

CS5530- Assignment 1 -Question 2

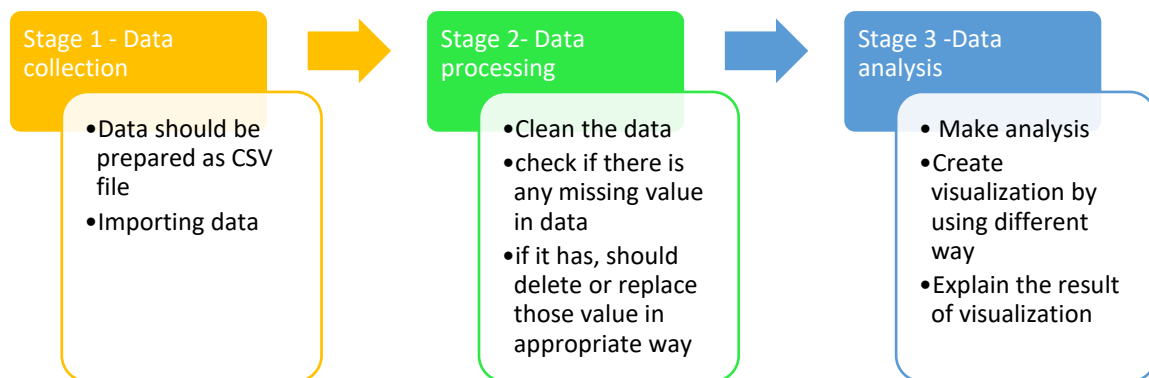
Question 2 Perform 5 data visualization tasks on the student performance dataset given in the link below (create 5 different visualizations). Explain what kind analysis has become easier with each of the visualizations. Create the folder structure for this question similar to question1. (15 points)

Answer:

In order to make an analysis and create data visualization based on given data, I have divided all processes into three stages including data collection, data processing, and data analysis.

The folder structure of the student performance project is as follows:

```
-- StudentPerformance_project
|-- data_raw
| |-- StudentsPerformance.csv
| |-- README.txt
|--data_clean
| |--clean_sp_data.csv
|-- results
| |-- results.pdf
|-- src
| |-- analysis.R
```



Stage 1 & Stage 2

I saved my raw data as CSV file in local directory and imported it successfully.

```
StudentsPerformance ← import("C:/Users/namuu/OneDrive/Documents/Namuun/UMKC/5530 /Assignment1_PDS/Question2/data_raw/StudentsPerformance.csv")
```

Then I converted the object "StudentsPerformance" to a data frame and saved it back to the object "StudentsPerformance" to ensure that my data is in the correct format to perform further analysis.

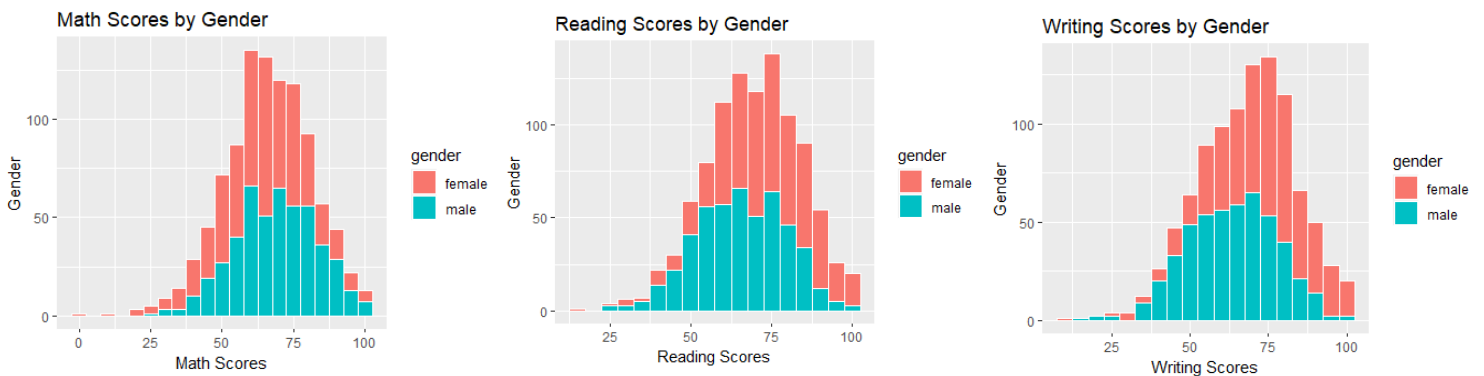
When I checked if there is any missing values in the raw data, there was not any missing value found. Therefore, I just converted the available non-numeric value "Test preparation" to numeric value 0 "Not completed" and 1 "Completed".

Stage 3

In this stage, I used 5 different data visualization techniques to make an analysis such as histogram, scatter plot (to see the correlation), pie chart, bar chart, and box plot.

Data visualization #1 Histogram

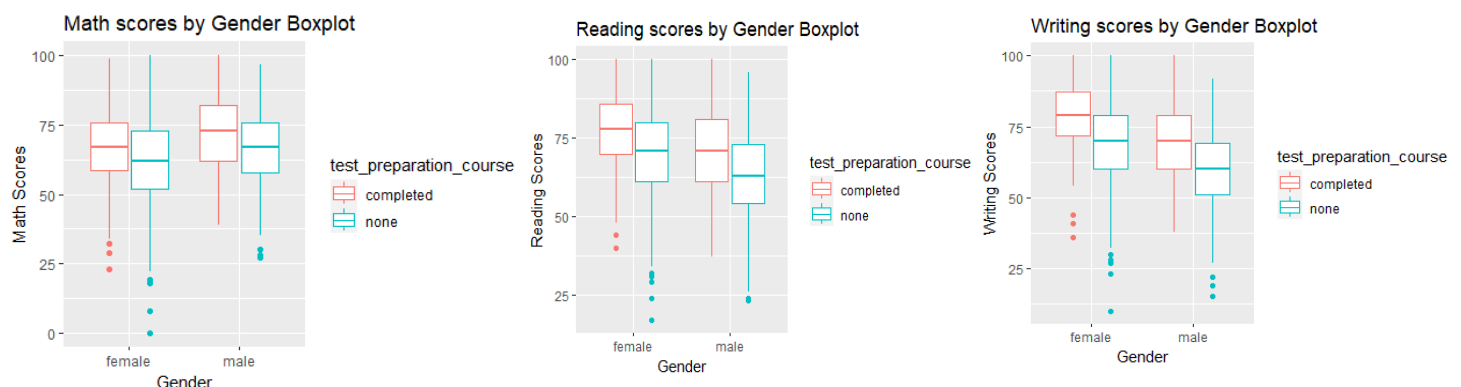
I used a histogram and distinguished by the color to see the difference in scores in genders. It helps to compare the two genders' scores. From this visualization, it can be concluded that female students tend to get higher scores in each exam compared to male students.



Visualization #2 Boxplot

This visualization provides with measures of central tendency, spread, and visual of outliers. I made the following conclusions based on the boxplot

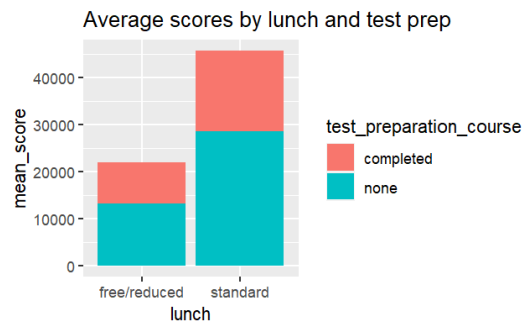
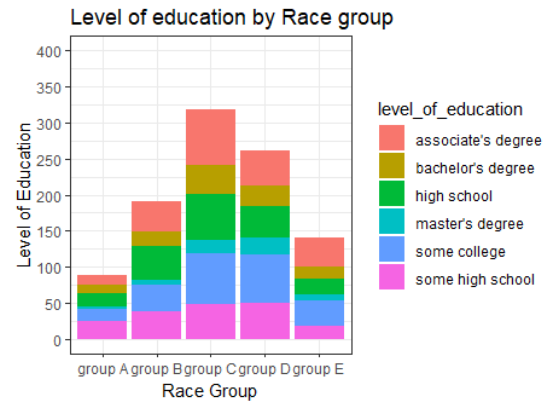
- Students who completed the test preparation got higher scores on all tests
- Male students had better scores in math while female students are good at the other two exams such as reading and writing
- Also, there are outliers in all three exams



Visualization #3 Bar plot

This visualization helps to determine if a particular race concerns a higher education level than others.

Based on this visualization, it can be concluded that in group A, students have less tendency to have associate degrees while in other groups, they prefer associate degrees.



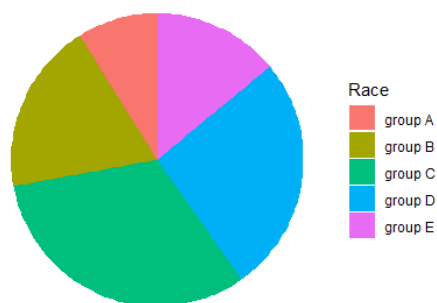
I analyzed average score for those three test score. Then I used bar plot to see the average scores of students who have standard lunch plan and completed test preparation

Visualization #4 Pie chart

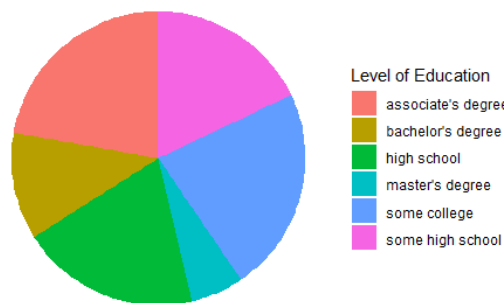
This pie chart visualization helps to see which group or which education level is dominant among all students.

- The left pie chart shows that group C and group D are dominant among the total students
- In the right pie chart, most of the students have an associate degree, college degrees, and high school while the lowest proportion of people have a master's degree

Proportion of Race ethnicity



Proportion of Education Level



Visualization #5 Scatter plot

This visualization helps to determine if there is a correlation between those three skills including math, writing, and reading. For example, based on the following scatter plot, I may

conclude that students good at math are good at both writing and reading skills. Also, students good at reading are good at writing since there is a positive correlation between the two skills.

