R- Group 13 Assignment

Gurjeet Singh, Priyasha Duvedi, Astha Dogra, Shar Spektor, Abhinandan Umakant Jawalekar

2023-02-12

```r
install.packages("r package", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## Warning: package 'r package' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```r
options(repos = list(CRAN="http://cran.rstudio.com/"))
install.packages("readxl")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## package 'readxl' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'readxl'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\gurje\AppData\Local\R\win-library\4.2\00LOCK\readxl\libs\x64\readxl.dll
## to C:\Users\gurje\AppData\Local\R\win-library\4.2\readxl\libs\x64\readxl.dll:
## Permission denied
```

```
## Warning: restored 'readxl'
```

```
##
## The downloaded binary packages are in
##   C:\Users\gurje\AppData\Local\Temp\RtmpKURLM8\downloaded_packages
```

```r
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\gurje\AppData\Local\Temp\RtmpKURLM8\downloaded_packages
```

```r
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\gurje\AppData\Local\Temp\RtmpKURLM8\downloaded_packages
```

```r
install.packages("rlang")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## package 'rlang' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'rlang'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\gurje\AppData\Local\R\win-library\4.2\00LOCK\rlang\libs\x64\rlang.dll
## to C:\Users\gurje\AppData\Local\R\win-library\4.2\rlang\libs\x64\rlang.dll:
## Permission denied
```

```
## Warning: restored 'rlang'
```

```
##
## The downloaded binary packages are in
##   C:\Users\gurje\AppData\Local\Temp\RtmpKURLM8\downloaded_packages
```

```r
install.packages("dplyr")
```

```
## Installing package into 'C:/Users/gurje/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'dplyr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\gurje\AppData\Local\R\win-library\4.2\00LOCK\dplyr\libs\x64\dplyr.dll
## to C:\Users\gurje\AppData\Local\R\win-library\4.2\dplyr\libs\x64\dplyr.dll:
## Permission denied
```

```
## Warning: restored 'dplyr'
```

```
##
## The downloaded binary packages are in
##   C:\Users\gurje\AppData\Local\Temp\RtmpKURLM8\downloaded_packages
```

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.2
## --
```

```
## v ggplot2 3.4.1      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.1.0
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 1.0.0
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
library(rlang)
```

```
##
## Attaching package: 'rlang'
##
## The following objects are masked from 'package:purrr':
##
##     %@%, flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
##     flatten_raw, invoke, splice
```

```
library(readxl)
Esports_Data_Filtered <- read_excel("C:/Users/gurje/OneDrive/Desktop/data/Esports_Data_Filtered.xlsx")
View(Esports_Data_Filtered)

#R PROGRAMMING

#Structure of the dataset
str(Esports_Data_Filtered)
```

```
## tibble [641 x 6] (S3: tbl_df/tbl/data.frame)
##  $ TeamId          : num [1:641] 760 588 584 608 857 ...
##  $ TeamName        : chr [1:641] "San Francisco Shock" "Finnish National Team" "Swedish National Tea
##  $ TotalUSDPrize   : num [1:641] 3105000 63000 58500 56500 54400 ...
##  $ TotalTournaments: num [1:641] 7 5 5 5 3 3 4 4 2 4 ...
##  $ Game            : chr [1:641] "Overwatch" "Overwatch" "Overwatch" "Overwatch" ...
##  $ Genre           : chr [1:641] "First-Person Shooter" "First-Person Shooter" "First-Person Shooter
```

```
#Variables in the dataset
names(Esports_Data_Filtered)
```

```
## [1] "TeamId"            "TeamName"            "TotalUSDPrize"    "TotalTournaments"
## [5] "Game"              "Genre"
```

```
#Top 15 rows of the dataset
head(Esports_Data_Filtered, n=15)
```

```
## # A tibble: 15 x 6
##     TeamId TeamName                 TotalUSDPrize TotalTournaments Game      Genre
##      <dbl> <chr>                            <dbl>            <dbl> <chr>     <chr>
##  1     760 San Francisco Shock            3105000                7 Overwatch Firs~
##  2     588 Finnish National Team            63000                5 Overwatch Firs~
##  3     584 Swedish National Team            58500                5 Overwatch Firs~
##  4     608 British National Team            56500                5 Overwatch Firs~
##  5     857 Toronto Defiant                  54400                3 Overwatch Firs~
##  6     342 Star Horn Royal Club             52328.               3 Overwatch Firs~
##  7     846 Nova eSports                     50278                4 Overwatch Firs~
##  8     581 Russian National Team            49500                4 Overwatch Firs~
##  9   24788 Lowkey Esports                   47500                2 Overwatch Firs~
## 10     579 German National Team             47000                4 Overwatch Firs~
## 11     599 Spanish National Team            47000                4 Overwatch Firs~
## 12     637 Taiwanese National Team          47000                4 Overwatch Firs~
## 13     663 eUnited                          45750                4 Overwatch Firs~
## 14     611 NGA Club                         45335.               7 Overwatch Firs~
## 15     500 REUNITED                         44989.              29 Overwatch Firs~
```

```
Total_of_any_three_teams_tournaments = function(x,y,z){x+y+z}
Total_of_any_three_teams_tournaments(7,5,5)
```

```
## [1] 17
```

```
#FILTER ROWS
newesports = as.data.frame(filter(Esports_Data_Filtered,Esports_Data_Filtered$TotalTournaments > 5))
```

```
#RESHAPING

data.frame1 <- data.frame(
  Teamname = c("Team Liquid", "ROOT Gaming", "Team Acer"),
  TotalTournaments = c(808,610,466)
)

data.frame2 <- data.frame(
  Teamname = c("Team Liquid", "ROOT Gaming", "Team Acer"),
  TotalUSDPrize = c(1498925,627435,587934)
)

Total <- merge(data.frame1,data.frame2, by = "Teamname")
print(Total)
```

```
##      Teamname TotalTournaments TotalUSDPrize
## 1 ROOT Gaming              610        627435
## 2   Team Acer              466        587934
## 3 Team Liquid              808       1498925
```

```r
#The following commands cover the Question 7-9 for the Assignment.


# The Following command will Omit/Remove the missing values from data set
# And create a new data set with omitted values.
#Original Data set values will remain unchanged.
Esports_Data_Filtered_without_na <- na.omit(Esports_Data_Filtered)
View(Esports_Data_Filtered)

#The Following command will identify the rows which have duplicates and will indicate whether the row h
duplicated_rows <- duplicated(Esports_Data_Filtered)
View(duplicated_rows)

#The following command will remove all the exact duplicates from the data set.
Esports_Data_Filtered_without_duplicates <- unique(Esports_Data_Filtered)
View(duplicated_rows)

# The following command will reorder multiple rows in descending order.
# And will create a new Data set with manipulated data.

Esports_Data_Filtered_sorted_descending <- Esports_Data_Filtered[order(Esports_Data_Filtered$TeamId, de
View(Esports_Data_Filtered_sorted_descending)

#The following commands cover the Question 10-12 for the Assignment.

#The following command will change the column names
View(Esports_Data_Filtered)
names(Esports_Data_Filtered)[1]="TeamNumber"
names(Esports_Data_Filtered)[4]="NumberOfTournamnets"
names(Esports_Data_Filtered)[4]="NumberOfTournaments"

#The following command will add new column
#with some mutation
Esports_Data_Filtered %>% mutate(TournamentsNew=NumberOfTournaments*2)
```

```
## # A tibble: 641 x 7
##     TeamNumber TeamName              TotalUSDPrize NumberOf~1 Game  Genre Tourn~2
##          <dbl> <chr>                         <dbl>      <dbl> <chr> <chr>   <dbl>
## 1          760 San Francisco Shock         3105000          7 Over~ Firs~      14
## 2          588 Finnish National Team         63000          5 Over~ Firs~      10
## 3          584 Swedish National Team         58500          5 Over~ Firs~      10
## 4          608 British National Team         56500          5 Over~ Firs~      10
## 5          857 Toronto Defiant               54400          3 Over~ Firs~       6
## 6          342 Star Horn Royal Club          52328.         3 Over~ Firs~       6
## 7          846 Nova eSports                  50278          4 Over~ Firs~       8
## 8          581 Russian National Team         49500          4 Over~ Firs~       8
## 9        24788 Lowkey Esports                47500          2 Over~ Firs~       4
## 10         579 German National Team          47000          4 Over~ Firs~       8
## # ... with 631 more rows, and abbreviated variable names
## #   1: NumberOfTournaments, 2: TournamentsNew
```

```r
install.packages("tidyverse")
```

```
## Warning: package 'tidyverse' is in use and will not be installed

Esports_Data_Filtered %>% mutate(TournamentsNew=NumberOfTournaments*2)
```

```
## # A tibble: 641 x 7
##    TeamNumber TeamName           TotalUSDPrize NumberOf~1 Game  Genre Tourn~2
##         <dbl> <chr>                      <dbl>      <dbl> <chr> <chr>   <dbl>
## 1         760 San Francisco Shock      3105000          7 Over~ Firs~      14
## 2         588 Finnish National Team      63000          5 Over~ Firs~      10
## 3         584 Swedish National Team      58500          5 Over~ Firs~      10
## 4         608 British National Team      56500          5 Over~ Firs~      10
## 5         857 Toronto Defiant            54400          3 Over~ Firs~       6
## 6         342 Star Horn Royal Club       52328.         3 Over~ Firs~       6
## 7         846 Nova eSports               50278          4 Over~ Firs~       8
## 8         581 Russian National Team      49500          4 Over~ Firs~       8
## 9       24788 Lowkey Esports             47500          2 Over~ Firs~       4
## 10        579 German National Team       47000          4 Over~ Firs~       8
## # ... with 631 more rows, and abbreviated variable names
## #   1: NumberOfTournaments, 2: TournamentsNew
```

```
library("tidyverse")
Esports_Data_Filtered %>% mutate(TournamentsNew=NumberOfTournaments*2)
```

```
## # A tibble: 641 x 7
##    TeamNumber TeamName           TotalUSDPrize NumberOf~1 Game  Genre Tourn~2
##         <dbl> <chr>                      <dbl>      <dbl> <chr> <chr>   <dbl>
## 1         760 San Francisco Shock      3105000          7 Over~ Firs~      14
## 2         588 Finnish National Team      63000          5 Over~ Firs~      10
## 3         584 Swedish National Team      58500          5 Over~ Firs~      10
## 4         608 British National Team      56500          5 Over~ Firs~      10
## 5         857 Toronto Defiant            54400          3 Over~ Firs~       6
## 6         342 Star Horn Royal Club       52328.         3 Over~ Firs~       6
## 7         846 Nova eSports               50278          4 Over~ Firs~       8
## 8         581 Russian National Team      49500          4 Over~ Firs~       8
## 9       24788 Lowkey Esports             47500          2 Over~ Firs~       4
## 10        579 German National Team       47000          4 Over~ Firs~       8
## # ... with 631 more rows, and abbreviated variable names
## #   1: NumberOfTournaments, 2: TournamentsNew
```

```
#The following command will generate random set of numbers
set.seed(1234)
Esports_Data_Filtered %>% sample_frac(0.75,replace = FALSE)
```

```
## # A tibble: 481 x 6
##    TeamNumber TeamName          TotalUSDPrize NumberOfTournaments Game   Genre
##         <dbl> <chr>                     <dbl>               <dbl> <chr>  <chr>
## 1       24746 Geek Fam               284938.                   20 Dota 2 Mult~
## 2          NA <NA>                        NA                   NA <NA>   <NA>
## 3         343 Royal Never Give Up    361620.                    4 Arena~ Mult~
## 4          NA <NA>                        NA                   NA <NA>   <NA>
## 5          NA <NA>                        NA                   NA <NA>   <NA>
## 6          NA <NA>                        NA                   NA <NA>   <NA>
```

```
##  7          593 Czech National Team          1500          1 Heart~ Coll~
##  8          711 Team Zenith              48217.          7 Dota 2 Mult~
##  9          NA <NA>                    NA          NA <NA>   <NA>
## 10          192 Gambit Esports          683780.          30 Dota 2 Mult~
## # ... with 471 more rows
```

```r
Esports_Data_Filtered_new=as.data.frame(Esports_Data_Filtered %>% sample_frac(0.75,replace = FALSE))
```

```r
#The following commands cover the Question 13-14 for the Assignment.

#The following command is used for summary stat of the data.
summary(Esports_Data_Filtered)
```

```
##     TeamNumber        TeamName          TotalUSDPrize        NumberOfTournaments
##  Min.   :  101   Length:641         Min.   :     175   Min.   :   1.00
##  1st Qu.:  224   Class :character   1st Qu.:   40000   1st Qu.:   4.00
##  Median :  494   Mode  :character   Median :  132490   Median :  13.00
##  Mean   : 2961                      Mean   :  817407   Mean   :  30.29
##  3rd Qu.:  706                      3rd Qu.:  416233   3rd Qu.:  36.00
##  Max.   :24987                      Max.   :33810636   Max.   : 808.00
##  NA's   :168                        NA's   :168        NA's   :168
##      Game               Genre
##  Length:641         Length:641
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```r
# For statistical functions, used na.rm - TRUE to avoid NA values from affecting the results
mean(Esports_Data_Filtered$TotalUSDPrize,na.rm = TRUE)
```

```
## [1] 817406.7
```

```r
median(Esports_Data_Filtered$TotalUSDPrize, na.rm = TRUE)
```

```
## [1] 132489.5
```

```r
#For mode, create a user defined function since standard function doesn't work
#Used a library modeest for mode calculation
mymode = function(x){
  library(modeest)
  mode = mfv(x,na_rm = TRUE)
  print(mode)
}
mymode(Esports_Data_Filtered$TotalTournaments)
```

```
## Registered S3 method overwritten by 'rmutil':
##   method         from
##   print.response httr
```

```
## Warning: Unknown or uninitialised column: 'TotalTournaments'.
```

```
## [1] NaN
```

```
view(Esports_Data_Filtered$NumberOfTournaments)
#The following command will show the range.
range(Esports_Data_Filtered$TotalTournaments,na.rm = TRUE)
```

```
## Warning: Unknown or uninitialised column: 'TotalTournaments'.
```

```
## Warning in min(x, na.rm = na.rm): no non-missing arguments to min; returning
## Inf
```

```
## Warning in max(x, na.rm = na.rm): no non-missing arguments to max; returning
## -Inf
```
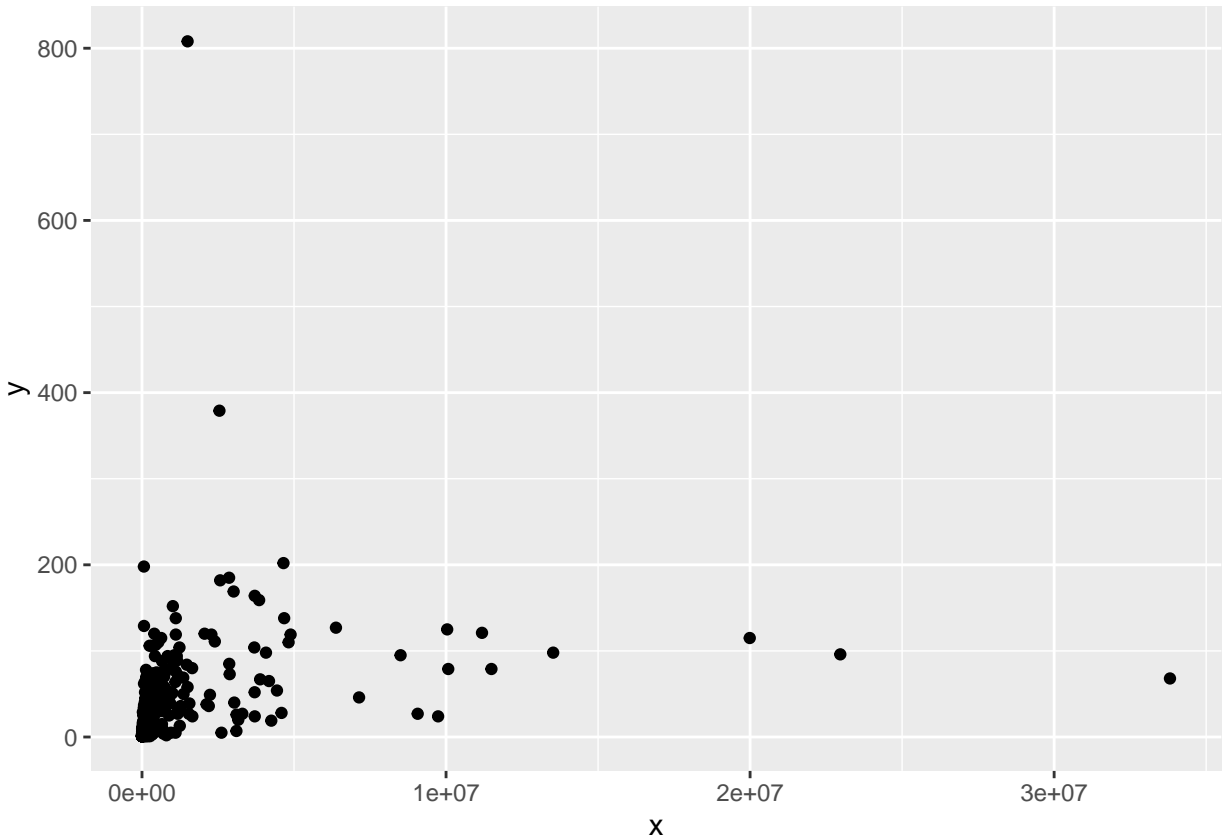
```
## [1]  Inf -Inf
```

```
view(Esports_Data_Filtered)
#The following commands cover the Question 15-17 for the Assignment.

# The following code will Plot a scatter plot for any 2 variables in data set
Esports_Data_Filtered
```

```
## # A tibble: 641 x 6
##    TeamNumber TeamName              TotalUSDPrize NumberOfTourname~1 Game  Genre
##         <dbl> <chr>                         <dbl>              <dbl> <chr> <chr>
## 1         760 San Francisco Shock         3105000                  7 Over~ Firs~
## 2         588 Finnish National Team         63000                  5 Over~ Firs~
## 3         584 Swedish National Team         58500                  5 Over~ Firs~
## 4         608 British National Team         56500                  5 Over~ Firs~
## 5         857 Toronto Defiant               54400                  3 Over~ Firs~
## 6         342 Star Horn Royal Club          52328.                 3 Over~ Firs~
## 7         846 Nova eSports                  50278                  4 Over~ Firs~
## 8         581 Russian National Team         49500                  4 Over~ Firs~
## 9       24788 Lowkey Esports                47500                  2 Over~ Firs~
## 10        579 German National Team          47000                  4 Over~ Firs~
## # ... with 631 more rows, and abbreviated variable name 1: NumberOfTournaments
```

```
x<-Esports_Data_Filtered$TotalUSDPrize
y<-Esports_Data_Filtered$NumberOfTournaments
ggplot(data.frame(x,y), aes(x=x,y=y))+ geom_point()
```

```
## Warning: Removed 168 rows containing missing values ('geom_point()').
```
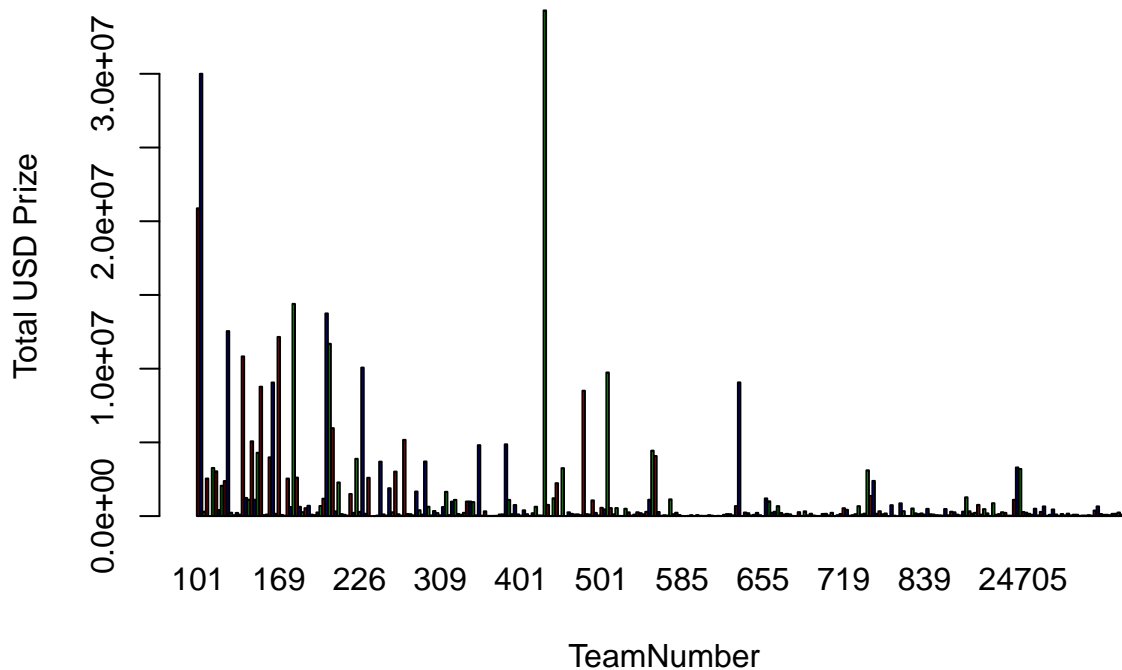
Esports_Data_Filtered

```
## # A tibble: 641 x 6
##    TeamNumber TeamName              TotalUSDPrize NumberOfTourname~1 Game  Genre
##         <dbl> <chr>                         <dbl>              <dbl> <chr> <chr>
## 1         760 San Francisco Shock         3105000                  7 Over~ Firs~
## 2         588 Finnish National Team         63000                  5 Over~ Firs~
## 3         584 Swedish National Team         58500                  5 Over~ Firs~
## 4         608 British National Team         56500                  5 Over~ Firs~
## 5         857 Toronto Defiant               54400                  3 Over~ Firs~
## 6         342 Star Horn Royal Club          52328.                 3 Over~ Firs~
## 7         846 Nova eSports                  50278                  4 Over~ Firs~
## 8         581 Russian National Team         49500                  4 Over~ Firs~
## 9       24788 Lowkey Esports                47500                  2 Over~ Firs~
## 10        579 German National Team          47000                  4 Over~ Firs~
## # ... with 631 more rows, and abbreviated variable name 1: NumberOfTournaments
```

```
agg_data <- as.data.frame(aggregate(TotalUSDPrize ~ TeamNumber, data = Esports_Data_Filtered, sum))
barplot(agg_data$TotalUSDPrize ~ agg_data$TeamNumber,
        xlab = "TeamNumber",
        ylab = "Total USD Prize",
        main = "Bar Plot of Total USD Prize by TeamNumber",
        col = c("red", "blue","green"))
```

Bar Plot of Total USD Prize by TeamNumber


```
# The following command will find the correlation between 2 variables via linear regression model
view(Esports_Data_Filtered)

Esports_Data_Filtered
```

```
## # A tibble: 641 x 6
##     TeamNumber TeamName              TotalUSDPrize NumberOfTourname~1 Game  Genre
##          <dbl> <chr>                         <dbl>              <dbl> <chr> <chr>
## 1          760 San Francisco Shock        3105000                  7 Over~ Firs~
## 2          588 Finnish National Team        63000                  5 Over~ Firs~
## 3          584 Swedish National Team        58500                  5 Over~ Firs~
## 4          608 British National Team        56500                  5 Over~ Firs~
## 5          857 Toronto Defiant              54400                  3 Over~ Firs~
## 6          342 Star Horn Royal Club         52328.                 3 Over~ Firs~
## 7          846 Nova eSports                 50278                  4 Over~ Firs~
## 8          581 Russian National Team        49500                  4 Over~ Firs~
## 9        24788 Lowkey Esports               47500                  2 Over~ Firs~
## 10         579 German National Team         47000                  4 Over~ Firs~
## # ... with 631 more rows, and abbreviated variable name 1: NumberOfTournaments
```

```
plot(Esports_Data_Filtered$TotalUSDPrize, Esports_Data_Filtered$NumberOfTournaments)
model <- lm(NumberOfTournaments ~ TotalUSDPrize, data = Esports_Data_Filtered)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = NumberOfTournaments ~ TotalUSDPrize, data = Esports_Data_Filtered)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -157.99  -22.18  -14.61    4.96  773.67
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.544e+01  2.453e+00  10.372  < 2e-16 ***
## TotalUSDPrize 5.931e-06  8.985e-07   6.601  1.1e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.9 on 471 degrees of freedom
##   (168 observations deleted due to missingness)
## Multiple R-squared:  0.08469,    Adjusted R-squared:  0.08274
## F-statistic: 43.58 on 1 and 471 DF,  p-value: 1.1e-10
```

```
abline(model, col = "red")
```