



Homework 0: Review

DUE Wednesday 9/9/20 by 10am EST

1. Open RStudio.


2. Create a new R Markdown document: *You will write all your R code in this file so that you can save it AND create documents! An R Markdown Document is abbreviated by “RMD”.*

- In the upper left corner, you will see a  symbol. Click on this to open a dropdown menu.
- Click on the third option, “R Markdown.” A window will appear where you can enter a title for your document (like “STAT 216 Homework 0: Review”) and the author (YOU!).
- For the Default Output Format, select “PDF” and click OK. You must have a TeX engine installed like MikTeX, MacTeX or TinyTeX.
- A new pane in the upper left side of your RStudio window should appear titled “Untitled 1”.
- Press the  (save symbol) under this tab to save the script. A pop-up window should open. Make sure you save the script to the folder you would like and name the RMD file like this “*Lastname_FirstInitial_HW0.Rmd*” and press save. Do not include any punctuation in the filename except for underscores or dashes. For example, mine would be Grzesik_K_HW0.Rmd.

Coding Tip!

R Markdown files allow us to blend R code AND explanations into one single file. Before this, we had to do a lot of copy/pasting from R into a Word document which can lead to many errors in an analysis. Statisticians use R Markdown to create a single analysis file that can be reproduced by compiling the same file.

3. Compile your first PDF document: *This is a template RMD file. Do not change anything yet, we will make some changes to it soon but let’s see what the template shows us!*

- At the top of the RMD file is a button to “knit” your R code together with your written text: .
- Since your default output is Word, just click that button to *compile* your document! (Compile means to create your document from the code.)
- Wait a moment until a PDF document pops up and then answer these questions by comparing the RMD file and the resulting document. Write your answers above where you see ## R Markdown .
 - What do the “#” symbols do to the text that follows it?
 - What does surrounding a word/phrase with two asterisks (**) do?
 - What does surrounding a word/phrase with backticks (very slanted mark, above the Tab key `) do?

Now, let’s try editing this template to complete this assignment!

4. Updating the title section. *In the template, the top section that starts and ends with three dashes (---) is called the YAML and makes a title section on your document.*

- Locate the YAML in your RMD file.
- Take a look at the title, author and date and make any changes you would like. Only edit the text in green in quotes! If you make any changes, click on  to compile again to see the changes in the resulting document.

Homework 0: Review

DUE Wednesday 9/9/20 by 10am EST


5. Meeting the first R Chunk! *R chunks are sections of the RMD file where R code goes. These start with three backticks, a set of braces and the letter r (``{r}``) and end with three more backticks (```). There are some other arguments that can go inside these curly braces, but we won't worry about it in this course.*

- Locate the first R chunk in the file. The first word after the letter r in the curly braces is the name of the chunk. This first chunk should be named “setup”.
- Now that you've located this R chunk, DO NOT CHANGE IT. Leave it alone! Move on to the next task.

6. Let's write about ourselves a little. *We can write paragraphs just like in any other text editor by typing ***not*** inside of an R chunk.*



- Locate the first header of the template. It looks like this: `## R Markdown`
- Change the text so that the headers will print “My Main Major Interest” (without quotation marks). Compile your document to make sure that it is still formatted like a header in the PDF document.
- Replace the next two paragraphs with at least 3 sentences about your major, why you enjoy that topic and how statistics can help you understand that subject more.
- Compile your document again to see your progress!

7. Meeting an R chunk you can edit. *You'll notice that the screen is shaded whenever there is an R chunk. This helps you to know where you can write code (in the gray area) and where you can write text (in the white area).*

- Locate the 2nd R chunk. In this template, this R chunk is named `cars`.
- Move your cursor to inside the curly braces and change the name of the chunk to `doingMath`.
- Move your cursor inside the R chunk, highlight `summary(cars)` and delete it.
- In this empty line, type: `2 + 2`
- Press Enter to get a new line and type: `10/2`
- To view the results of these lines of code while you are coding, try each of the following:
 - Select/highlight one line of code you want to run and press  (above the RMD file). The resulting output will display in either the Console (lower-left) or Plots pane (lower-right). This time, you will see the output (4) in the Console (lower-left).
 - Select/highlight BOTH lines of code and press again. Locate where the output is.

Coding Tip!

- With your cursor on a line of code, you can ‘run’ the line by using the keyboard shortcut Ctrl+return (Windows/Linux) or Cmd+return (Mac).


- To run the entire R chunk, locate these buttons   and click on the last green arrow. This output should show up right below the R chunk.
- Now try a few more basic math calculations of your own inside that same R chunk! The asterisk (*) is multiplication, the forward slash (/) is for division, and the carat (^) is for exponents.

Homework 0: Review

DUE Wednesday 9/9/20 by 10am EST

8. Importing a dataset. *There are many ways to import data, here is one of them!*

- a. Locate the header named “Including Plots” and change that to say “Importing a Dataset”.
- b. Replace the text between the header and the next R chunk with a description about how the following is the code for importing a CSV file. You should delete the text underneath this chunk
- c. Let’s prepare this next R chunk by doing the following:
 - i. Change the name to be `importData`
 - ii. Remove the comma and the option `echo=FALSE`
 - iii. Delete the R code `plot (pressure)`
- d. The dataset ‘menu.csv’ was provided to you via Blackboard. Download this CSV file and put it where your “Lastname_FirstInitial_HW0.Rmd” file is located. These files must be together! This dataset contains nutrition information for menu items at McDonalds.
- e. Obtain the computer’s pathway to the folder containing the Rmd and data files:
 - If you have a WINDOWS computer, right-click on the data file, click Properties, and then copy what you see for the “Location”.
 - If you have a MAC computer, right-click on the data file, press and hold the Option key, select “Copy as Pathname”. ([Source](#))
- f. In the R chunk, type `setwd (“ ”)` and then paste the pathway in between the quotation marks. (Do NOT copy/paste any quotation marks from this file. You must type them out from your keyboard!)
- g. Next, hit Enter to get a new line and type out:

```
menu <- read.csv (“menu.csv”, header=TRUE)
```
- h. Run that code using either the Highlight these three lines of code all together and press . A new window should appear showing you the dataset. As long as your “menu.csv” file is in the same folder on your computer, this code will import that data!

8. Create some summaries. *Create a new section named “Exploring the Data” and make a new R chunk.*


Complete the following tasks. Feel free to use your notes from previous courses!

- a. Barchart of the menu categories.
- b. Boxplot of calories.
- c. Histogram of calories.
- d. Calculate the correlation between protein content and calorie content.
- e. Scatterplot of calories versus protein.
- f. Side-by-side boxplots of sugar content by the category. Make sure you can see all of the group names!
- g. Calculate the mean sugar content for each group using as few lines of code as possible!

9. Discuss your observations! *Under the graphics chunk, discuss the following in separate paragraphs:*

- a. Describe the distribution of the calories variable using statistical vocabulary.
- b. Explain how you decided to assign the variables to the axes of the scatterplot.
(Which did you choose to be on the X and Y axes and WHY?)
- c. What kind of relationship do you see in the scatterplot? Do we expect a linear model to work? Explain.
- d. Comment on any unusual features – are they outliers or influential points? Explain your reasoning.
- e. How do the categories differ in sugar content? Is what you're seeing making sense? Why?

10. Complete and submit your assignment.

- a. Compile your PDF document again by clicking  to update your document!
- b. In Blackboard, submit your **PDF** document and your RMD file as individual files. Please do NOT submit a zipped folder.