Toronto Business College

Data Visualization

Assignment (1)

How to read data in R.from an excel file and then use ggplotgraphs for quick analysis

Hand Out Date: 29/09/2021

Hand In Date:06/10/2021

Assignment Type: Individual

Due Marks: 10%

**Student Name: - Rohan Bhatt**

**Student ID: - 500187633**

Objective: -

Detect and understand the stories from Datasets and extract insights from the datasets. Work with XLS (external datasets file)

Requirements: -

How to read an excel file and use ggplot to display data-graph. You have given the following XLS file Data Structure. Use R function to read the contents of the file then display them as graph to compare among all the salaries all over the department.

**Data File🡪 Transfer the file format into CSV or XLS**

EMPID,EMPNAME,SALARY,HireDate,dept

1,"Gamal",623.3,"2010-01-01","IT".

2,"Kumar",515.2,"2020-09-23","Operations".

3,"Varpan",611,"2016-11-15","IT".

4,"Raj",729,"2011-05-11","HR".

5,"Kamal",843.25,"2019-03-27","Finance".

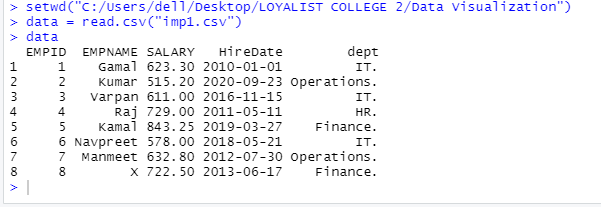
6,"Navpreet",578,"2018-05-21","IT".

7,"Manmeet",632.8,"2012-07-30","Operations".

8,"X",722.5,"2013-06-17","Finance".

***ANSWER : -***

There can be several ways to save the data in .csv or lets say excel file but as at this very point in time my function is not taking the excel file commands as library(“.xls”),so I would be going with the csv format only.



Now, that I am only able to use the csv format as it is not taking my xls code .

Guidelines: -

You need to read and data from a excel file into a data frame in R environment. once you have the data read in then you can use ggplotor ggplot2. read in external data (excel files) with readr and readxl. initial data exploration. At this point you can build several common types of graphs (scatterplot, column, line) in ggplot2. customize gg graph aesthetics (color, style, themes, etc.) update axis labels and titles.

**Deliverable: -**

***Word document that shows: -***

1. ***- The program you developed to read, load, generate the graph***

ANS 1**>**

**getwd()**  # To get the working directory started

**data = read.csv("imp1.csv")** #To read the csv file and load the data into ‘data’ Variable.

**Data** #Printing the data that is inserted in the variable

**library(ggplot2)** #calling the ‘ggplot2’ Library to draw beautifull graphs on it.

**ggplot(data, aes(x = dept, y=SALARY)) + geom\_line()** # this is taking in data and that plotting the two variables on line and trying to find the relationship.

**ggplot(data , aes(y=dept,x=SALARY)) + geom\_point()**

# this is also trying to find between salary and department of it.

**ggplot(data= data, aes(x=SALARY, y=dept,group=1)) + geom\_line( color="green", size=2, alpha=0.9, linetype=2) + geom\_point()**

**#** for more in-depth analysis

**aggregate(x= data$SALARY,by= list(data$dept),FUN=max)**

**aggregate(x= data$SALARY, by= list(data$dept),FUN=min)**

**# Save the file.**

**dev.off()**

1. ***- Screen shots of the generated graph/graphs***

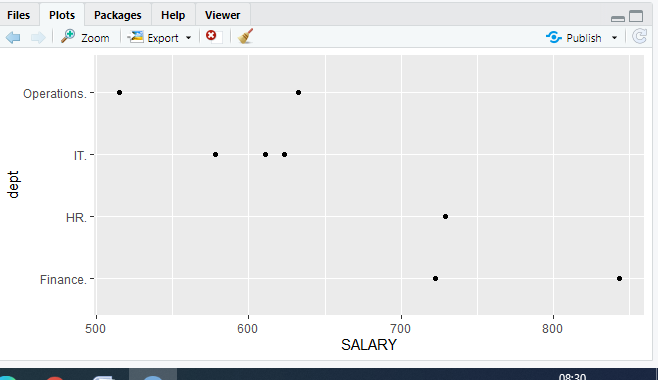
ANSWER 2>

2.1

It is showing the graph between department and salary of each department.



2.2



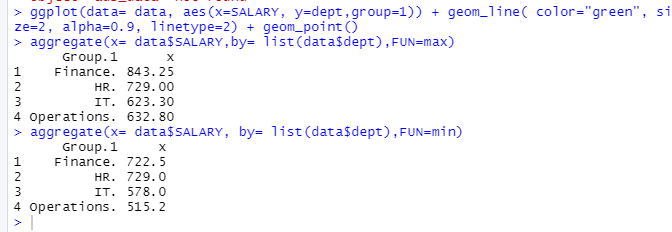
It is showing the graph between department and salary of each department using the helps of points on it.

2.3 ggplot(data= data, aes(x=SALARY, y=dept,group=1)) + geom\_line( color="green", size=2, alpha=0.9, linetype=2) + geom\_point()



3- Explain, Whereas the highest salaries and lowest salaries (which departments).

ANSWER 3>



***4- From technical perspectives, The role or purpose of the used R-Packages.***

ANSWER 4>

The plot that I created for were used in it and created by ggplot2 : -

So explaining about it : -

GGplot2 is a plotting package that makes **it simple to create complex plots from data in a data frame**. It provides a more programmatic interface for specifying what variables to plot, how they are displayed, and general visual properties.