

COMPSCIX 415.2 Homework 4

Robert Clements

DUE DATE: Feb 27, 2018 @ 6:15PM

In this assignment...

...we will finish up with data munging and data wrangling in R using the `dplyr` and `tidyr` packages. You'll learn how to use the `summarize()` *verb* and how to reshape data.

Remember to make your document look good, which means you may need to change some settings on the figure sizes and locations, use markdown syntax to create headings or to format your text (use the cheatsheet), and you may want to play with the different themes.

Use complete sentences, and **divide your work in a logical way**. Remember that the whole point of doing reproducible analysis in R Markdown is so that a complete stranger can take your results, *understand them*, and reproduce them.

Remember to save and knit often. Commit when you've completed a big chunk of work or when you are done for the day and will be resuming later.

What to Turn In

For this assignment you have two choices:

You can upload a pdf document (you will have to install latex);
You can upload a standalone html file.

Prerequisite

Basic R Markdown knowledge
R and RStudio
`tidyverse`, and **nycflights13** package installed
Access to internet
git and Github

To start your assignment

1. Go to File -> Recent Projects
2. Click on the Project that you created during Homework 1. This project should be the one that is already under git version control.
3. RStudio will switch to that project and the Git pane should appear.
4. Go to File -> New File and choose either R Markdown or R Notebook.

5. Change the title (“COMPSCIX 415.2 Homework 4”) and add your name and the date to the YAML header.
6. Save the file in the same folder (or create a subfolder) with your other HW assignments.
7. Knit your document into an html or pdf document.
8. Go to the Git pane and commit both your Rmd and html (or pdf) files by clicking on the checkboxes next to the file names and hitting the Commit button. Write a useful message, and hit the commit button.

Exercises

Here are the exercises for you to complete.

Go to R for Data Science at <http://r4ds.had.co.nz/index.html>.

Complete these Exercises:

Section 5.6.7: #2, #4, #5 and #6 only

Section 10.5: #1, #2, #3 and #6 only

Section 12.3.3: #2, #3 and #4 only

Section 12.4.3: #1 and #2 only

Answer these questions:

Follow these steps:

- Download the baby_names.txt file from Canvas which is in the Homework 4 assignment section.
- Load this file into R *correctly* and take a glimpse of the output.
- Export this file as an rds file and call it ‘baby_names.rds’.
- Reload the baby_names.rds file and take another glimpse.
- **Show all of your code and the output.** There should be two data import lines of code, one data export line of code, and two glimpses of the data.

Turn in your completed assignment

Commit your changes with the comment “finished assignment 4” and push your R Markdown file and your html or pdf file to Github.

This week you should turn in your assignment by uploading it to Canvas.