*# Assignment: Week 2 Exercise*

*# Name: Wiltse, Daniel*

*# Date: 12-08-2019*

*A professor has recently taught two sections of the same course with only one difference between the sections. In one section, he used only examples taken from sports applications, and in the other section, he used examples taken from a variety of application areas. The sports themed section was advertised as such; so students knew which type of section they were enrolling in.*

*The professor has asked you to compare student performance in the two sections using course grades and total points earned in the course.*

*You will need to import the Scores.csv dataset that has been provided for you.*

Use the appropriate R functions to answer the following questions:

**1. What are the observational units in this study?**

The observational units are groups of students in the different groups, whether they are in the "Sports" or "Regular" sessions

**2. Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?**

Variables in the study:

Section - Categorical

Score - Quantitative

Count of students’ scores -Quantitative

**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.**

See R script below at end

**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:**

See R script below at end

**a. 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.**

Looking at the data, it appears the Regular group outperformed the Sports Group. Looking at the plot below, you can see most of the Regular sections scores are further to the right (i.e higher) than the Sports only sections. In addition, the average performance for the 290 students in the Regular section was 335, where the Sports section average for the 260 students was only 306.9.

(I originally used the summary() function in R but realized I was doing an average by Score of bucketed students, not by student, so did separate calculation to get sum of points divided by total students, so I could make sure I accounted for the difference in students in each group (290 vs 260). But in both calculations, the Regular section was higher than the Sports section.)

**b. 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.**

No, not every student in one section scored more than the other section. You can see blue dots interspersed with the red dots, meaning some of the Sports section students scored higher than the Regular section students. Interestingly enough, the bucket of students with the highest score (400) were in the Sports section.

The statistical tendency in this context means that choosing the Regular section led to students being more likely to score higher than students with the Sports only examples, but doesn’t mean that 100% of the time the students in the Regular section would do better than the Sports section, as reflected in the plot below.

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

Time of day during the class could impact performance, if a class was at 7 am compared to one at 3 pm, that could influence performance.

**R Script and Output for Questions 3 and 4**

#Loading Data

library(readr)

scores <- read\_csv("C:/Users/Dan/Desktop/DSC 520/scores.csv")

Parsed with column specification:

cols(

Count = col\_double(),

Score = col\_double(),

Section = col\_character()

)

#get subset of sports section

sport <- subset(scores, Section=="Sports")

#get subset of regular section

reg <- subset(scores, Section=="Regular")

#load ggplot library and create scatterplots

library(ggplot2)

ggplot(sport, aes(x = Score,y = Count, col = Section)) +

geom\_point()

ggplot(reg, aes(x = Score,y = Count, col = Section)) +

geom\_point()

ggplot(scores, aes(x = Score,y = Count, col = Section)) +

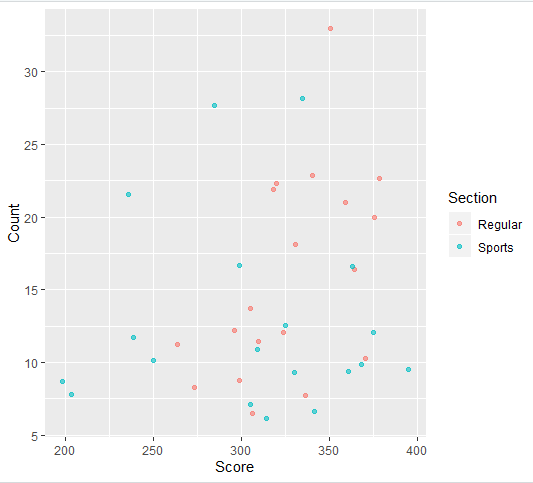
geom\_jitter(alpha = 0.6)

#get summary data for both subsets above

summary(sport)

summary(reg)

Comparison Plot:



Subset Plots:

