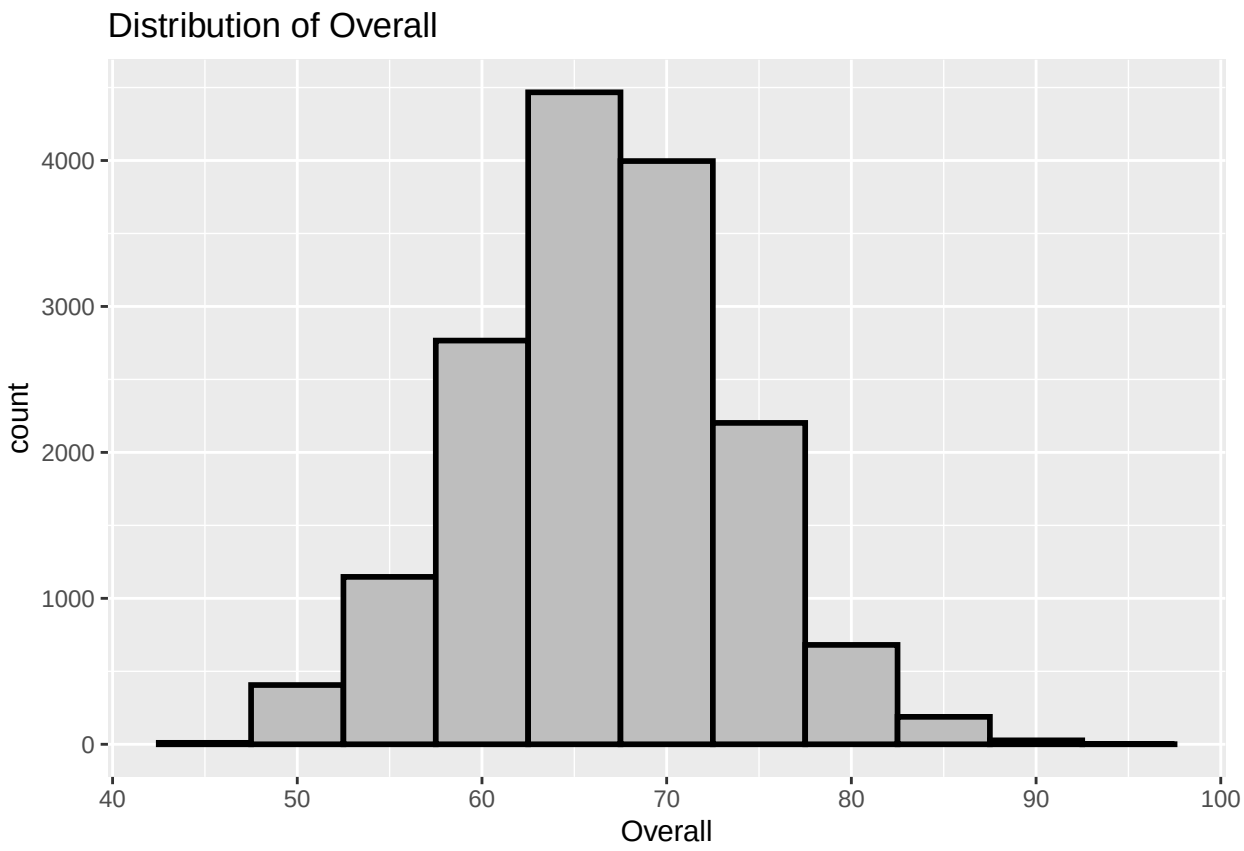
```
# Barplot of preferred foot
ggplot(fifa, aes(as.factor(International.Reputation),
                 fill=as.factor(International.Reputation))) +
               geom_bar(width = 1) + coord_polar(start=0) +
                 scale_fill_brewer(palette="Dark2")
```

Distribution of Wages

```
# Distribution of wage
wg = ggplot(fifa, aes(Wage)) + geom_histogram(binwidth=100000,fill="gray",color = "black",size=1) +
  labs(title="Distribution of Wage")
wg + scale_x_continuous(labels = comma)
```

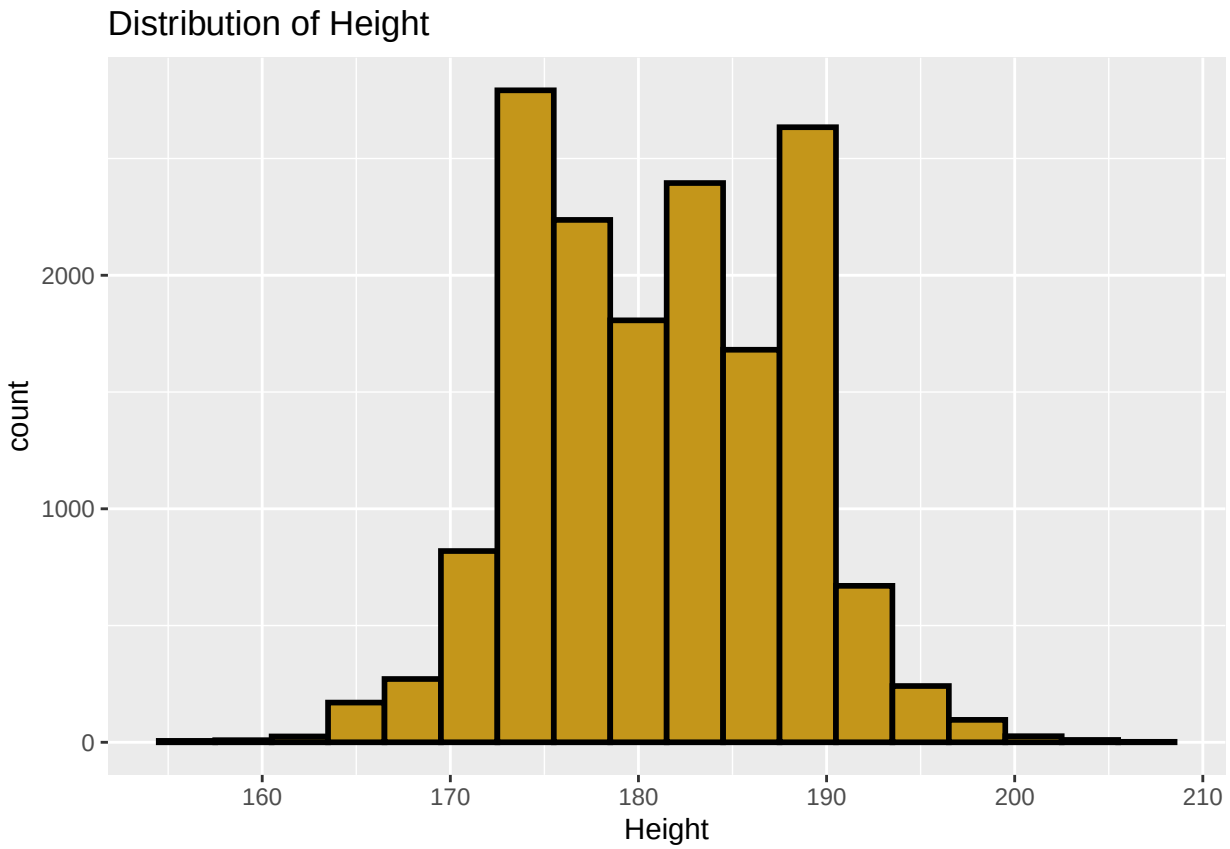
Distribution of Overall Rating

```
# Distribution of Overall
ggplot(fifa, aes(Overall)) + geom_histogram(binwidth=5,fill="gray",color = "black",size=1) +
  labs(title="Distribution of Overall")
```

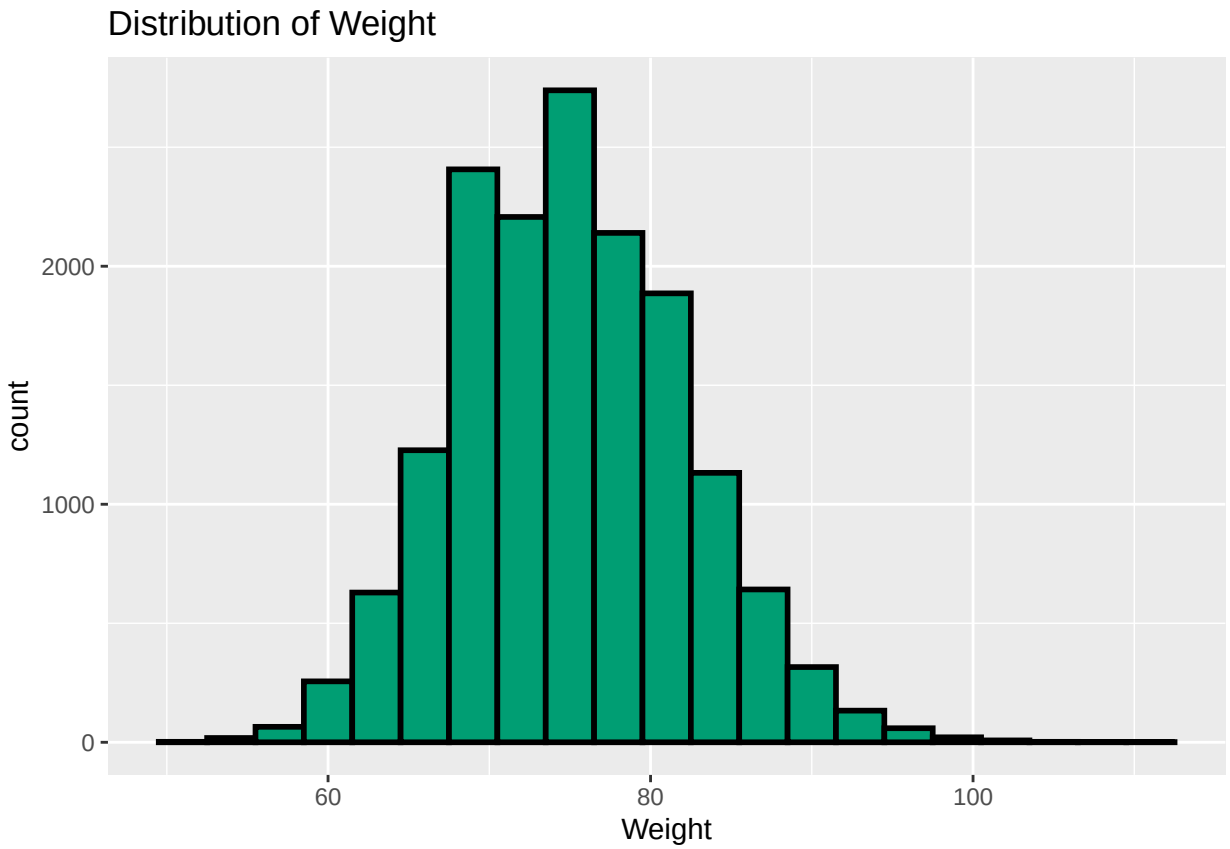


Distribution of Height and Weight

```
# Distribution of Height and weight
par(mfrow=c(2,1))
ggplot(fifa, aes(Height)) + geom_histogram(binwidth=3,fill="#C4961A",color = "black",size=1) +
  labs(title="Distribution of Height")
```

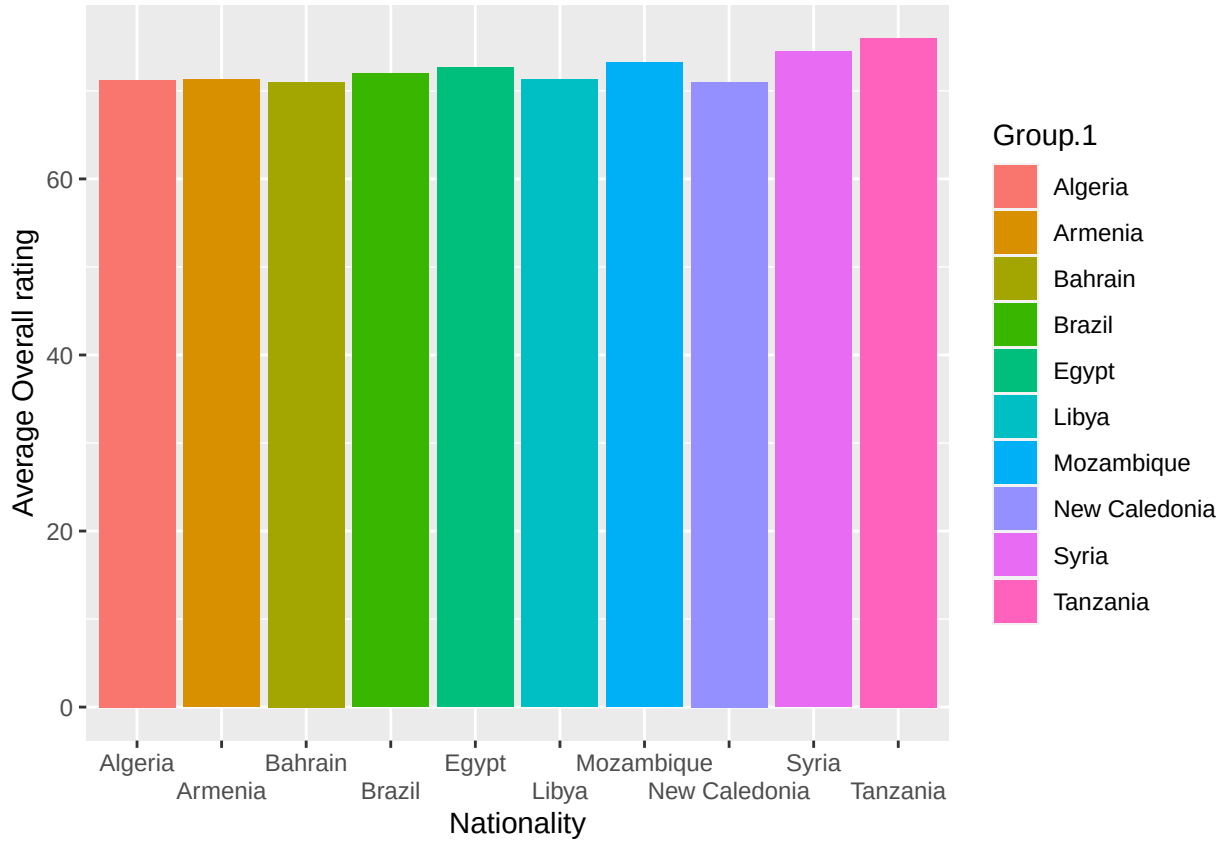



```
ggplot(fifa, aes(Weight)) + geom_histogram(binwidth=3,fill="#009E73",color = "black",size=1) +
  labs(title="Distribution of Weight")
```



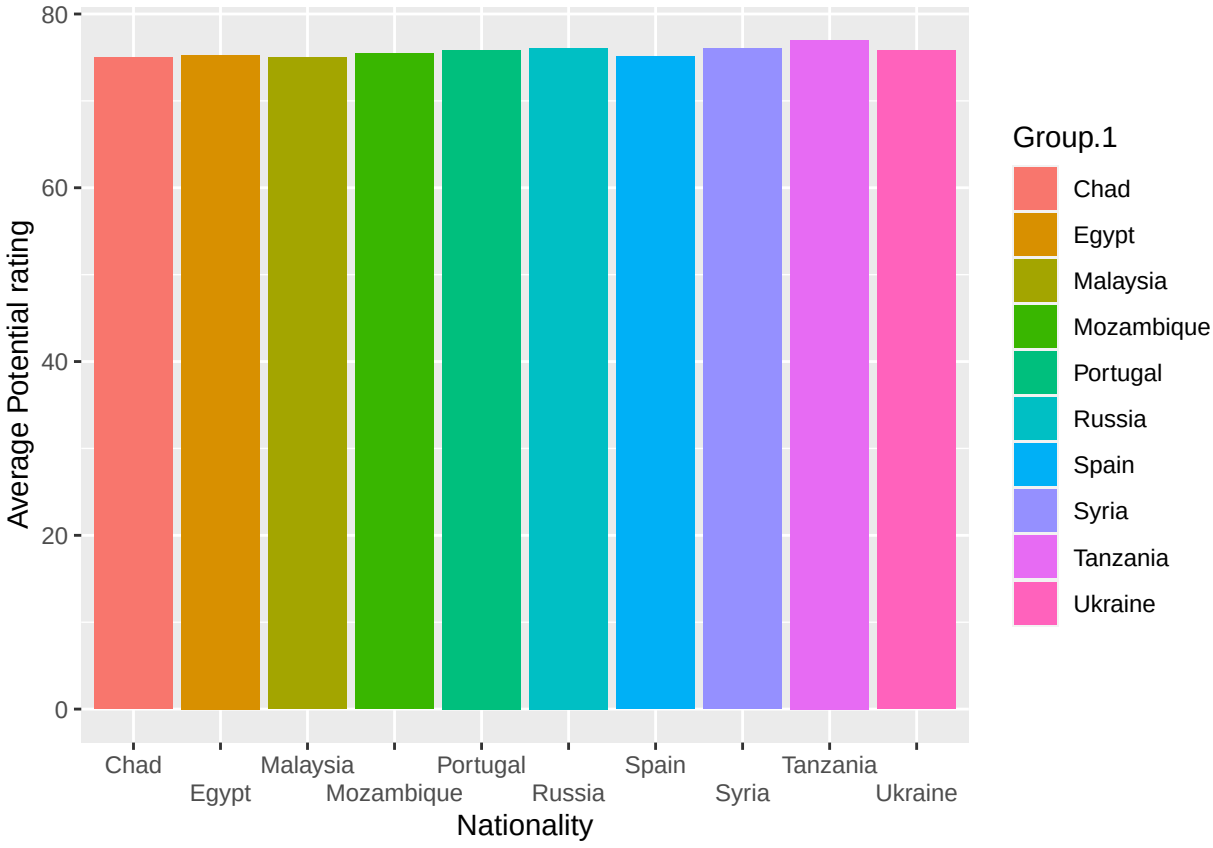
Average overall w.r.t Nationality

```
# Plot of average overall w.r.t Nationality
national.overall = aggregate(fifa$Overall, list(fifa$Nationality), mean)
national.overall = national.overall[order(-national.overall[,2]),]
national.overall = national.overall[1:10,]
ggplot(national.overall, aes(Group.1, x, fill=Group.1)) +
  geom_col() + labs(y="Average Overall rating", x="Nationality") +
  scale_x_discrete(guide = guide_axis(n.dodge=2))
```



Plot of average potential w.r.t Club

```
national.potential = aggregate(fifa$Potential, list(fifa$Nationality), mean)
national.potential = national.potential[order(-national.potential[,2]),]
national.potential = national.potential[1:10,]
ggplot(national.potential, aes(Group.1, x, fill=Group.1)) +
  geom_col(aes(fill = Group.1)) + labs(y="Average Potential rating", x="Nationality") + scale_x_discret
```



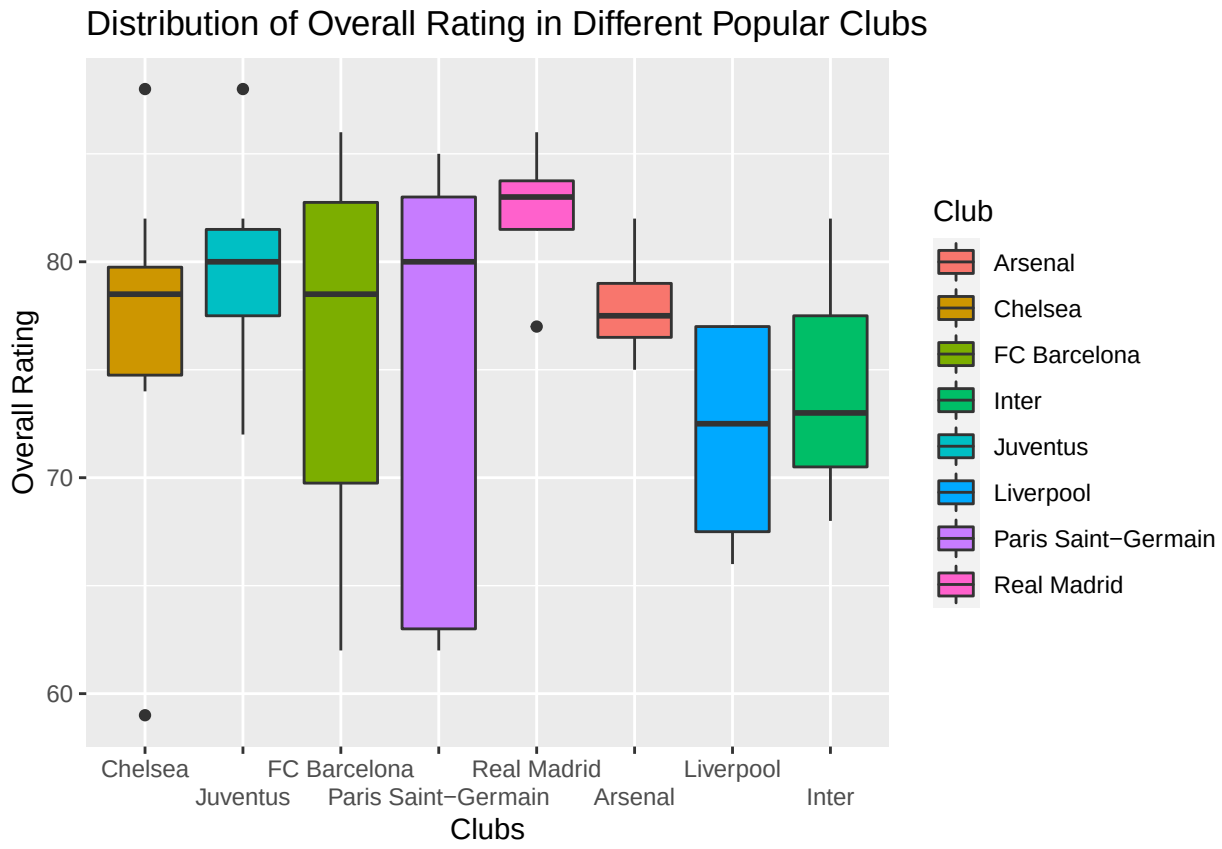
Plot of boxplot of overall w.r.t some Clubs that we may be interested in

```
clubs.interested.in = subset(fifa, Club == c("Chelsea","Liverpool","Juventus",
                                             "Inter","Totterham Hotspur",
                                             "FC Barcelona","Real Madrid",
                                             "Paris Saint-Germain","Arsenal"))
```

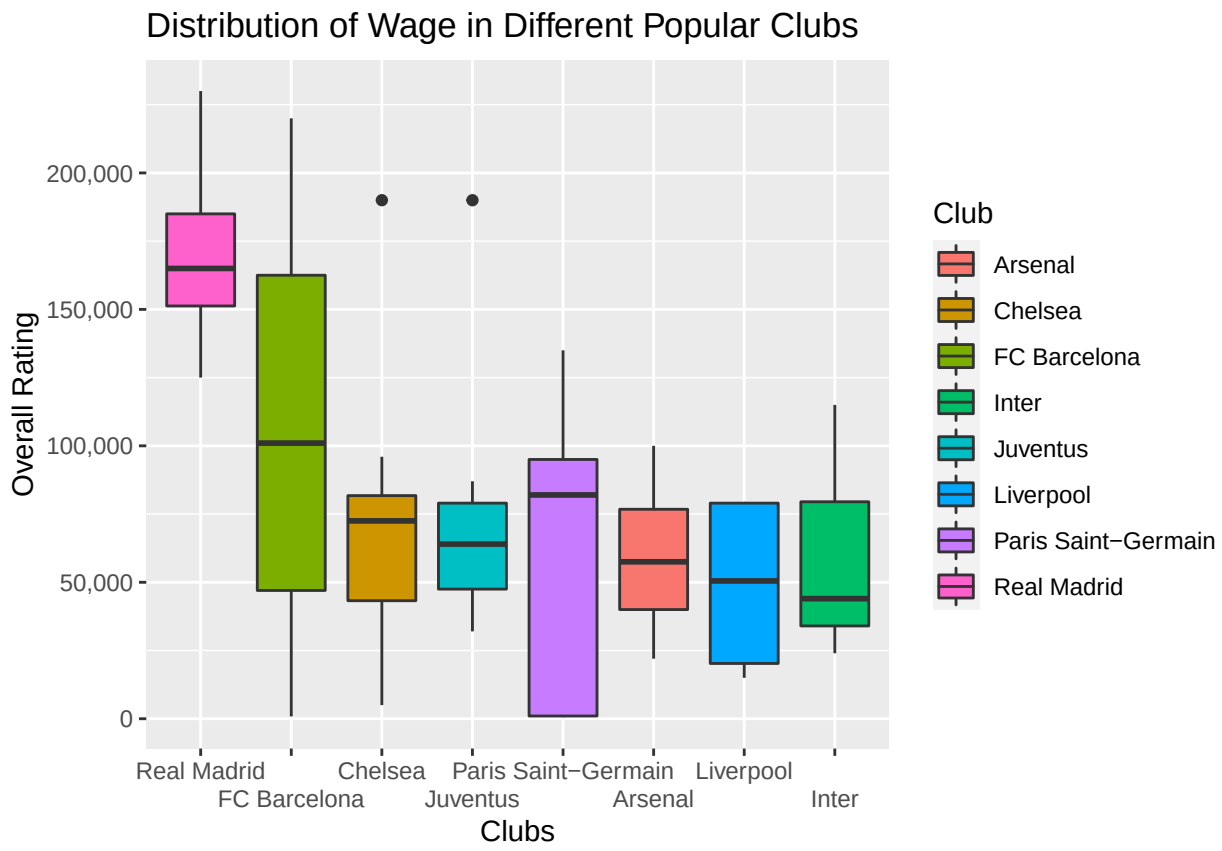
```
## Warning in `==.default`(Club, c("Chelsea", "Liverpool", "Juventus", "Inter", :
## longer object length is not a multiple of shorter object length
```

```
## Warning in is.na(e1) | is.na(e2): longer object length is not a multiple of
## shorter object length
```

```
ggplot(clubs.interested.in, aes(reorder(Club, -Overall, sum), Overall, fill=Club)) + geom_boxplot() + la
  scale_x_discrete(guide = guide_axis(n.dodge=2))
```

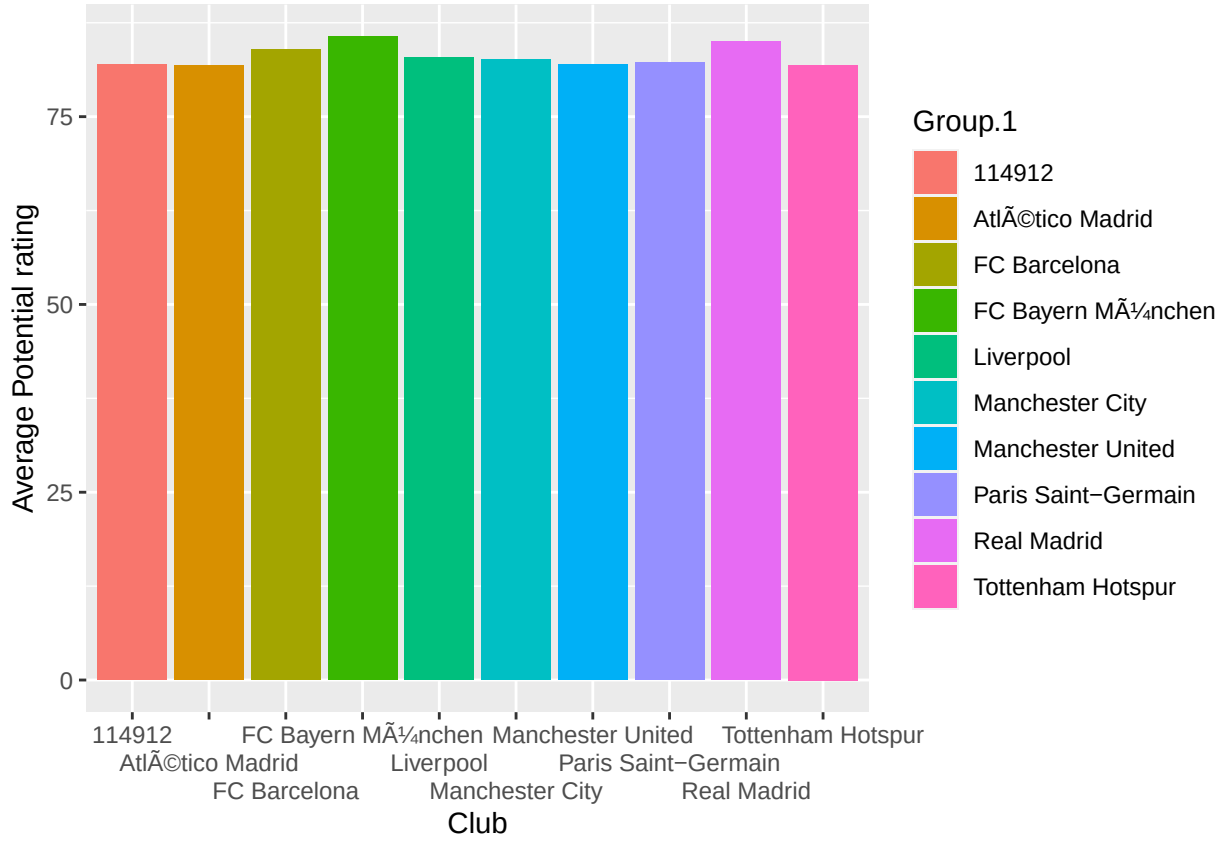



```
ggplot(clubs.interested.in, aes(reorder(Club, -Wage, sum), Wage, fill=Club)) +
  geom_boxplot() + labs(y="Overall Rating", x="Clubs", title="Distribution of Wage in Different Popular
```



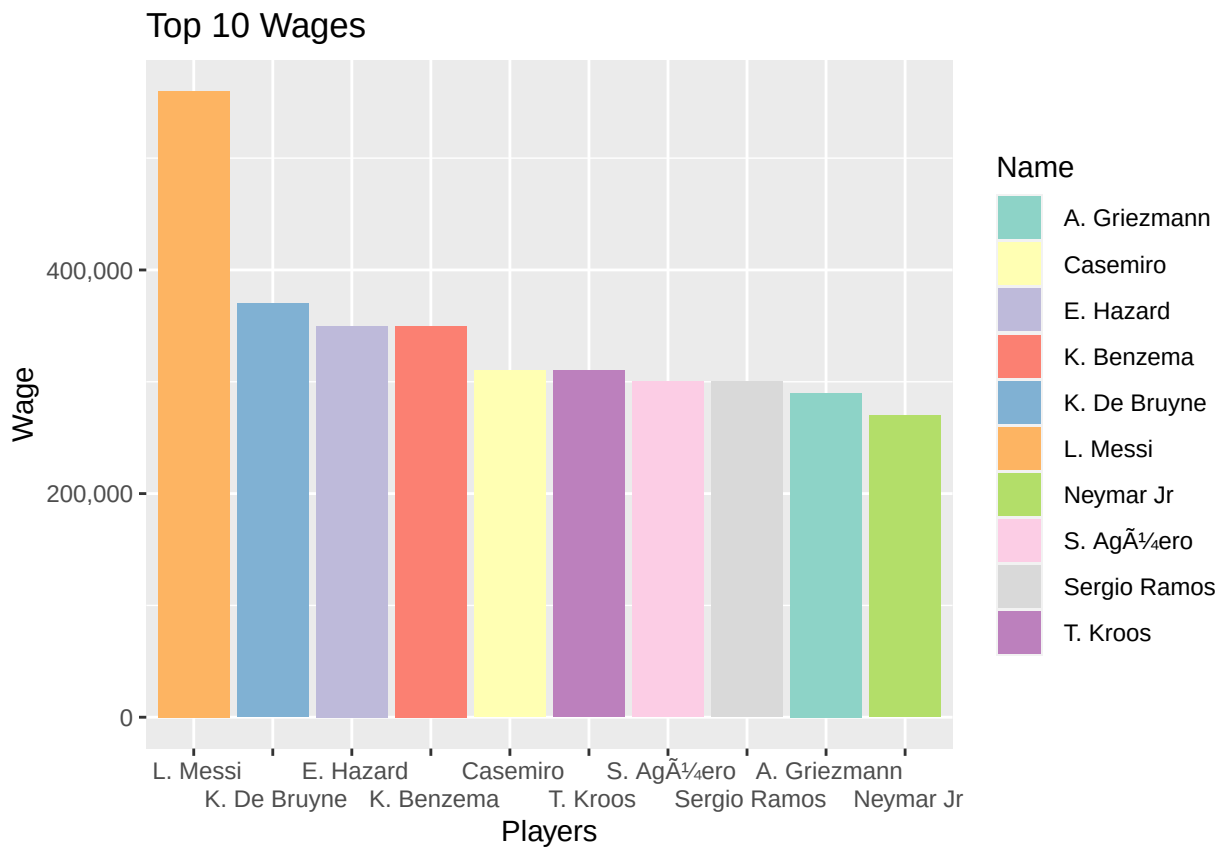
Plot of average potential w.r.t Clubs

```
club.potential = aggregate(fifa$Potential, list(fifa$Club), mean)
club.potential = club.potential[order(-club.potential[,2]),]
club.potential = club.potential[1:10,]
ggplot(club.potential, aes(Group.1, x, fill=Group.1)) +
  geom_col() + labs(y="Average Potential rating", x="Club") + scale_x_discrete(guide = guide_axis(n.dod
```



Top 10 earning wages of players currently

```
top.wages = fifa[order(-fifa$Wage),][1:10,]
ggplot(top.wages, aes(reorder(Name, -Wage, sum), Wage, fill=Name)) +
  geom_col() + scale_y_continuous(labels = comma) + scale_fill_brewer(palette="Set3") +
  labs(y="Wage", x="Players", title="Top 10 Wages") +
  scale_x_discrete(guide = guide_axis(n.dodge=2))
```



Top 10 left footed players and right footed players

```
# Top 10 left footed players
top.leftfoot = fifa[fifa$Preferred.Foot == "Left",]
top.leftfoot = top.leftfoot[order(-top.leftfoot$Overall),][1:10,]
top.leftfoot.table = data.frame(top.leftfoot$Name, top.leftfoot$Age, top.leftfoot$Club, top.leftfoot$Na
colnames(top.leftfoot.table) = c("Name","Age","Club","Nationality")
top.leftfoot.table
```

```
##              Name Age               Club Nationality
## 1         L. Messi  33       FC Barcelona   Argentina
## 2         M. Salah  28          Liverpool       Egypt
## 3      T. Courtois  28        Real Madrid     Belgium
## 4        P. Dybala  26           Juventus   Argentina
## 5          Ederson  26    Manchester City      Brazil
## 6     A. Griezmann  29       FC Barcelona      France
## 7     A. Di MarÃa  32 Paris Saint-Germain   Argentina
## 8     A. Robertson  26          Liverpool    Scotland
## 9   Bernardo Silva  25    Manchester City    Portugal
## 10      A. Laporte  26    Manchester City      France
```

```r
# Top 10 right footed players
top.rightfoot = fifa[fifa$Preferred.Foot == "Right",]
top.rightfoot = top.rightfoot[order(-top.rightfoot$Overall),][1:10,]
top.rightfoot.table = data.frame(top.rightfoot$Name, top.rightfoot$Age, top.rightfoot$Club, top.rightfo
colnames(top.rightfoot.table) = c("Name","Age","Club","Nationality")
top.rightfoot.table
```

```
##                  Name Age               Club Nationality
## 1   Cristiano Ronaldo  35           Juventus    Portugal
## 2        K. De Bruyne  29     Manchester City     Belgium
## 3      R. Lewandowski  31  FC Bayern MÃ¼nchen      Poland
## 4           Neymar Jr  28 Paris Saint-Germain      Brazil
## 5             J. Oblak  27    AtlÃ©tico Madrid    Slovenia
## 6             S. ManÃ©  28           Liverpool     Senegal
## 7           K. MbappÃ©  21 Paris Saint-Germain      France
## 8          V. van Dijk  28           Liverpool Netherlands
## 9        M. ter Stegen  28        FC Barcelona     Germany
## 10             Alisson  27           Liverpool      Brazil
```

Creating a spider plot for different players statistics by wage

Lionel Messi Statistics

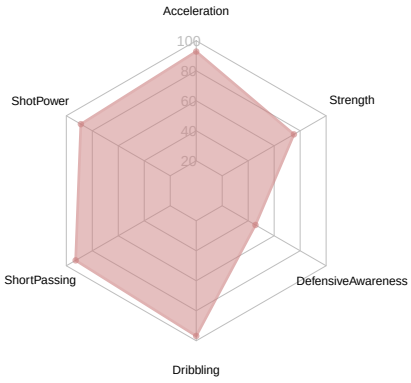


Figure 1: ...

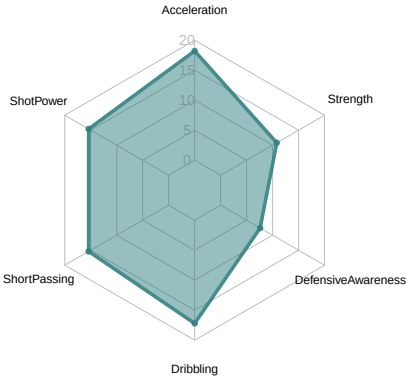


Figure 2: ...

C.Pulisic Statistics

Son Heung Min Statistics

Sergio Ramos Statistics

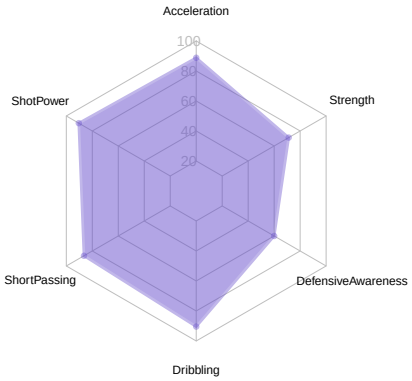


Figure 3: ...

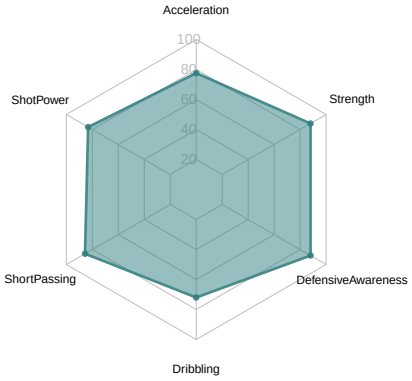


Figure 4: ...