**ASSIGNMENT-1**

**ANSWER - 1:**

**Stage 1: Data Acquisition**

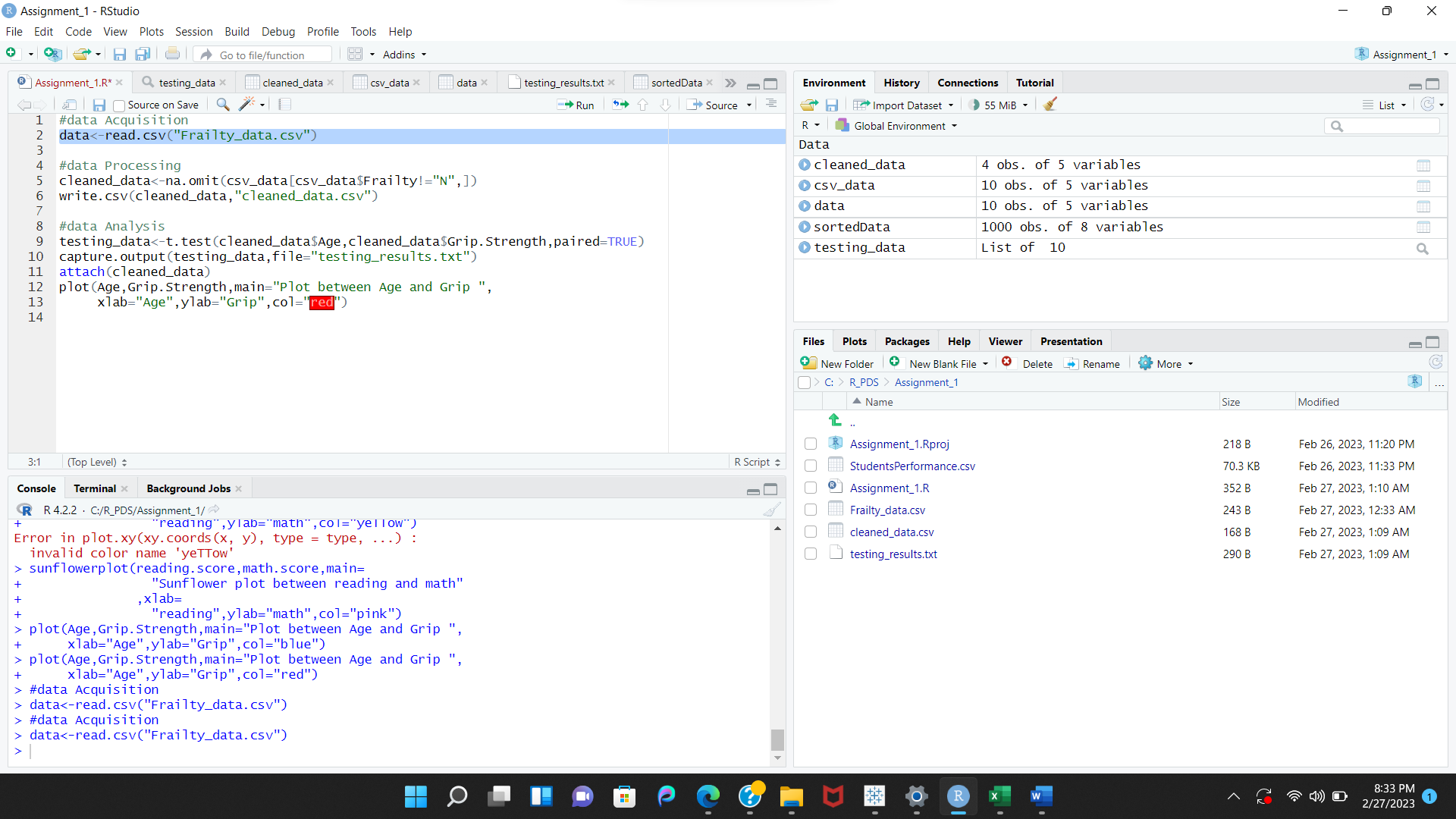
Data extraction and data gathering, which are ongoing activities, are required in the first step.

• Excel file is first transformed into a CSV file.

• The project structure is constructed in the following phase to house the data collection and project script in an organized approach.

**#Data Acquisition**

**data<-read.csv("Frailty\_data.csv")**



Table

Description automatically generated

**Stage 2: Data Processing**

Create a script to process the raw data

#Reading the raw data into an object called RemitDATA or Frailty\_data

**#Data Processing**

**cleaned data<-na.omit(csv\_data[csv\_data$Frailty!="N",])**

**write.csv(cleaned\_data,"cleaned\_data.csv")**

A screenshot of a computer

Description automatically generated

Graphical user interface, application, table

Description automatically generated

**Stage 3: Data Analysis**

Read the clean data into a variable.

Run a T-Test analysis on the clean data field-by-field to determine whether there is a statistically significant difference in age and grip.

**#Data Analysis**

**testing\_data<-t.test(cleaned\_data$Age,cleaned\_data$Grip.Strength,paired=TRUE)**

**capture.output(testing\_data,file="testing\_results.txt")**

**attach(cleaned\_data)**

**plot(Age,Grip.Strength,main="Plot between Age and Grip ",**

**xlab="Age",ylab="Grip",col="red")**

A screenshot of a computer

Description automatically generated with medium confidence

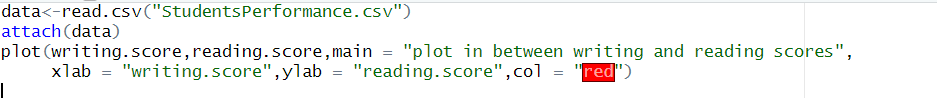
**Chart, scatter chart

Description automatically generated**

**ANSWER – 2**

1. **PLOT**

The relationship between two or more variables is shown graphically via a plot, which often takes the shape of a graph. The storyline could be created manually or digitally. The two main categories of plotters in the past were mechanical and electronic**.**



Chart, scatter chart

Description automatically generated

1. **Sunflowerplot**

Use a scatterplot with a sunflower plot to reduce overplotting. Sunflower plots only show one point when multiple points with the same x, y values exist; however, the number of points that are actually present is indicated by the small edges or petals radiating from the point.

A screenshot of a computer

Description automatically generated with medium confidence

Chart, scatter chart

Description automatically generated

1. **Pairs**

The coordinates of one point are included in an ordered pair. To find a location, use an ordered pair of coordinates (x, y). The x-coordinate is the first value, while the y-coordinate is the second. Put a dot on the graph where the ordered pair's locations are.

Graphical user interface, text

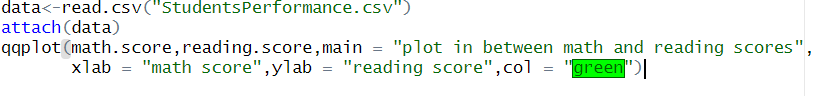
Description automatically generated

Chart

Description automatically generated

**4.Q-QPlot**

A value of two variables is drawn corresponding to each other in a (Q-Q Plot), which is used to determine whether or not the distributions of the two variables are similar with respect to their positions.



Chart

Description automatically generated with low confidence

1. **Histogram**

In addition to graphically displaying the frequency distribution of a small number of data points from a single variable, a histogram is a visual representation of the frequency distribution of classified continuous classes. Data is typically divided into "bins" or "range groups" in histograms, where each bin's number of data points is counted.

Text

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

**Explain what kind analysis has become easier with each of the visualizations.**

* **Histogram** :

The histograms are flexible and straightforward, in my opinion. It can be applied in a variety of situations to offer an instructive look into frequency distribution.

* **Histogram vs plot:**

In contrast to a density plot, which is a continuous and smoothed representation of the data-estimated histogram, a histogram can be used to illustrate vast amounts of data. The values of the set are displayed using plots.

* **Histogram vs pairs:**

The scatter plot and the histogram are the two main graphs on which the pairs plot is built. The top and bottom triangular scatter plots show the association between two variables, whereas the diagonal histogram shows the distribution of a single variable.

* **Histogram vs sunflowerplot:**

A histogram categorizes or organizes the probable values of a data set, assisting in the detection of outliers or missing data. Yet, it can be challenging to identify significant trends in sizable data sets and perplexing for viewers who are not familiar with the idea.

* **Histogram vs Q-QPlot:**

Two methods for visually determining if a dataset follows a normal distribution are Q-Q plots and histograms. The histogram will display a bell-shaped pattern if the data is regularly distributed.