

Assignment 01

Introduction to R

This assignment is intended to give you experience working with the R program. Please submit your responses to each of the questions below in a printed document. Label the sections as indicated below within your printed document. All graphics should be resized so that they do not take up more room than necessary and all should have an appropriate caption. Any equations should be appropriately typeset within the document. There are 14 points possible for the assignment (each question is worth one point unless otherwise noted).

Use RStudio to open the *GoodReads-2015.csv* dataset and assign it into an object called **read**. Use the data to answer the following questions. Again, copy the syntax you used to carry out the computation, as well as any output from the computation, into your word-processed document.

1. Enter the following syntax. (Note: If you have not already done so, install the **dplyr** package prior to running these commands.) Copy the resulting output into your word-processed document.

```
library(dplyr)
p = read %>%
  filter(bookshelf == "read") %>%
  group_by(month_read) %>%
  summarize(pages = sum(pages))
p
```

2. Explain (in no more than 5 sentences) what the `filter()`, `group_by()` and `summarize()` syntax does in the syntax above.

Lineplot of Total Pages Read per Month

The next several questions will have you create a lineplot displaying the total number of pages read by month of the year. While each question contains more instruction to add detail to your plot, there is no need to put a separate plot in your document for each question. Include one plot in your document that has all of the elements asked for.

3. Using `ggplot()`, create a lineplot (similar to the plot seen [here](#). This plot should show the total number of pages read per month. The plot should include points and lines as in the example. It should also be displayed chronologically (e.g., January should be the first month).
4. **Axes:** Read the *Axes* page of the [Cookbook for R website](#). Add to the syntax above to change the label on the *x*- and *y*-axis of the plot.
5. **Color:** Change the line color to some non-default color of your choice. (Note that if you are printing the document using a non-color printer that the color choice you make should print well in greyscale.)
6. **Labels:** Read the *scale_x_continuous* page of the [ggplot2 documentation website](#). Add to the syntax you wrote for the previous question to add a break and label for each of the 12 months.
7. **Figure Number and Caption:** Add a figure number and caption using the appropriate APA style to do so. Note that this is often easier to do using your word-processor than using R.
8. **Syntax:** Copy the R syntax you used to create the plot into your document. Be sure that this syntax is appropriately commented and indented to make for easy reproducibility and understanding. (Note: Use a mono-spaced font to display all R syntax.)

Mean-and-Error-Bar Plot of Average Pages Read per Month

9. Change the **dplyr** syntax from above to compute (1) the average number of pages read per month—rather than the total number read per month; and (2) the standard deviation of the pages read per month. (Hint: The output should be a data frame with three columns rather than two columns.) Copy the R syntax you used, and the output, into your document.
10. Using `ggplot()`, create a mean-and-error-bar plot (i.e., Tie-Fighter plot) similar to the [plot seen here](#). Be sure the plot has appropriate axes and labels, and is numbered and has a caption. Include this plot in your word-processed document. **(2pts.)**

Lineplot of Average Pages Read per Month by Year

11. Again change the **dplyr** syntax from above to compute the average number of pages read per month—rather than the total number read per month (you do not need the standard deviation of the pages read per month this time). This time however, also group the output by year. (Hint: The output should again be a data frame with three columns.) Copy the R syntax you used, and the output, into your document.
12. Using `ggplot()`, create a lineplot by month where each year's data is displayed in a separate line (six lines total). Read the *Mapping variable values to colors* section of the *Colors* page of the [Cookbook for R website](#). Be sure the plot has appropriate axes and labels, and is numbered and has a caption. Include this plot in your word-processed document. **(2pts.)**