

Final Project

Ellen Bledsoe

2025-04-17

Final Project

Purpose

The purpose of the final project is to synthesize material across the semester.

Task

Create a fully computationally reproducible project that uses the skills we have learned in the class. Apply them to a dataset of your choosing.

Criteria for Success

- Project is in a GitHub repository with effective project structure.
- Project is fully computationally reproducible, meaning your instructor can clone your GitHub repo, open the RStudio Project, and successfully recreate your project without having to make any changes to the code.
- Project contains code or material from the appropriate number of weeks, as described below.
- RMarkdown file contains written descriptions of the tasks being performed and reaches the minimum code length.

Due Date

May 15 at midnight MST

Final Project Guidelines

For your final project in Data Wrangling in R, you will be synthesizing the concepts we have learned throughout the semester and applying them to a dataset of your choosing.

In many ways, the content of the final project is up to you. However, there are a few guidelines that need to be met for your project to be considered complete. The guidelines differ based on whether you are in WFSC 496B or 596B, so be sure you are looking at the correct section.

Week 1: Spreadsheets

Week 2: Intro to R

Week 3: Data Tables

Week 4: Groups and Joins

Week 5: Data Visualization
Week 6: Tidy Data
Week 7: Dates and Strings
Week 8: Problem Decomposition
Week 9: Reproducibility in RStudio
Week 10: Version Control
Week 11: Functions
Week 12: Making Choices
Week 13: Iteration
Week 14: Spatial Data

General Requirements for the Project

Based on what I am asking for your project to be, your project will cover content from a number of weeks by default.

1. **Week 8:** Your project should accomplish a larger task with your dataset. In your project's RMarkdown file, you should include text:
 - At the beginning of the document to introduce your dataset and the general tasks that you will be completing in the following code chunks (problem decomposition).
 - Text between the code chunks outlining the tasks that each code chunk is performing.
2. **Week 9:** Your project should exist within an RStudio Project.
 - The RStudio Project should have sub-directories (at a minimum, **data** and **figures** or **outputs** folders).
 - All file paths in your project should be relative file paths (not absolute).
3. **Week 10:** Your project should be in a “Final Project” GitHub repo. You will submit the link to this repo as your submission for the final project.
 - The README file should give a brief description of the data, project, and file structure.
 - If there are multiple scripts or files, explain what they each contain and the order in which they should be run.

*Note regarding Week 1: If you choose to include content from Week 1: **Spreadsheets**, it should be in the context of (a) creating your own spreadsheet for future data entry that includes data validation or (b) discussion of why you have chosen to save your cleaned data in multiple different “tidy” tables. Reading a spreadsheet file into R does not count. If you have questions, feel free to ask!*

Demonstrating You’ve Met the Requirements

You will need to explicitly note somewhere in the final project where content from your required number of weeks is present in your project.

This could take a few forms:

- a README file (could be the repo’s README file) that outlines the weeks covered and where to find each week’s material in your RMarkdown file (hint: the source code has line numbers)
- within your RMarkdown file and the text accompanying your code, explicitly stating the weeks included in the following code chunk
- another way that makes the most sense for you!

WFSC 496B Requirements

Your final project is worth a total of 300 points and should contain content from at least 6 weeks. Each week of content included is worth 50 points.

In addition to including content from Weeks 8, 9, and 10, as stated above, your final project should include content we learned in at least 3 other weeks, with the following guidelines.

1. You must include some type of data visualization in your project (Week 5)
2. You must include content from at least one of the following weeks:
 - Week 11: Functions (writing your own)
 - Week 12: Making Choices (if/else statements)
 - Week 13: Iteration (for loops)
3. Your .Rmd file should have a minimum 75 lines (this includes everything in the document: the YAML, any text, lines of code, lines creating code chunks, etc.). For context, assignment answer keys with little to no text range between about 100 and 300 lines of code).
4. Your .Rmd file should have `github_document` as the output, and the knitted and updated .md file should be in your GitHub repo.

WFSC 596B Requirements

Your final project is worth a total of 500 points and should contain content from 10 different weeks throughout the semester. Each week of content included is worth 50 points.

In addition to including content from Weeks 8, 9, and 10, as stated above, your final project should include content we learned in at least 3 other weeks, with the following guidelines.

1. You must include some type of data visualization in your project (Week 5)
2. You must include content from at least one of the following weeks:
 - Week 11: Functions (writing your own)
 - Week 12: Making Choices (if/else statements)
 - Week 13: Iteration (for loops)
3. Your .Rmd file should have a minimum 150 lines (this includes everything in the document: the YAML, any text, lines of code, lines creating code chunks, etc.). For context, assignment answer keys with little to no text range between about 100 and 300 lines of code).
4. Your .Rmd file should have `github_document` as the output, and the knitted and updated .md file should be in your GitHub repo.

WFSC 596B Deadlines

We will be performing a peer code review as part of the final project process, so you will need to have a “draft” ready a bit earlier to give each other time to conduct the peer code review.

- *Sunday, May 4 at midnight:* >50% of project due
- *Sunday, May 11 at midnight:* Peer code review due
- *Thursday, May 15 at midnight:* Completed final project due

The more you have done by the draft due date, the more helpful feedback you can receive! You can also continue to work on your project and push changes up to your repo during the week that the code review is occurring; you do not have to halt work on it!

Making a New RMarkdown File

For this project, you'll need to create a new RMarkdown file.

As a reminder (we did this together in Week 10), here are the steps to create a new RMarkdown file.

1. Go to File > New File > RMarkdown
2. From the option list on the left-hand side, make sure Document is selected.
3. Add a title and put your name in the "Author" section. You can either leave the date as is or check the box that will automatically fill in the current date.
4. Choose any output (we are going to change it, anyway).
5. Click "OK" (for reasons I won't go into, do NOT choose "Create Empty Document").
6. You should now have a .Rmd file with some stuff in it. The stuff is giving you a brief rundown on how .Rmd files work. You can delete everything from line 11 (**## R Markdown**) and below. You need to leave the YAML (the stuff between the --- lines) and I recommend leaving the very first code chunk.

You should be good to go!