HW1_dle0819

David Edwards

September 1, 2019

Problem 2

Part A

There are several skills that I am hoping to get out of this class such as learning about statistical software and how to write code in a group setting. However, the 3 main objectives I have in this course are:

1. Learn good practices for presenting my findings and about the software that can help create this presentation.
2. Start to build a foundation of statistical programming knowledge that I will use throughout my degree and beyond.
3. Learn about the various computing resources available through the university and how I can best utilize them.

Part B

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

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

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

Problem 3

The following is a summary of the 10 rules for creating reproducible computational research.

1.Keep track of how results were created, including any assumptions made and the sequence of commands, scripts or programs used.

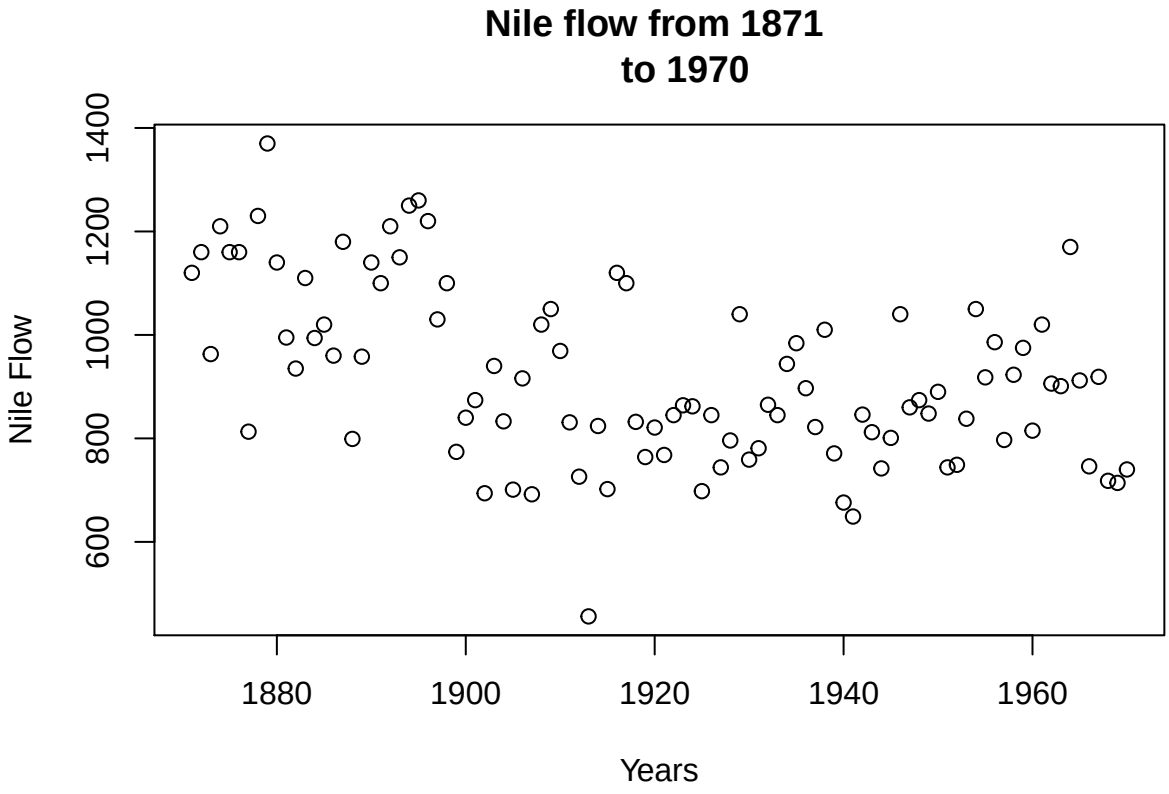
2. When possible do not change the data manually, this includes formatting the data. If manually changing the data is necessary then one should at the least keep a record of what was manipulated and how.
3. Keep a copy of all programs used in the analysis. This helps to make sure you and others have easy access to the exact version of each program used. This could be as simple as saving a single executable or script to as complicated as keeping a full virtual image of the OS used and the programs run.
4. Version control all programs and scripts created for the project. This is usually accomplished with a version control manger like Git or Mercurial.
5. Record the results of intermediate processes. This will allow you to easily retry alternate strategies and allows for easier tracking of inconsistencies.
6. When using randomness be sure to include the seed so that results can be reproduced exactly and now just approximately.
7. Keep all data used to create plots and tables and keep the code used to create them. That way if a modification needs to be made the date is easily accessible and you won't need to recreate the plots from scratch.
8. When giving a generalization be sure allow access to the details that the generalization is based on.

9. Keep a record for each statement or claim that points to the underlying data or result that supports that statement or claim.
10. Make your data, scripts, programs and results publicly available. This will allows others the opportunity to reproduce your work and if nothing else give a backup for what you have.

Problem 4

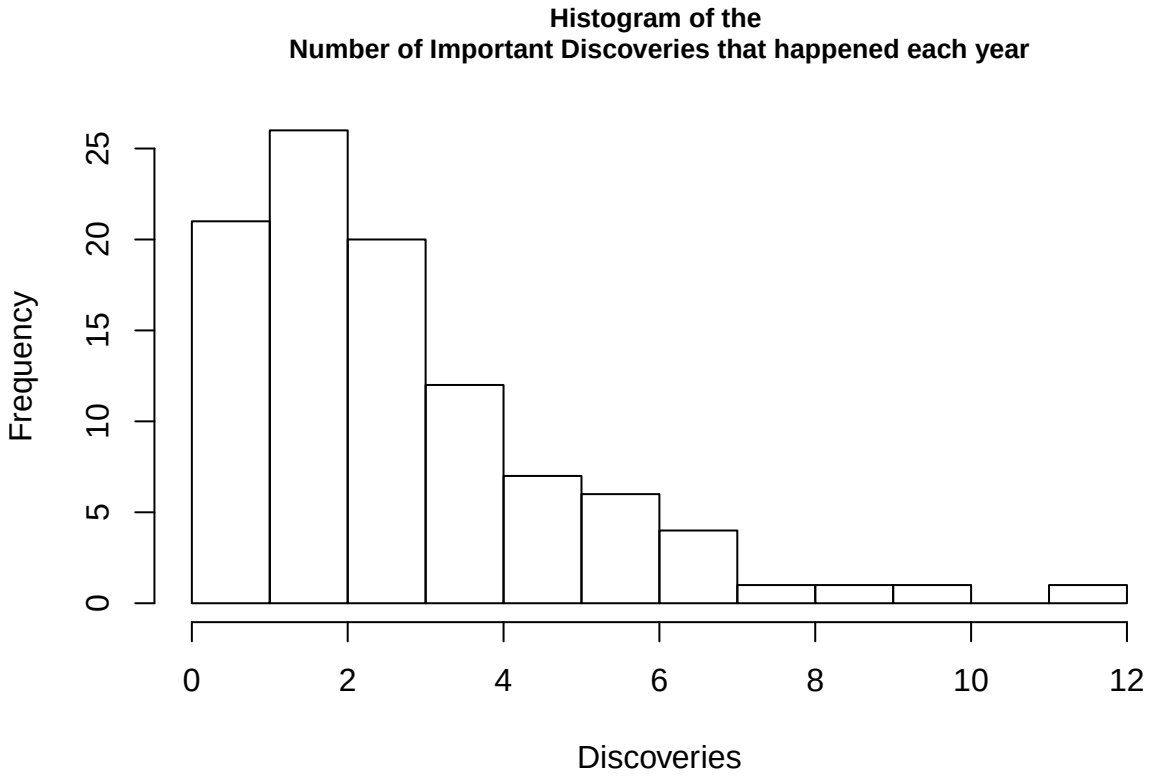
For my scatter plot I will be looking at the flow of the Nile river of the years 1871 to 1970.

```
years<-seq(from = 1871, to = 1970, by =1)
plot(Nile[1:100]~years[1:100], main = "Nile flow from 1871
     to 1970", xlab = "Years", ylab="Nile Flow")
```



For my histogram I will be looking at the number of important discoveries each year from 1860 to 1959.

```
hist(discoveries, breaks = 10, main = "Histogram of the
Number of Important Discoveries that happened each year",
xlab = "Discoveries", cex.main = .8)
```



This shows that most years there are relatively few major discoveries (usually 0-3).