Unique ID: E7321008

M.Sc. Degree

Continuous Assessment -2

CSC:540: Data Science with R

Algorithm:

1) Step 1 : Start

Step 2 : Import the covid-19 confirmed, recovered & deaths cases as raw data using read.csv().

Step 3 : Importing library 'tidyverse' & 'dplyr'.

Step 4 : Gather and group the imported data separately by country/Region and date using gather(), group_by() & summarize().

Step 5 : Combine all three datasets using full_join().

Step 6 : Assigning a variable to date & converting the character to date using sub() and as.date().

Step 7 : Combine the datasets assigned for each country to look at the world level of covid cases using group_by(), summarize(), sum() and mutate() & assign it to variable 'world'.

Step 8 : Importing library 'ggplot2' for plotting graphs.

Step 9 : Plotting a barchart with 'world' as data name for Date & Daily confirmed cases using ggplot function geom_bar().

Step 10 : Plotting a line plot with 'world' as data name

for cases over the period of time using ggplot function geom_line().

Step 11 : Filtering confirmed cases in Italy from column country/Region using filter() and assigning it to variable 'italy'.

Step 12 : Plotting a barchart with 'italy' as data name for the confirmed cases in Italy using geom_bar() in ggplot.

Step 13 : Repeat steps 11 & 12 for country/Region 'USA' and plot the barchart.

Step 14 : Repeat step 11 for country/Region 'Australia' and plotting a barchart for cumulative incidences of covid cases in 'Australia'.

Step 15 : Plotting a line chart for the world confirmed deaths and recovery over a period of time using geom_line().

Step 16 : Stop

2) a) Step 1 : Start

Step 2 : Finding mean, median and 50% trimmed mean for the given set of data using mean(), median() & mean(variable_name, trim=0.5) functions.

Step 3 : Import library 'modeest' for finding the mode of the given set of data.

Step 4 : mode = mfv(variable_name).

Step 5 : Stop.

b) Step 1 : Start

Step 2 : Creating a data frame for the given set of cholesterol data using data.frame().

Step 3 : Finding the median & IQR of cholesterol measurements for the patients before treatment using median(column_name) & IQR(column_name).

Step 4 : Repeat step 3 for patients cholesterol measurements After treatment using 'After' as column_name.

Step 5 : Stop

c) Step 1 : Start

Step 2 : Install packages "ggpubr" for hypothesis testing.

Step 3 : Importing library 'ggpubr'.

Step 4 : Install package "MASS" for importing dataset 'birthwt'.

Step 5 : Importing library 'MASS'.

Step 6 : Using one way anova to do hypothesis testing because we ~~use~~ are comparing independent & dependent variables. aov (dependent variable ~ independent variable)

Step 7 : Finding the results of anova created using summary( ).
Null hypothesis is rejected because p-value is less than 0.5.

Step 8 : Importing library "ggplot" for visualizing the data.

Step 9 : Plotting a scatterplot with 'birthwt' as data name using geom_point( ) in ggplot( ).

Step 10 : Stop

3) A) Step 1 : Start

Step 2 : Importing library 'tidyverse' for accessing 'mpg' cars dataset.

Step 3 : Assigning the 'mpg' dataset to variable 'data'.

Step 4 : Displaying the information about the dataset using: str(data).

Step 5 : Stop

B) Step 1 : Start

Step 2 : Creating a scatterplot for 'hwy' & 'cyl' in 'mpg' dataset using geom_point() in ggplot().

Step 3 : Stop

C) Step 1 : Start

Step 2 : Creating a scatter plot for 'class' & 'drv' in 'mpg' dataset using geom_point() in ggplot().

Step 3 : Stop.

D) Step 1 : Start

Step 2 : Mapping a continuous variable to color, size and shape using 'mapping', 'color', 'size' & 'shape' attributes in geom_point using ggplot.

Step 3 : Stop

E) Step 1 : Start

Step 2 : Performing a scatterplot with 'displ' on x-axis, 'hwy' on y-axis with regression line using geom_point() & geom_smooth() in ggplot().

Step 3 : Stop

F) Step 1 : Start

Step 2 : Creating a scatterplot for 'displ' & 'hwy' & a regression line using geom_point() & geom_smooth().

Step 3 : Repeat step 2 and change use the required attributes for different types of plots.

Step 4 : Stop.