**Step 1:**[Importing the Data](https://github.com/heyodai/cs5530-assignment1/blob/ffd6bbee25a93db6df87f3e654a337b31fd166b5/#step-1-importing-the-data)

**Step 2:** [Data Visualization](https://github.com/heyodai/cs5530-assignment1/blob/ffd6bbee25a93db6df87f3e654a337b31fd166b5/#step-2-data-visualization)s

**Visualization 1:** Histogram - Distribution of reading scores from Students performance.

**Visualization 2:** Box Plot –Comparison of students who completed test completion course with their reading score and writing score results.

**Visualization 3:** Bar Plot – to display different groups of race/ethnicity.

**Visualization 4:** Pie chart – Charts to evaluate of parental education.

**Visualization 5:**[3D Scatter Plot](https://github.com/heyodai/cs5530-assignment1/blob/ffd6bbee25a93db6df87f3e654a337b31fd166b5/#visualization-5-3d-scatter-plot) - Comparison of the correlation between math, reading, and writing scores.

**Importing Data (R lang):**

This dataset is provided as part of assignment1 in Principles of Data Science Course. The dataset is clean and contains 8 columns and 1000 rows. Below are the columns and their description

|  |  |
| --- | --- |
| **Column** | **Description** |
| gender | Listed as male or female |
| race/ethnicity | Listed as group A, group B, group C, or group D |
| Parent level of education | Listed as some high school, high school, some college, associate's degree, bachelor's degree, or master's degree |
| lunch | Either standard or free/reduced |
| test preparation course | Either none or completed |
| math score | Numerical value between 0 and 100 |
| reading score | Numerical value between 0 and 100 |
| writing score | Numerical value between 0 and 100 |

Below is code used to import the file, to attach and to view it.

library(readr)

StudentsPerformance <- read\_csv("Desktop/StudentsPerformance.csv")

View(StudentsPerformance)

attach(StudentsPerformance)

Table

Description automatically generated

**Data Visualizations:**

**1.Histogram:**

Let’s start with Histogram, as it is a useful tool for displaying the distribution of a single continuous variable, here we are creating one specifically for the reading score variable.

**Code:**

library(readr)

StudentsPerformance <- read\_csv("Desktop/StudentsPerformance.csv")

View(StudentsPerformance)

attach(StudentsPerformance)

hist(`reading score`,col = "green")

**Chart, histogram

Description automatically generated**

**2.Box Plot:**

Now let’s try Box plot as a common method of showing data distribution based on a five-number summary.Here we are comparing the students who completed test completion course and their writing score and reading score.

**Code:**

tpc <- c(none=10,completed=90)

boxplot(`writing score`,`reading score`,`tpc`,main="box plot for comparison",col="blue")

Chart, box and whisker chart

Description automatically generated

**3.Barplot:**

Let’s go with bar plot to display the different groups of race/ethnicity variable.

**Code:**

race <- c(group1=10,group2=15,group3=20,group4=25,group5=30,group6=35)

barplot(race,col=c("red2", "green3", "slateblue4", "yellow2", "olivedrab2", "orange"),main="Bar plot for race/ethinicity")

Chart, bar chart

Description automatically generated

**4.Piechart:**

Let’s go for pie chart now for Parents level of education variable.

**Code:**

ple <- c(119,227,60,197,180,223)

lbl <- c("bachelor's degree","some college","master's degree","high school","some high school","associate's degree")

pie3D(ple,labels = lbl,main="Pie chart for Parent Level of Education")

Chart, pie chart

Description automatically generated

Piechart for race/ethnicity variable

**Code:**

rac <- c(group1=10,group2=15,group3=20,group4=25,group5=30)

pie(rac,main="Pie chart for race/ethnicity")

Chart, pie chart

Description automatically generated

**5.Scatter Plot:**

The last plot is 3D Scatter plot to compare the correlation between math, reading, and writing scores which are representing x-axis,y-axis and z-axis respectively.

**Code:**

scatterplot3d(`math score`,`reading score`,`writing score` , main = "3D Scatterplot of Student Performace in Math,Reading & Writing Scores"

Chart, scatter chart

Description automatically generated

**R files:**

****