Elementary Statistics for Public Administration: An Applied Perspective - Ch. 5: R Software and the R Commander

Daniel S. Scheller

July 22, 2022

Ch. 5: R Software and the R Commander

5.1 Introduction

The previous four chapters have focused on the principles of robust research design and measurement of variables. By now, as a student, you should have some working knowledge of how to devise research questions and hypotheses, including how to operationalize concepts into observable variables. The purpose of this chapter is to describe the R Commander software and how to access it - how to download it, open the software, upload data into the software, and to make sense of the software for beginners. From there, the focus of the chapter turns to the basics of the R Commander software and how to use it for recoding and cleaning data.

5.2 Description of R

The R software for statistical computing is an open source software, available for free download from the Internet. Being open source, any computer programmers and statisticians can create their own programs for running statistical tests, known as packages. In fact, the package used in this textbook is the R Commander package, which was created by John Fox and turns the R software interface into a more user-friendly drop-down or Graphical User Interface (GUI) program for students and researchers. When one downloads the R Software and opens it on their computer, it is a largely nondescript interface. In fact, to get the R software to make calculations or to conduct statistical tests, R code needs to be entered into the interface. Therefore, it becomes necessary to understand the R coding language to run your analyses in the R interface. With the R Commander package, running statistics becomes much easier, much in the same vein as the popular SPSS software.

5.2.1 Comparing R with Other Software

The most often used software in public administration courses tends to be SPSS. It is seen as the most user-friendly and accessible software. To be sure, SPSS has a fairly intuitive GUI with drop-down menus. Its popularity has led to the publication of many resources and tutorials from books to YouTube videos. Certainly, the benefit of SPSS is its level of intuitive use and availability of tutorial resources. However, it is a fairly expensive software to own and can be pricey to rent. Despite these drawbacks, this textbook recognizes the importance of SPSS In public administration research and its

continued use. As such, each chapter with software instruction in this textbook includes the relevant SPSS steps to conduct the statistical tests. There is no debate regarding the ease of use for many students with SPSS and its popularity in public administration programs and thus the reason for its inclusion in this textbook. The one downside of SPSS Is its cost and it lack of use in the public sector. This lack of use increases the importance of free and easily-downloadable software like the R Commander.

Another software that is gaining popularity in public administration programs is STATA, especially at the Ph.D. level. While STATA has drop-down GUI options, it is more commonly used as a code-based program, where analysts create do-files of STATA code to run statistical tests. For many, the STATA coding language is much more intuitive than the R interface coding language. However popular STATA may be among some public administration programs, it is still a fairly expensive program to purchase and is not widely-used in the public sector compared to other software like Excel. However, given the increasing popularity of STATA in public administration programs, I have decided to include STATA instruction alongside SPSS instruction in the Appendices of chapters. The focus of the STATA instruction is on the use of code in do-files as this is usually the main method of STATA use.

The most commonly used data software in the public sector is Microsoft Excel. Most students and practitioners are at least somewhat familiar with Excel or have used it in some manner in their studies or profession. Excel is certainly a great tool/spreadsheet for data entry. However, beyond basic graphing and descriptive statistics, Excel is quite difficult to use. In fact, some of the statistical output with Excel may contain errors. This textbook utilizes Excel in the manner for which the author sees its strength - data organization/entry. Excel is an excellent and often preferred software for data entry. It can be easily uploaded to nearly every statistics software. Therefore, its use is highlighted in this text as a tool for data entry and organization to be uploaded into the R Commander for analysis.

5.3 Installation of R and R Studio

There are a few ways to download and access the R Commander software program. The general R platform can be downloaded from the Comprehensive R Archive Network (CRAN). It is a code-based platform and can be used to call up the R Commander. However, a more user-friendly platform to call up the R Commander is RStudio. The RStudio is free to download. Follow these steps to easily download RStudio:

1. Go to `https://www.rstudio.com/`

2. There should be a "Products" menu towards the top of the webpage. Hover over "Products" with your cursor. Doing so should reveal a popup box.

3. Slide your cursor to the words "RStudio The premier IDE for R" under the "Open Source" column and click.

4. A new page should appear. Scroll down and you will see two versions for downloading R. Click on the box with the words "RStudio Desktop."

5. The screen should then scroll down further on the page for two more options. The **free** Open Source Edition is the correct version. Click the blue box that says "DOWNLOAD RSTUDIO DESKTOP."

6. A new page will appear showing different versions of the RStudio software. It is not necessary to pay for RStudio. Click the blue "DOWNLOAD" button under the Free version of RStudio Desktop.

7. The resulting page is the download page. The website should be able to detect the respective operating system on your computer. Click the blue download button, which should download the software download file to where downloaded files go on your computer.

8. Double-click on the download file and follow any relevant instructions for downloading the RStudio to your computer. If you are using a Mac, you will need to drag the RStudio icon to the Applications folder to have it installed on your computer.

5.3.1: For Mac Users - Downloading X Quartz

Mac users need to download an additional piece of software for the R Commander to run and successfully display results and graphics. This software is called X Quartz, and it is also free. To download X Quartz:

1. Go to `https://www.xquartz.org/`. The page should have a ".dmg" file to download.

2. Click on the ".dmg" file, which will be downloaded to your computer.

3. Double-click the downloaded ".dmg" file where download files are stored on your computer, which will open a window containing the XQuartz software file to install to your computer (XQuartz.pkg). Double-click this file.

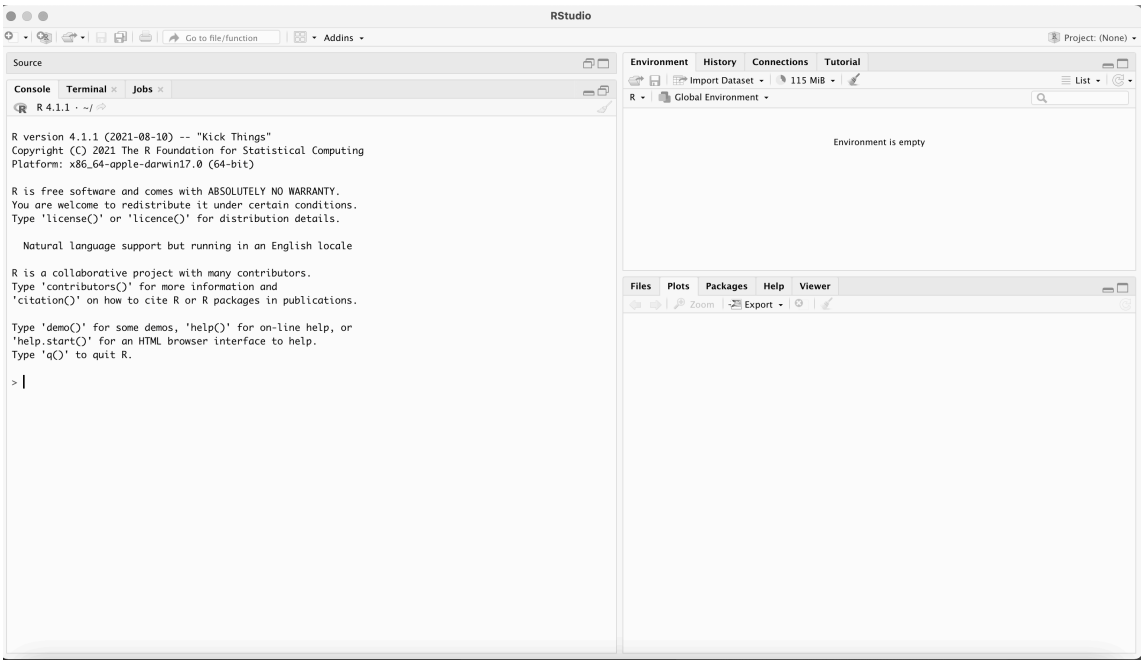
4. Several approval windows/popups may appear. Allow all approvals, and continue with the installation instructions on the computer. There should be no need to change any installation location on your computer. Default settings should work.

Your computer may ask if you want to send the Installer to the Trash. That is perfectly fine to allow, since the file is simply the installation file and not the actual

software. At this point, you can proceed to the next section of accessing and installing the R Commander package.

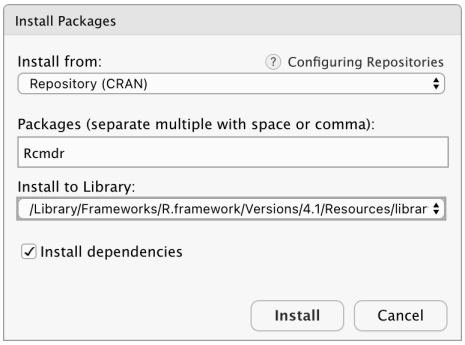
5.4 Installing and Accessing the R Commander

Once RStudio is installed, simply double-clicking its icon from its stored location on the computer will open the RStudio. Depending upon your computer, a popup security-type box may alert you that the RStudio is an application downloaded from the Internet. If this box appears, simply agree to open the application. The RStudio should open and look like the following image:



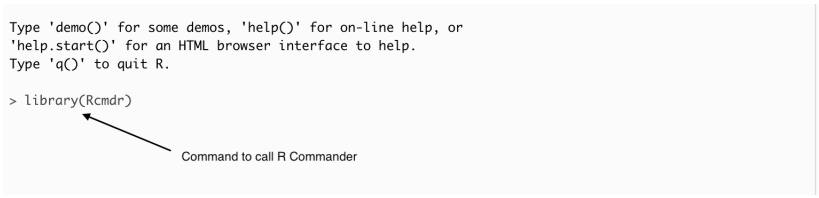
At this point, one could begin using the RStudio as a general R interface if one knows R Code to upload and analyze data. Since this is not our purpose in this text, and we want to use a drop-down version of R, called the R Commander, we must install the R Commander package. To install the R Commander package:

1. In the R menu at the top of the screen, click Tools → Install Packages. A popup box should appear like below:

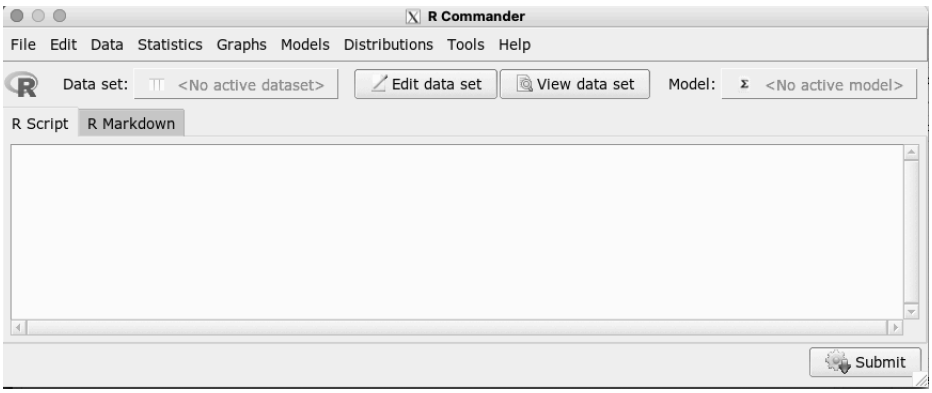


2. In the blank "Packages" box, enter Rcmdr as shown in the box above. A list of packages may appear. If so, simply click Rcmdr in the list. It will populate Rcmdr into the blank "Packages" box. There is no need to change the Library where the package is stored. Also, make sure that the "Install Dependencies" box is checked. It should be checked by default. Click Install.

3. The installation process should be completed fairly quickly. When the cursor appears again, type "library(Rcmdr)" and click Return/Enter. See the image below to ensure you have correctly written the code. This is really the only code that you need to know.



It calls up the R Commander software from the RStudio interface. The R Commander interface should look like:

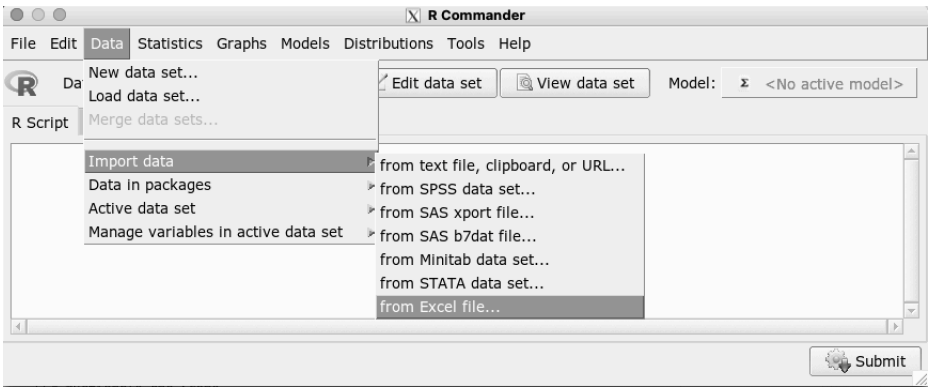


This interface is the main application you will use for all of your data analysis and visualization and will be referred throughout the text as the R Commander or R Commander interface.

5.5 Accessing Data in the R Commander

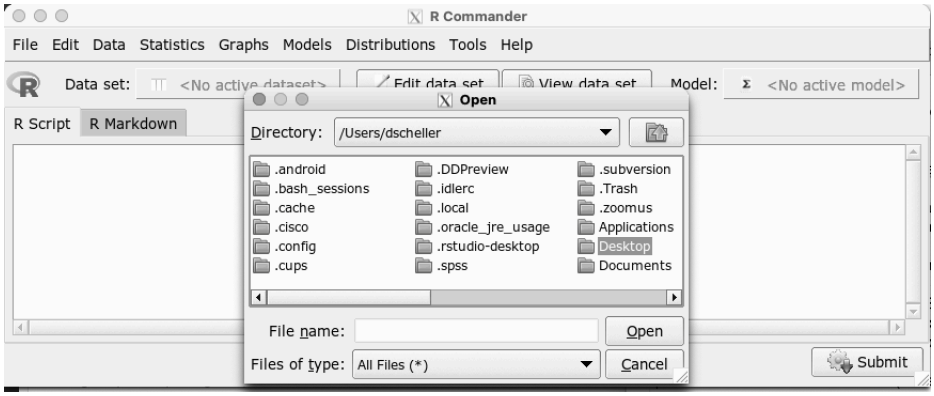
With the R Commander software open, the next logical step is having it read data from a dataset. It can call up different types of data formats: Excel, STATA, SPSS, SAS, delimited files, etc. As a reminder of some of the best practices from data collection and storage from the last chapter, it is important to have your data stored in a central location, like a folder on your computer, or other easily-remembered and accessible location. If you followed along in the previous chapter, we created a folder on our computer desktop called "Research Data." To continue to follow along with the next example, place the Excel file called "income" (in Canvas module) into the Research Data folder. This dataset contains median household income values for various Census tracts. We will simply use it as a small dataset to ensure that we have successfully downloaded the R Commander and can call up data. To call up this dataset, in the R Commander menu at the top of the screen, do the following:

1. Click Data → Import data → from Excel file... as seen below. If your dataset was stored in STATA format, you would obviously use the "from STATA file" option.



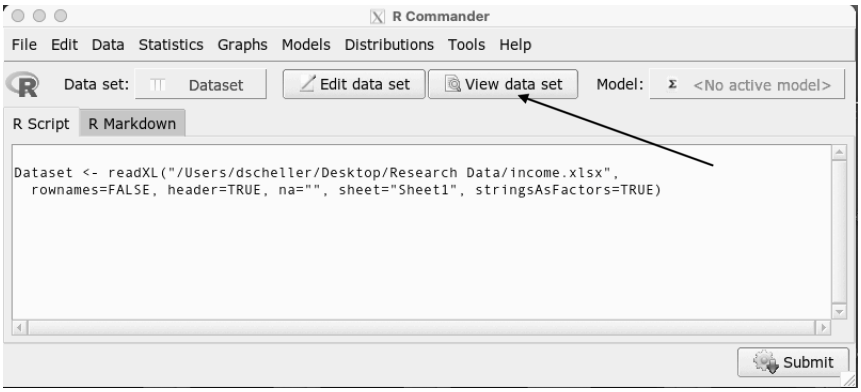
2. The next window will ask you to name your dataset. The default "Dataset" and other defaults should work if your data is organized according to the principles discussed in Chapter 4. Therefore, simply click OK.

3. The R Commander will then bring up a window for you to select the file to upload. If you have followed along and created a "Research Data" file on your desktop, this is the folder for which you will look for the "income" Excel file. In this particular example, the following image depicts the author's screen.

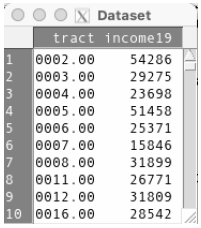


4. Double-clicking this folder should then reveal all files and folders on your Desktop. If you have followed along and created a "Research Data" folder, it should also appear. Double-click its icon. In the folder should be the "income" dataset if you have stored the "income" dataset in this folder.

5. Double-click the "income" file icon. This action should upload the data into the R Commander. However, there will be no bells/whistles or confirmation window. Most likely, you will see some complicated-looking code in the R Commander window. Not to worry. Simply click the "View data set" button in the R commander. The screenshot below shows what you may see, along with an arrow pointing where to click.



6. When you click the "View data set" button, another window should appear with the dataset. Confirm with the image below:



This nondescript window allows you to check that your data is correctly uploaded and allows you you to double-check that it is in a readily-analyzable format for data analysis.