Project RD

John Gleason

5/4/2022

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

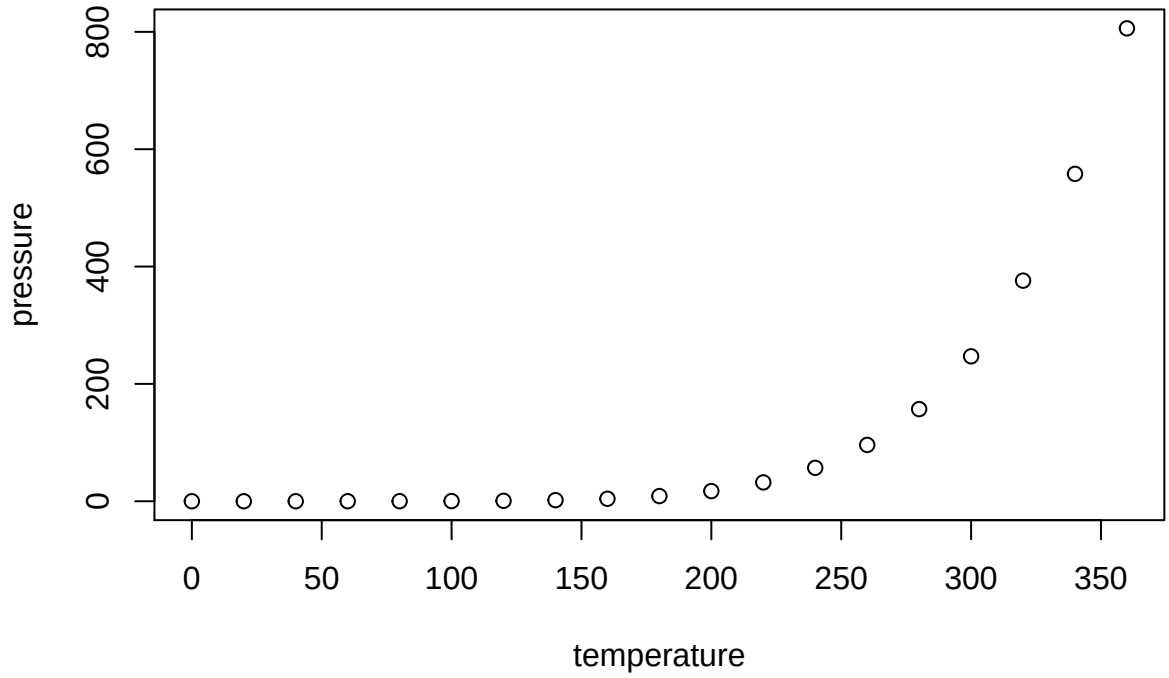
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Intro

What:

I would like to look through and examine PGA Tour stats from 2019 to 2015 to shed light on what area of the game effective the player score and world rankings the most.

Why:

A new approach to the game of golf on the PGA tour has caught some momentum in the last few years. Historically, the thought was to score well on tour players must be extremely efficient on the putting green. In today game, there are a hand full of players that are beginning to focus more on driving distance than putting efficiency. I want to dig deeper into which approach is backed by the numbers.

How:

I am going to pull 5 years of data from pgatour.com site for the top 100 players in the world. I am going to be looking Driving Distance, Green in Regulation and putting strokes per round.

Body

Topics From Class

Topic 1:

Installing and Cleaning of Data

```
library(readxl)
library(tidyverse)
```

```
## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'pillar'
```

```
## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'hms'
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.6
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

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

```
library(stringr)
```

```
pgafulldata_1_copy <- read_excel("~/Desktop/pgafulldata (1) copy.xlsx")
```

```
pga_data_2 <- pgafulldata_1_copy %>% na.omit()
```

```
pga_data_2$Year.Date <- str_sub(pga_data_2$'PLAYER NAME', -6)
```

```
pga_data_2$PLAYER.NAME <- str_sub(pga_data_2$'PLAYER NAME',0 ,-7)
```

```
PGA <- pga_data_2[-c(1,3,6,8,9,10,11)]
```

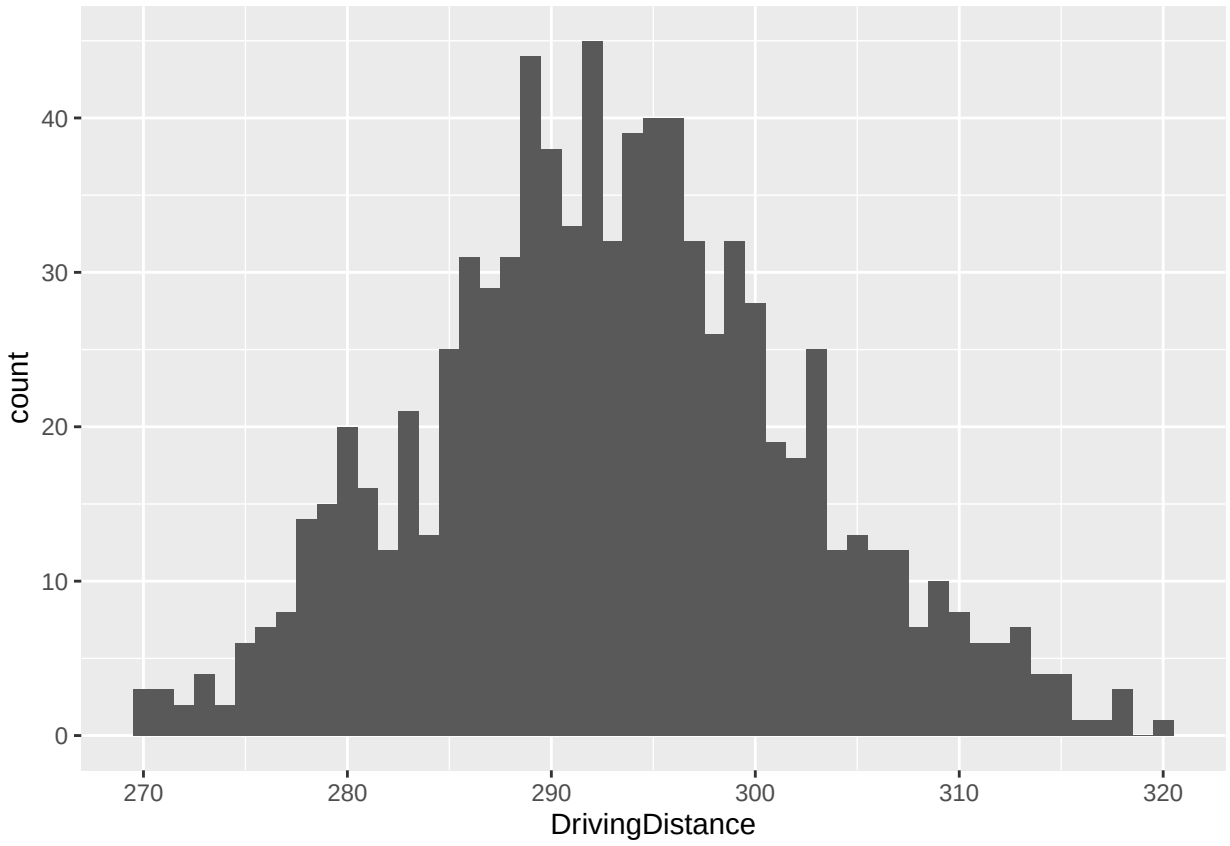
Topic 2:

Understanding the histograms of the variable and if their is linear regression

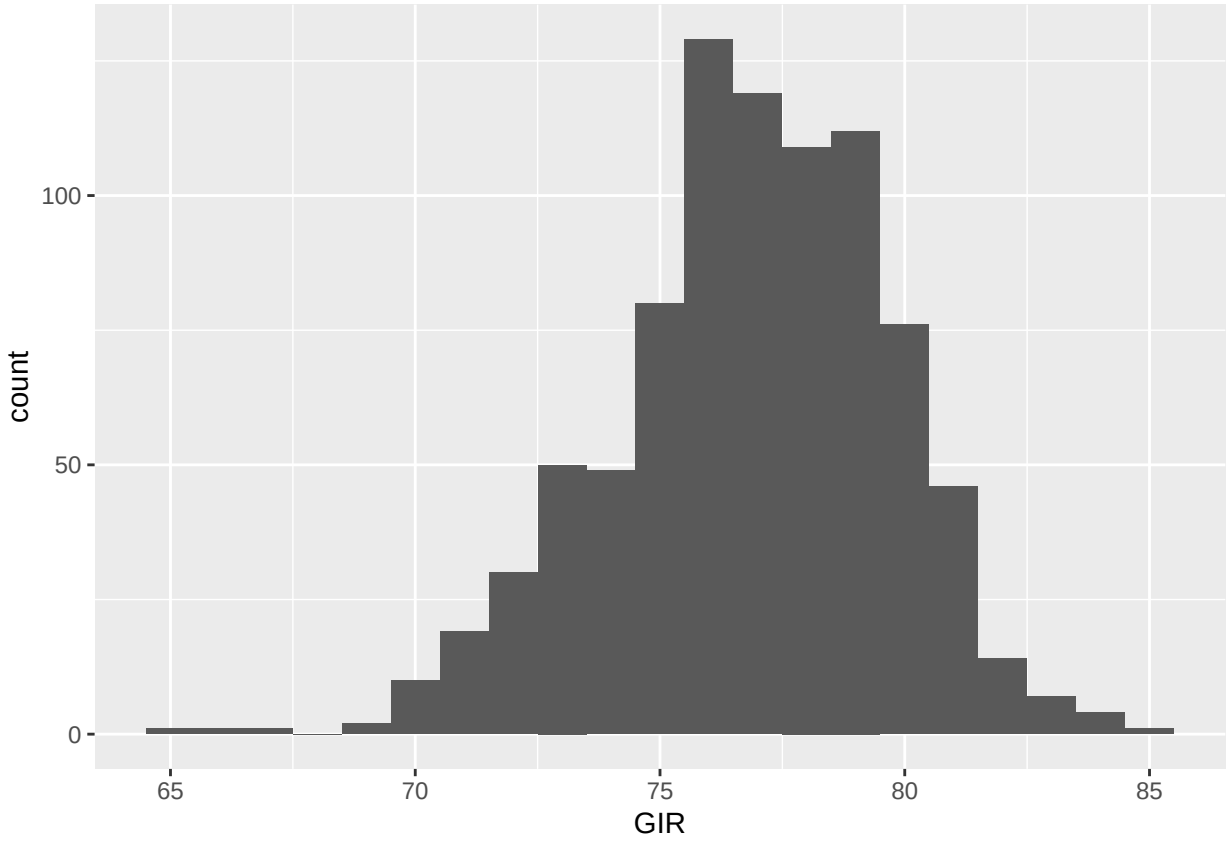
```
summary(PGA)
```

```
##  DrivingDistance PuttsPerRound   FedexCupPoints        GIR
##  Min.   :269.7   Min.   :27.71   Min.   :   9.0   Min.   :65.15
##  1st Qu.:287.1   1st Qu.:28.72   1st Qu.: 322.8   1st Qu.:75.16
##  Median :292.7   Median :29.07   Median : 529.0   Median :77.00
##  Mean   :293.1   Mean   :29.07   Mean   : 644.2   Mean   :76.85
##  3rd Qu.:298.9   3rd Qu.:29.38   3rd Qu.: 833.5   3rd Qu.:78.95
##  Max.   :319.7   Max.   :30.54   Max.   :4169.0   Max.   :84.86
##   Year.Date         PLAYER.NAME
##  Length:860         Length:860
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
ggplot(PGA) + geom_histogram(aes(x=DrivingDistance), binwidth = 1)
```

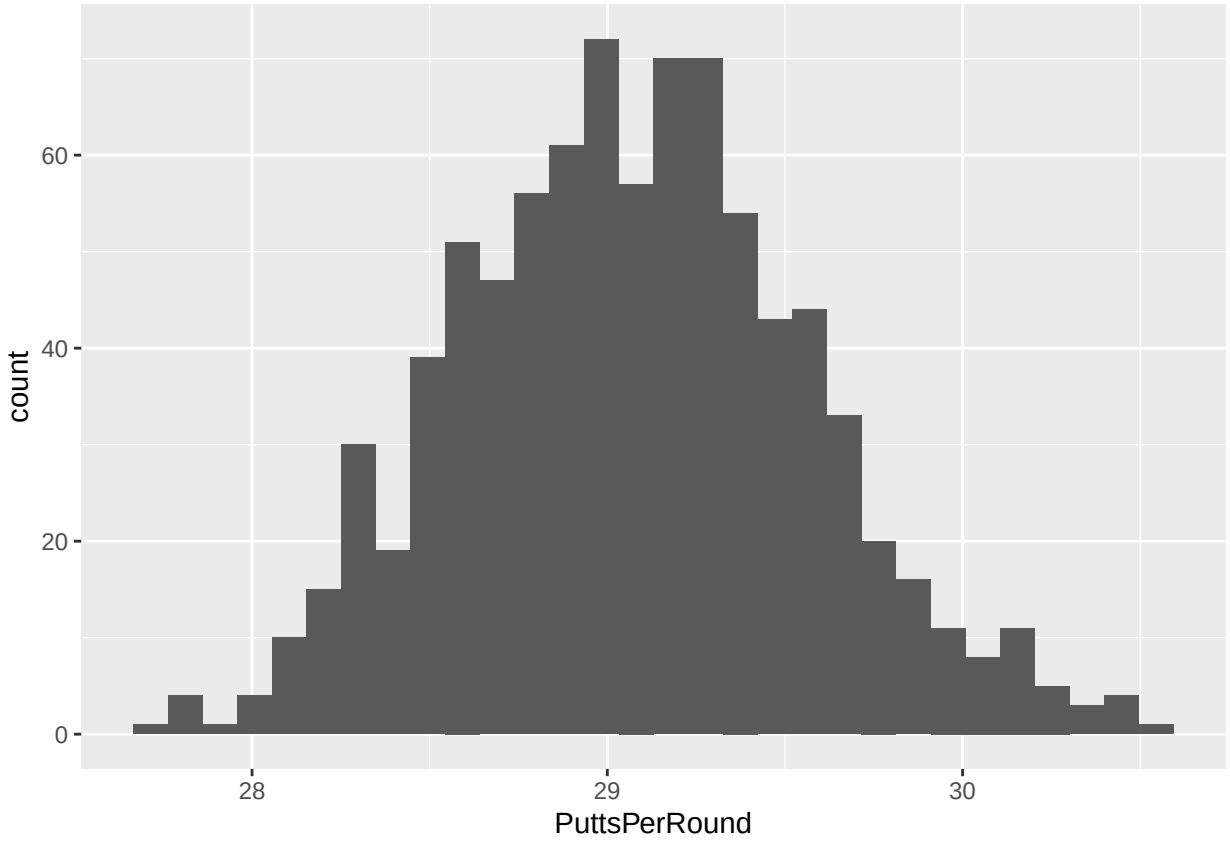



```
ggplot(PGA) + geom_histogram(aes(x=GIR), binwidth = 1)
```

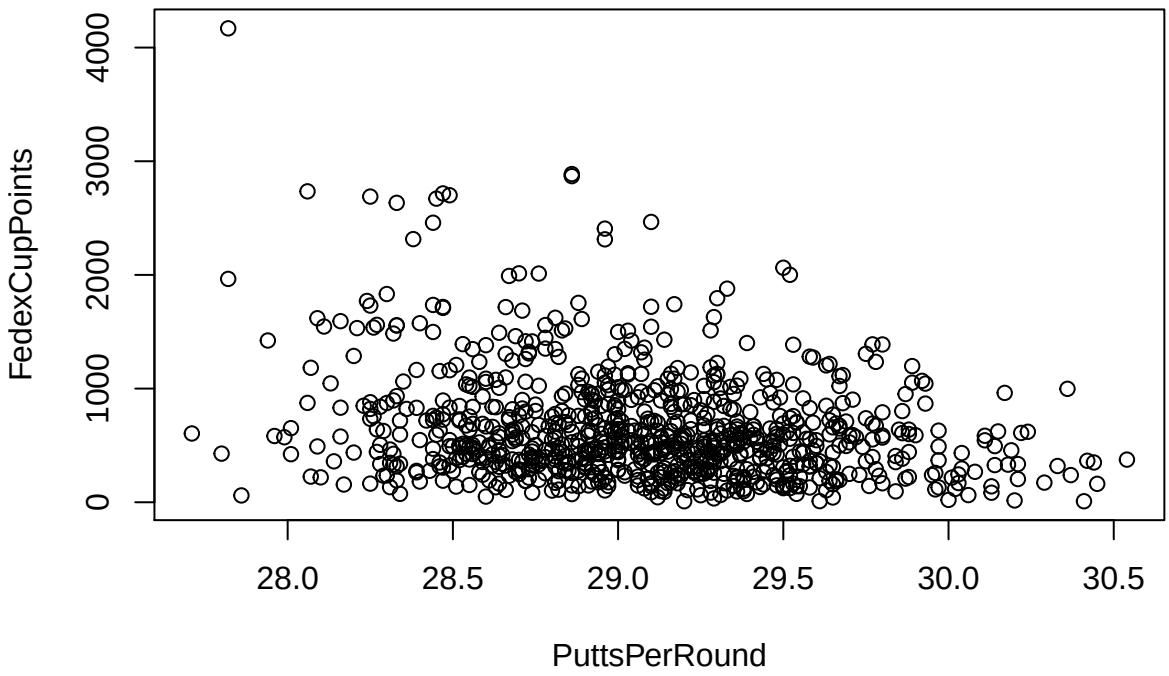


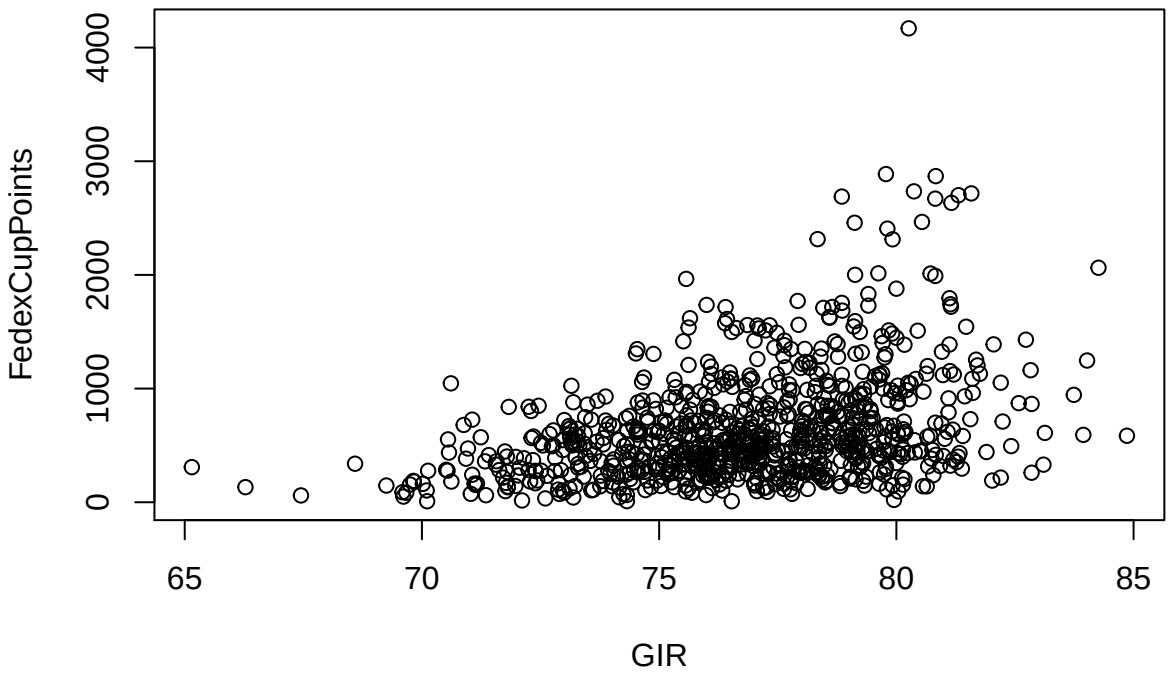

```
ggplot(PGA) + geom_histogram(aes(x=PuttsPerRound))
```

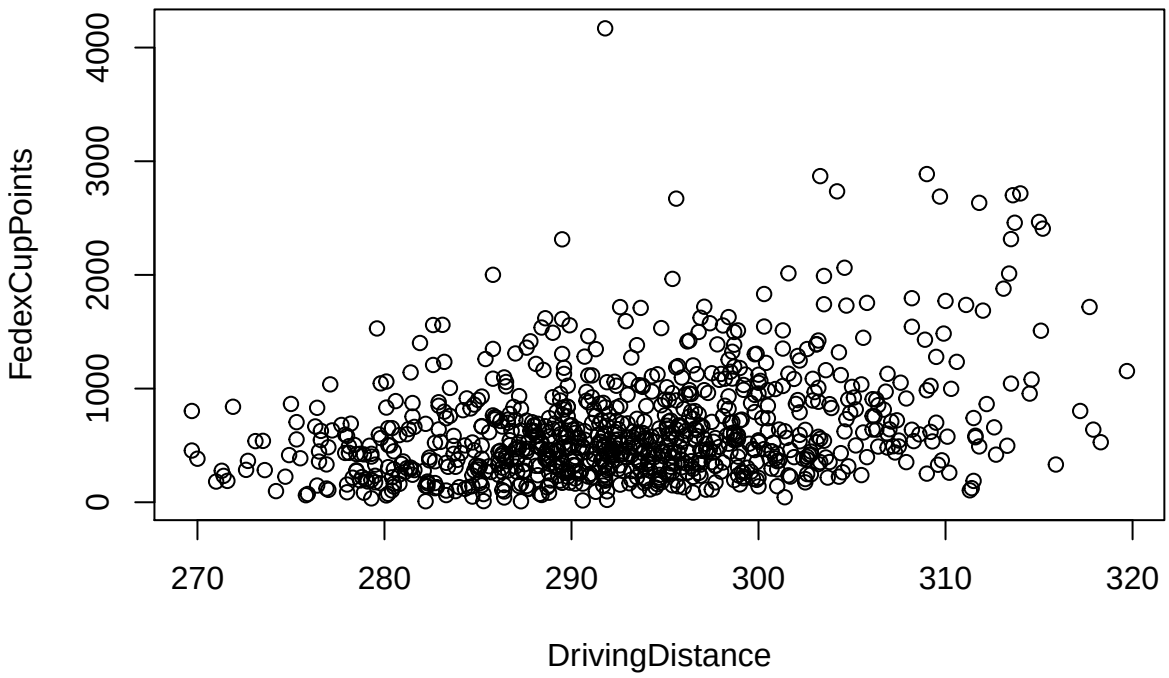
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
plot(FedexCupPoints ~ PuttsPerRound + GIR + DrivingDistance, data = PGA)
```


FedexCupPoints
0
1000
2000
3000
4000
28.0
28.5
29.0
29.5
30.0
30.5
PuttsPerRound


FedexCupPoints
0
1000
2000
3000
4000
65
70
75
80
85
GIR



Topic 3:

Filtering and using Mean to Compare

```
Q3 <- PGA %>% filter(FedexCupPoints > 833)
```

```
Q1 <- PGA %>% filter(FedexCupPoints < 323)
```

```
summary(Q3)
```

```
##  DrivingDistance PuttsPerRound   FedexCupPoints      GIR
##  Min.   :271.9   Min.   :27.82   Min.   : 835   Min.   :70.61
##  1st Qu.:290.9   1st Qu.:28.59   1st Qu.: 961   1st Qu.:76.82
##  Median :297.4   Median :28.96   Median :1127   Median :78.55
##  Mean   :297.3   Mean   :28.94   Mean   :1293   Mean   :78.34
##  3rd Qu.:303.2   3rd Qu.:29.26   3rd Qu.:1488   3rd Qu.:79.90
##  Max.   :319.7   Max.   :30.36   Max.   :4169   Max.   :84.26
##   Year.Date          PLAYER.NAME
##  Length:215         Length:215
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
summary(Q1)
```

```
##  DrivingDistance PuttsPerRound   FedexCupPoints       GIR
##  Min.   :271.0   Min.   :27.86   Min.   :  9.0   Min.   :65.15
##  1st Qu.:283.3   1st Qu.:28.88   1st Qu.:143.0   1st Qu.:73.26
##  Median :289.4   Median :29.22   Median :211.0   Median :75.79
##  Mean   :289.5   Mean   :29.21   Mean   :201.7   Mean   :75.51
##  3rd Qu.:295.2   3rd Qu.:29.52   3rd Qu.:268.0   3rd Qu.:77.80
##  Max.   :311.5   Max.   :30.45   Max.   :322.0   Max.   :82.85
##   Year.Date         PLAYER.NAME
##  Length:215         Length:215
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
DDPC<- ((297.4-289.4)/289.4)*100
```

```
DDPC
```

```
## [1] 2.76434
```

```
PPRPC <-((28.94 - 29.22)/29.22)*100
```

```
PPRPC
```

```
## [1] -0.9582478
```

```
GIRPC <- ((78.51 - 75.51)/ 75.51)*100
```

```
GIRPC
```

```
## [1] 3.972984
```

Conclusion