Homework 1

Jie Min

9/2/2019

Problem 1

```
install.packages("swirl")
library(swirl)
install_course("R_Programming_E")
swirl()
```

Problem 2

Part A

My desired learning objectives are:

1)Get used to ggplot2. Learn how to do fancy data visualization in R

2)Learn about useful R packages.

3)Get a sense of programming in Python.

Part B

1)Exponential(β) pdf:

$$f(x|\beta) = \frac{1}{\beta}e^{-x/\beta}, \quad 0 \le x < \infty$$

2)Gamma(α,β) pdf:

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

3)Normal(μ,σ^2) pdf:

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

Problem 3

Rule 1

Summary:Record the full sequence of pre- and post-processing steps when doing analysis. Record every detail that may influence the execution of each step.

Challenge:Difficult to recognize which detail may influence the final result in previous steps, especially in beginning steps.

Rule 2

Summary: Use the execution of programs but not manual procedures when modifying data as much as you can. If manual operations can't be avoided, record which data files were modified and for what purpose.

Rule 3

Summary:Archive the exact versions of programs you used. Store a full virtual machine image of the operation system and program. At least, record the exact names and versions of the main programs you used.

Rule 4

Summary:Use a version control system to track evolution of your code. At least, archive copies of your scripts.

Rule 5

Summary: Record all intermediate results and make the intermediate results be easily accessible. At lest, archive any intermediate result files when running an analysis.

Rule 6

Summary: Record your random seed when running random experiments.At least record which analysis steps involve randomness.

Rule 7

Summary: When generating plots, record both underlying data and processed values(if any) related to the plots. Also record codes used to make the plot(if applicable).

Rule 8

Summary:In order to be able to inspect the main result, record all underlying data when a main result is generated and use a systematic naming convention when saving. At least once generate, inspect, and validate the detailed values underlying the summaries.

Rule 9

Summary:Connect the textual statements to underlying results from the time the statements are initially formulated using, for example, a simple file path. Other tools such as Sweave, GenePattern Word add-in, etc. can also be used. At least, provide full details along with textual statements.
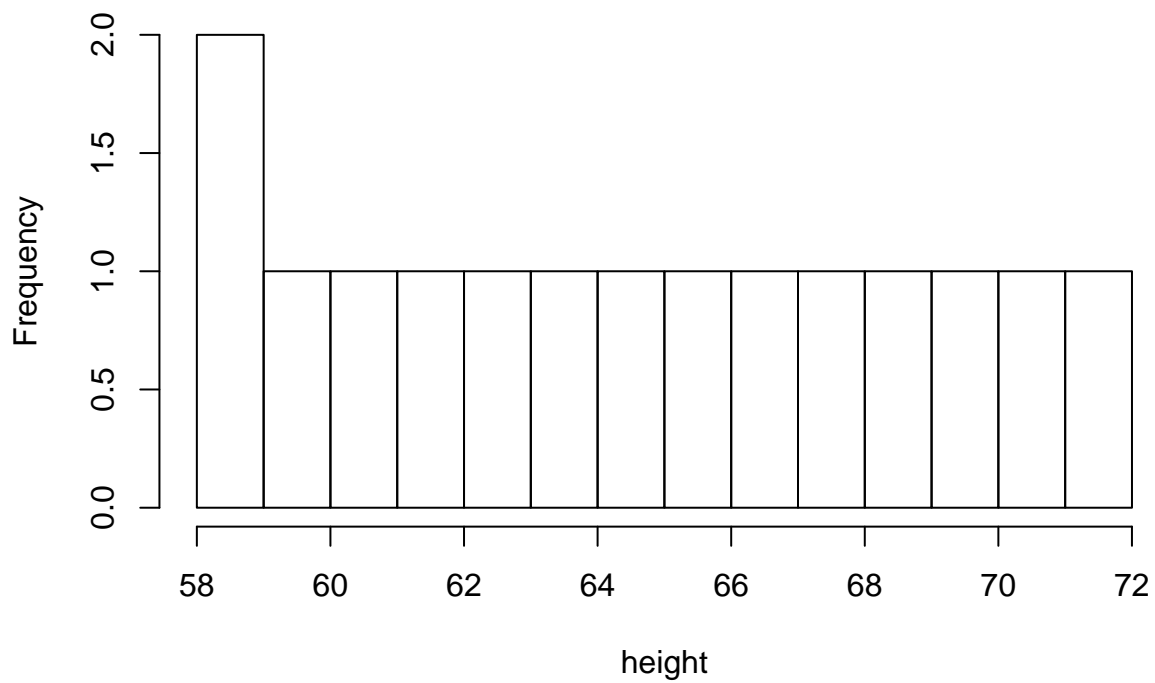
Rule 10

Summary: provide public access to all input data, versions, parameters,and intermediate results. At least, submit the main data and source code as supplementary material.

Problem 4

Scatterplot of Height vs Weight
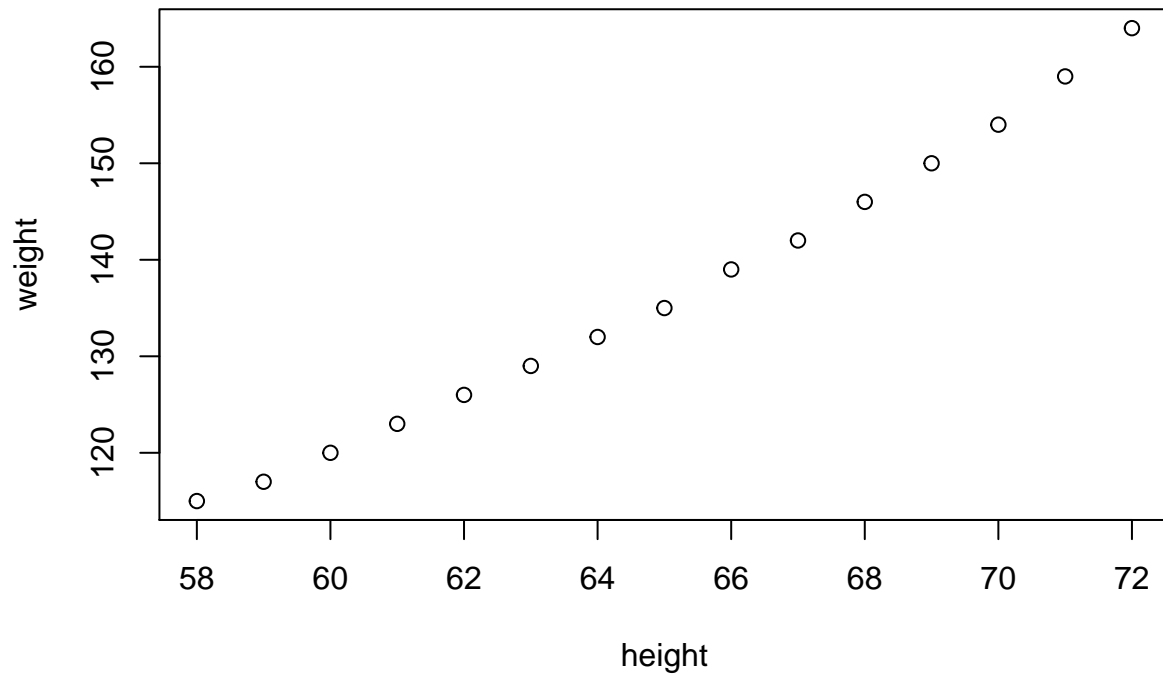
Histogram of Height

Histogram of Weight

Frequency

weight