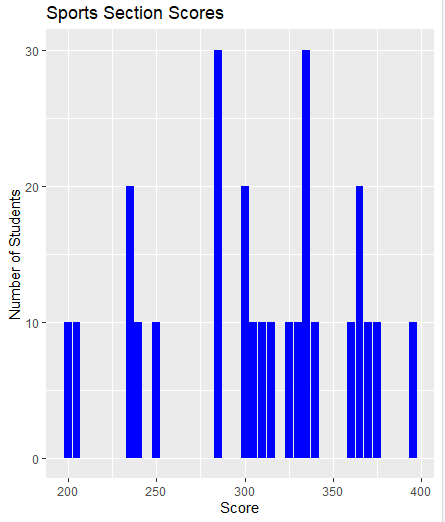
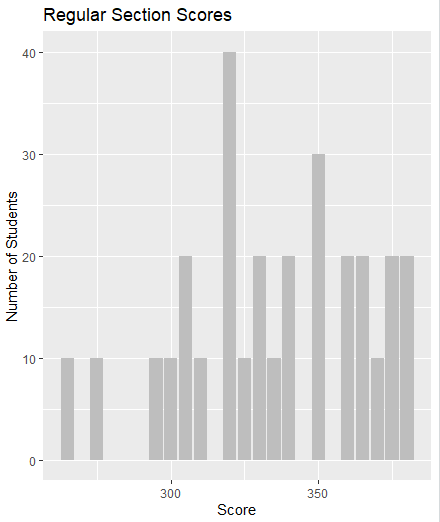
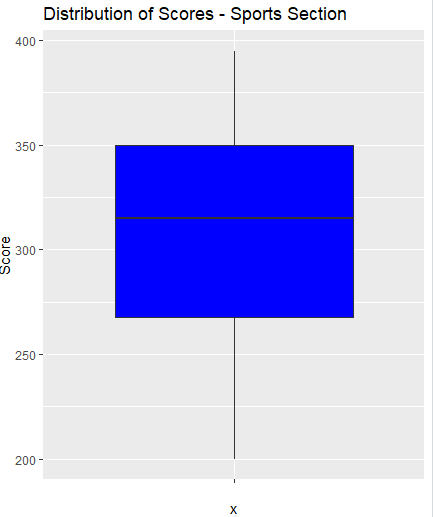
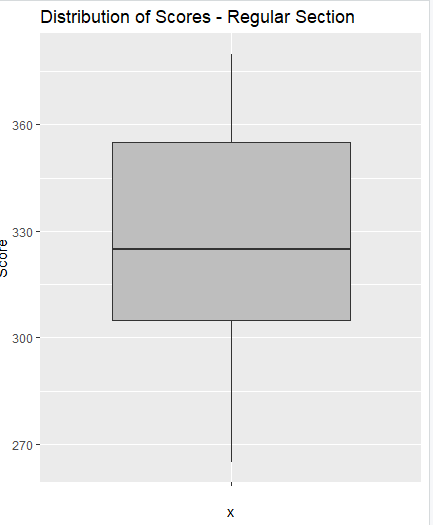
Assignment 4 – Exercise 1  
Kalyan Pothineni

04/04/2023

1. What are the observational units in this study  
   Observational units in this study are individual students score and section type. Each row represents the performance of students in the course. Student score, course they took (Sports or Regular) and count of students
2. Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?  
   There are three variable mentioned in the dataset  
   **Score** : This variable is Quantitative, represents the numerical score earned of students in the course  
   **Section**: This variable is Categorical, it indicates the section of the source students was enrolled in (Sports or Regular)  
   **Count:** This variable is Quantitative and represents the number of students in the dataset
3. Create one variable to hold a subset of your data set that contains only the Regular Section and one variable for the Sports Section  
   To create variables containing subsets of the dataset for the sections (Sports or Regular). We use subset(). Below code will create two data frames, containing data by section.   
   sports\_section <- subset(df, Section == "Sports")  
   regular\_section <- subset(df, Section == "Regular")
4. 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 sections, looking at both tendency and consistency: Can you say that one section tended to score more points than the other? Justify and explain your answer.  
  To compare two sections and determine if one section tends to score more points, firstly we calculate mean and standard deviation  
  Based on mean and standard deviation calculations, we can see that the mean score for the "Regular" section is 327.6316 with a standard deviation of 33.26528, while the mean score for the "Sports" section is 307.3684 with a standard deviation of 58.0318. This suggests that, on average, the "Regular" section scored higher than the "Sports" section, and that the scores in the "Regular" section were somewhat less variable than in the "Sports" section  
    
    
  We can use the above plots to compare the spread of scores between two sections.
* 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.  
  It might be true. Mean score of one section is more than the other, there can be individual students who has low scoring section. The test shows one section tends to score higher than other. Based on the data, there is a higher change of a student from a higher scoring section having a higher score compared to students from the lower scoring 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?

One additional variable that could influence the point distributions between the two sections is the teaching style/strategies or different course materials or different level

Other potential variables that could be influence the distributions could include student prior knowledge, student motivation or random stuff