# qin_alan_report1_28

Alan Qin

2/10/2020

Project Introduction

My project is about players in the game FIFA 20 and how wages are affected by rating. This is interesting to me because I want to see how big the gap in wage between the highest rated players in the game. The source of the data is from Kaggle.com, specifically (https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset). The Y (outcome) in this data are the wages of the players while the X's are the ratings of the player. This includes overall rating, potential rating, shooting rating, and many of other ones.

Histogram and Box Plots for my X's and Y's

```
library("ggplot2")
data = read.csv("/Users/alanqin/Downloads/fifa-20-complete-player-dataset/players_20.csv")
```

```
ggplot(data = data, aes(x = wage_eur)) + geom_histogram(binwidth = 10000) + ggtitle('Histogram of Player
```

Histogram of Players Wages
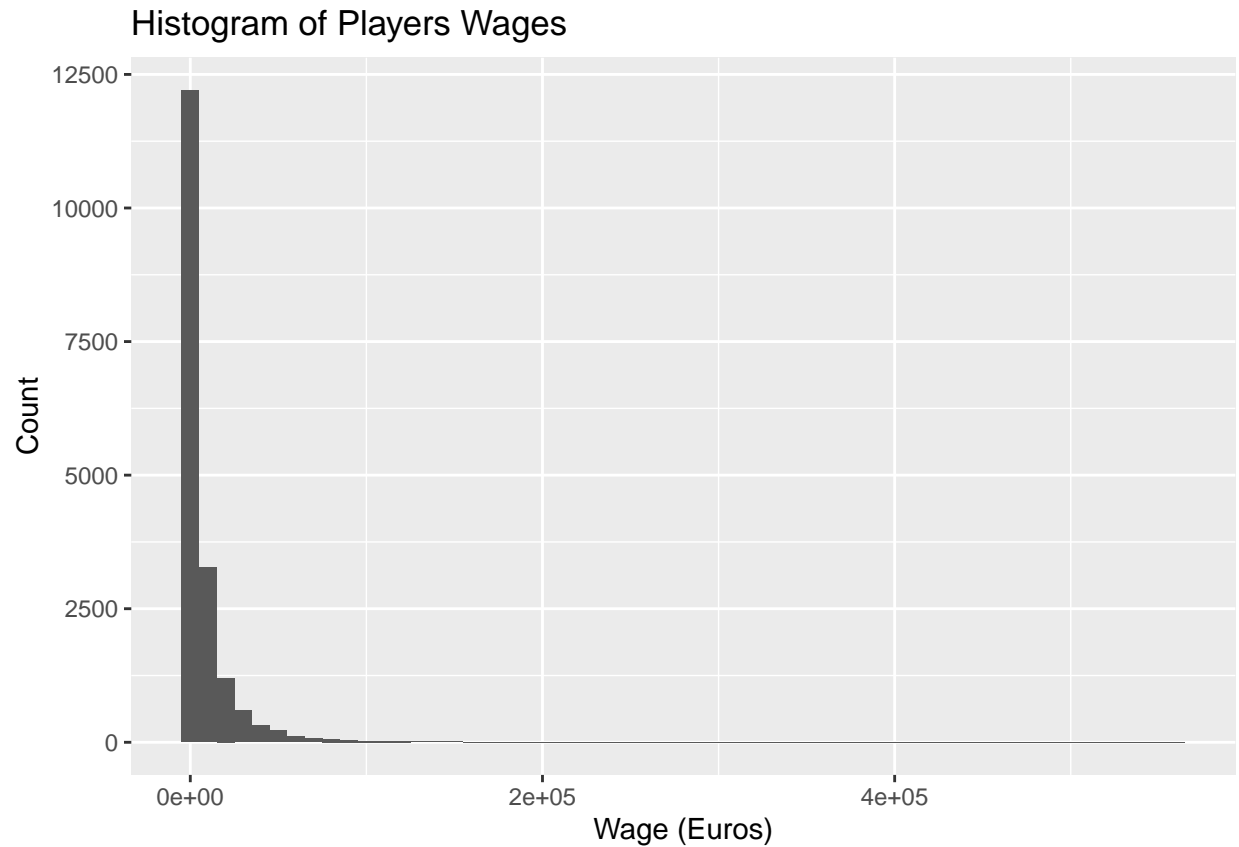
This histogram shows the wages of all the players in fifa. As you can see, there are over ten thousand players with wages of around 0. As we go up, there are less and less playesr with one player making around 500000 euros a week.

```
ggplot(data = data, aes(x = as.factor(overall), y = wage_eur))+ geom_boxplot() + ggtitle('Box Plot of O
```

Box Plot of Overall Rating vs Wage
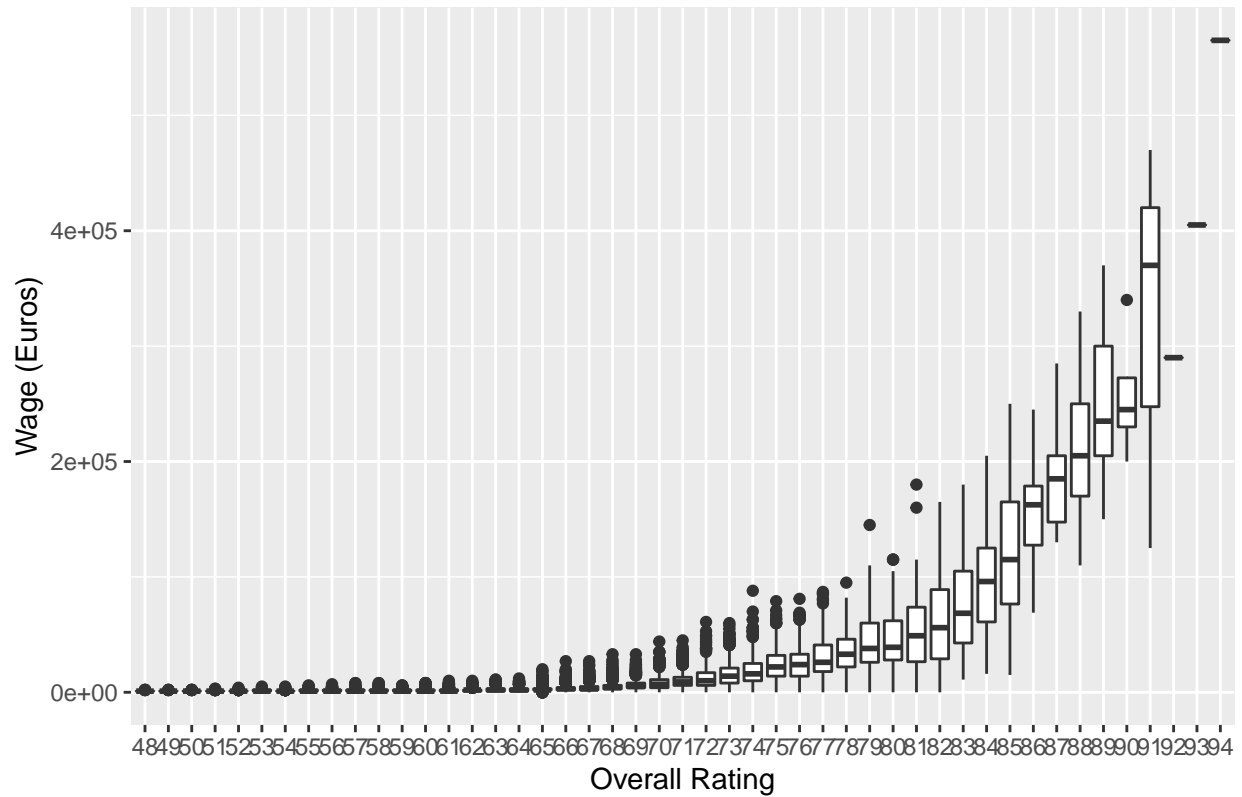
This box plot shows the ranges, averages, and quartiles of wage for each rating in fifa 20. As we can see from this box plot, at lower ratings, the range of the player wages are very low and as the rating creeps up, both the player wages go up significantly and there is a much larger range for players wages.

```
ggplot(data = data, aes(x = as.factor(potential), y = wage_eur))+ geom_boxplot()  +  ggtitle('Box Plot
```

3

Box Plot of Potential Rating vs Wage
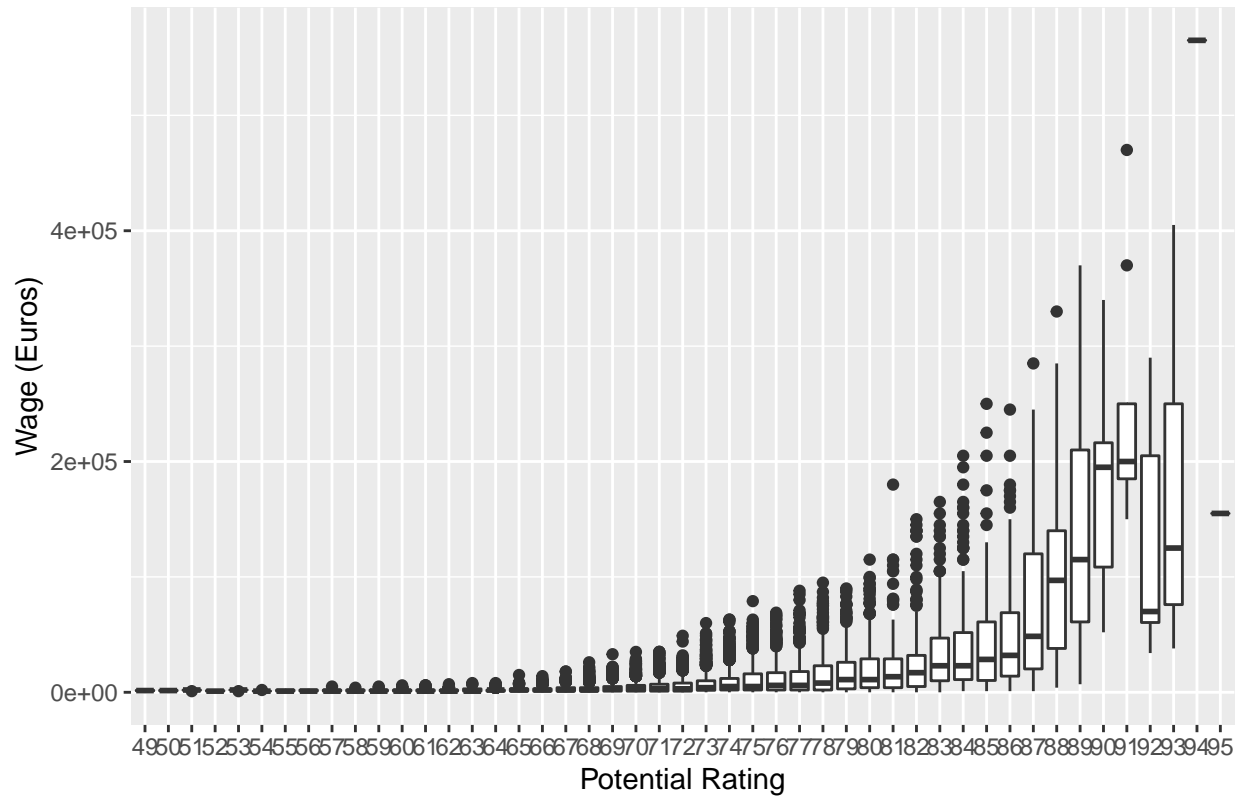
From this box plot, I learned that potential rating is not the best representation for wages because there is an outlier at the 95 rating. Also for some reason the highest potential rated players in the game strangely has a lower rating than the 91's and 92's.

```
ggplot(data = data, aes(x = as.factor(international_reputation), y = wage_eur))+ geom_boxplot() + ggtit
```

Box Plot of International Repuation vs Wage
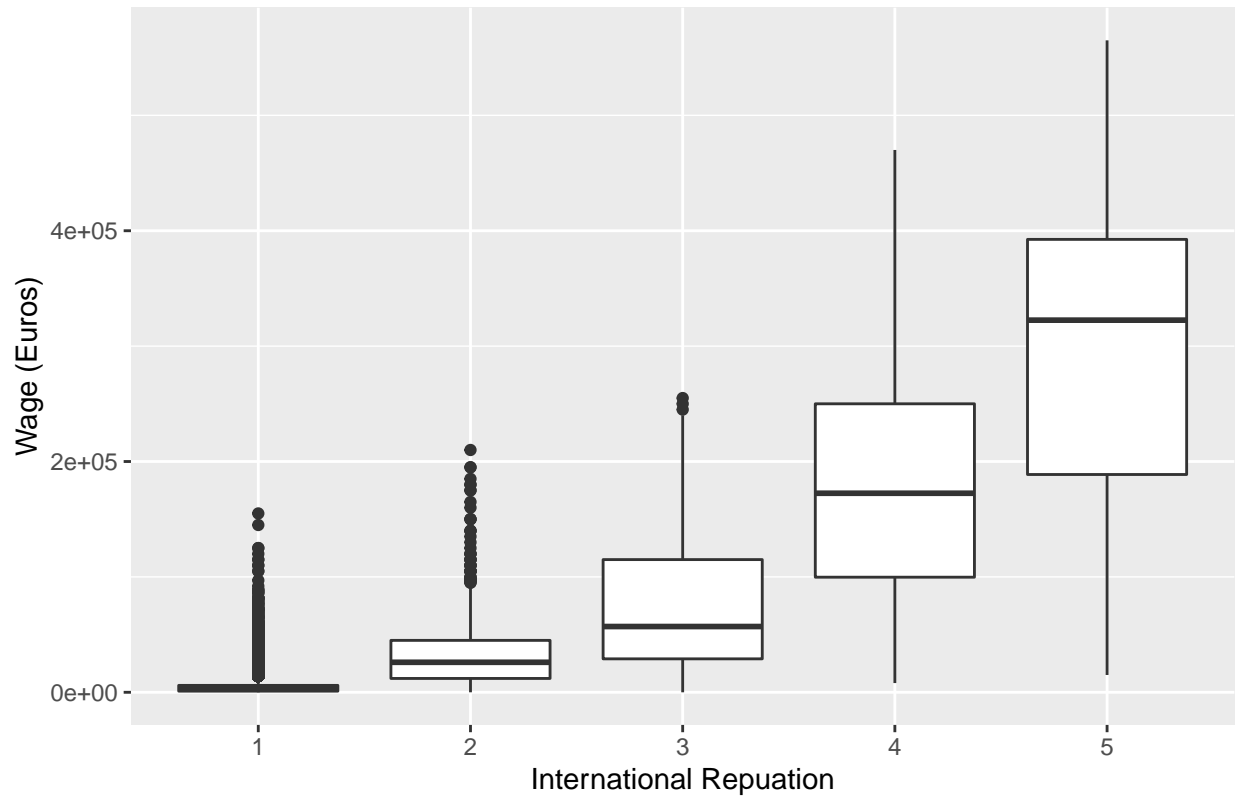
This box plot shows that there are many players with an international reputation of 3 or less but they do not make that much money. Compared to the players with 4 or above, the wages are much higher with larger ranges. I think this is because in Fifa, overall rating is sometimes based on international reputation which makes sense looking at this data.

```
ggplot(data = data, aes(x = as.factor(shooting), y = wage_eur))+ geom_boxplot() + ggtitle('Box Plot of
```

Box Plot of Shooting Rating vs Wage
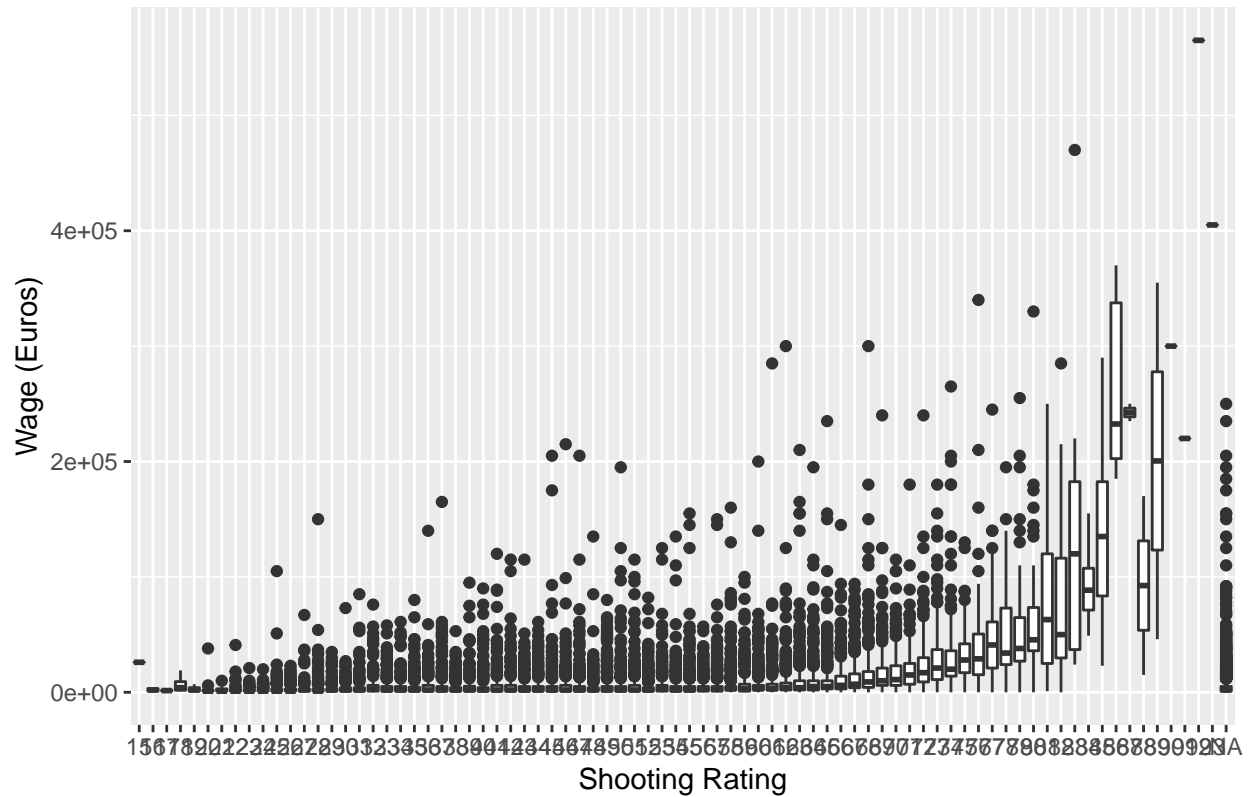
As we see from this box plot, we can see that shooting rating is not that great of indicator of wages. I think that this is the case because there are really high paid players that are defenders. For example, Matthijs de Ligt is the highest paid defender in the game with only a 58 shooting rating.

```r
ggplot(data = data, aes(x = as.factor(age), y = wage_eur))+ geom_boxplot() + ggtitle('Box Plot of Age v
```

Box Plot of Age vs Wage
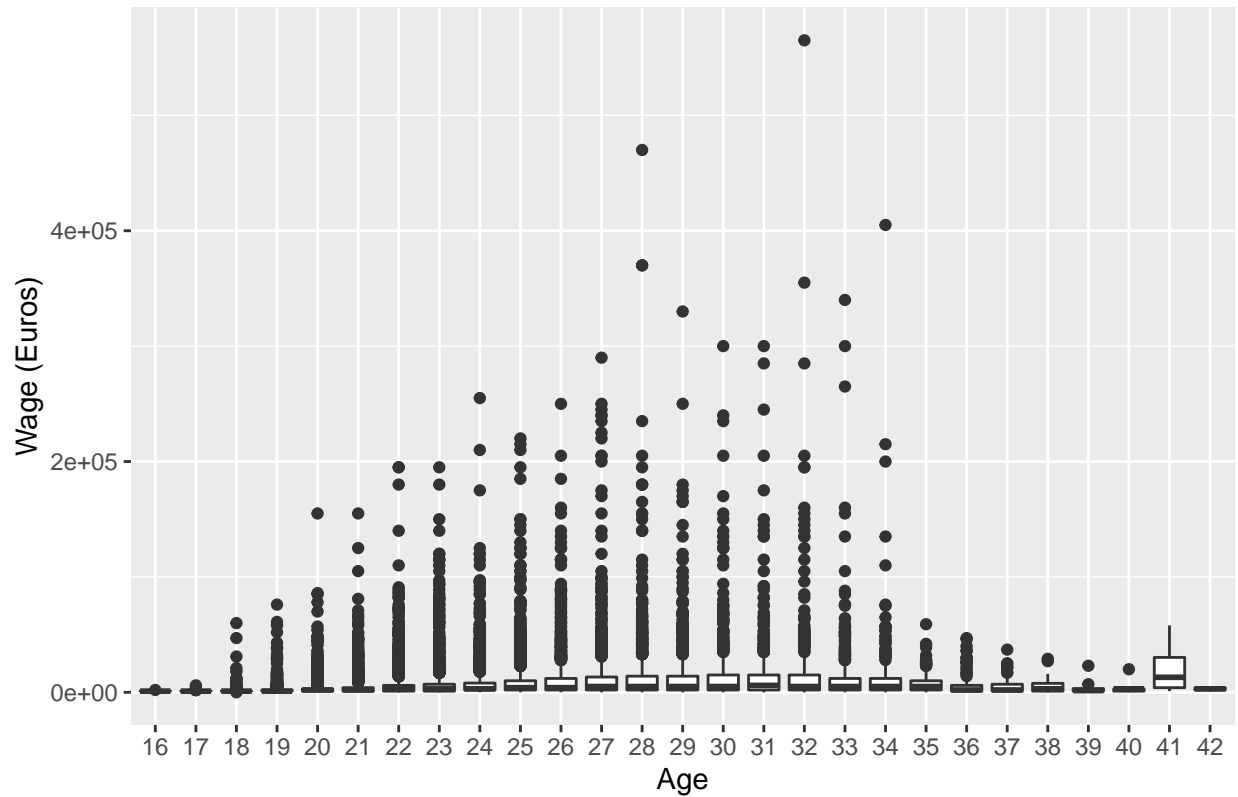
As we can see from this box plot, as long as you are not very young or very old, you will generally get paid around an equal wage. I learned from this box plot that the players that get paid the most are around 28-32. I think this is because that is when their physical peak is.

Plots of X's vs Y's

```
plot(x = data$overall, y = data$wage_eur, ylab = "Wage (Euros)", xlab = "Overall Rating", main = 'Wage
```
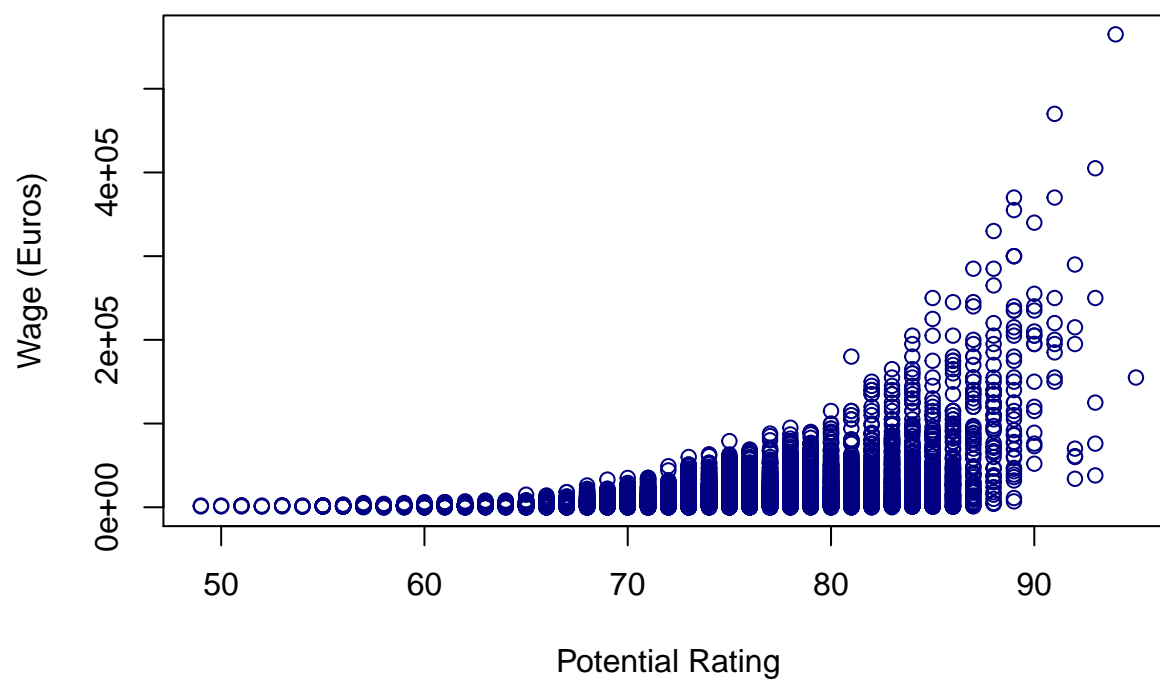
Wage vs Overall Rating
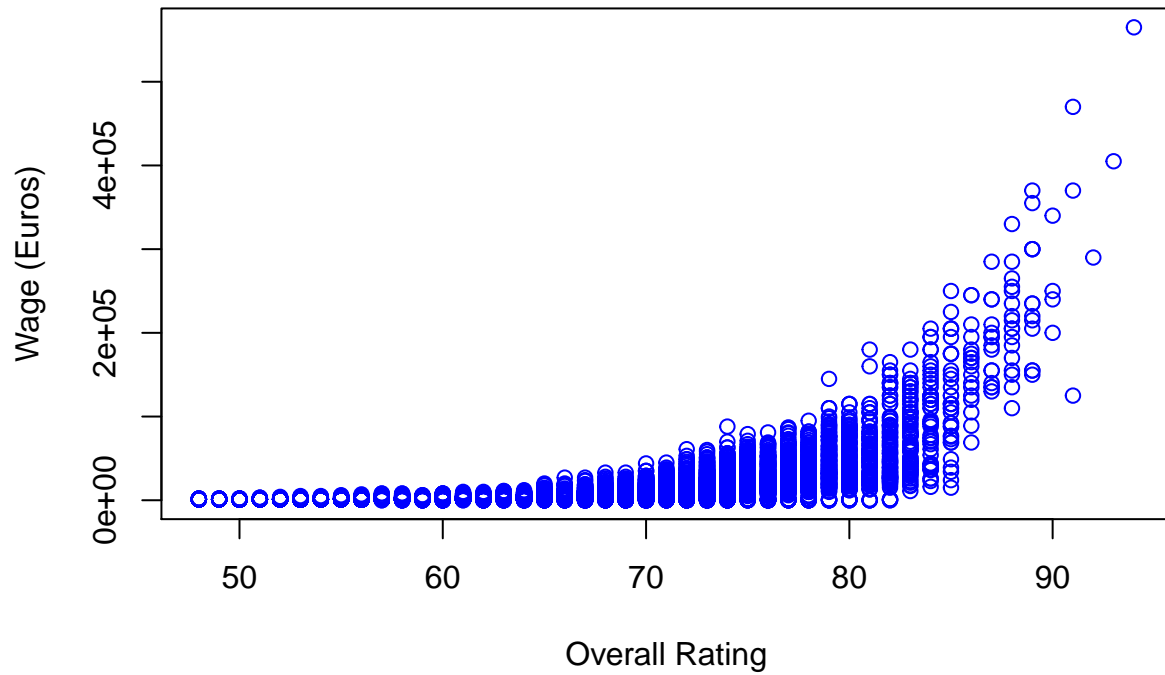
In this plot, you can see that overall rating has a clear correlation with wage. As overall rating goes up, wage also goes up. This plot also looks like an exponential function

```r
plot(x = data$potential, y = data$wage_eur, ylab = "Wage (Euros)", xlab = "Potential Rating", main = 'Wa
```
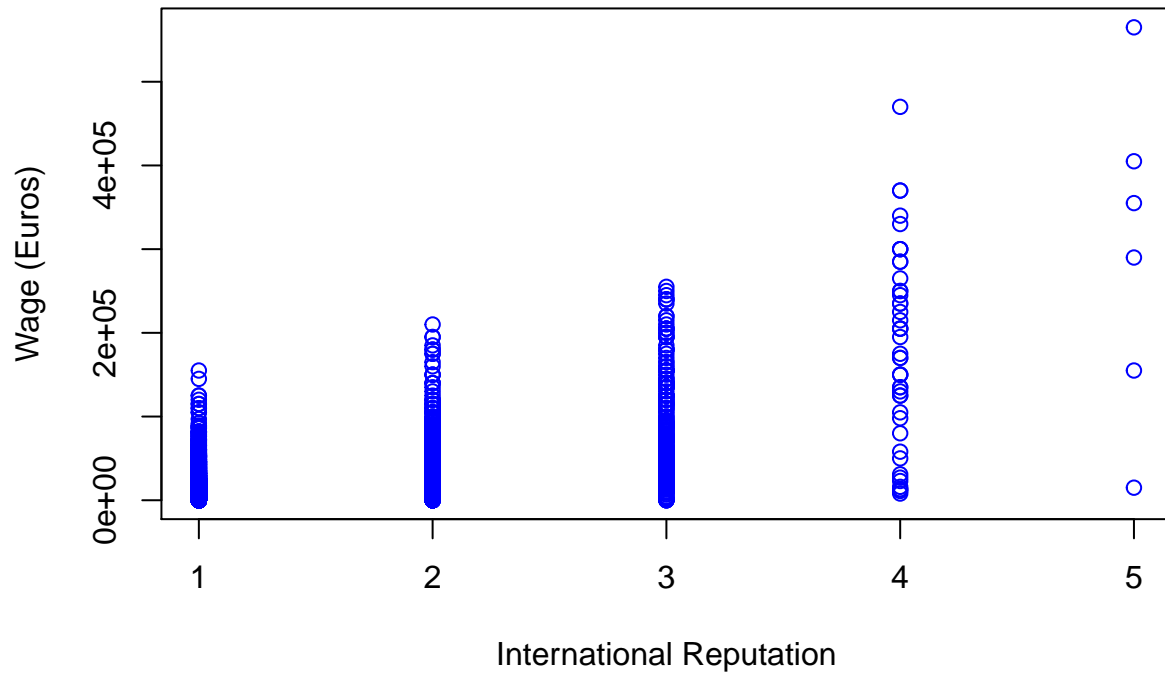
Wage vs Potential Rating

In this plot, you can also see that it is very similar to the results of the overall rating. This is because I think that overall rating is almost the same as potential rating on the players with the highest wages.

```
plot(x = data$international_reputation, y = data$wage_eur,ylab = "Wage (Euros)", xlab = "International
```

Wage vs International Reputation

Again, this box plot shows that there are many players with an international reputation of 3 or less but, they do not make that much money. Most of the players with an international repuation of 4 or higher earn a much larger salary than those with a rating of 3 or less.

```
plot(x = data$shooting, y = data$wage_eur, ylab = "Wage (Euros)", xlab = "Shooting Rating", main = 'Wag
```

Wage vs Shooting Rating
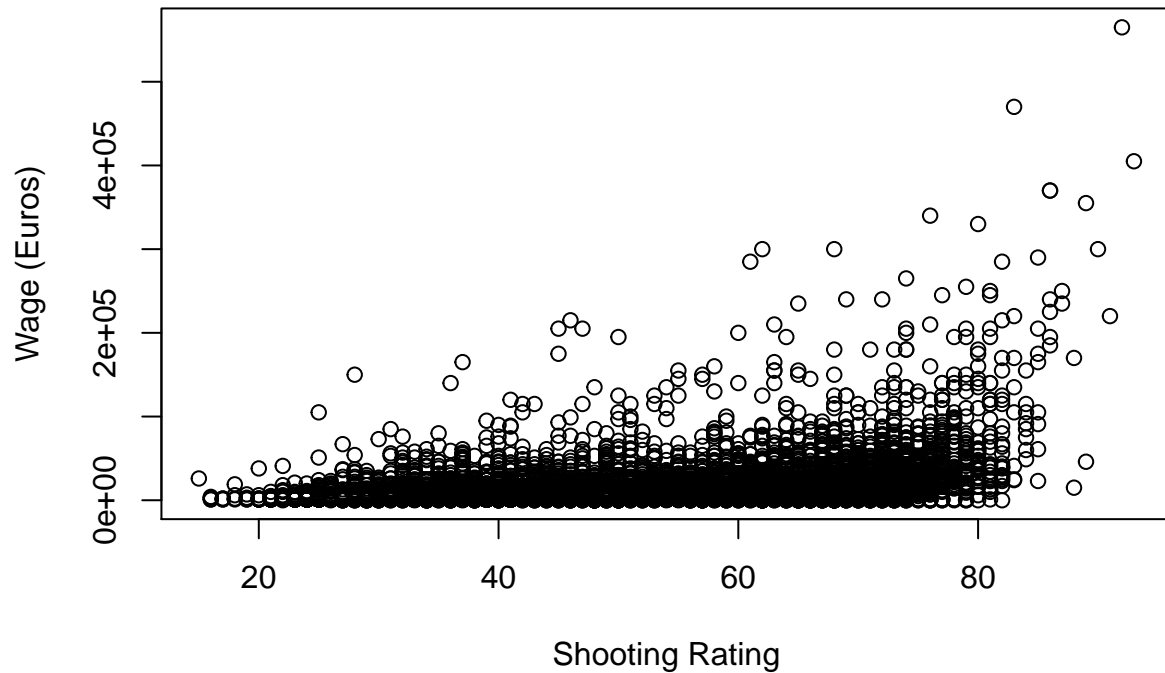
In this plot, shooting rating does not tell us much about the wage of the players unless their shooting stat is very high compared to other players. As we can see, the players with 90 shooting or above has high salaries compared to 85 and below. I think this is because attackers in soccer are paid more than defenders.

```r
plot(x = data$age, y = data$wage_eur, ylab = "Wage (Euros)", xlab = "Age", main = 'Wage vs Age')
```

Wage vs Age

As we can see from this plot, most of the fifa players are between 23 and 33 years old. This is also where the players get paid the most because of their combination of physical form and experience. These are peak years for athletes and they have the wages to back that up.