HW1_wliu3

Wei Liu

8/24/2020

Problem 1

Finished in the RStudio Cloud

Problem 2

Part A

My 3 desired learning objectives are as follows:

1. have a better understanding of tools that data scientists work with
2. be able to use these tools to explore the process of data analysis
3. be able to communicate with my collaborators efficiently using these tools

Part B

y_1: uniform distribution
y_2: exponential distribution
y_3: pareto distribution

$$y_1 = \frac{1}{2}, \quad 2 \leq x \leq 4 \tag{1}$$

$$y_2 = \frac{1}{2} exp\left(\frac{-x}{2}\right), \quad 0 \leq x \leq \infty \tag{2}$$

$$y_3 = \frac{2}{x^3}, \quad 1 \leq x \leq \infty \tag{3}$$

Problem 3

Summary of the steps in performing Reproducible Research

1. For Every Result, Keep Track of How It Was Produced Comment: It's a good way to comment the meaning or purpose of important code to make it clear and understandable.
2. Avoid Manual Data Manipulation Steps Comment:For problem 4, there is no such problem since an internal R dataset is used.

3. Archive the Exact Versions of All External Programs Used Comment: Good suggestion since there might be different functions in different version.
4. Version Control All Custom Scripts Comment: Github is an efficient example tool to keep track all changes and scripts we made, so we can check back anytime and anywhere.
5. Record All Intermediate Results, When Possible in Standardized Formats Comment: Intermediate results are important because they might provide insights for our final results and they allow parts of process to run firstly!
6. For Analyses That Include Randomness, Note Underlying Random Seeds Comment: Nice point! Record the seed number to make sure get reproducible results.
7. Always Store Raw Data behind Plots Comment: Good habit to store raw data behind plots in case further investigation.
8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected Comment: Put as much details since all of them maybe useful in future.
9. Connect Textual Statements to Underlying Results Comment: Textual statements is a good way to communicate with others.
10. Provide Public Access to Scripts, Runs, and Results Comment: Set the access as public such as in Github to make the resources accessible to readers.

Problem 4

Scatter plot

```
data(mtcars)
plot(mtcars$wt, mtcars$mpg,
     col = "green",
     main = "MPG as a Function of Weight",
     xlab = "Weight",
     ylab = "MPG")
```

Histogram

```
hist(mtcars$hp,
     main = "HP of cars",
     xlab = "hp",
     col = "pink")
```

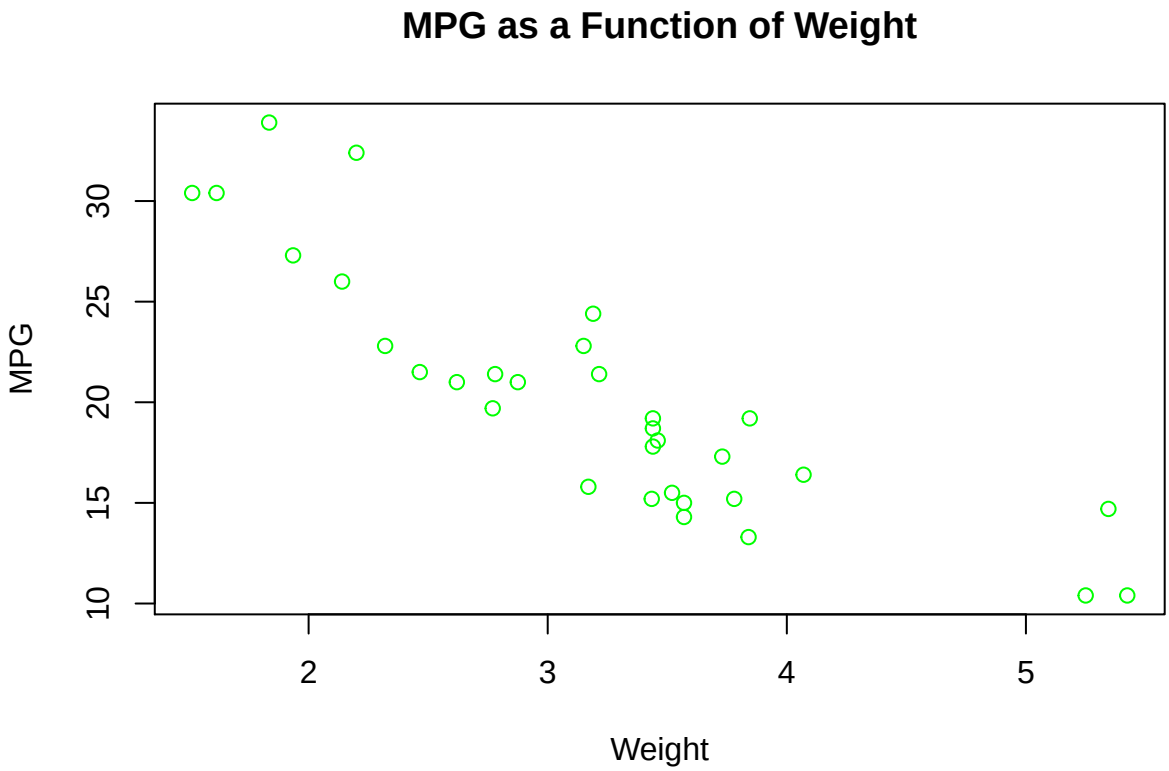

Figure 1: A simple scatter plot.

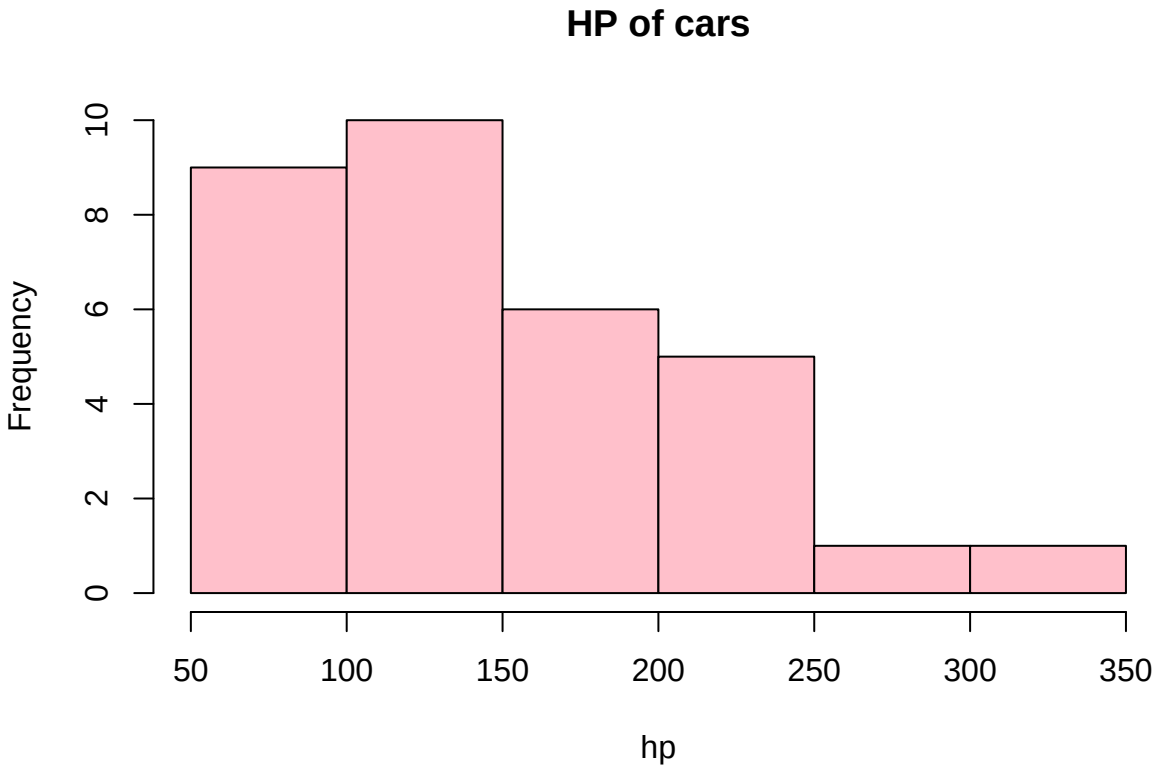

Figure 2: A simple histogram.

Appendix 1: R code

Problem 4

RStudio version: R 4.0.2

```
data(mtcars) # loading the whole data set for mtcars
plot(mtcars$wt, mtcars$mpg,
     col = "green",  # color of the points
     main = "MPG as a Function of Weight", # main title of the plot
     xlab = "Weight",
     ylab = "MPG") #a simple scatter plot

hist(mtcars$hp,
     main = "HP of cars", # main title of the histogram
     xlab = "hp",
     col = "pink")  # A simple histogram
```