

Exploratory Data Analysis

This is a file for explorative data analysis of the premier league players.

```
season17 <- read.csv("~/DSI-SRP1/season17.csv", encoding="UTF-8")
season18 <- read.csv("~/DSI-SRP1/season18.csv", encoding="UTF-8")
season19 <- read.csv("~/DSI-SRP1/season19.csv", encoding="UTF-8")
```

Checking out the datasets

```
View(season17)
View(season18)
View(season19)
```

Loading tidyverse

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.0    v purrr   0.3.4
## v tibble  3.0.1    v dplyr  0.8.5
## v tidyr   1.0.3    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

Calculating correlation coefficients between goal scored and total fpl points

```
message("Correlation between goal scored and total fpl points for 2016/17: ", with(season17, cor(goals_scor, total_points)))
```

```
## Correlation between goal scored and total fpl points for 2016/17: 0.700148307138786
```

```
message("Correlation between goal scored and total fpl points for 2017/18: ", with(season18, cor(goals_scor, total_points)))
```

```
## Correlation between goal scored and total fpl points for 2017/18: 0.679586851190284
```

```
message("Correlation between goal scored and total fpl points for 2018/19: ", with(season19, cor(goals_scor, total_points)))
```

```
## Correlation between goal scored and total fpl points for 2018/19: 0.703249889002981
```

Editing the datasets

```

season17$position_index <- as_factor(season17$position_index)
season18$position_index <- as_factor(season18$position_index)
season19$position_index <- as_factor(season19$position_index)

```

Creating a new metric for calculation total fpl points to minutes played

```

fpl_to_gaming <- function(df) {
  df %>%
    mutate(fpl_to_game = (total_points/minutes.played)*90)
}
season17 <- fpl_to_gaming(season17)
season18 <- fpl_to_gaming(season18)
season19 <- fpl_to_gaming(season19)

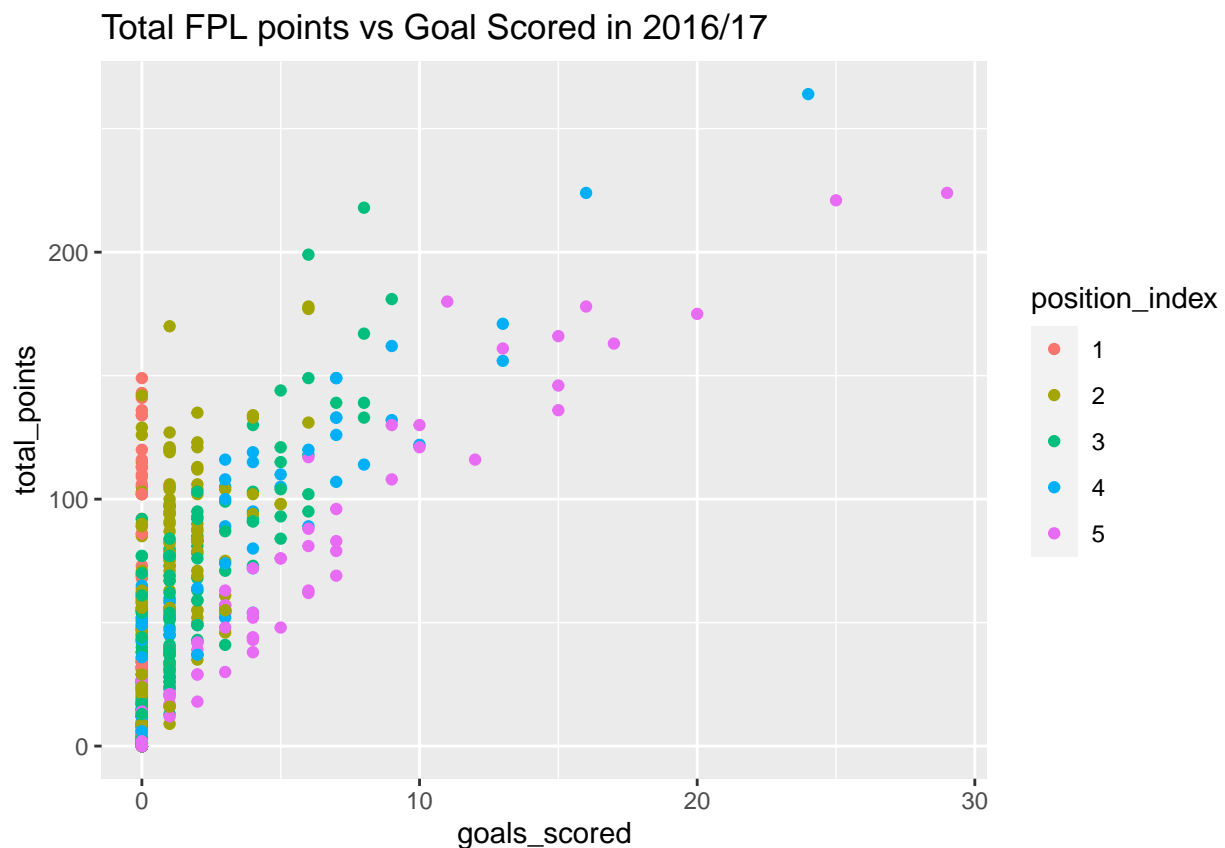
```

Scatter plots graph to highlight how the goals scored vary with fpl points

```

ggplot(season17, aes(goals_scored, total_points)) + geom_point(aes(color = position_index)) + labs(title = "Total FPL points vs Goal Scored in 2016/17")

```

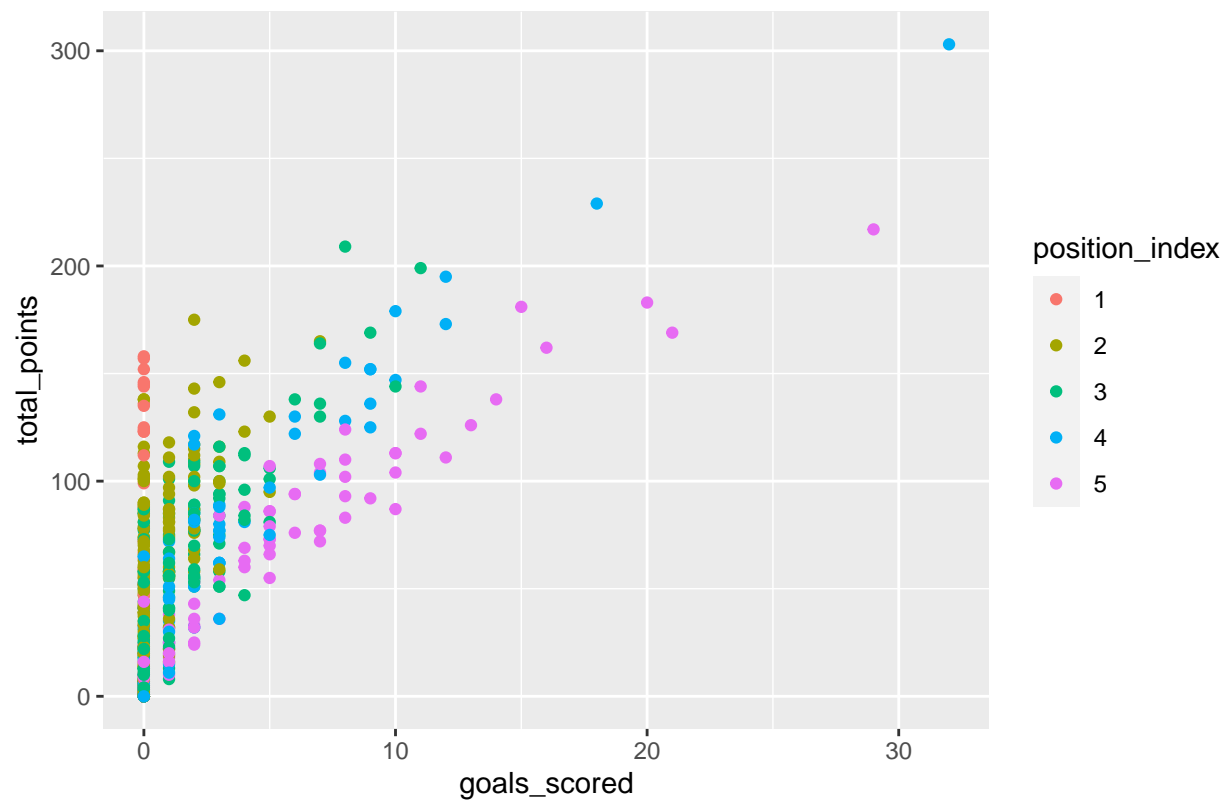


```

ggplot(season18, aes(goals_scored, total_points)) + geom_point(aes(color = position_index)) + labs(title = "Total FPL points vs Goal Scored in 2017/18")

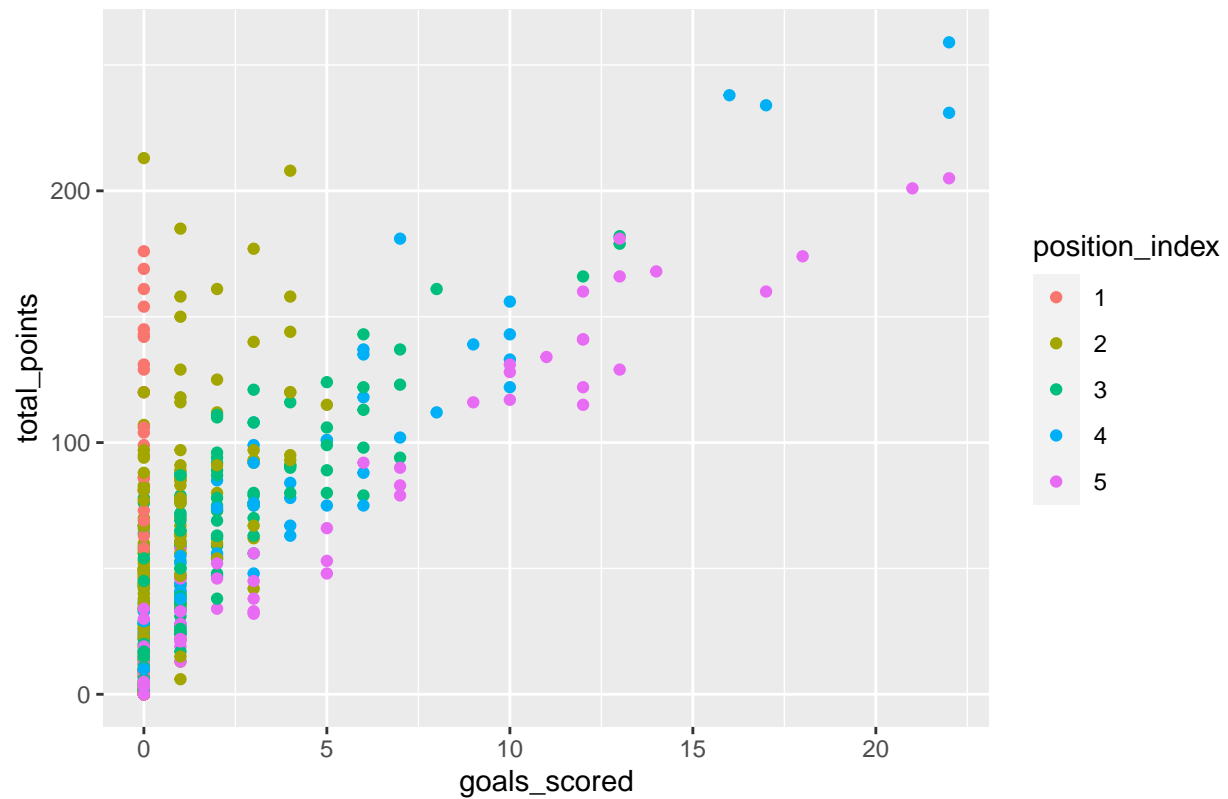
```

Total FPL points vs Goal Scored in 2017/18



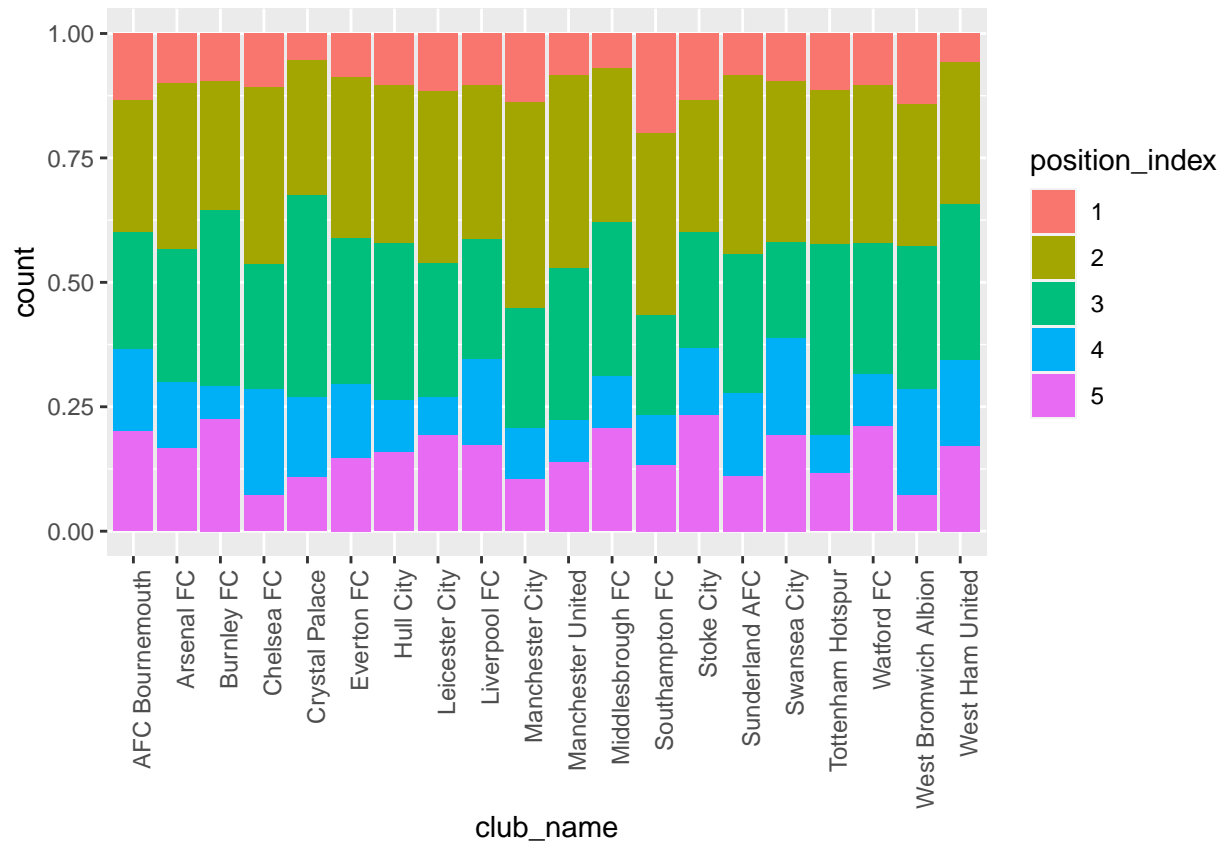
```
ggplot(season19, aes(goals_scored, total_points)) + geom_point(aes(color = position_index)) + labs(title = "Total FPL points vs Goal Scored in 2017/18")
```

Total FPL points vs Goal Scored in 2018/19



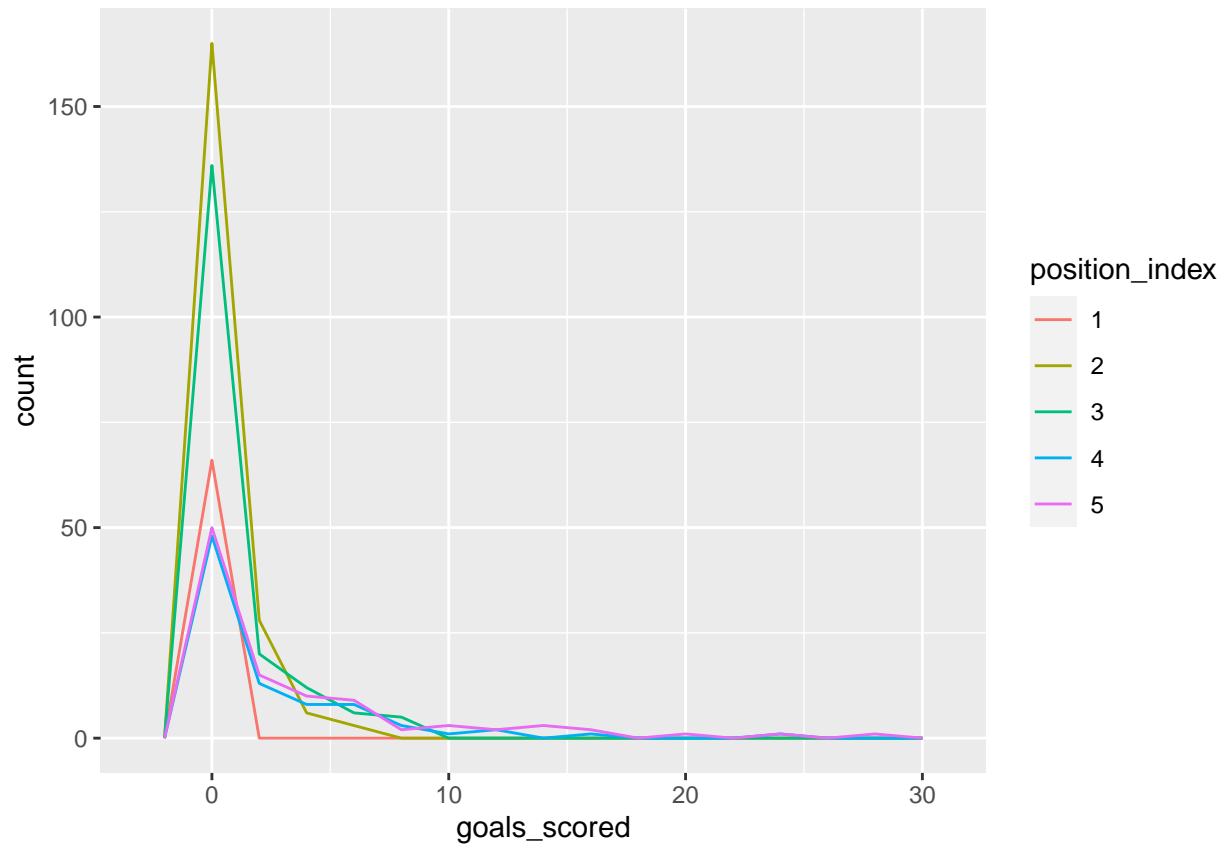
Bar graphs to show how the number of players vary per team with regards to position

```
ggplot(season17, aes(club_name)) + geom_bar(aes(fill = position_index), position = "fill") + theme(axis.
```



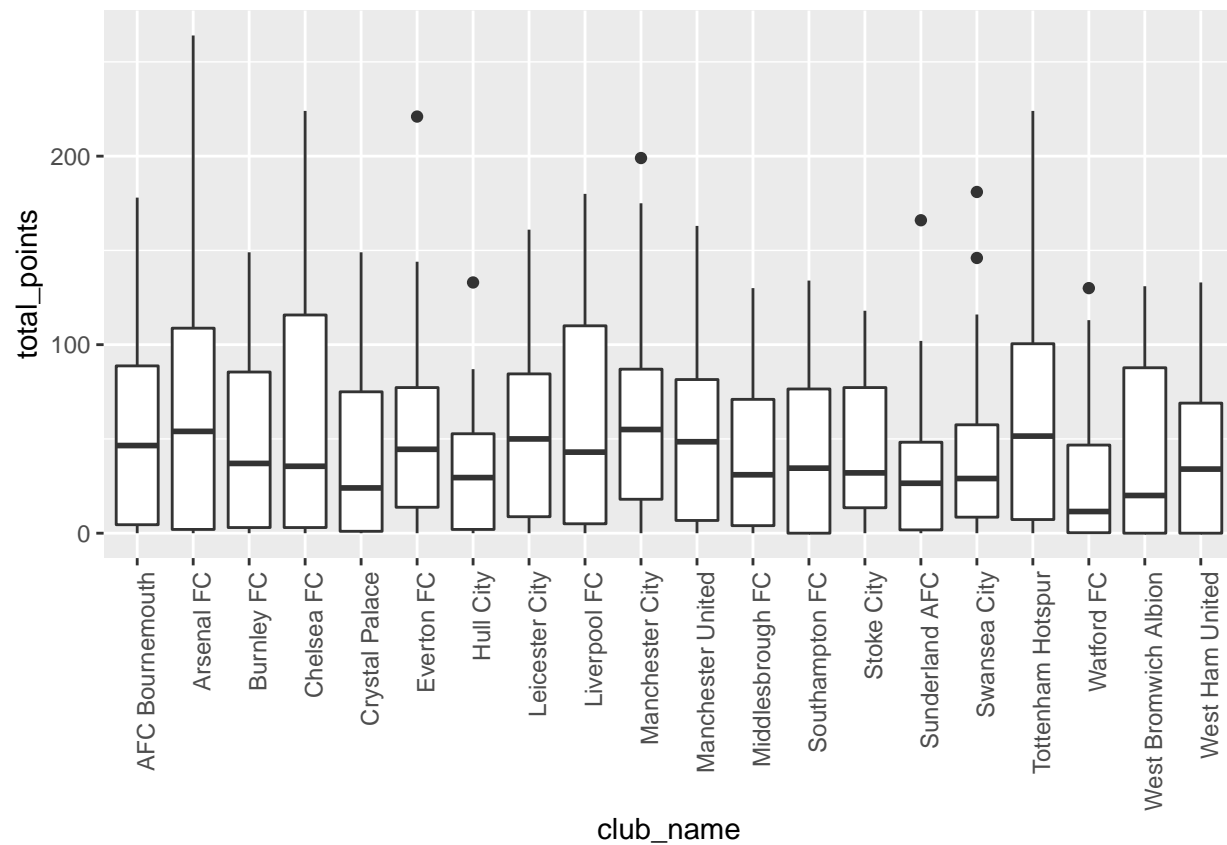
Creating frequency plot for goal scored and how it varies for the players' position

```
ggplot(season17, aes(x = goals_scored)) + geom_freqpoly(aes(color = position_index), binwidth = 2)
```



Creating boxplots to show the distribution of the teams' fpl points according to their players

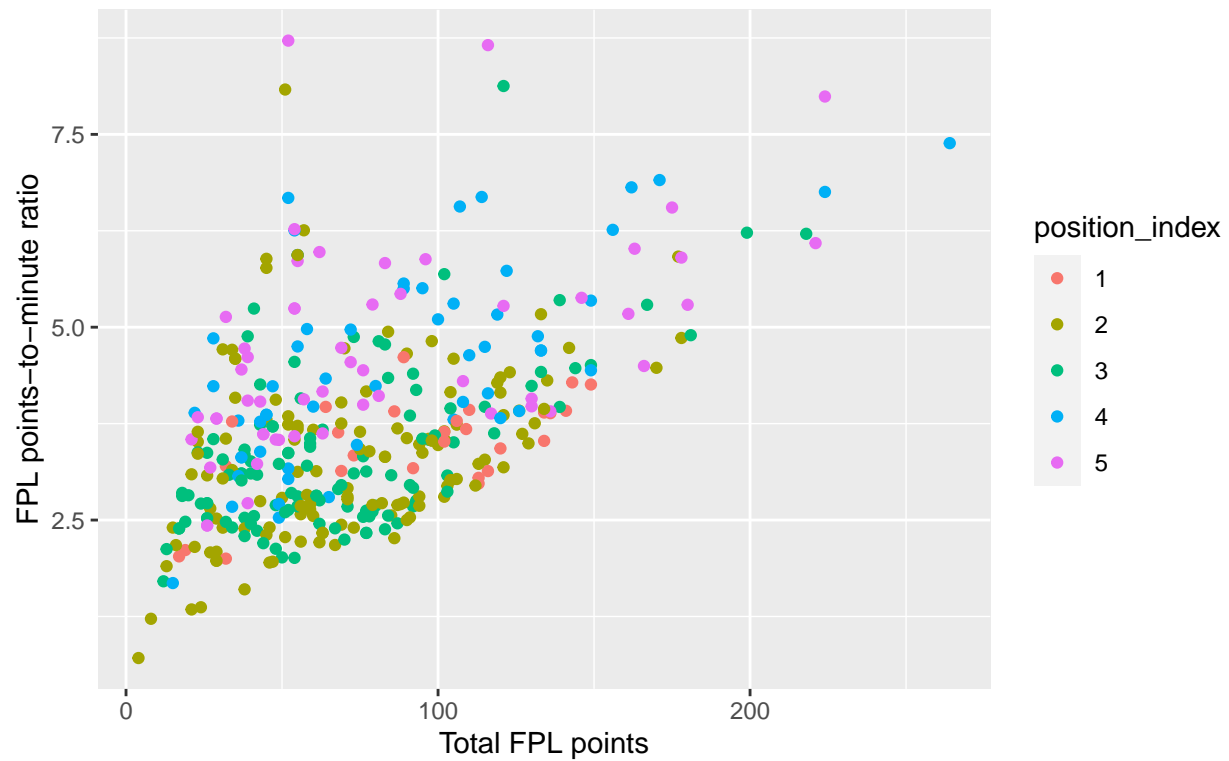
```
ggplot(data = season17) + geom_boxplot(aes(x = club_name, y = total_points), na.rm = TRUE) + theme(axis
```



Scatter plotsd showing the FPL points to minute ratio vs the total FPL points

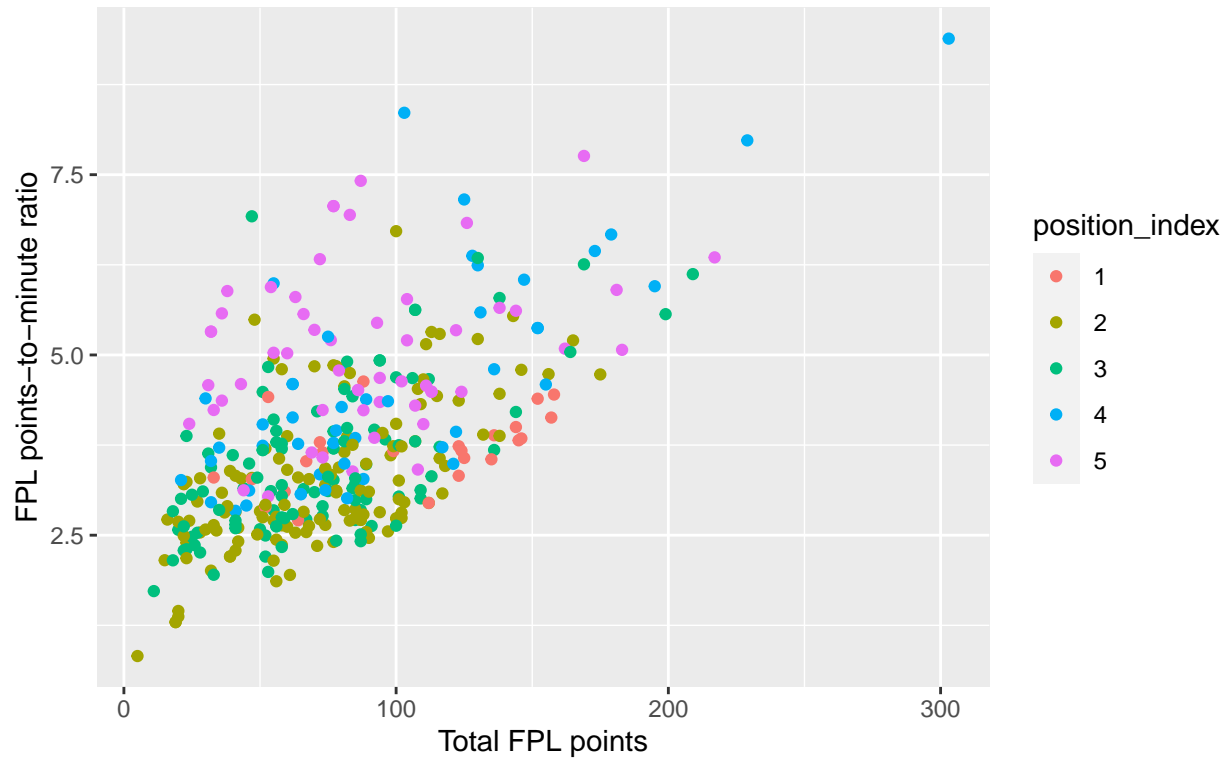
```
season17 %>%
  filter(minutes.played >= 500) %>%
  ggplot() + geom_point(aes(total_points, fpl_to_game, color = position_index), na.rm = TRUE) + labs(x = "total_points", y = "fpl_to_game", color = "position_index")
```

FPL points to minutes played ratio vs Total FPL points for players who have played more than 500 minutes



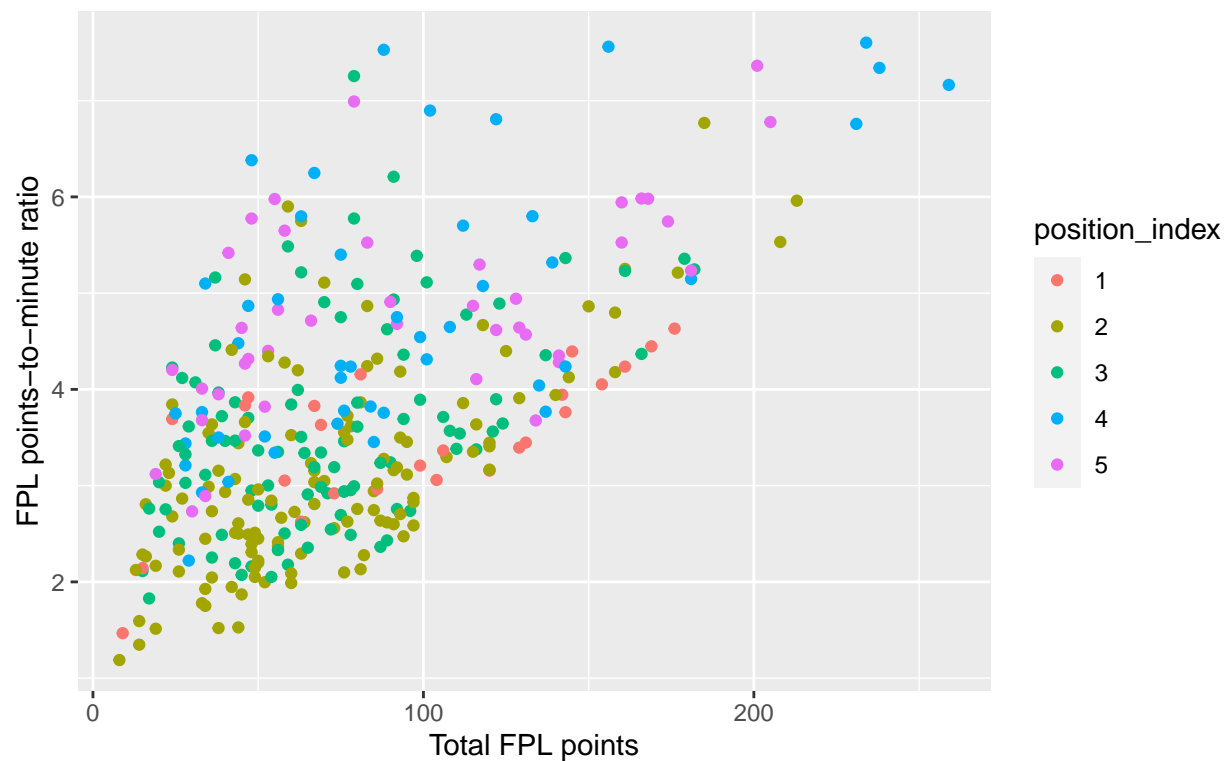
```
season18 %>%
  filter(minutes.played >= 500) %>%
  ggplot() + geom_point(aes(total_points, fpl_to_game, color = position_index), na.rm = TRUE) + labs(x = "Total FPL points", y = "FPL points-to-minute ratio")
```


FPL points to minutes played ratio vs Total FPL points for players who have played more than 500 minutes



```
season19 %>%
  filter(minutes.played >= 500) %>%
  ggplot() + geom_point(aes(total_points, fpl_to_game, color = position_index), na.rm = TRUE) + labs(x = "Total FPL points", y = "FPL points-to-minute ratio")
```

FPL points to minutes played ratio vs Total FPL points for players who have played more than 500 minutes

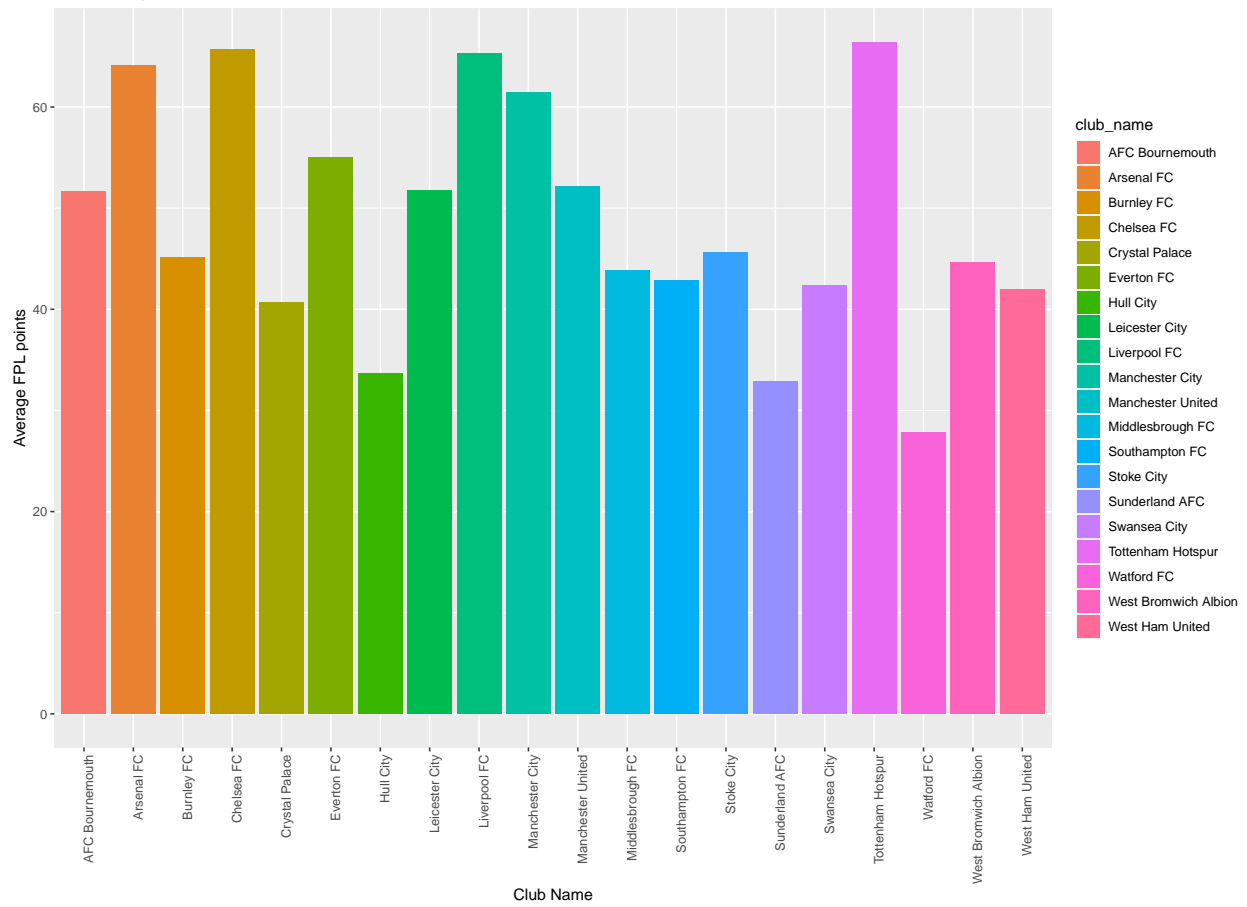


Calculating the average fpl points per team

```
avg_fpl_point <- function(df) {
  df %>%
    group_by(club_name) %>%
    summarize(average = mean(total_points, na.rm = TRUE)) %>%
    ggplot() + geom_bar(aes(x = club_name, y = average, fill = club_name), stat = "identity") + theme(a
}

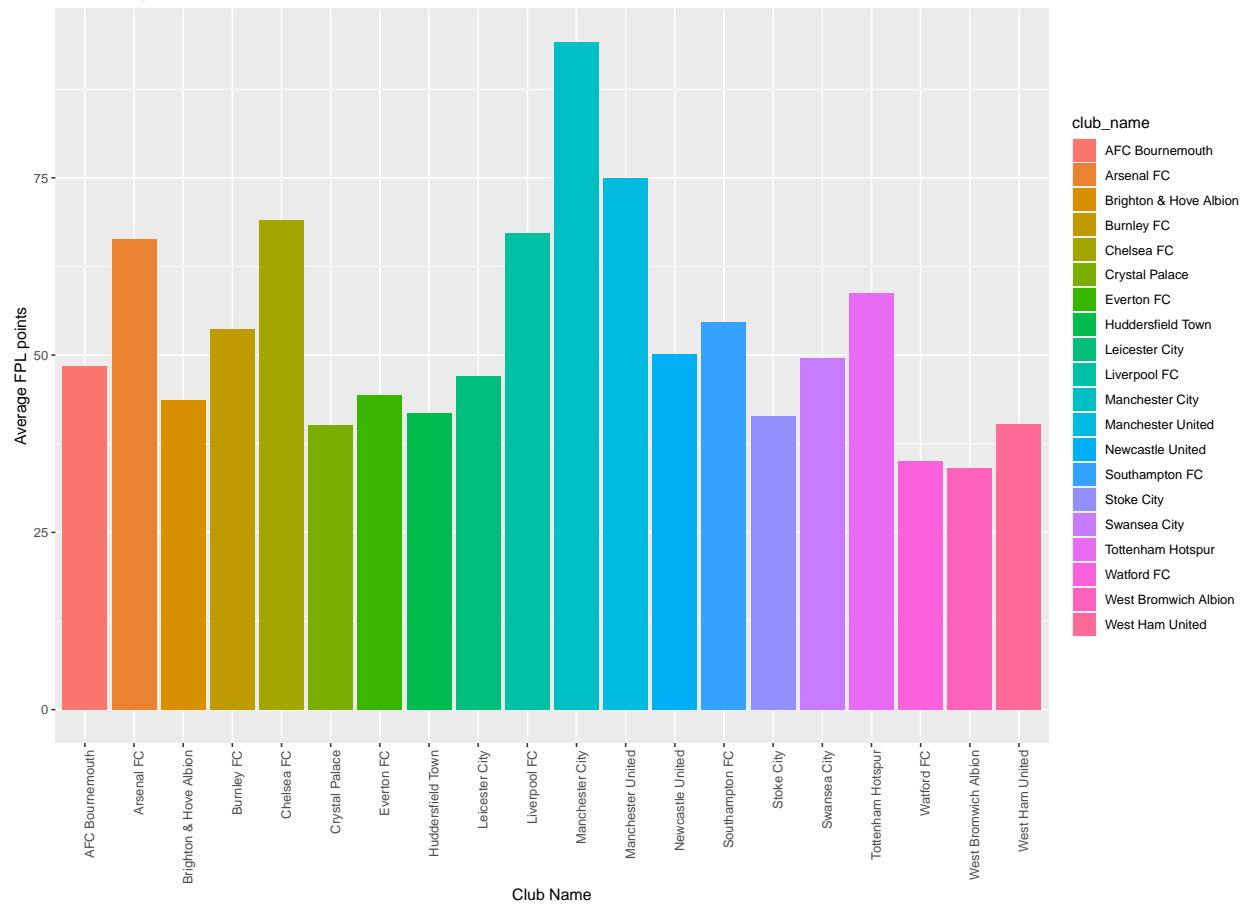
avg_fpl_point(season17)
```

The average FPL points achieved by each team in the season 2016/2017

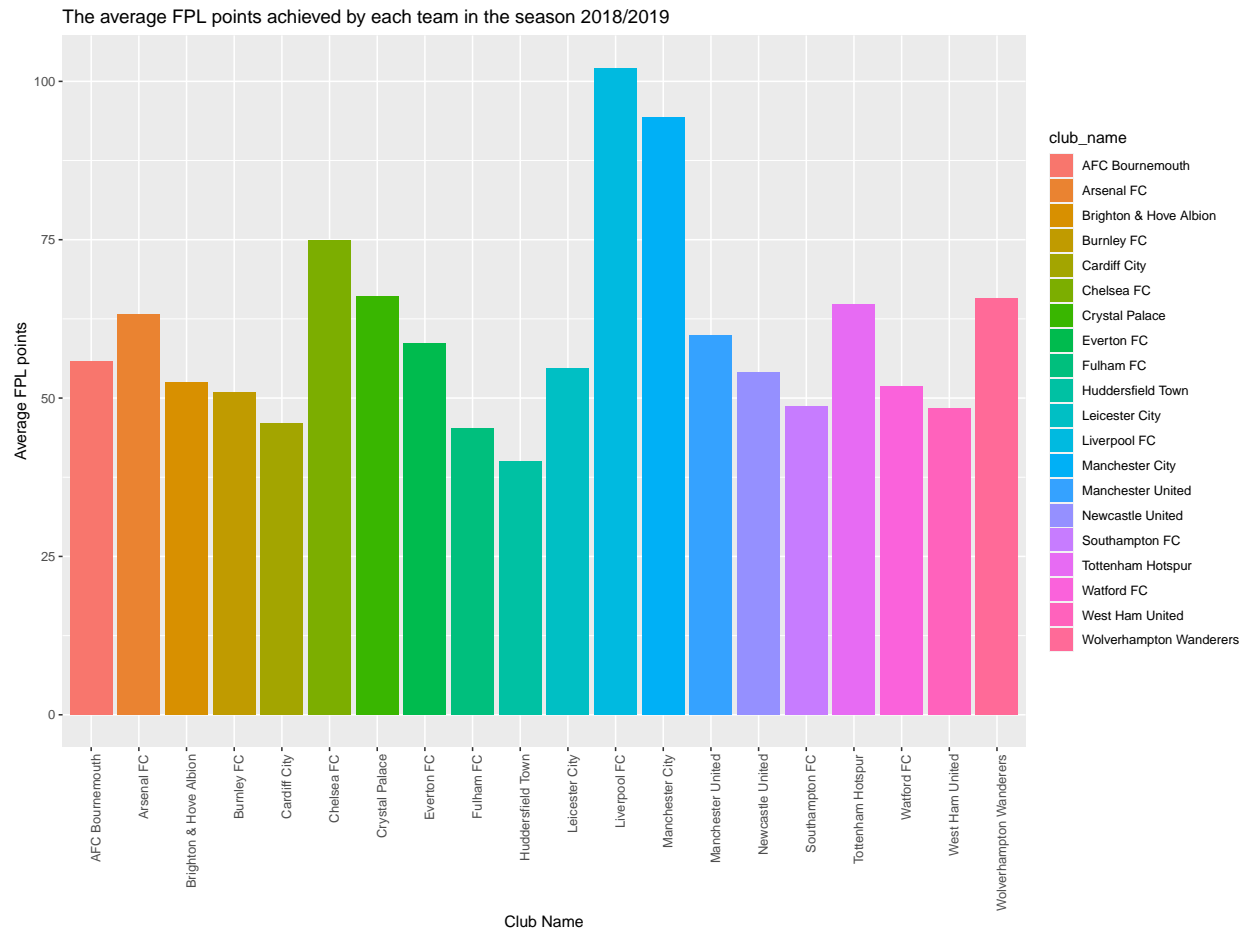


```
avg_fpl_point(season18)
```

The average FPL points achieved by each team in the season 2017/2018



```
avg_fpl_point(season19)
```

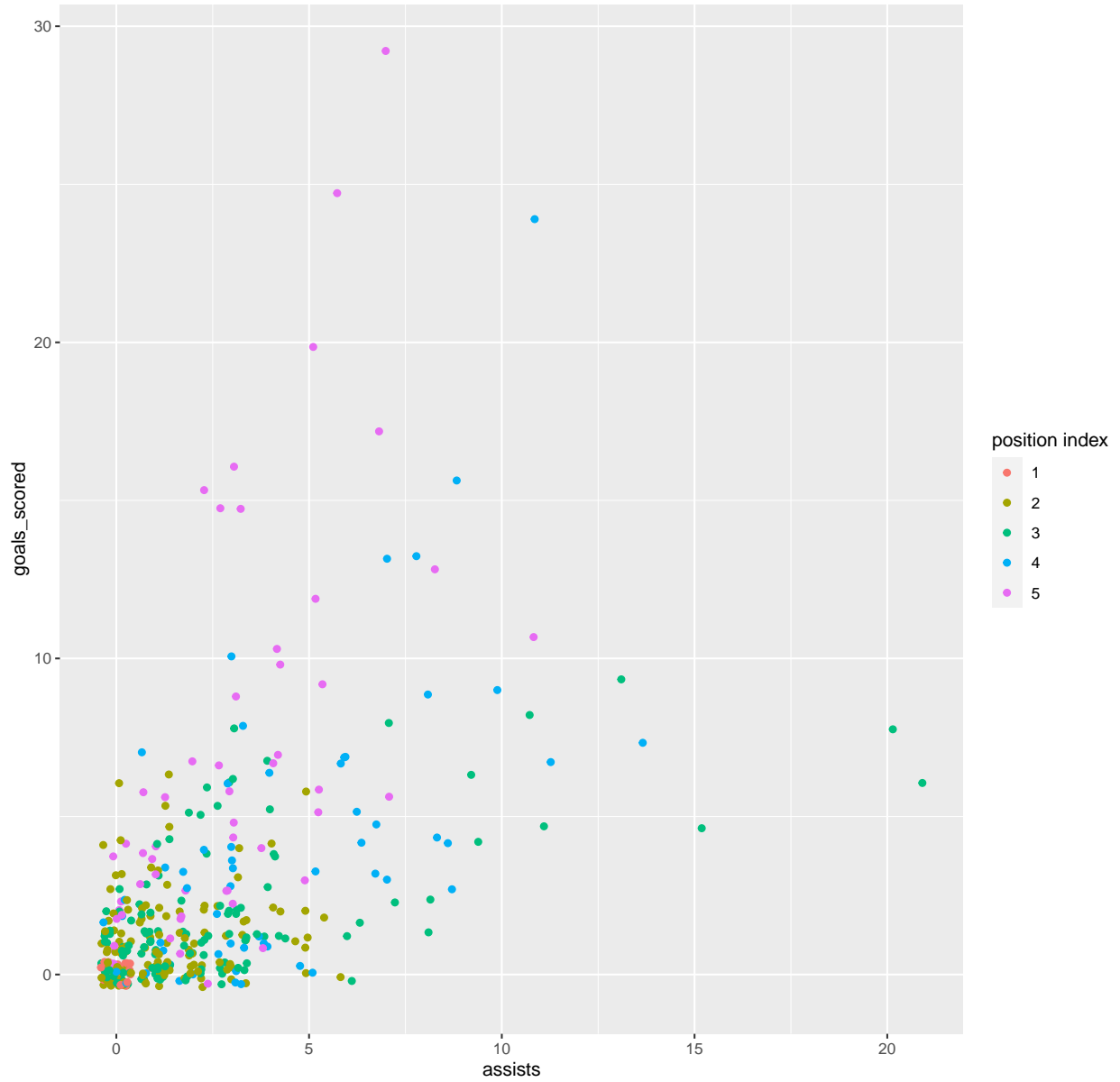


Plotting graphs of goals scored vs assists made for every season

```
goals_vs_assist <- function(df) {
  df %>%
    filter(minutes.played >= 500) %>%
    ggplot(aes(assists, goals_scored)) + geom_point(aes(color = position_index), position = "jitter") +
}
```

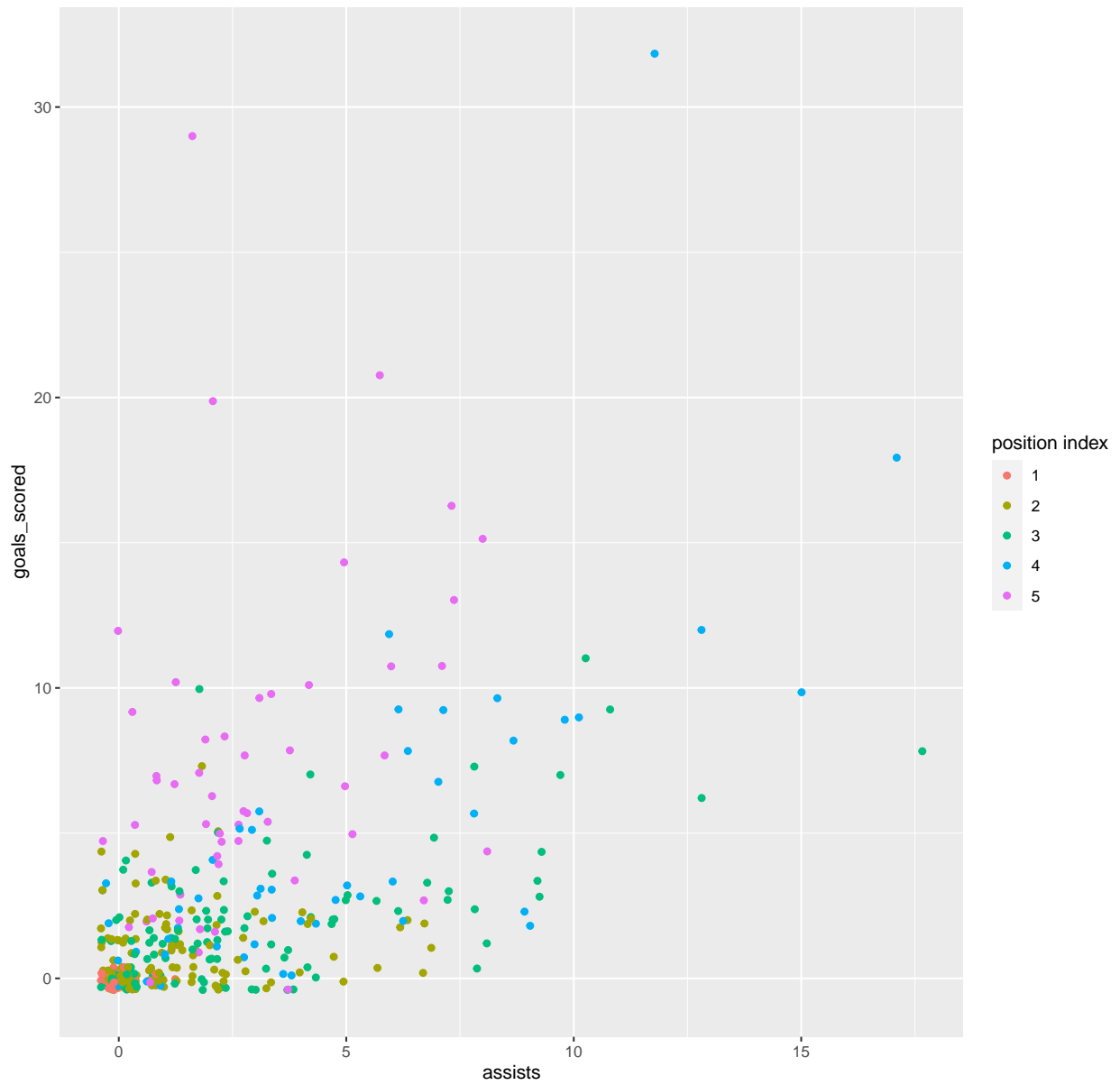
```
goals_vs_assist(season17)
```

Goals scored vs Assists during the 2016/2017 season

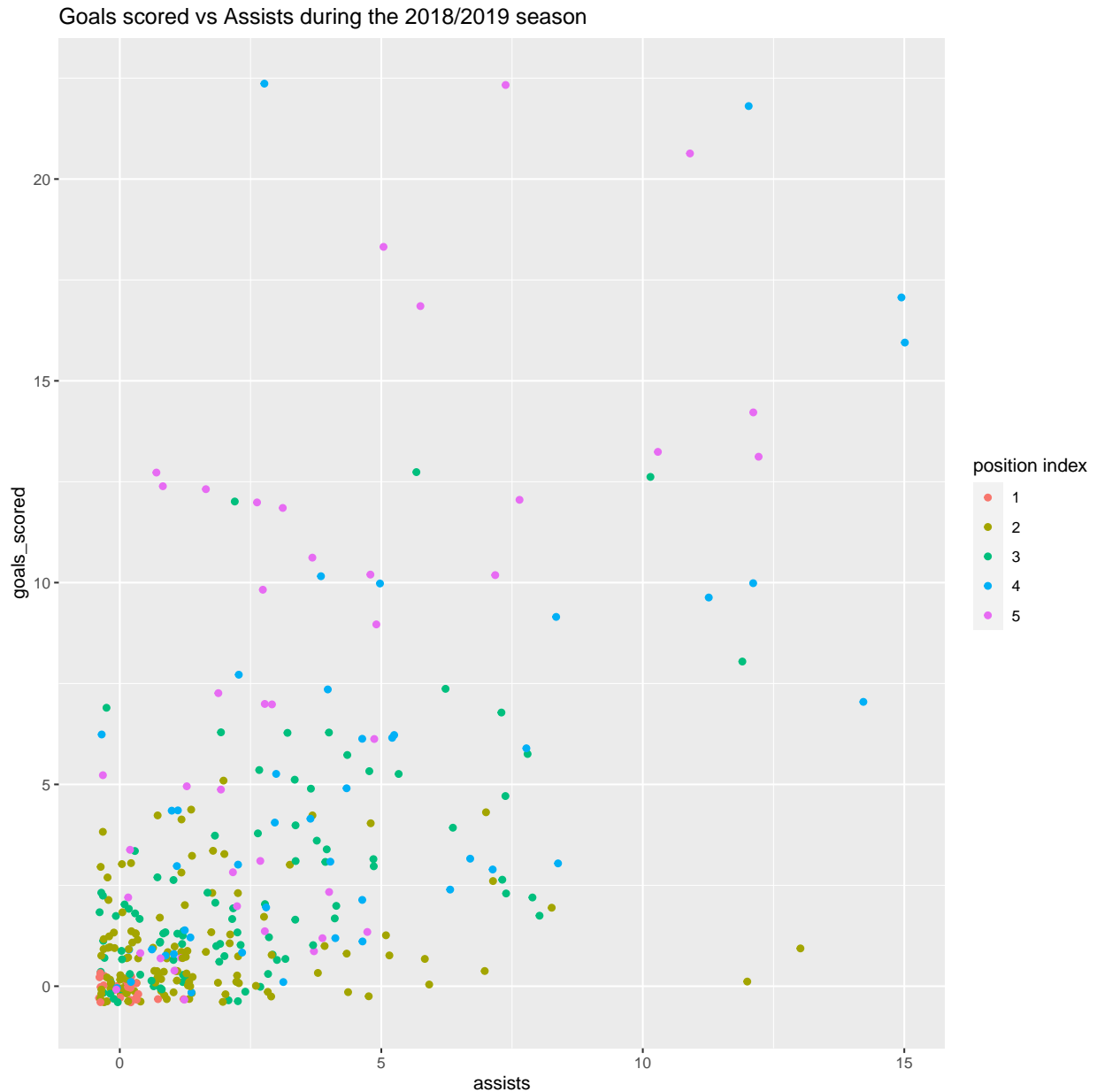


```
goals_vs_assist(season18)
```

Goals scored vs Assists during the 2017/2018 season



```
goals_vs_assist(season19)
```



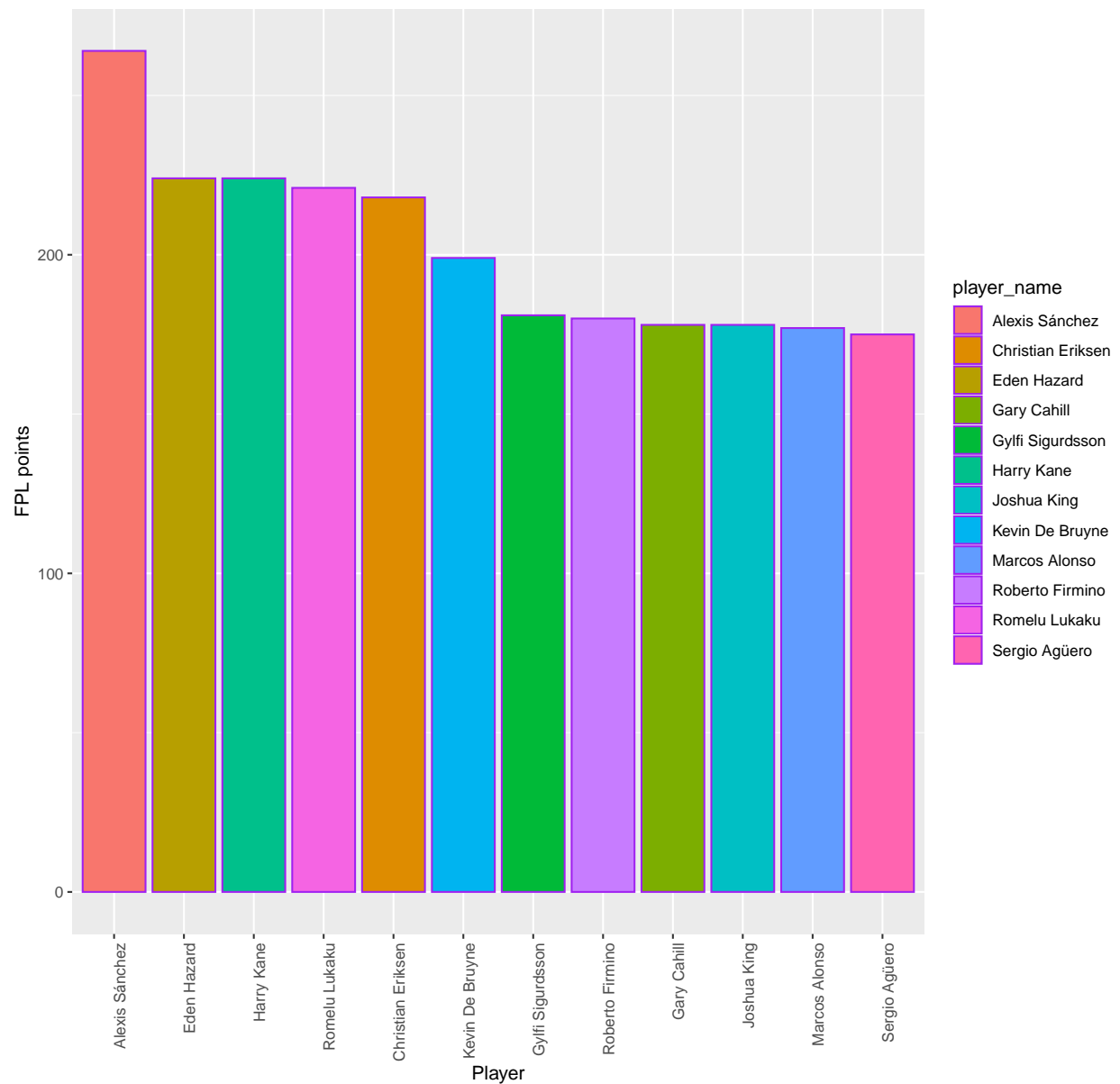
Ranking the players with the top 10 fpl points per season

```
fpl_point_rank <- function(df) {
  df %>%
    mutate(ranking = dense_rank(desc(total_points))) %>%
    filter(ranking <= 10) %>%
    arrange(desc(total_points)) %>%
    select(ranking, player_name, total_points) %>%
    print() %>%
    ggplot(aes(reorder(player_name, -total_points), total_points, fill = player_name)) + geom_bar(stat = "identity")
}
```

```
fpl_point_rank(season17)
```


##	ranking	player_name	total_points
## 1	1	Alexis Sánchez	264
## 2	2	Eden Hazard	224
## 3	2	Harry Kane	224
## 4	3	Romelu Lukaku	221
## 5	4	Christian Eriksen	218
## 6	5	Kevin De Bruyne	199
## 7	6	Gylfi Sigurdsson	181
## 8	7	Roberto Firmino	180
## 9	8	Gary Cahill	178
## 10	8	Joshua King	178
## 11	9	Marcos Alonso	177
## 12	10	Sergio Agüero	175

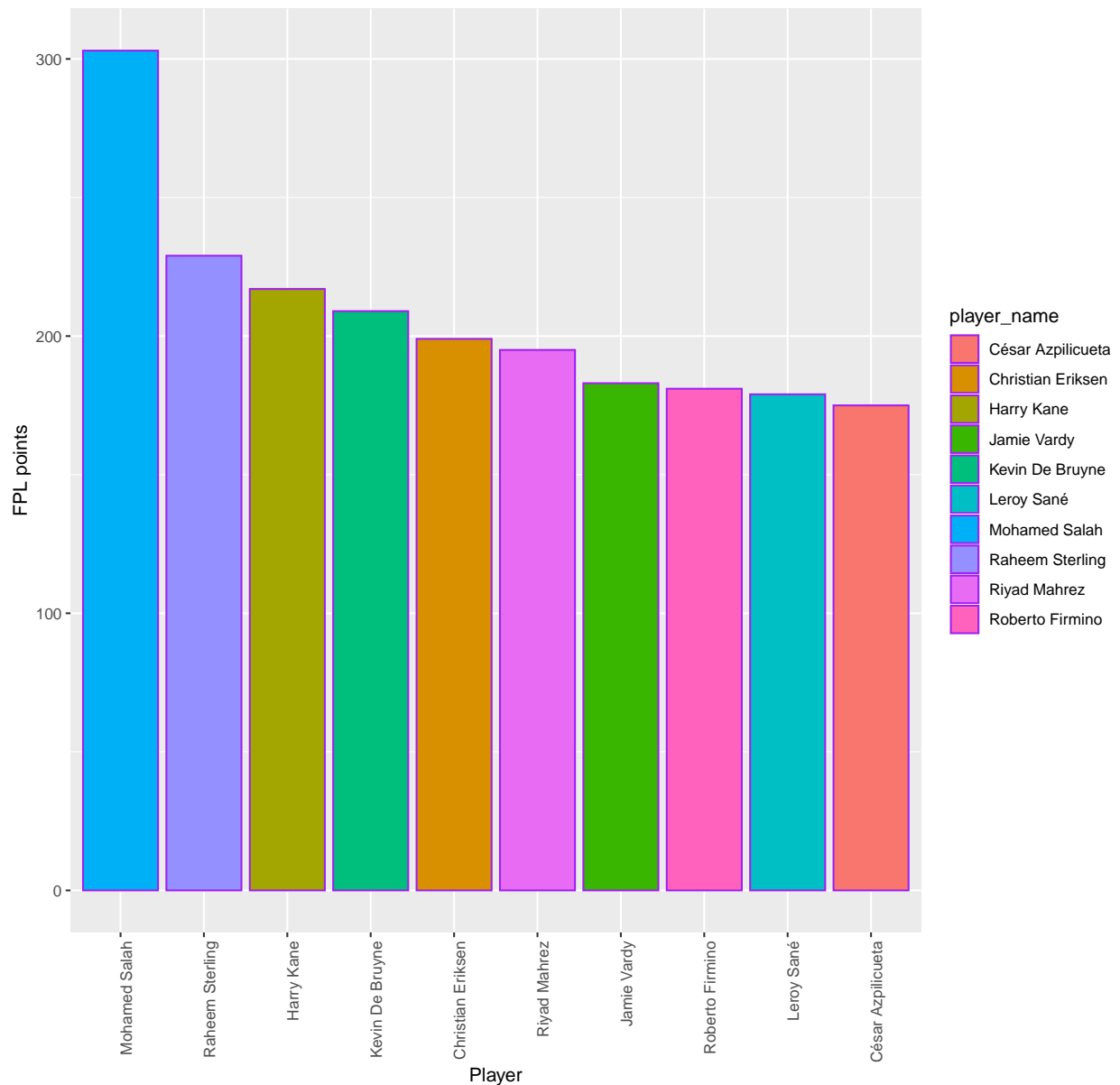
The top 10 players with the highest FPL points in 2016/2017



```
fpl_point_rank(season18)
```

##	ranking	player_name	total_points
## 1	1	Mohamed Salah	303
## 2	2	Raheem Sterling	229
## 3	3	Harry Kane	217
## 4	4	Kevin De Bruyne	209
## 5	5	Christian Eriksen	199
## 6	6	Riyad Mahrez	195
## 7	7	Jamie Vardy	183
## 8	8	Roberto Firmino	181
## 9	9	Leroy Sané	179
## 10	10	César Azpilicueta	175

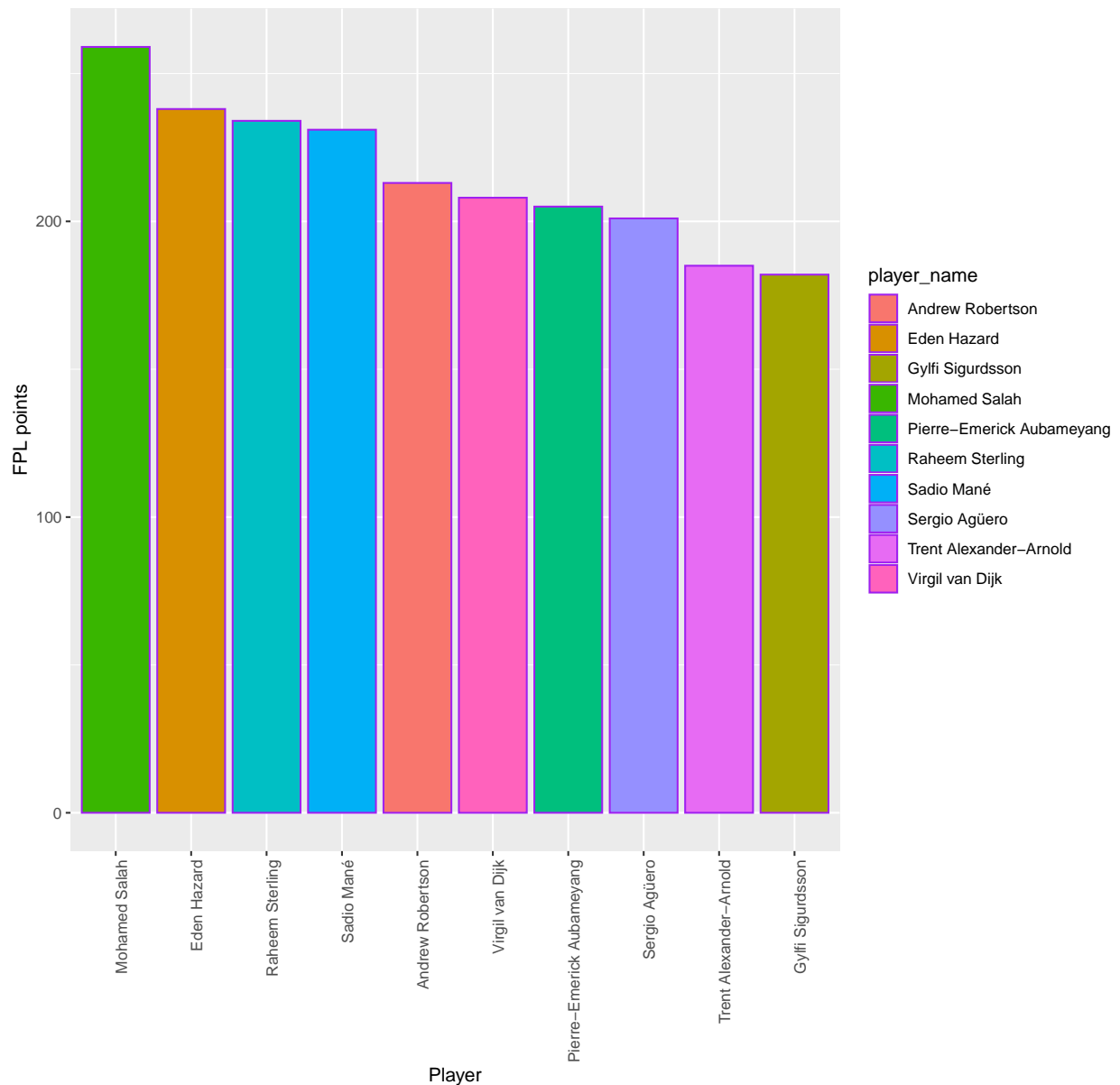
The top 10 players with the highest FPL points in 2017/2018



```
fpl_point_rank(season19)
```

##	ranking	player_name	total_points
## 1	1	Mohamed Salah	259
## 2	2	Eden Hazard	238
## 3	3	Raheem Sterling	234
## 4	4	Sadio Mané	231
## 5	5	Andrew Robertson	213
## 6	6	Virgil van Dijk	208
## 7	7	Pierre-Emerick Aubameyang	205
## 8	8	Sergio Agüero	201
## 9	9	Trent Alexander-Arnold	185
## 10	10	Gylfi Sigurdsson	182

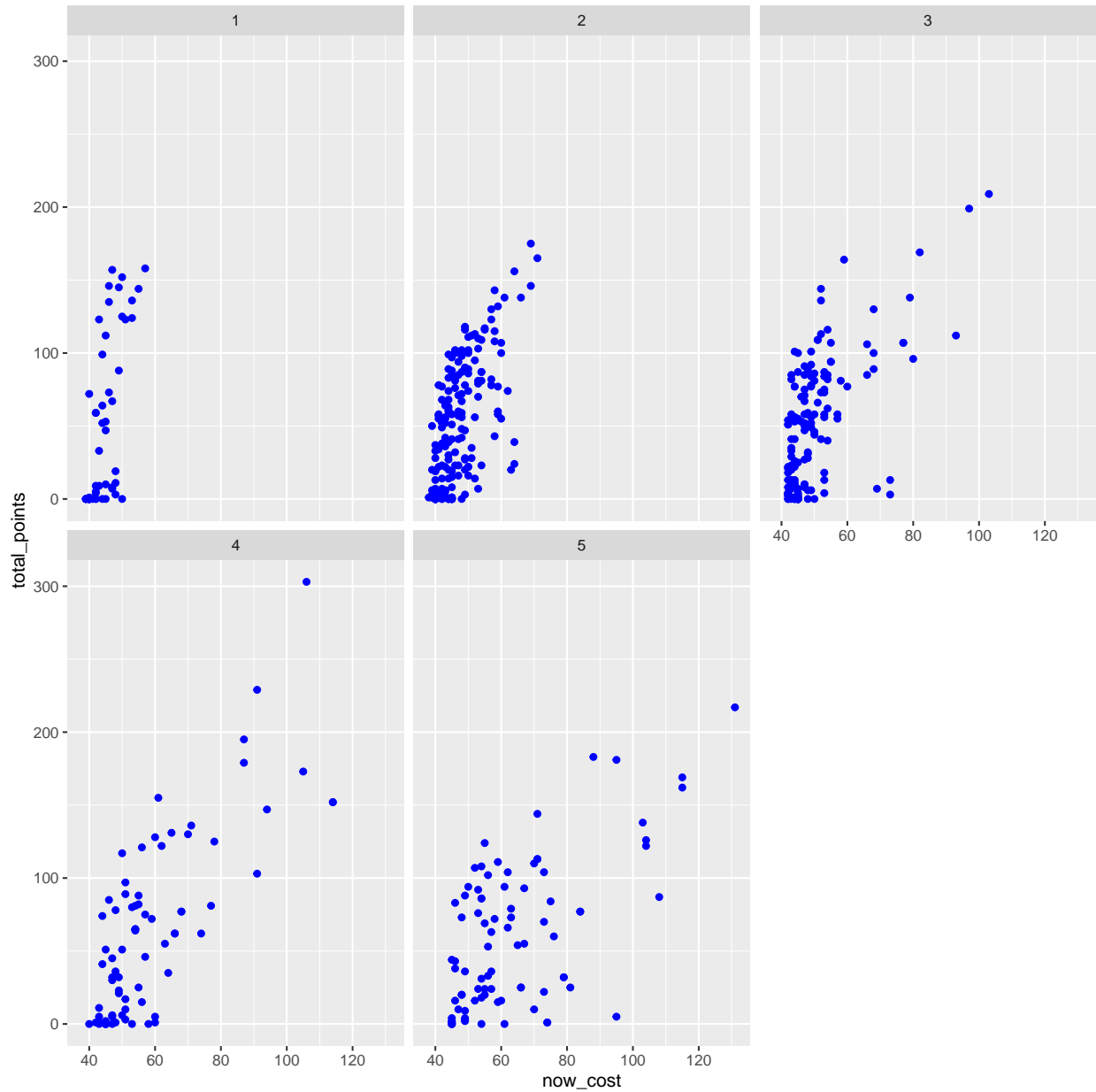
The top 10 players with the highest FPL points in 2018/2019



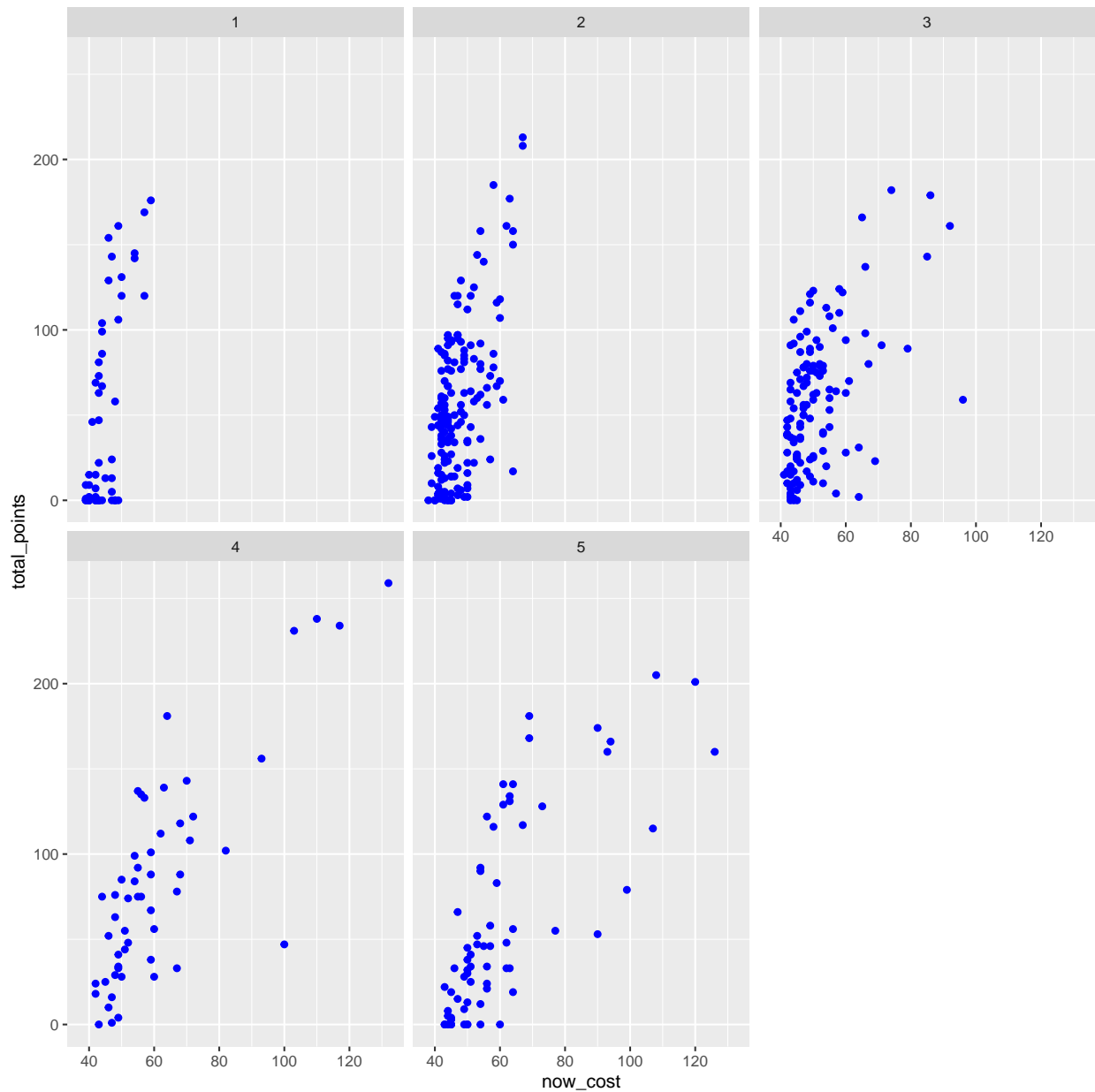
Plotting the total FPL points vs players' cost for every season

```
library(modelr)
options(na.action = na.warn)
value <- function(df) {
  ggplot(df, aes(x = now_cost)) + geom_point(aes(y = total_points), color = "blue") + facet_wrap(~positi
}
```

```
value(season18)
```



```
value(season19)
```

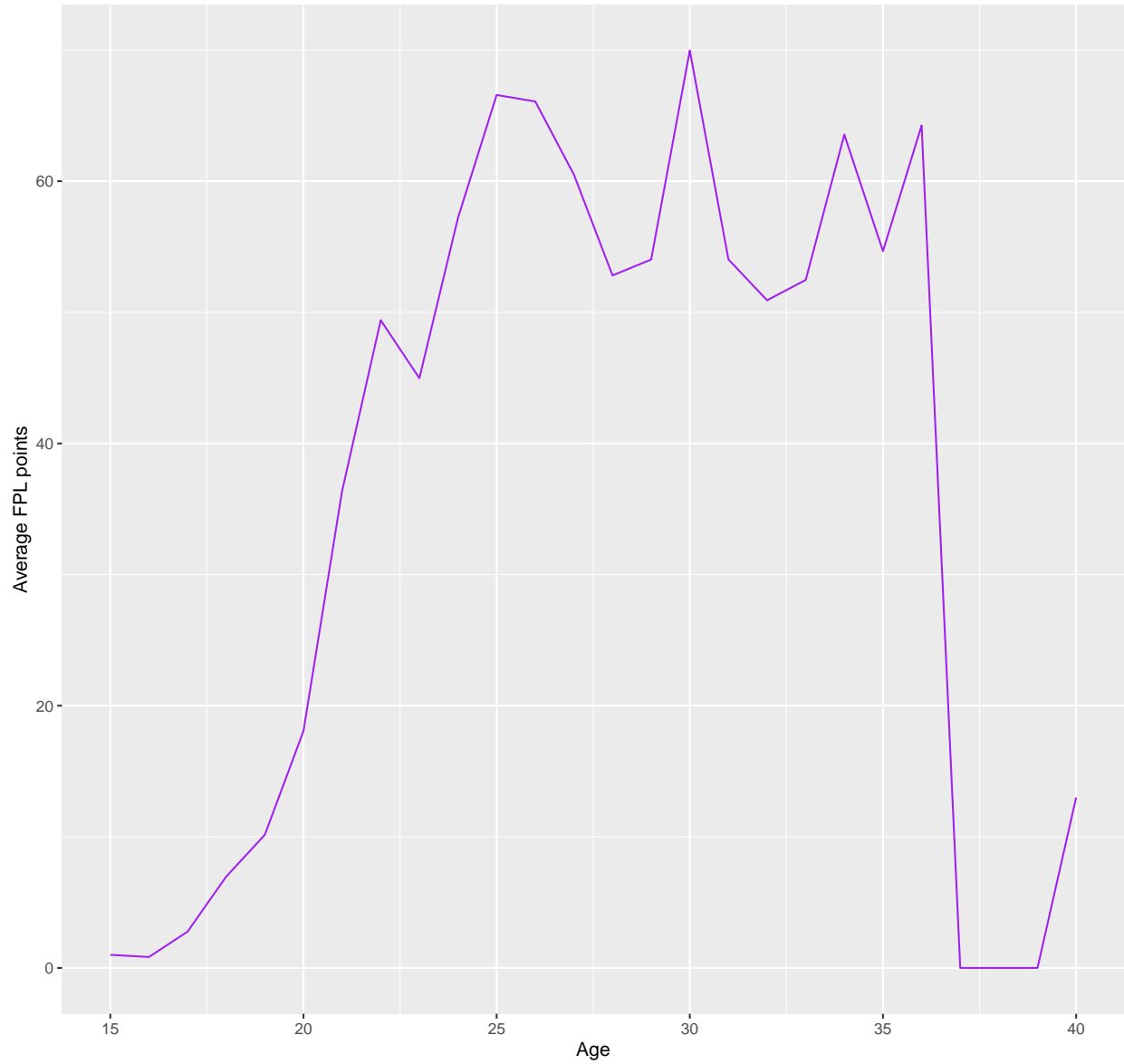


Analyzing how age has an impact on FPL points

```
age_freq <- function(df) {
  df %>%
    group_by(age) %>%
    summarize(mean = mean(total_points, na.rm = TRUE)) %>%
    ggplot(aes(x = age, y = mean)) + geom_freqpoly(stat = "identity", color = "purple") + labs(title = "Age vs FPL Points")
}
```

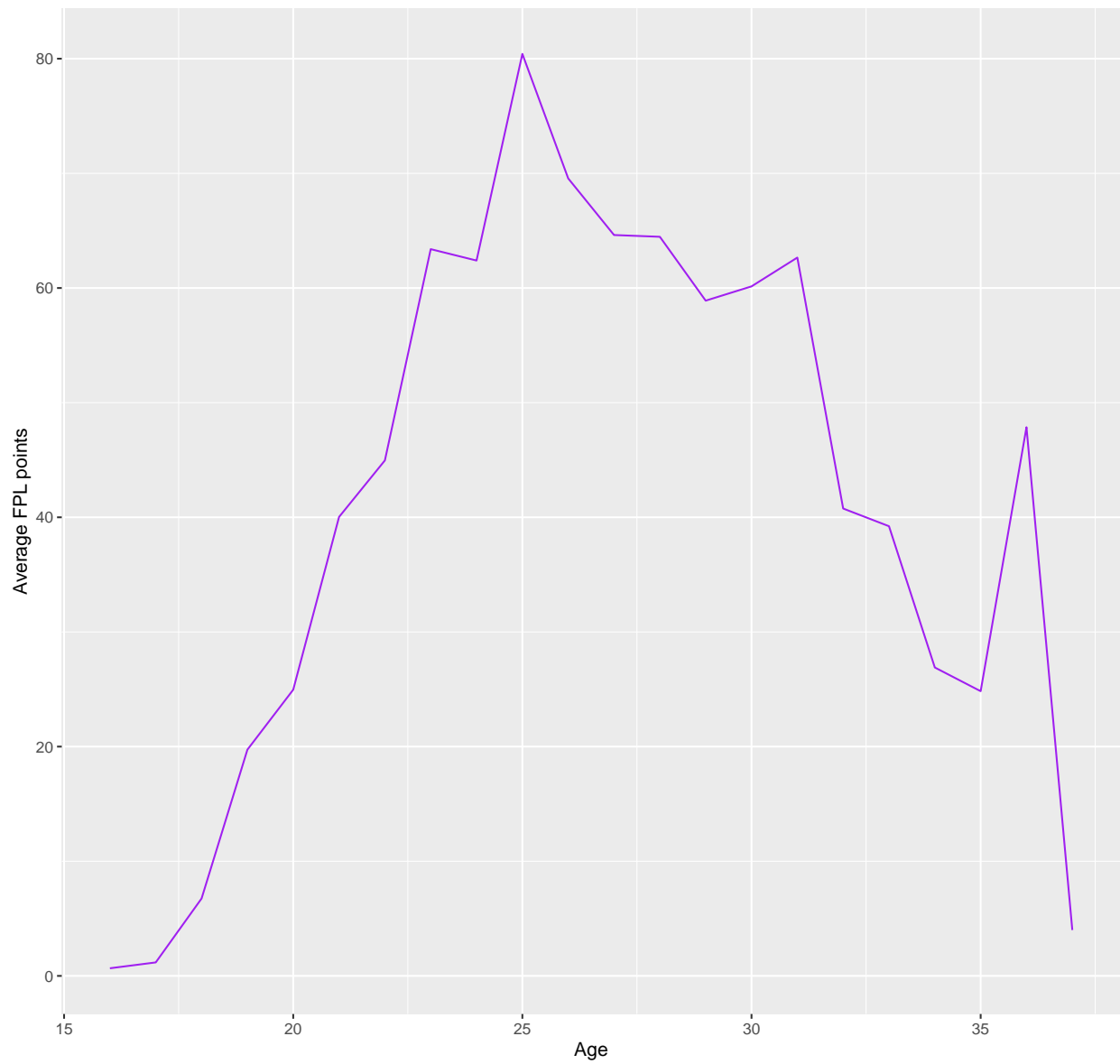
```
age_freq(season17)
```

A frequency polygon graph of the ages of Premier League players with respect to their mean FPL points during the season 2016/2017



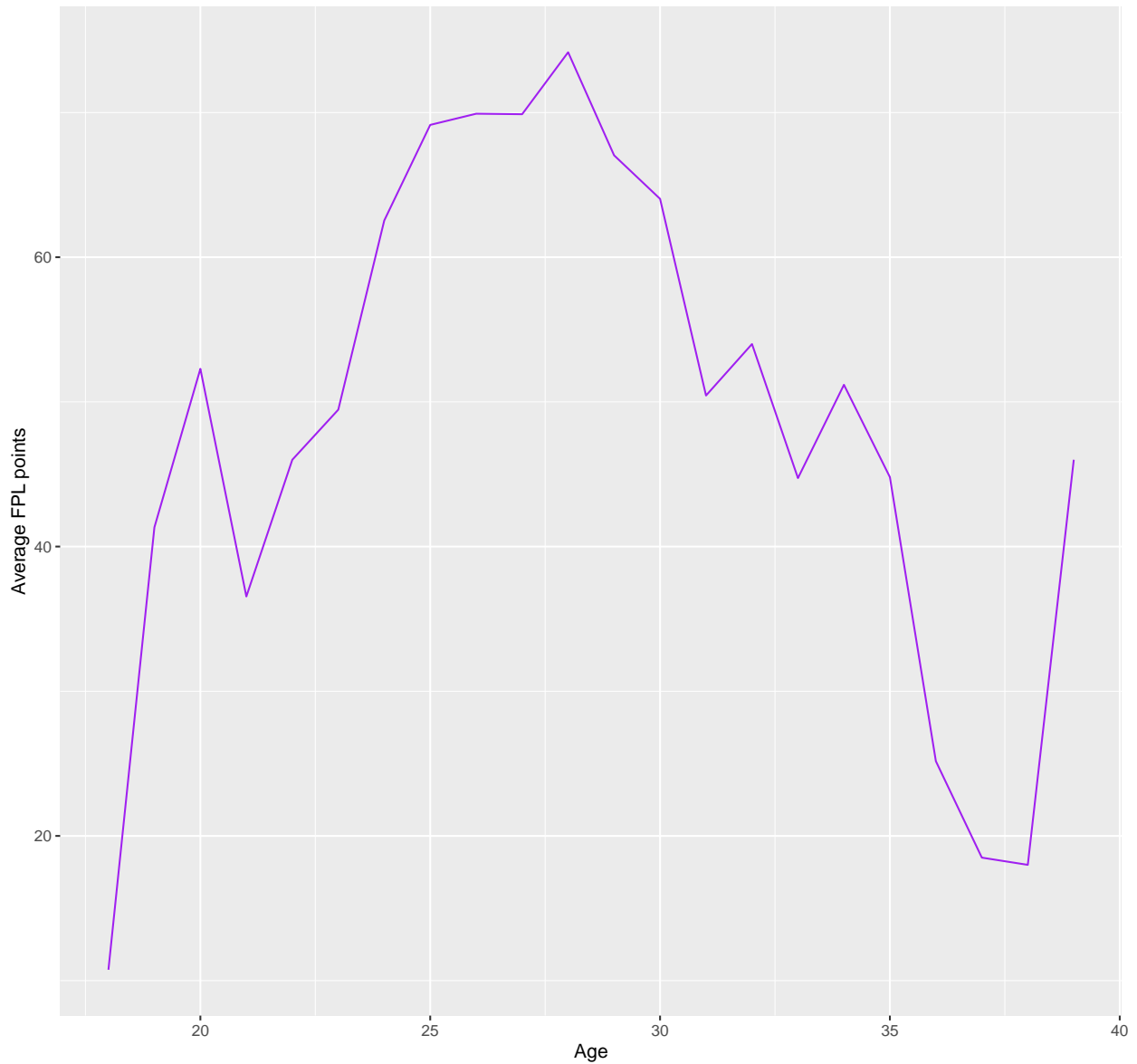
```
age_freq(season18)
```

A frequency polygon graph of the ages of Premier League players with respect to their mean FPL points during the season 2017/2018



```
age_freq(season19)
```

A frequency polygon graph of the ages of Premier League players with respect to their mean FPL points during the season 2018/2019

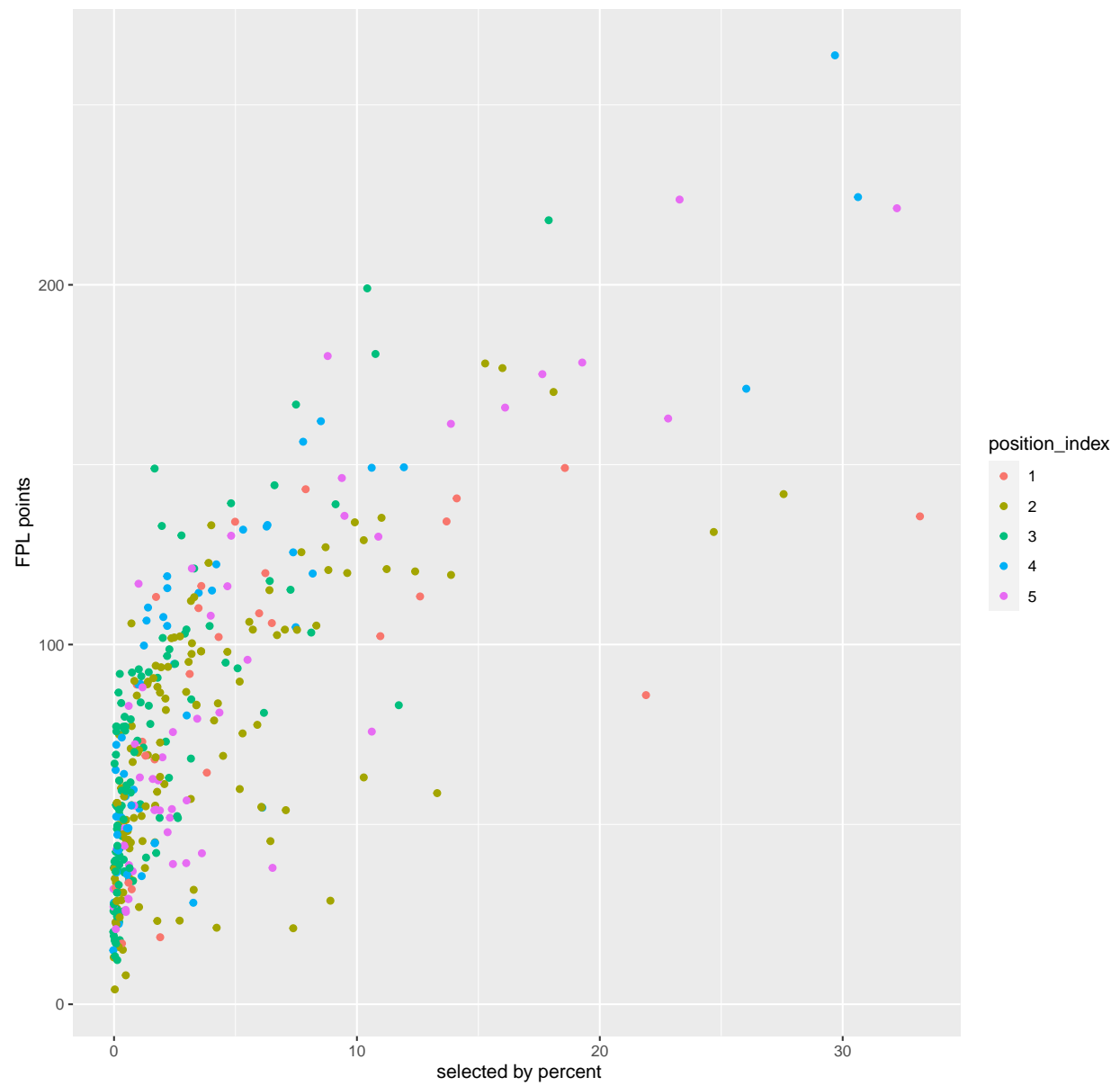


How selected by percent affects the FPL points

```
sel_fpl <- function(df) {
  df %>%
    filter(minutes.played >= 500) %>%
    ggplot(aes(selected_by_percent, total_points)) + geom_point(aes(color = position_index), position =
}
```

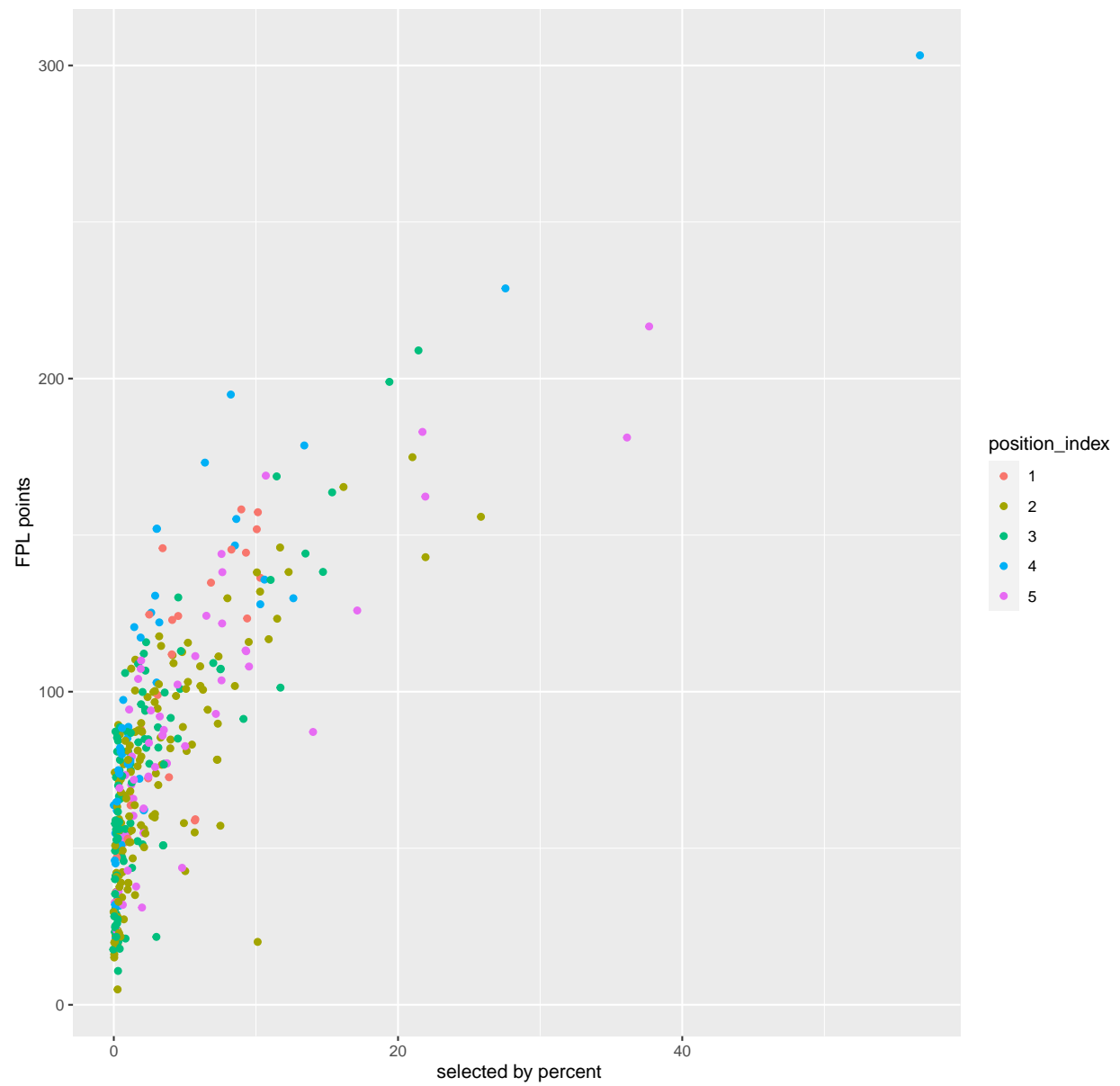
```
sel_fpl(season17)
```


A graph of FPL points vs selected by percent in the season 2016/2017



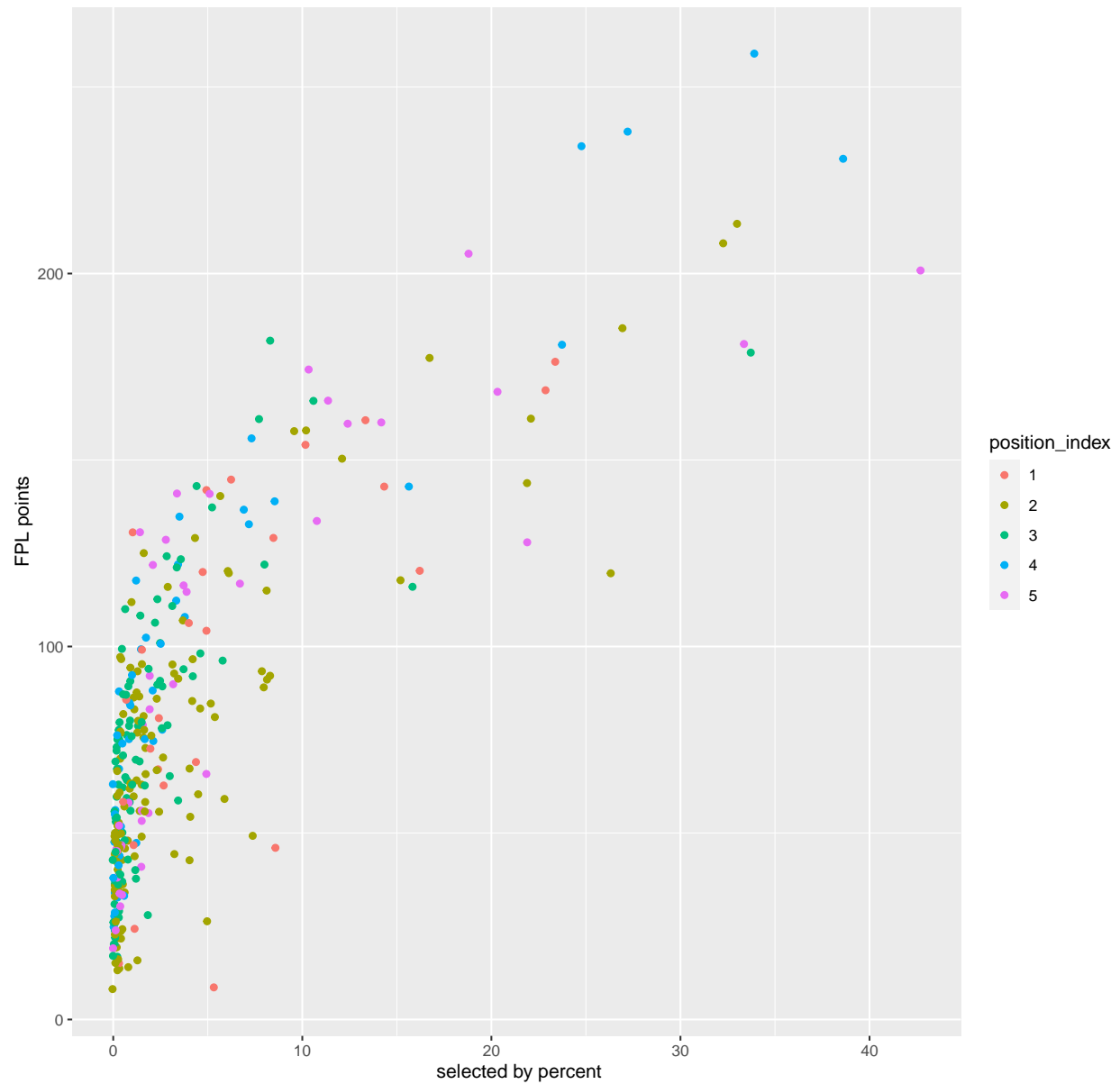
```
sel_fpl(season18)
```

A graph of FPL points vs selected by percent in the season 2017/2018



```
sel_fpl(season19)
```

A graph of FPL points vs selected by percent in the season 2018/2019



writing to csv

```
s17 <- file('season2017.csv', encoding = "UTF-8")
write.csv(season17, file = s17)
s18 <- file('season2018.csv', encoding = 'UTF-8')
write.csv(season18, file = s18)
s19 <- file('season2019.csv', encoding = "UTF-8")
write.csv(season19, file = s19)
```