Hi Sajad

What follows is a detailed description of your first assignment. Don't hesitate to ask questions if my description is unclear.

Chapter 3 of my textbook will describe the data preparation process. The Rmarkdown file for Chapter 3 is "data_cleaning_bookdown.Rmd" and the rendered html file is "data-preparation.html". Note that the rendered file shows as Chapter 1. This will change automatically when all of the chapters are combined and the full book is rendered.

I have decided that this current version of Chapter 3 is too long. I will shorten it considerably, mainly to give the reader a sense of the standard methods of data preparation, and to provide a few examples. I will make the full data preparation code for all applications available on the textbook website. Your job is to create a R file (not a Rmd file) which does the full set of data preparation and will eventually be posted. The file you create should be set up as an R project and the "here" function used so that users can run the file you will create without having to change the path to the raw data files (i.e., do the same as what I did in the Rmd file). The file should contain short descriptions embedded in the code to make understanding the code easier for a user with beginner to intermediate R skills. Refer to other on-line textbooks to determine the appropriate level of coding comments to use.

Following my lead, you will create a series of tsibble objects, typically one object for each data set which is imported into R. The objects are typically merged into a single object for the specific application. Think of an application as a case study. One chapter may have three separate applications, and each application may have one or more data sets which have been merged into a large tsibble object for that application. The main body of the "data_cleaning_bookdown.Rmd" file creates the merged objects. At the end of "data_cleaning_bookdown.Rmd" there is the option to uncomment the "save RDS" code in order to save the final tsibble data frame as an RDS file. These RDS files are later imported into the Markdown file for the various chapters. The Word document "Data Sets by Chapter" shows the various RDS files which are used in each chapter. You don't have to worry about the applications at this point. For this current assignment you will focus on creating a R document which replicates the code contained in "data_cleaning_bookdown.Rmd" for the purpose of posting on the textbook website. After this assignment is complete you will begin working on the individual chapters. However, if you are curious you can see a html draft of the chapters at https://s3.us-west-1.amazonaws.com/www.vercammen.cc/_book_ch2/time-series-programming.html

After you have created the new R file for preparing all of the data you will generate a final set of RDS files which will replace the existing RDS files. Obviously the new RDS files you create should be identical to the ones I currently created using "data_cleaning_bookdown.Rmd" unless we both agree that I previously made a mistake, or I wish to add additional data to the application. It is important for you to design a procedure for ensuring your RDS files and my RDS files are identical. I am not concerned if the column order is different but otherwise the data in the files should be identical.

I don't believe I have made mistakes during the data preparation stage but please keep an eye out for possible problems.

The final textbook (and also the copy I will send to the publisher for review) will not import RDS files as is currently being done. Rather, those who wish to execute the code in the textbook will import a user-

created package at the beginning of Chapter 1 -- the package will contain all data sets for the textbook. The package will reside on my Github website – I notice that other online textbooks use this approach. I suggest that you create a Github website for hosting the package of data. This will allow you to test to ensure the package procedure works properly for loading all of the textbook data at the beginning of Chapter 1 without having to explicitly import RDS files.

The beginning of the "data_cleaning_bookdown.Rmd" file contains the URLs to all of the raw data sites. You should create a new list in Excel, Word or R which is the master list that we will both use when moving forward. Similarly, I have included a folder with the csv and Excel files of the non-API data. I expect there are some extra files in this folder which I never ended up using. I would like you to create a new folder which only has those Excel and csv files which you used within your R program (and my current Rmd program). Your folder will be part of the R project that will be downloaded by users who wish to replicate the data preparation outcomes. It is important that the web links to the Excel and csv files which you will include in the master list are able to generate the Excel and csv files which are in the folder. Please check each one.

A big part of your current assignment is to improve the formatting of the code I used to prepare the data. Your coding is very good already but you may wish to consult https://style.tidyverse.org/syntax.html for additional ideas. Pay particular attention to the creation of variable names. I have generally used underscores rather than dots to connect different parts of a variable name. I am fine with which ever approach is being commonly used. You will soon see that my approach to naming variables has been very ad hoc. If I was to write these chapters again I would have spent more thinking about naming conventions and best practices before starting.

It is also important that the coding is consistent. For example, the FRED API typically involves selecting columns, filtering, slicing, mutating and date conversion. Your code should do these steps in the same order throughout the date preparation process whenever possible. In your R code it is okay to use the head() function at key points in the code so that the user who runs the code can have an idea of what the data looks like as the final objects are being created. I have not used the head function very often in "data_cleaning_bookdown.Rmd" in order to prevent excessive length of the chapter.

After you have completed your R file I will use parts of your code to rewrite a much shorter and concise versions of Chapter 3 on data preparation.

I hope this all makes sense.

Good luck

Jim