Installing R and RStudio

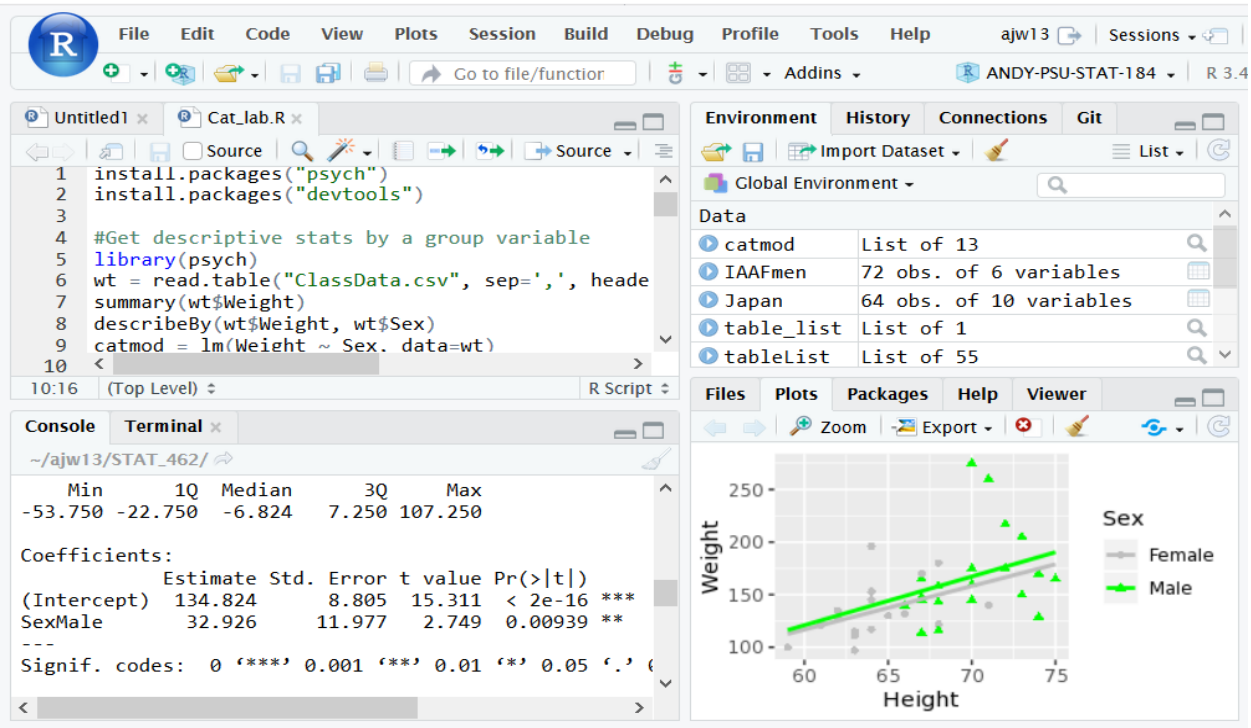
We will use the statistical software program RStudio which runs off the R program. Both need to be installed. You are encouraged to download and install them on your personal computer; each are free. You must install R first (http://lib.stat.cmu.edu/R/CRAN/). Then after installing R, you can install RStudio (https://rstudio.com/products/rstudio/) – choose the free desktop version. Both programs are supported by Windows and MAC machines. **NOTE: I will use the terms R and RStudio interchangeably. Just realize that we will do ALL our work in RStudio so when I reference R, I am referring to RStudio.**

After Opening RStudio

After you open RStudio, the window should show four panes. If only three appear, you are not seeing the R Script pane in upper left – just the console pane. To open this pane, go to File > New File > R Script.

RStudio is a four-pane workspace for 1) creating file containing R script, 2) typing R commands, 3) viewing command histories, 4) viewing plots and more. An image of these four panes is given on top of page 2.

1. Top-left panel: Code editor allowing you to create and open a file containing R script. The R script is where you keep a record of your work. You can also type R commands here which we will do.
2. Bottom-left panel: Console tab: Will display output and can also be used for typing R commands.
3. Top-right panel:
    - Environment tab: shows the list of R objects you created during your R session
    - History tab: shows the history of all previous commands
4. Bottom-right panel:
    - Files tab: show files in your working directory
    - Plots tab: show the history of plots you created. From this tab, you can export a plot to a PDF or an image file
    - Packages tab: show external R packages available on your system. If a listed package is checked, the package is loaded in R. More on this latter, but in brief many functions we/you will use in R are not built into the base R program. To run such functions, you need to install a package that houses such functions. You only need to install a package once, but you need to call for it (this is what "checks" the package) in any R session where you will use that function.



Getting Data into RStudio

Working with data in R is easy…once you get the hang of it! For instance, you can import data from website, call in a data set stored on your computer, or write the data directly into R. We will work mainly with reading data into R from your computer. That is, I will provide you with the data set (usually as a csv or txt file) that you will save locally to your computer where RStudio is installed and then read into RStudio.

This process is aided by knowing about the Working Directory (WD) in R. Think of the WD as the fixed (i.e. home) location of where R stores files. When you start any R session the session defaults to your WD. You can change the location of the WD during a session. In the lower right pane of RStudio click the Files tab you will see the word "Home". To find exactly where this Home working directory is located on your computer you can use the getwd() function in one of two ways:

1. In the Console pane after the '>' type the following and then click Enter: getwd()
2. In the Script pane type in getwd() then with cursor at beginning or end of statement click Run.

My suggestion is to create a new folder in this WD by clicking the tab New Folder in menu ribbon in the s lower right pane. Call this new folder STAT501. This will create a folder with this name inside your Working Directory. Then any data sets, R code used for assignments, etc. related to this course you can save in this folder. This will make working with R much easier as each time you want to use R in this course you can just open RStudio, click on the STAT501 folder, then set this folder as your working directory for that R session by clicking "More" in the lower right panel and select "Set as Working Directory". After setting this, you should see just above the lower right pane something like Home > STAT501.

Once you have this set up, try the following to practice reading a data set into RStudio.

1. In Canvas on the Intro to RStudio page, click and download the two files: ClassData.csv and ClassData.txt. Save both to the STAT501 folder your created in previous steps. This will require you to go to the R Working Directory (e.g. this might be "C:/Users/Documents").

2. In RStudio in the lower right pane click Files > STAT501. You should see the two files downloaded from Step 1.

3. Set this STAT501 as your working directory by clicking the More and choosing Set as Working Directory. You can verify that you have done this by re-running the "getwd()" function. What returns will be the extension from Step 1 plus /STAT501 at the end e.g. "C:/Users/Documents/STAT501"

4. To read in a data set, copy/paste either one of the following codes into the R script window. With cursor at either end of the code, click the 'Run' tab found in ribbon directly above R script pane. This will "call in" the ClassData set into an R object (think data frame) called 'cdt' or 'cdc'**. You will see this object in the upper right pane under the Environment tab with information that it contains 37 observations of 4variables.

```
cdt <- read.table("ClassData.txt", header=T)

cdc <- read.table("ClassData.csv", sep=',', header=T)
```

**These object names are arbitrary – you can name them whatever you want. In my case, I used 'cdt' to represent Class Data from text and 'cdc' for Class Data from csv. There are basically two constraints when naming an object in R: The name cannot begin with a number and there cannot be any spaces in the name. For example, 2data would not be acceptable but data2 would be.

SPECIAL NOTE: Like with most programming, R is case, punctuation, and spelling sensitive!

Saving your R code

If you want to save your R script code in your RScript window (the top left panel), click the Save icon in the ribbon above that panel. The save file directory location that opens will be your current working directory e.g. your STAT501 folder location on your computer. Name the file and click Save. You should then see in the STAT501 folder in RStudio the name of that file with a .R extension.