DSC520_Week1_Assignment00

Reenie Christudass

2022-07-03

```
df <- read.csv("C:/Users/chris/dsc520/data/scores.csv")
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
## 7     10   275 Regular
## 8     30   285  Sports
## 9     10   295 Regular
## 10    10   300 Regular
## 11    20   300  Sports
## 12    10   305  Sports
## 13    10   305 Regular
## 14    10   310 Regular
## 15    10   310  Sports
## 16    20   320 Regular
## 17    10   305 Regular
## 18    10   315  Sports
## 19    20   320 Regular
## 20    10   325 Regular
## 21    10   325  Sports
## 22    20   330 Regular
## 23    10   330  Sports
## 24    30   335  Sports
## 25    10   335 Regular
## 26    20   340 Regular
## 27    10   340  Sports
## 28    30   350 Regular
## 29    20   360 Regular
## 30    10   360  Sports
## 31    20   365 Regular
## 32    20   365  Sports
## 33    10   370  Sports
## 34    10   370 Regular
## 35    20   375 Regular
## 36    10   375  Sports
## 37    20   380 Regular
## 38    10   395  Sports
```

```
print(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
## 7     10   275 Regular
## 8     30   285  Sports
## 9     10   295 Regular
## 10    10   300 Regular
## 11    20   300  Sports
## 12    10   305  Sports
## 13    10   305 Regular
## 14    10   310 Regular
## 15    10   310  Sports
## 16    20   320 Regular
## 17    10   305 Regular
## 18    10   315  Sports
## 19    20   320 Regular
## 20    10   325 Regular
## 21    10   325  Sports
## 22    20   330 Regular
## 23    10   330  Sports
## 24    30   335  Sports
## 25    10   335 Regular
## 26    20   340 Regular
## 27    10   340  Sports
## 28    30   350 Regular
## 29    20   360 Regular
## 30    10   360  Sports
## 31    20   365 Regular
## 32    20   365  Sports
## 33    10   370  Sports
## 34    10   370 Regular
## 35    20   375 Regular
## 36    10   375  Sports
## 37    20   380 Regular
## 38    10   395  Sports
```

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

```
## Installing package into 'C:/Users/chris/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\chris\AppData\Local\R\win-library\4.2\00LOCK\dplyr\libs\x64\dplyr.dll
## to C:\Users\chris\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\chris\AppData\Local\Temp\RtmpmsqSL9\downloaded_packages
```

```
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
```

```
sport_df <- df %>% filter(Section =='Sports')
sport_df <- sport_df[1:2]
sport_df
```

```
##    Count Score
## 1     10   200
## 2     10   205
## 3     20   235
## 4     10   240
## 5     10   250
## 6     30   285
## 7     20   300
## 8     10   305
## 9     10   310
## 10    10   315
## 11    10   325
## 12    10   330
## 13    30   335
## 14    10   340
## 15    10   360
## 16    20   365
## 17    10   370
## 18    10   375
## 19    10   395
```

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

```
##    Count Score
## 1     10   265
## 2     10   275
## 3     10   295
## 4     10   300
## 5     10   305
## 6     10   310
## 7     20   320
## 8     10   305
## 9     20   320
## 10    10   325
## 11    20   330
## 12    10   335
## 13    20   340
## 14    30   350
## 15    20   360
## 16    20   365
## 17    10   370
## 18    20   375
## 19    20   380
```

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

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

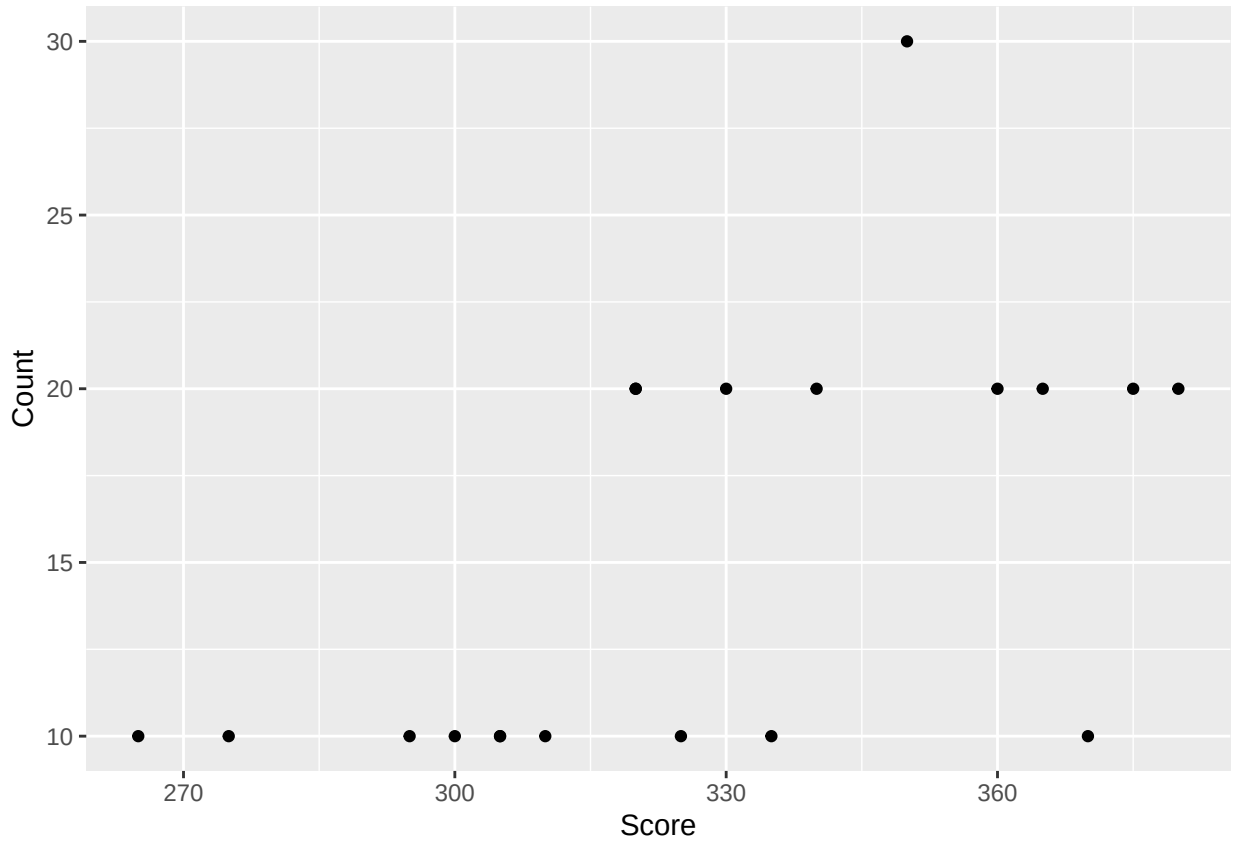
```r
print("Categorical variables is Section ie Regular and Sport")
```

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

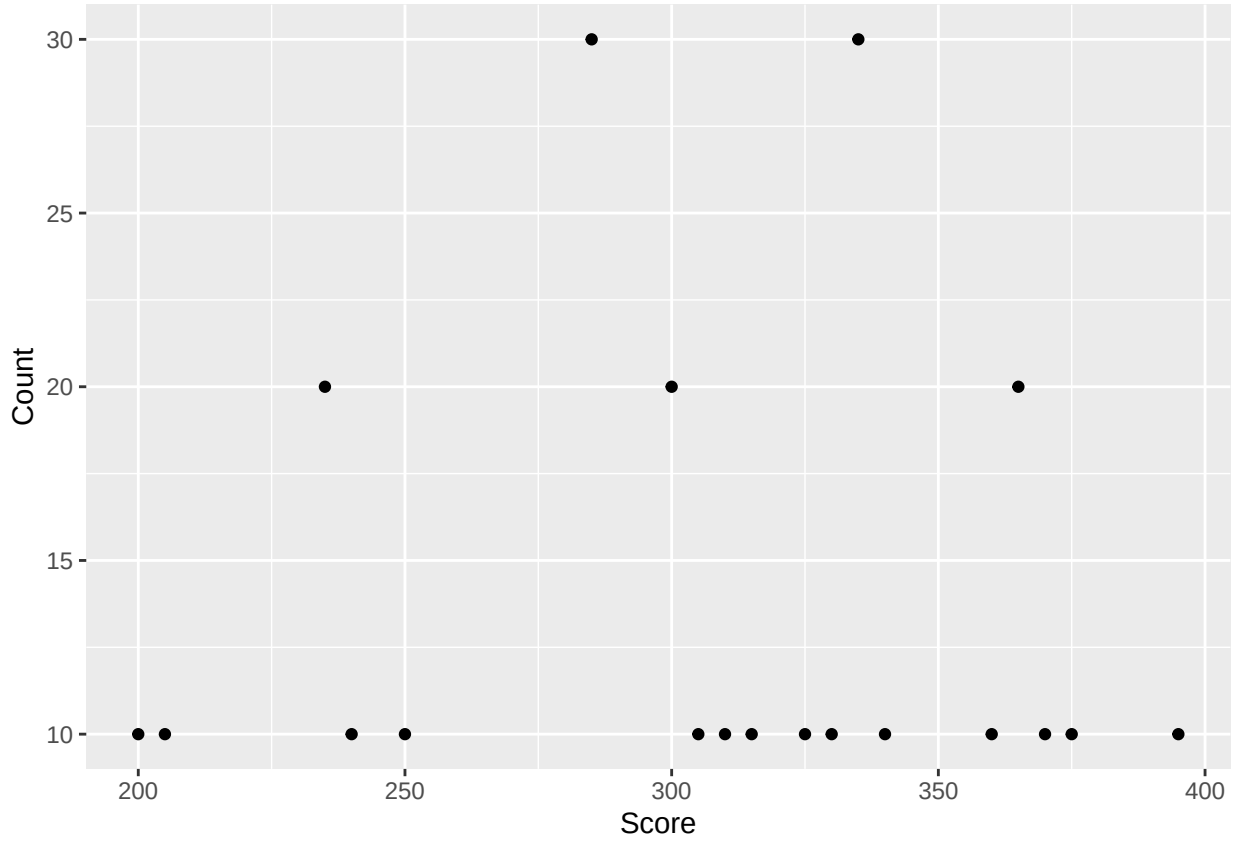
```r
print("Quantitative are Count and Score")
```

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

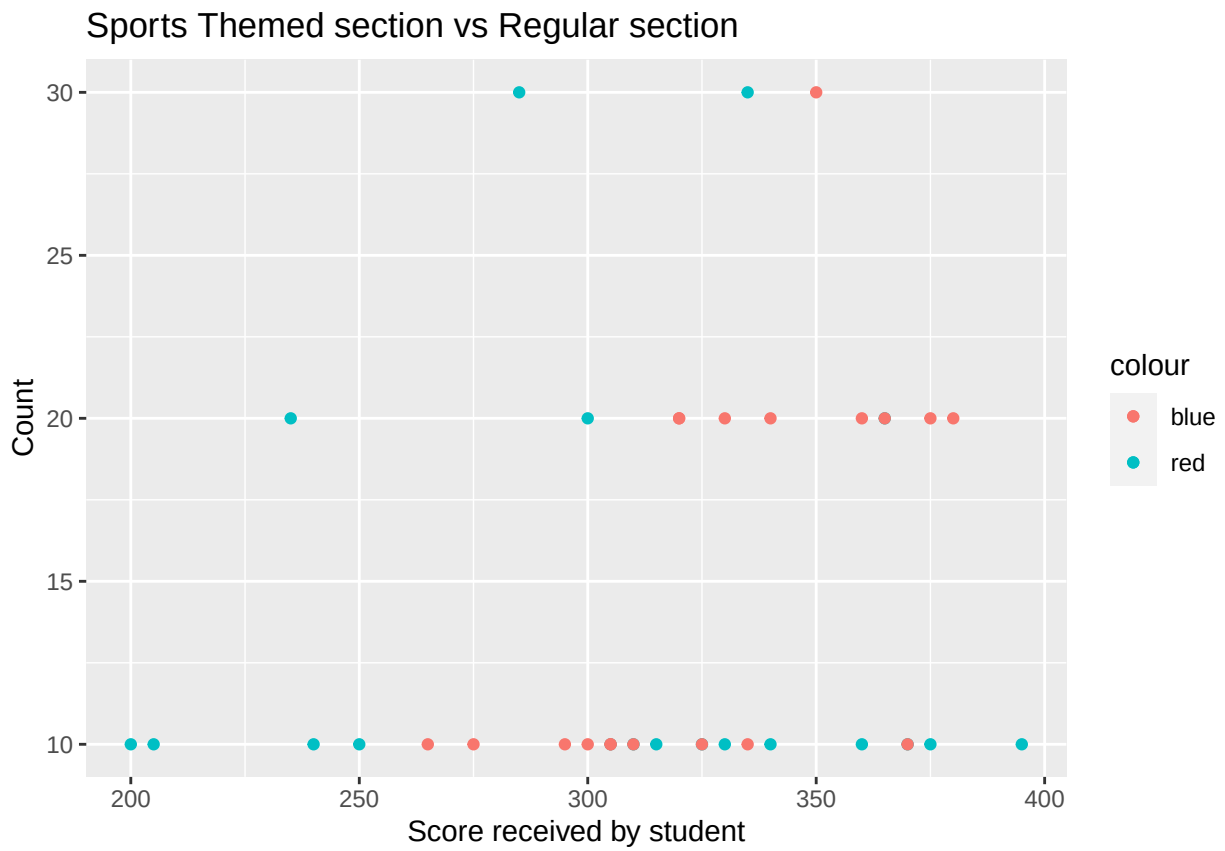
```r
library(ggplot2)
ggplot(regular_df, aes(x = Score , y = Count)) +
    geom_point()
```




```
library(ggplot2)
ggplot(sport_df, aes(x = Score , y = Count)) +
    geom_point()
```




```
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, e
"The Regular section had 30 more student enrolled in the program compared to Sport section. 307.368 is m
```

```
## [1] "The Regular section had 30 more student enrolled in the program compared to Sport section. 307.3
```

```
##What could be one additional variable that was not mentioned in the narrative that could be influenci
"The professor who taught these two sections can greatly influenced the distribution."
```

```
## [1] "The professor who taught these two sections can greatly influenced the distribution."
```