1. Create a new project in RStudio. Save the empty directory in your class folder for this class (if you haven’t already created a folder for this class, do so now).
2. Find a data set to store in your project.
   1. Try the spam data set in the kernlab package. (The use of this data set is inspired by the RWDSR book by Roger Peng).
3. Create a new R Markdown document with html output
   1. Intermediate: have the document create both html and MS-Word output.
   2. Intermediate: would it work to also have the document create a presentation (e.g. beamer) from your Rmd file? Why or why not?
4. Create a “setup” code chunk at the beginning of your R markdown document, if it does not already exist. Install and load the following packages, but make sure the output from this code will not appear in the final document (loading packages can get messy).
   1. Install.load
   2. mosaic
   3. ggplot2
   4. devtools
   5. tidyverse
5. Create a section heading in your Rmd file called “Intro”. In this section:
   1. Write a couple sentences or a short paragraph describing the data. Use the data documentation in R to get this information. (Note: this data is not well documented; part of the assignment is to list some questions you have about the data).
   2. Load the data into your RMarkdown document using the correct code. (If you aren’t sure how to do this, use the Import Data button, and then copy and paste the code it creates).
   3. Make a table of the first few rows of the data set.
   4. Intermediate: Find an image online that seems to fit with your data set. Include this image in your Rmd file, with a caption showing the source.
6. Create a section heading in your Rmd file called “Summary of Data”. In this section:
   1. Identify three variables you think will be most helpful, and summarize them using the summary function. Comment on the results.
      1. Intermediate: Try creating grouped summaries (e.g. the compare the means of one group vs another).
   2. Suppose you want to mention one particular number from the summary in your paragraph. (e.g. “the mean is 5”). Look up how to use inline R code in order to do this. Incorporate some inline R code to discuss the results of your summary.
   3. Explore your data with a minimum of 2 graphs.
      1. Beginner: use the plot function, or ask for help.
      2. Intermediate: try using the ggplot graphics.
      3. Advanced: Treat this as you would an actual data analysis project for EDA.
7. Advanced (optional): If you have time, add a “Data Analysis” section.
   1. Split the data into a training set and a test set.
   2. Try to use the training set to classify whether an email is spam or not.
   3. For more guidance, check the RWDSR book by Roger Peng.
8. Add a short conclusion section summarizing what you found in the previous section.
9. Add a table of contents in the header. Look around for other YAML options and add anything you are interested in.
10. Customize R Studio with add-ins.
    1. Note: you need to install the package “devtools” to do this. This may require some work outside of class if you run into issues.
    2. Check here: <https://rstudio.github.io/rstudioaddins/> and here: <https://github.com/daattali/addinslist> to find useful addins.
    3. At a minimum, you should plan to install remedy. I have also found splitChunk and insertImage to be mostly helpful.
    4. You may also want to look up TODOr, Plot Colour Helper, Hist Add-in, Snake caser, Word count add-in, Add GIFs, gitgadget, Align Assign, etc.
    5. Intermediate: Discuss common keyboard shortcuts for add-ins of interest, and start assigning keyboard shortcuts to improve your efficiency (e.g. Ctrl-B for bold text).
11. Save and knit your document.
12. Three options for submitting your work:
    1. Zip your project and submit the compressed file in Canvas.
    2. Submit your Rmd file and output file in Canvas.
    3. Add to your Github account and share with me (paste link to repo in Canvas).