assignment_04-01_MunjewarSheetal.R

sheetal

2023-01-08

```r
# Assignment: ASSIGNMENT 4.1
# Name: Munjewar, Sheetal
# Date: 2023-01-08

## Load the ggplot2 package
library(ggplot2)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
##setwd("/home/jdoe/Workspaces/dsc520")
setwd("E:\\Data_Science_DSC510\\DSC520-Statistics\\dsc520")

## Load the `data/r4ds/heights.csv` to
score_df <- read.csv("data/scores.csv")
head(score_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
```

```r
summary(score_df)
```

```
##      Count           Score          Section
##  Min.   :10.00   Min.   :200.0   Length:38
##  1st Qu.:10.00   1st Qu.:300.0   Class :character
##  Median :10.00   Median :322.5   Mode  :character
##  Mean   :14.47   Mean   :317.5
##  3rd Qu.:20.00   3rd Qu.:357.5
##  Max.   :30.00   Max.   :395.0
```

```r
str(score_df)
```

```
## 'data.frame':    38 obs. of  3 variables:
##  $ Count  : int  10 10 20 10 10 10 10 30 10 10 ...
##  $ Score  : int  200 205 235 240 250 265 275 285 295 300 ...
##  $ Section: chr  "Sports" "Sports" "Sports" "Sports" ...
```

```
Sports_df <- score_df[ which( score_df$Section == "Sports"), ]
head(Sports_df)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 8    30   285  Sports
```

```
Regular_df <- score_df[ which(score_df$Section == "Regular"),]
head(Regular_df)
```

```
##    Count Score Section
## 6     10   265 Regular
## 7     10   275 Regular
## 9     10   295 Regular
## 10    10   300 Regular
## 13    10   305 Regular
## 14    10   310 Regular
```

```
##  Use the Plot function to plot each Sections scores and the number of students achieving that score.
##  Use additional Plot Arguments to label the graph and give each axis an appropriate label.
##  Once you have produced your Plots answer the following questions:
```
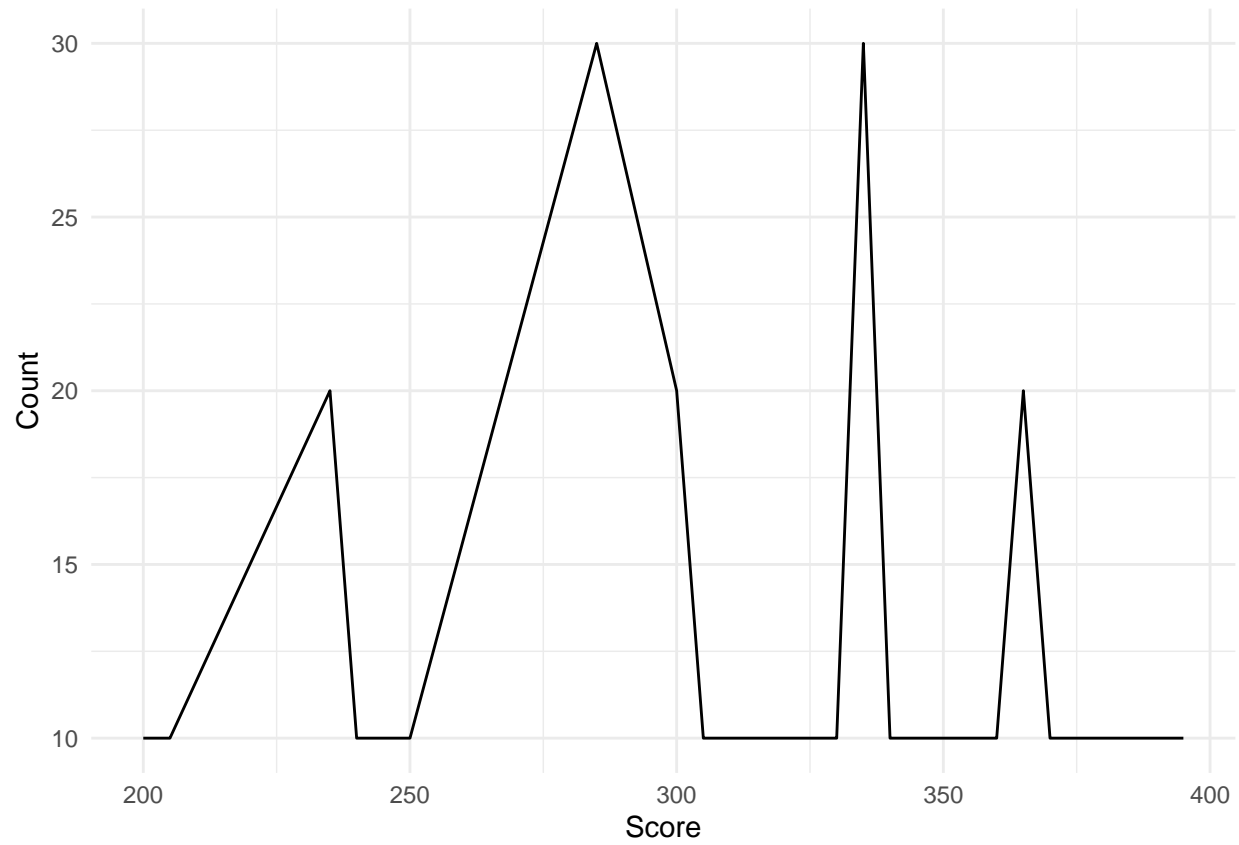
```
head(Sports_df)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 8    30   285  Sports
```
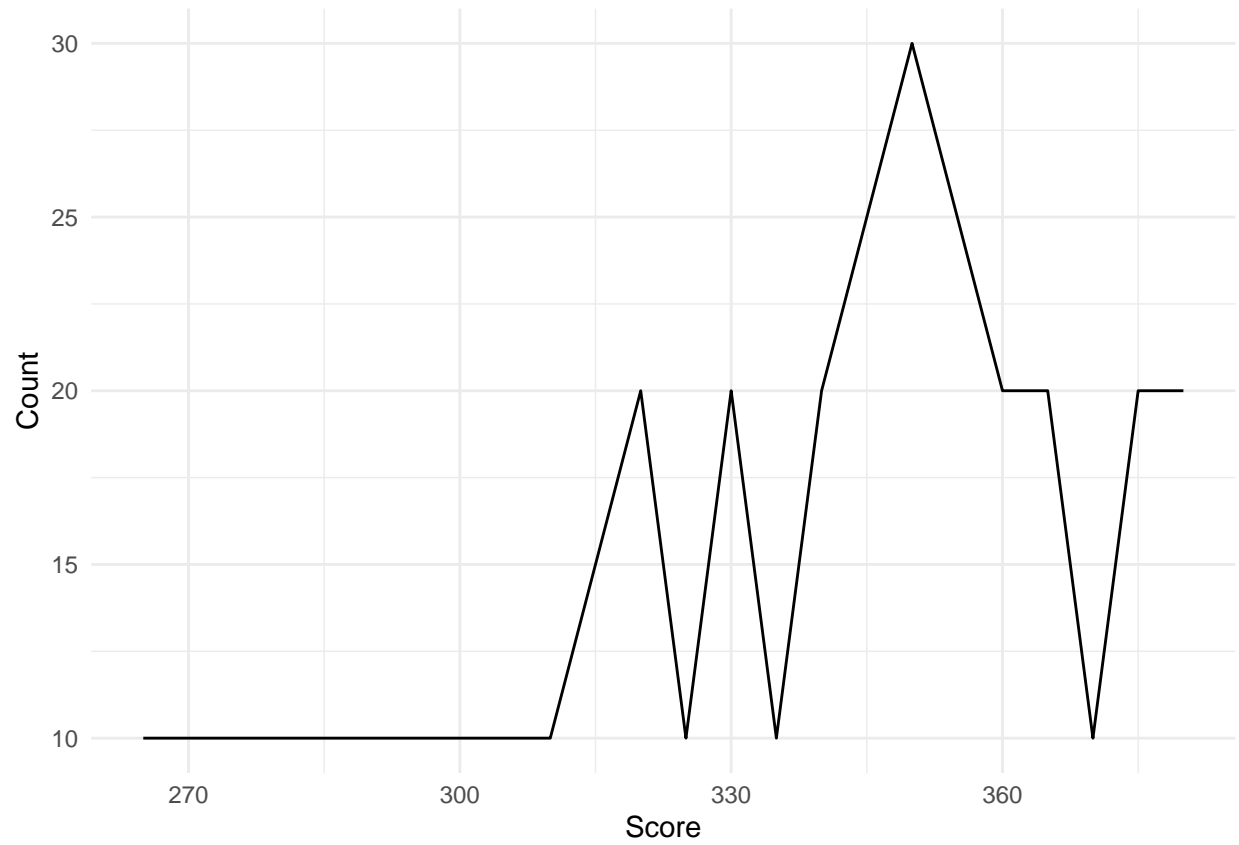
```
str(Sports_df)
```

```
## 'data.frame':    19 obs. of  3 variables:
##  $ Count  : int  10 10 20 10 10 30 20 10 10 10 ...
##  $ Score  : int  200 205 235 240 250 285 300 305 310 315 ...
##  $ Section: chr  "Sports" "Sports" "Sports" "Sports" ...
```
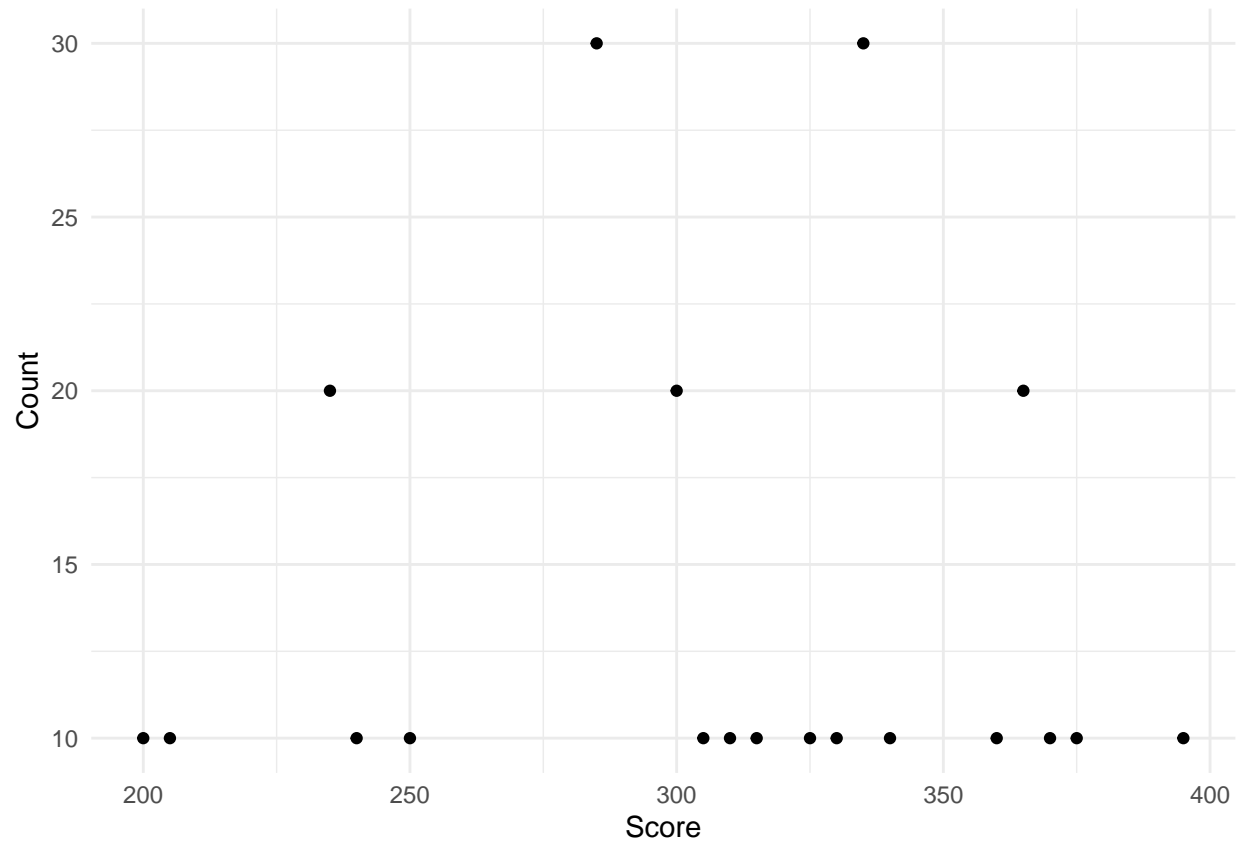
```
ggplot(data=Sports_df, aes(x=Score, y=Count)) + geom_line()
```
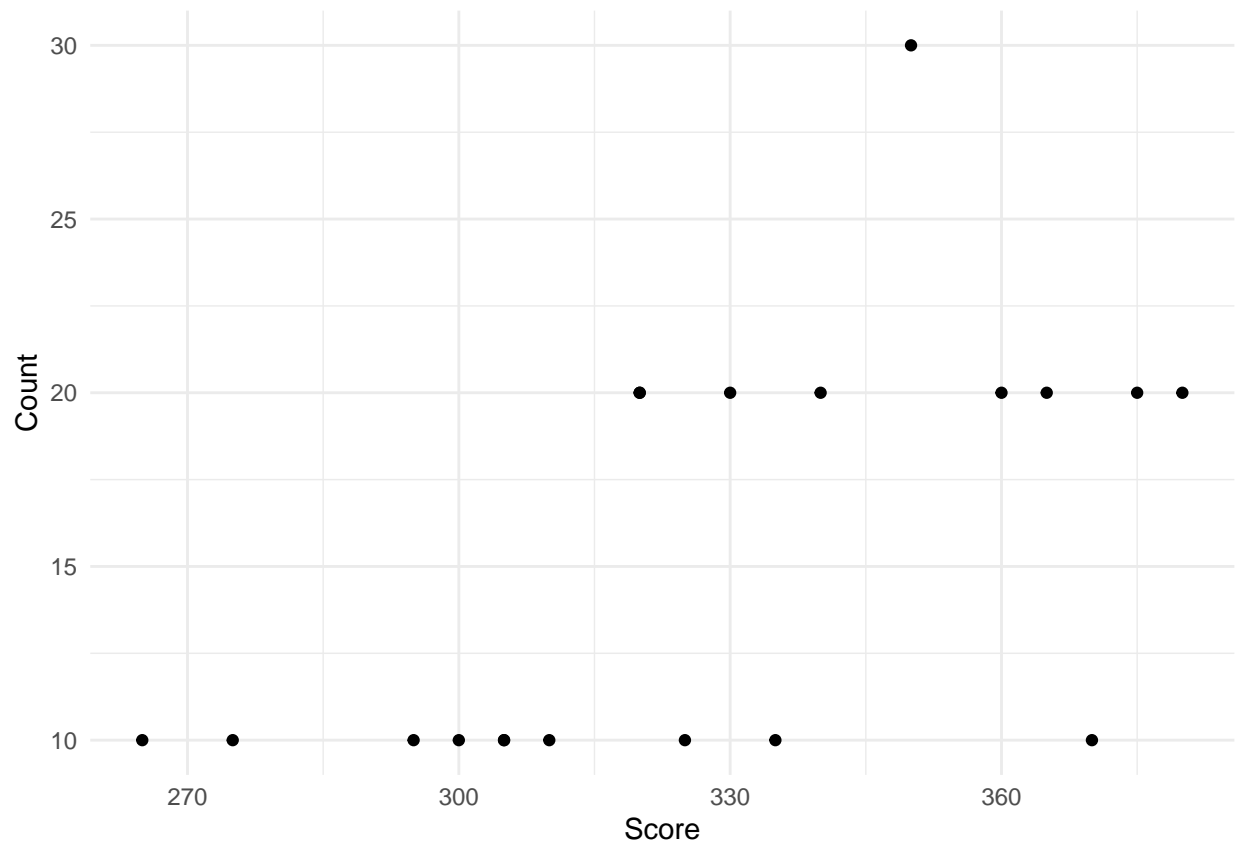
```
ggplot(data=Regular_df, aes(x=Score, y=Count)) + geom_line()
```

```
ggplot(data=Sports_df, aes(x=Score, y=Count)) + geom_point()
```
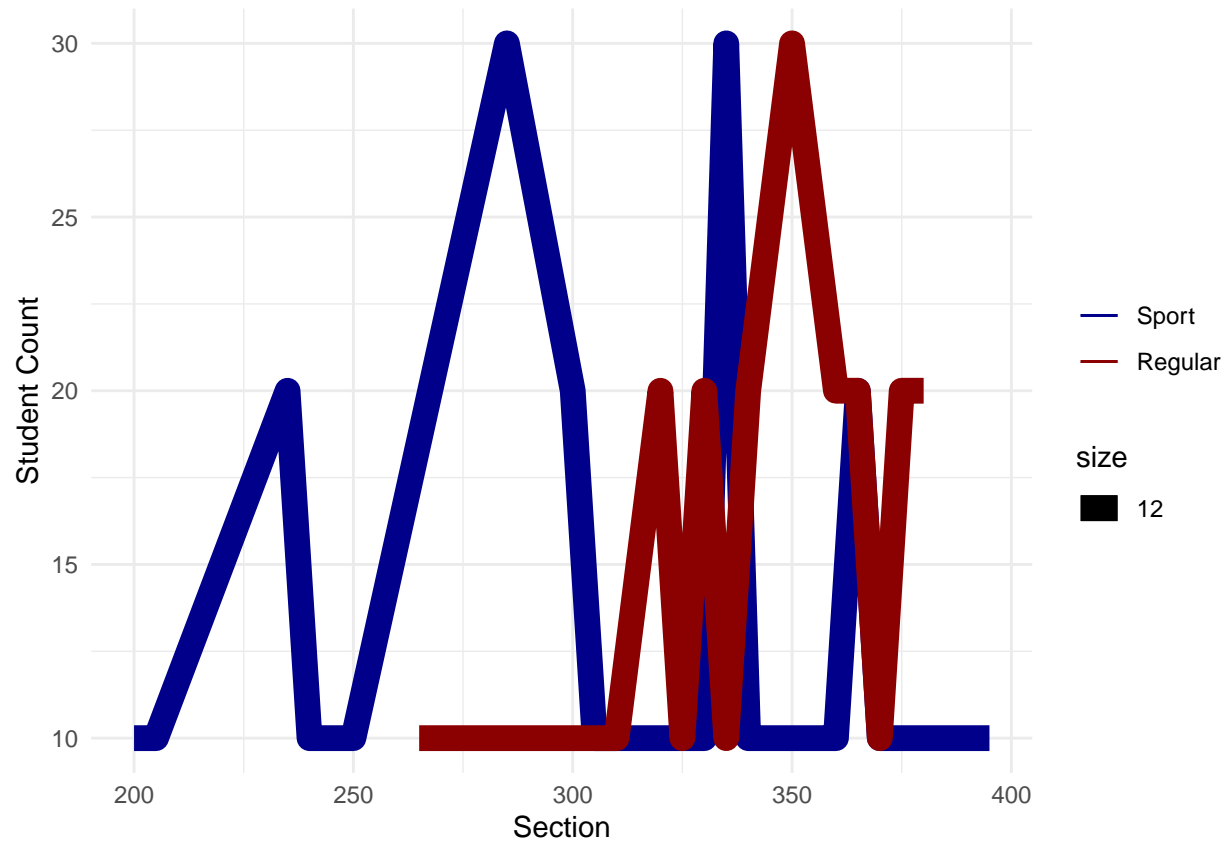
```
ggplot(data=Regular_df, aes(x=Score, y=Count)) + geom_point()
```

```
ggplot(data=Sports_df, aes(x=Score, group=1)) +
  geom_line(aes(y = Count, colour="Sport", size=12)) +
  geom_line(data=Regular_df, aes(y = Count, colour="Regular",size=12)) +
  scale_colour_manual("",
                      breaks = c("Sport", "Regular"),
                      values = c("darkblue", "darkred")) +
  xlab("Section") + ylab("Student Count")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

```
ggplot(data=Sports_df, aes(x=Score, group=1)) +
  geom_point(aes(y = Count, colour="Sport", size=12)) +
  geom_point(data=Regular_df, aes(y = Count, colour="Regular",size=12)) +
  scale_colour_manual("",
                      breaks = c("Sport", "Regular"),
                      values = c("darkblue", "darkred")) +
  xlab("Section") + ylab("Student Count")
```
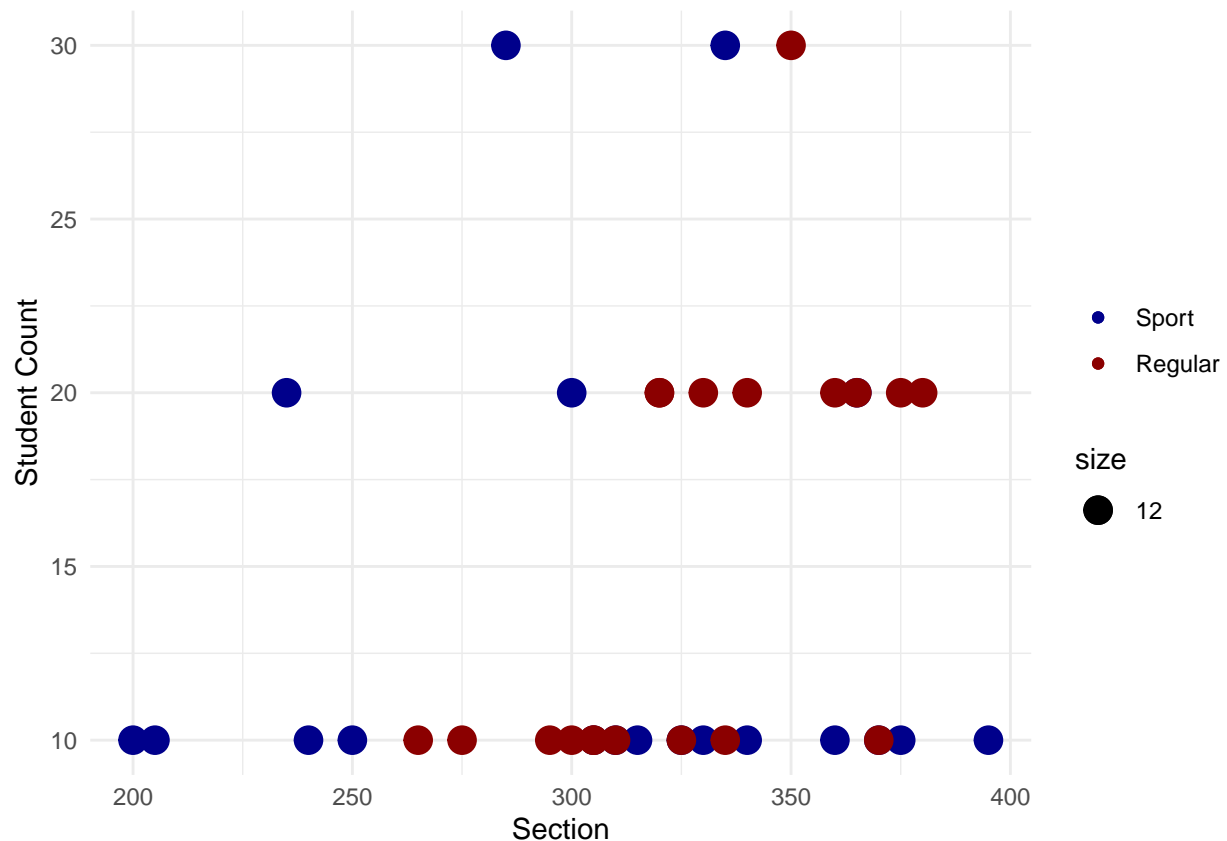
```
# Assignment :

#Use the appropriate R functions to answer the following questions:
#   What are the observational units in this study?
#   Answer - Students with scores and sections ( regular and Sports.)

#   Identify the variables mentioned in the narrative paragraph and determine which are categorical and
#   Answer - Section ( Regular and Sport are categorical ), however student score and counts are quantit

#   Create one variable to hold a subset of your data set that contains only the Regular Section and one
#   Answer - Sports_df and Regular_df, Two separate variables are created to hold subset of data.

#Use the Plot function to plot each Sections scores and the number of students achieving that score.
#Use additional Plot Arguments to label the graph and give each axis an appropriate label.
#Once you have produced your Plots answer the following questions:

# Comparing and contrasting the point distributions between the two section, looking at both tendency a
# Can you say that one section tended to score more points than the other? Justify and explain your ans
# Answer -> Based on plotted graph, regular section student highlighted in "Red" consistently doing bet
#           than Sport Section.

# 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.
# Answer -> Not every student from sport section score less score compared to regular section, however
#           prospective, regular section  students score are higher overall.
```

```
# What could be one additional variable that was not mentioned in the narrative that could be influenci
# distributions between the two sections?
# Answer -> mean score and student counts.
```