Northeastern University

Course: DA5020

Assignment: Module 1 - Creating Graphs in R

Total Points: 100

Date Due: Posted on Blackboard

Learning Objectives

In this assignment, you will learn how to:

- Create and save an R script
- Create simple charts

Tasks

Before working through the content in Chapter 3, “R for Data Science”, make sure your meet the prerequisites as laid out at the beginning of the chapter. Install the libraries you need and call `library(..)` function to require them. Use the mtcars dataset in R to answer the below questions:

1.What does the gear variable describe? Read the help for ?mtcars to find out. (5 points)

2. Make a scatterplot of mpg vs cyl. (5 points)

3.Which variables in mtcars are categorical? Which variables are continuous? (Hint: type ?mtcars to read the documentation for the dataset). How can you see this information when you run mtcars? (10 points)

4. What plots does the following code make? What does . do? (10 points)

```
ggplot(data = mtcars) +
geom_point(mapping = aes(x = disp, y = mpg)) +
facet_grid(cyl ~ .)
```

```
ggplot(data = mtcars) +
geom_point(mapping = aes(x = disp, y = mpg)) +
facet_grid(. ~ am)
```

5.Read ?facet_wrap. What does nrow do? What does ncol do? What other options control the layout of the individual panels? Why doesn't facet_grid() have nrow and ncol argument? (10 points)

6. Review the code below and predict what the output will look like. Then, run the code in R and check your predictions. (10 points)

```
ggplot(data = mtcars, mapping = aes(x = disp, y = mpg, color = cyl)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

7.What is the default geom associated with stat_summary()? How could you rewrite the below plot to use that geom function instead of the stat function? (10 points)

```
ggplot(data = mtcars) +
  stat_summary(
    mapping = aes(x = disp, y = mpg),
    fun.ymin = min,
    fun.ymax = max,
    fun.y = median
  )
```

8/ Use the mpg dataset from the ggplot2 package to answer the further questions: (10 points)

For the following code, compare geom_jitter() with geom_count()

```
 ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +
    geom_point()
```

9.What does the below plot code tell you about the relationship between cty and hwy in mpg? Why is coord_fixed() important? What does geom_abline() do? (15 points)

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +
  geom_point() +
  geom_abline() +
  coord_fixed()
```

10.Work through all the examples (excluding those from exercises) within Chapter 3 in "R for Data Science", write the code down into a single.R file, add appropriate commenting lines describing each code snippet and make sure they all work for you. Submit code for this task in a separate file. (15 points)

CODING STYLE: Part of the assignment is to practice good coding standards: consistent naming, formatting with indentation, testing input values, dealing with errors, commenting on the code, etc.. You need to submit an R script, which is a text file having an .R extension, i.e., To help you and us read the code you must follow the coding standards mentioned above. This applies to ALL assignments going forward. If your code is not readable and/or does not run 20% will be automatically deducted from your assignment grade.

Deliverables

You need to submit either an R or an .Rmd extension file. Be sure to state all the assumptions and give explanations as comments in the .Rmd file wherever needed to help us assess your submission. **Please name the submission file LAST_FirstInitial_1.R or (rmd)** for example for John Smith's 1st assignment, the file should be named **Smith_J_1.R**. Name the separate file for Task 8 as **Smith_J_1_8.R.**