

HW1__nichb15

Nick Brown

9/4/2019

R Markdown

Problem 2-

I hope to gain experience in coding R. The three learning objectives I seek are as follows:

- 1) Be able to process datasets with thousands of records to help business executives make meaningful decisions.
- 2) Learn how to become a masterful R coder who is familiar with tens of libraries and can answer data-related questions rapidly
- 3) Learn to use latex to publish journals in a variety of formats very simply and easily.

$$f(y|N, p) = \frac{N!}{y!(N-y)!} \cdot p^y \cdot (1-p)^{N-y} = \binom{N}{y} \cdot p^y \cdot (1-p)^{N-y}$$

source: <https://smac-group.github.io/ds/rmarkdown.html>

$$F(x) = 1 - e^{-\lambda x}$$

source: <https://www.stat.cmu.edu/~cshalizi/statcomp/14/labs/01/lab-01.Rmd>

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

source: https://www.academia.edu/21823064/Creating_Dynamic_Mathematical_Documents_in_RStudio_by_Unifying_Computing_with_R_and_Typesetting_with_LaTeX

Problem 3-

Rule 1: For Every Result, Keep Track of How It Was Produced comment: Some projects are one-offs and are short-lived and researchers may be in a rush to publish. This may take too long.

Rule 2: Avoid Manual Data Manipulation Steps comment: Some software such as JMP or excel may require manual steps. Some researchers may not know programming, so it would be difficult for them to script.

Rule 3: Archive the Exact Versions of All External Programs Used comment: Some researchers may work in teams and not everyone may follow the correct protocols or their work will be in multi-platforms.

Rule 4: Version Control All Custom Scripts comment: Researchers would need git or bitbucket for proper version control and not every researcher knows how to use these tools.

Rule 5: Record All Intermediate Results, When Possible in Standardized Formats comment: This can be too much of a time hassle, especially if journals want to see the end results only.

Rule 6: For Analyses That Include Randomness, Note Underlying Random Seeds comment: Not all randomness can be accounted for outside of a laboratory setting. Some research, such as organizational research, cannot be done in a lab, however.

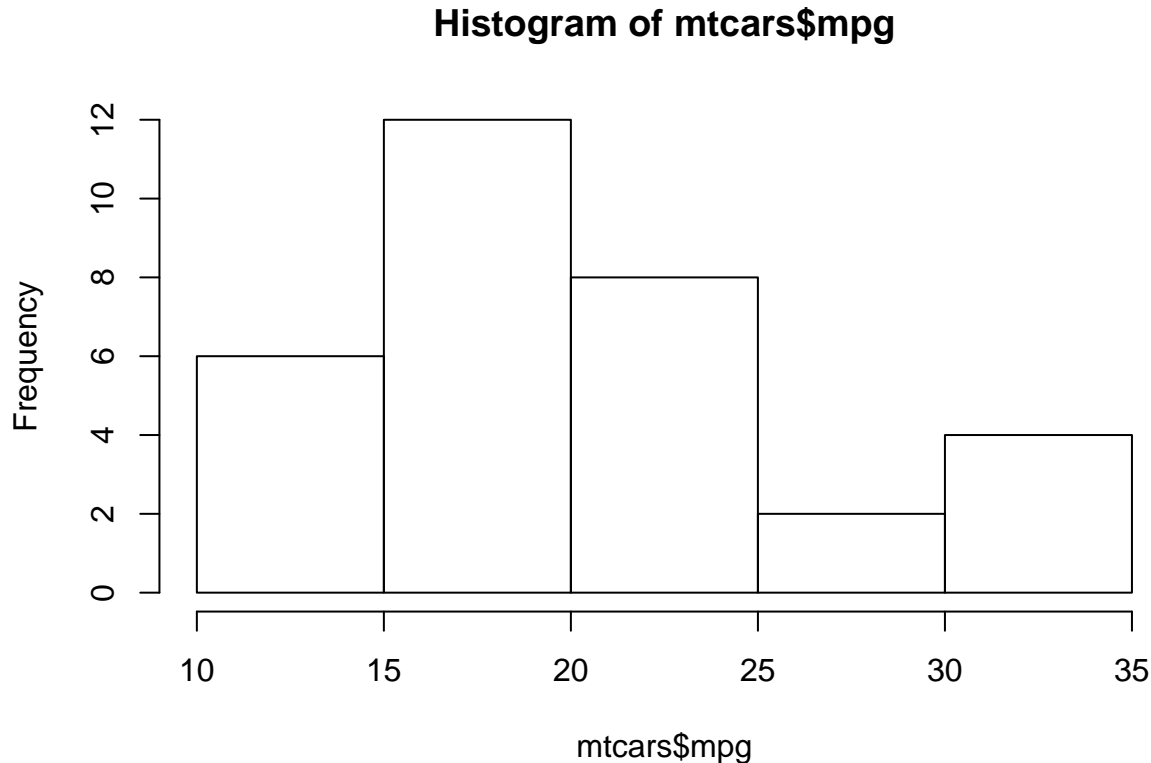
Rule 7: Always Store Raw Data behind Plots comment: Data may be manipulated and cleaned in order to achieve certain plots. The original data may be saved, but the actual data producing the results may have come about serendipitously and therefore may be more difficult to recreate.

Rule 8: Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected comment: This can be very time-consuming, especially if teams of researchers have their own working methodologies and focus solely on creating the final output. They may not document steps properly during the process.

Rule 9: Connect Textual Statements to Underlying Results comment: This can be too cumbersome if there are a lot of data computed. Accompanying qualitative narratives with quantitative results may be helpful, but can consume massive time allotments.

Rule 10: Provide Public Access to Scripts, Runs, and Results comment: Some company and government data are confidential and should not be provided to the public.

```
hist(mtcars$mpg)
```



Including Plots

You can also embed plots, for example:

