

HW1_youngjin

Youngjin Cho

2020 8 28

Problem 1

I finished both Visualization Basics and Programming Basics in “The Basics” of the “RStudio Primers”.

Problem 2

Part A

The following are parts of course objectives listed in the beginning of lecture 1.

- R programming
- Reproducible research concepts
- Good programming practices

Part B

The following are exponential, gamma, and normal distribution from Appendix in “Statistical Inference” written by Cassella & Berger.

Exponential Distribution

$$f(X = x|\beta) = \frac{1}{\beta}e^{-\frac{x}{\beta}}; 0 \leq x < \infty; \beta > 0 \quad (1)$$

Gamma Distribution

$$f(X = x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}; 0 \leq x < \infty; \alpha, \beta > 0 \quad (2)$$

Normal Distribution

$$f(X = x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}; -\infty < x < \infty; -\infty < \mu < \infty, \sigma > 0 \quad (3)$$

Problem 3

1. Record the way it was produced for every result. It might be challenging for me when the project is huge and complicated.
2. As manual procedures are hard to reproduce as well as inefficient and error-prone, it is recommended to avoid manual data manipulation steps. In order to avoid such steps, it is important to enhance my programming skills.
3. Record exact versions of every external programs used in the research. If there are lots of external programs in the project, it should be thoroughly examined.
4. Use a version control system for all custom scripts. It is important to know how to run such systems.
5. In standardized format, record every intermediate results. It could be challenging when there are lots of results in the research.
6. When there is a randomness in the analysis, record random seeds for it. It is important to track all randomness parts of the project to record their random seeds.
7. For visual consistency between figures, store raw data behind plots. If one data is connected to multiple plots, it is important to store it for all of connected plots.
8. To validate and understand the research result, it is good to inspect the summaries and generating hierarchical analysis output is useful for doing it. It is critical to know each steps of the hierarchical analysis.
9. Write down textual statements to underlying results. It would be critical to write text for results such as interpretation or conclusion efficiently, i.e, to make readers easy to understand the text.
10. Give public access to all scripts, runs, and results in the research. It is crucial to make public to easily access to the research.

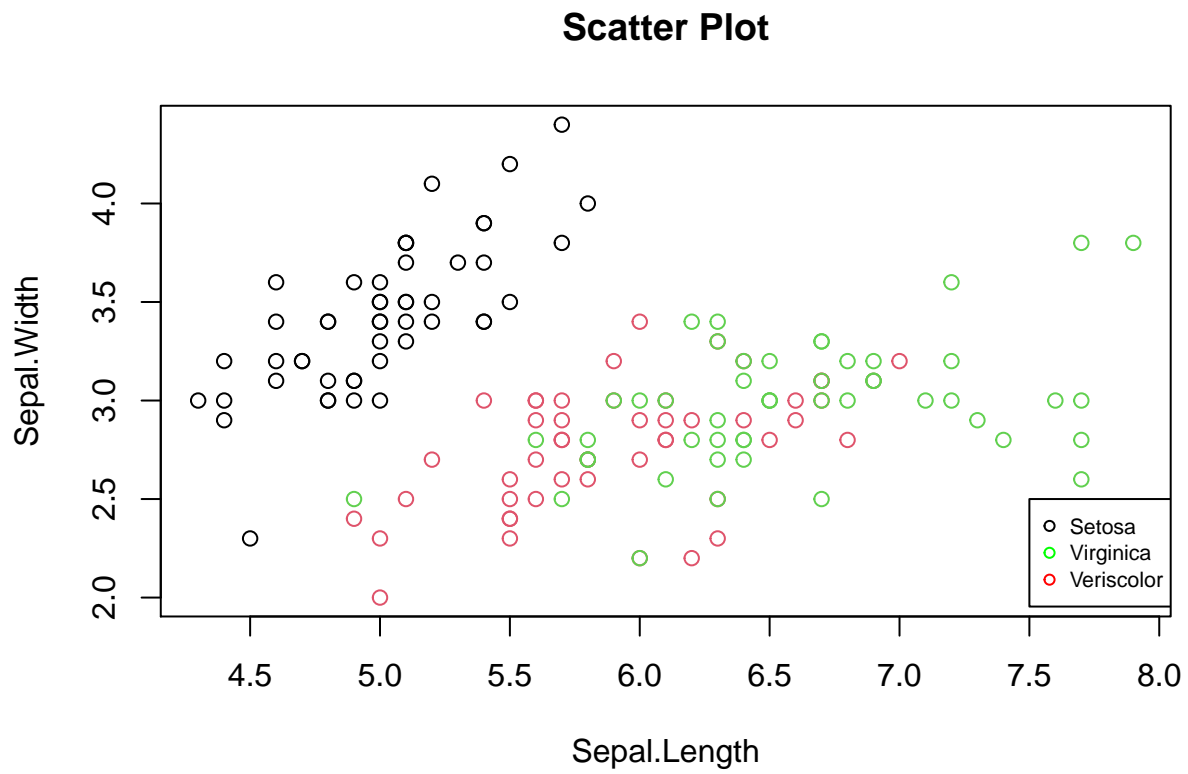
Problem 4

Data

| ## | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|------|--------------|-------------|--------------|-------------|---------|
| ## 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| ## 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| ## 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| ## 4 | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| ## 5 | 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| ## 6 | 5.4 | 3.9 | 1.7 | 0.4 | setosa |

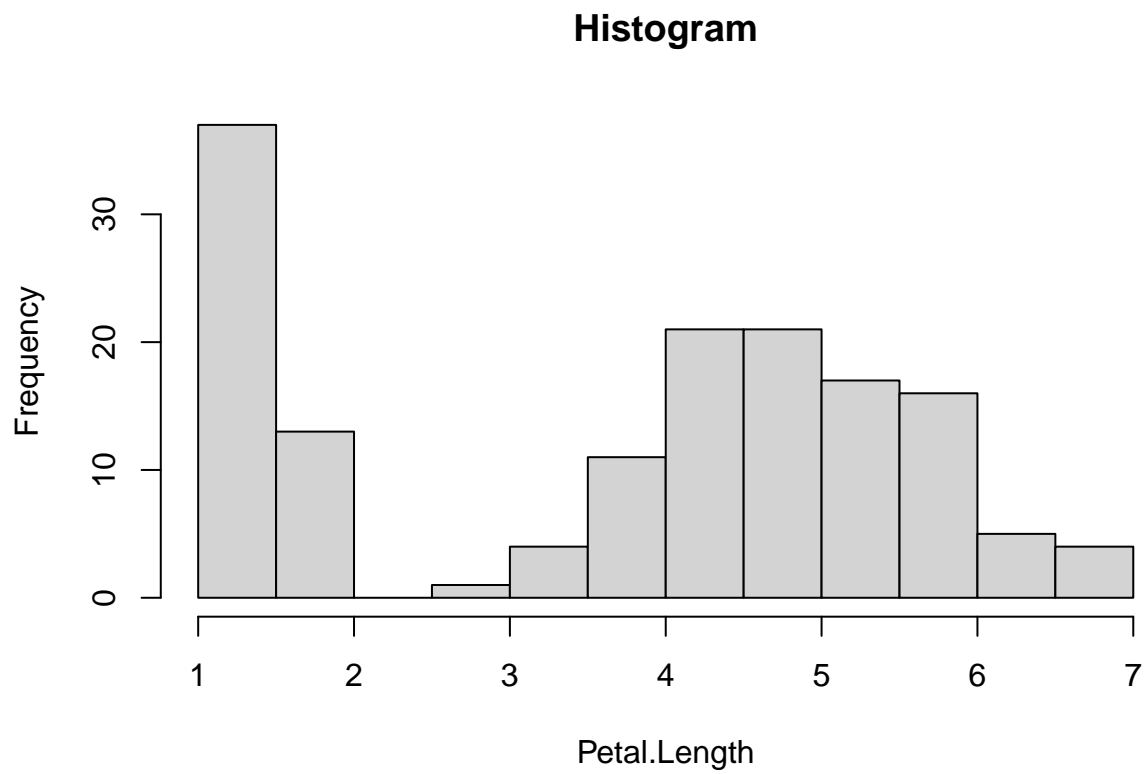
I use 'iris' data in R dataset.

Scatter Plot



I plot 'Sepal.Length' and 'Sepal.Width' of observations in the data.

Histogram



I draw histogram with 'Petal.Length' in the data.