DSC520_Week1_Assignment00

Reenie Christudass

2022-07-11

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.1
```

```
library(ggplot2)
```

```
## Read the file scores
df <- read.csv("C:/Users/chris/dsc520/data/scores.csv")
head(df)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 6    10   265 Regular
```

```
## Split the dataframe by section
sport_df <- df %>% filter(Section =='Sports')
sport_df <- sport_df[1:2]
head(sport_df)
```

```
##   Count Score
## 1    10   200
## 2    10   205
## 3    20   235
## 4    10   240
## 5    10   250
## 6    30   285
```

```
regular_df <- df %>% filter(Section =='Regular')
regular_df <- regular_df[1:2]
head(regular_df)
```

```
##   Count Score
## 1    10   265
## 2    10   275
## 3    10   295
## 4    10   300
## 5    10   305
## 6    10   310
```

```
print("Observational unit is - Professor teaching the student")
```

```
## [1] "Observational unit is - Professor teaching the student"
```

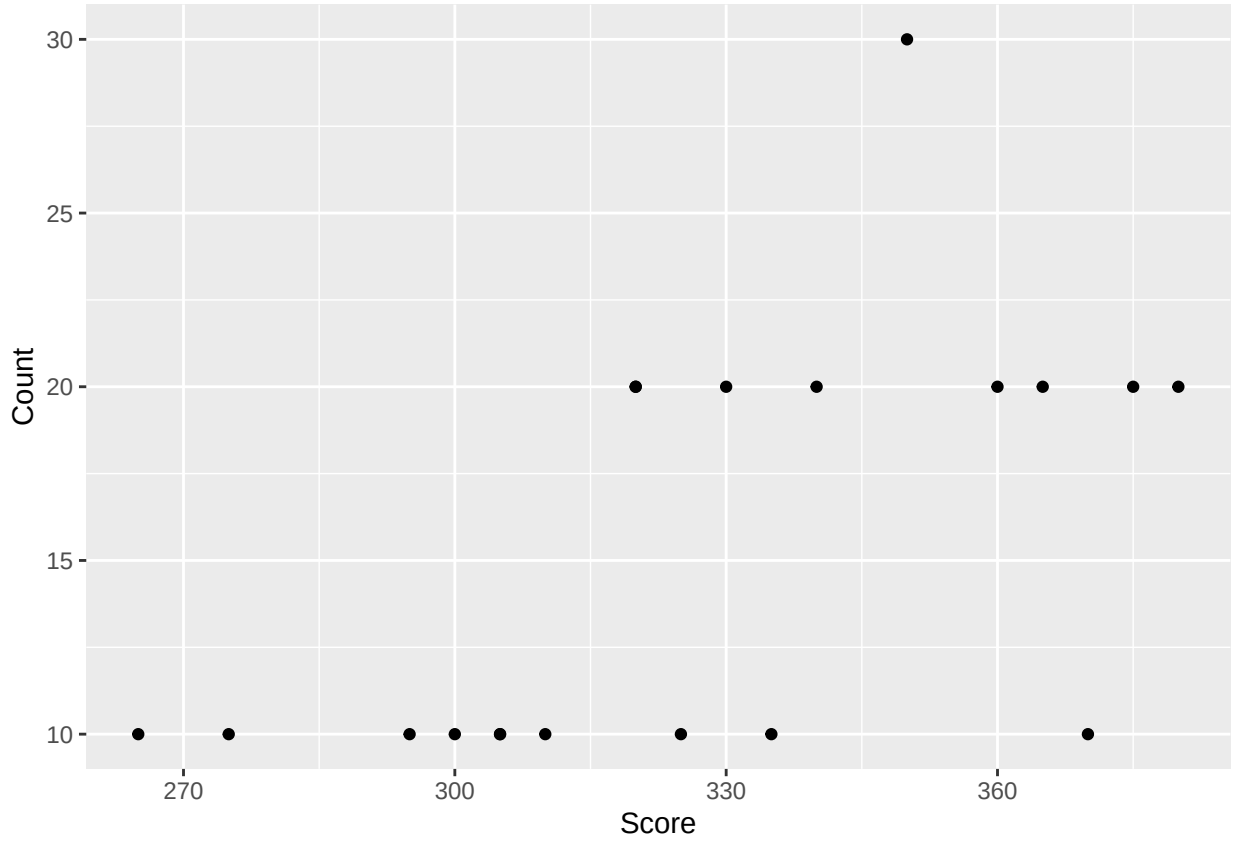
```
print("Categorical variables is Section ie Regular and Sport")
```

```
## [1] "Categorical variables is Section ie Regular and Sport"
```

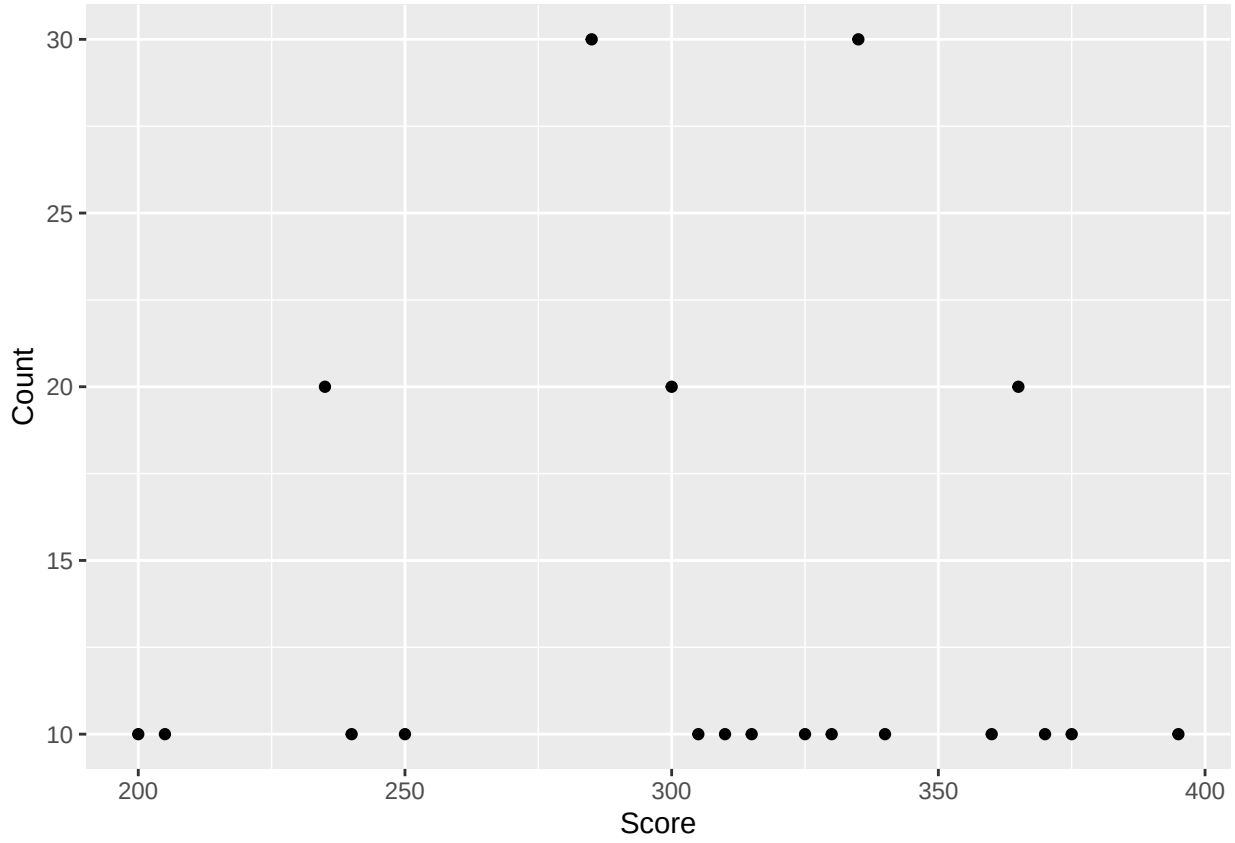
```
print("Quantitative are Count and Score")
```

```
## [1] "Quantitative are Count and Score"
```

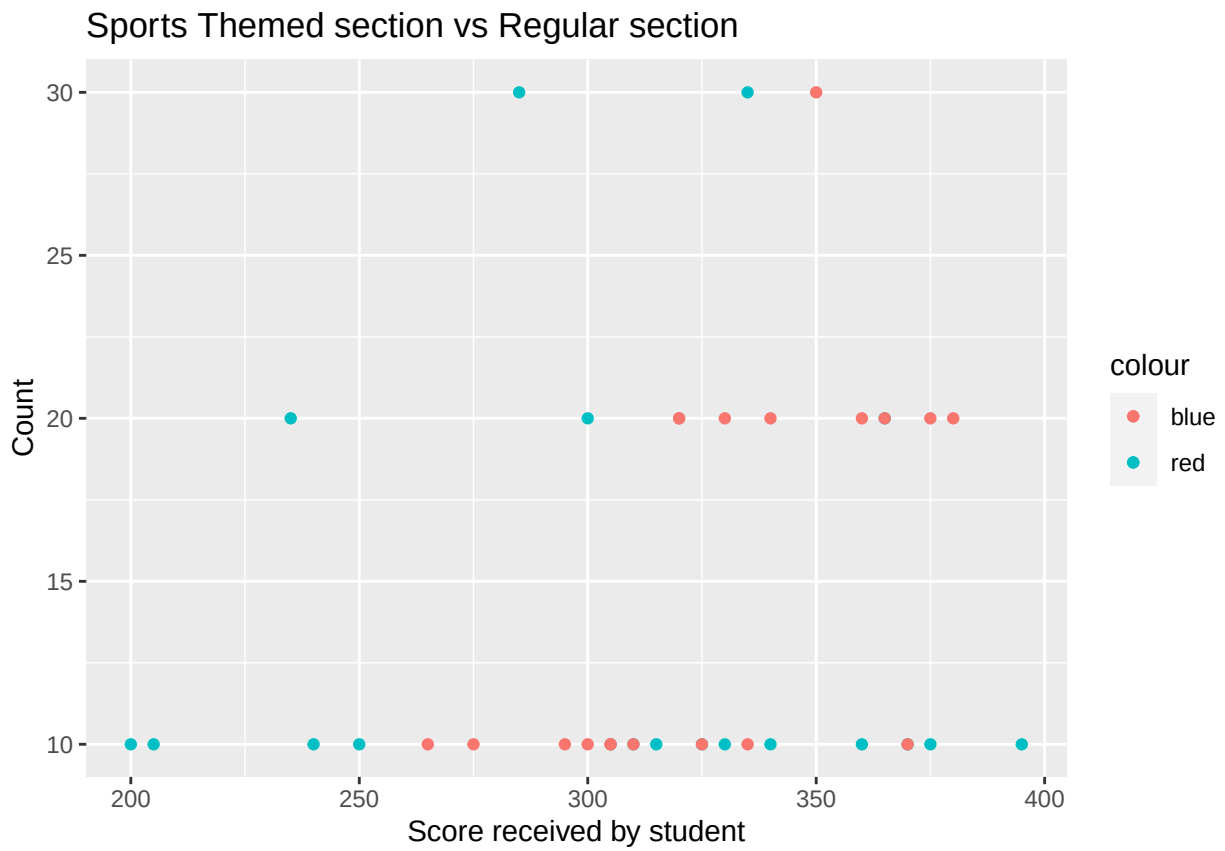
```
## Create Scatter lot of students attending Regular session
ggplot(regular_df, aes(x = Score , y = Count)) +
    geom_point()
```




```
## Create Scatter lot of students attending Sport session
ggplot(sport_df, aes(x = Score , y = Count)) +
    geom_point()
```




```
## Compare Sports vs Regular session students performance
ggplot() +
  geom_point(data=sport_df, aes(Score, Count, color='red')) +
  geom_point(data=regular_df, aes(Score, Count, color='blue'))+
     labs(
       x = "Score received by student",
       y = "Count",
       title = "Sports Themed section vs Regular section")
```




```
cat("Total Score in Sport Section =",sum(sport_df$Score),"\n")
```

```
## Total Score in Sport Section = 5840
```

```
cat("Total Count in Sport Section =",sum(sport_df$Count),"\n")
```

```
## Total Count in Sport Section = 260
```

```
cat(sum(sport_df$Score)/sum(sport_df$Count),"\n")
```

```
## 22.46154
```

```
mean(sport_df$Score)
```

```
## [1] 307.3684
```

```
cat("Total Score in Regular Section =",sum(regular_df$Score),"\n")
```

```
## Total Score in Regular Section = 6225
```

```
cat("Total Count in Regular Section =",sum(regular_df$Count),"\n")
```

```
## Total Count in Regular Section = 290
```

```
cat(sum(regular_df$Score)/sum(regular_df$Count),"\n")
```

```
## 21.46552
```

```
mean(regular_df$Score)
```

```
## [1] 327.6316
```

```
##Did every student in one section score more points than every student in the other section?
##If not, explain what a ##statistical tendency means in this context.


##"The Regular section had 30 more student enrolled in the program compared to Sport section.
##307.368 is mean for Sport section vs 327.632 is the mean for Regular Section "
```

```
##What could be one additional variable that was not mentioned in the narrative that could
##be influencing the point ##distributions between the two sections?


##"The professor who taught these two sections can greatly influence the distribution."
```