Review of R

Therri Usher

Thursday, January 29, 2015

Announcements

- Submit verification of completion of the Try R course by Friday at 11:59 EST.

Upcoming events:

- Information session on the Center for Global Health Established Field Placement Program today from 1:30 to 2:30 in Shriver Hall, 104.
- The "Health and Wellbeing in Baltimore" course, in Remsen 1, will have the following guest speakers:
  - Tuesday, February 3rd, 4:30 - 5:45pm - Dr. Gregory E. Thornton, CEO of Baltimore City Public Schools.
  - Tuesday, March 3rd, 4:30 - 5:45pm - Dr. Anthony Batts, Commissioner of the Baltimore City Police Department; and Dwaine Johnson, Baltimore Area Director for the Maryland Department of Juvenile Services.
  - Tuesday, April 28th, 4:30 - 5:45pm - Dr. Leana Wen, Commissioner of the Baltimore City Health Department.

R

What is R?
"A software environment for statistical computing and graphics."

Why do we use R?

- It's FREE.
- It's open-sourced.
- It's growing in popularity.
- It's good at what it does.

How Can I Learn R?

- Practice, practice, practice
- Academic courses (JHSPH's Statistical Computing course)
- Training modules
  - DataCamp
  - Code School
- Coursera
- Helpful websites

What Helpful Websites?

- R and RStudio's websites
- UCLA's R Website
- R Tutor
- Quick R
- R's Wikibook

And last but not least,

- Google!

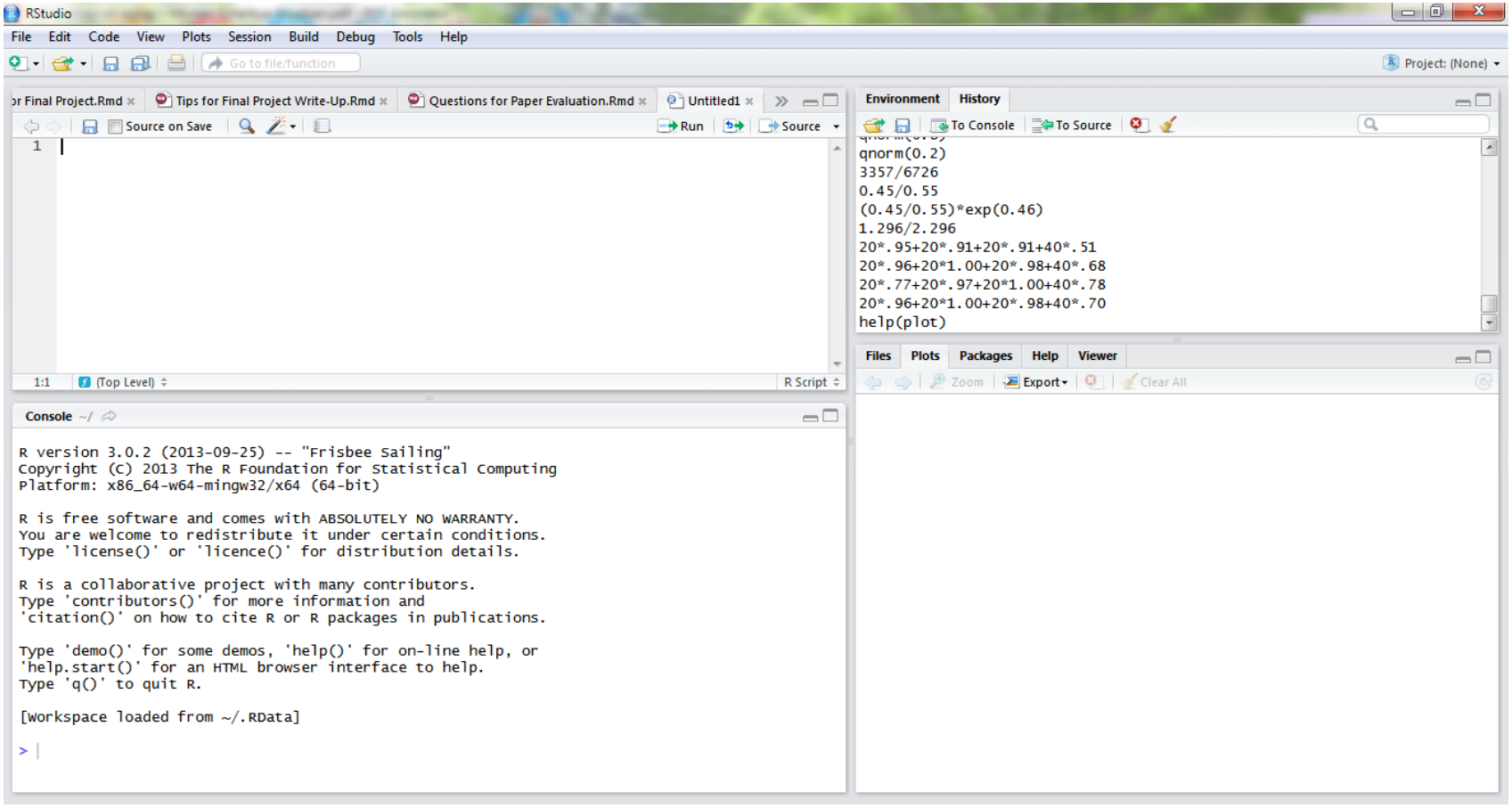
RStudio

RStudio is one of many user interfaces for R.

Why use RStudio?

- It's FREE.
- It's popular.
- It combines R with other programs, such as LaTeX, knitr, and Shiny.
- The interface is very organized.

RStudio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Project: (None)
Final Project.Rmd
Tips for Final Project Write-Up.Rmd
Questions for Paper Evaluation.Rmd
Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Environment
History
To Console
To Source
qnorm(0.2)
3357/6726
0.45/0.55
(0.45/0.55)*exp(0.46)
1.296/2.296
20*.95+20*.91+20*.91+40*.51
20*.96+20*1.00+20*.98+40*.68
20*.77+20*.97+20*1.00+40*.78
20*.96+20*1.00+20*.98+40*.70
help(plot)
Files
Plots
Packages
Help
Viewer
Zoom
Export
Clear All
Console ~/
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
>

RStudio Tips

- Take advantage of the interface. Use it, particularly while you are learning the code.
- Keyboard shortcuts can be useful. For instance, to run code from the editor, highlight it then press Ctrl + Enter.

Today's Activity

You will be completing Code School's Try R course. Be sure to create a free account before starting the course. There are 8 chapters to the course but it should not take very long. Once you have completed the course, go to your report card and print it as a PDF. Submit the report card showing that you have completed the course through Blackboard to earn full credit. If you do not finish the review in the time allowed, complete it as homework.

DUE DATE: January 30th, 2015 at 11:59 PM EST

Additional Info: Installing and Using Packages

Not every command you will need will be offered in the basic R packages.

To download an R package:

- Give the command install.packages("packagename") OR
- go to Tools, Install Packages. . . , enter package name, then click Install (make sure that "Install dependencies" is checked)

To load the R package for use:

- Use the commands library(packagename) OR require(packagename)

Then, you can proceed to use commands that are in the loaded package.

Additional Info: Help

To access the help page for a command, use the following commands:

- ?commandname
- help(commandname)

If you want a more general search, use

- ??searchterm

Additional Info: Directories and Files

- To find out the current directory: getwd()
- To set the working directory:
  - setwd(path)
  - go to Session, Set Working Directory, Choose Directory...
  - Press Ctrl + Shift + H
- Downloading a file from the internet: download.file(url)
- List files in a directory: list.files()

Additional Info: Opening Data

- Read data from a txt file: read.table(filename, header)
  - header = TRUE if the data set has a header of variable names
- Read data from a csv (Excel) file: read.csv(filename, header)
- Read data from an xlsx (Excel) file: read.xlsx(filename, sheetnumber)
  - requires R package xlsx
- Read data from a Stata dta file: read.dta(filename)
  - requires R package foreign
- Load available data sets in R: data(dataname)
- Load .Rdata files: load(filename)

Additional Info: Saving Data

- Save data in a txt file: write.table(x, filename, sep)
  - x is the matrix or data frame you want to save
  - filename is the name you want to give the file
  - sep is the character used to separate the data (ex: “,” or " “)
- Save data in a csv file: write.csv(x, filename)
- Save data in an xlsx file: write.xlsx(dataframe, filename)
- Write files in Stata dta format: write.dta(dataframe, filename)
- Creating .Rdata files: save(list, filename)
  - list is the list of variables you want to save

Tips For Using R

- If you get stuck, seek help!
  - Google is your friend.
  - Use the previously listed websites
  - I am here to help
- Have a good filing system. Keep related things in the same directory.
- Remember to comment your code! Use "#" at the beginning of a line of code to indicate that it is a comment. It will be useful if you ever need to look at the code some time from now - you will remember why you did what you did.
- It takes time to master R. Be patient with yourself. But be sure to practice.

Examples

1. Download the test.dta file from Blackboard. Load the dta file into R and assign it to the variable name test.data.
2. Create a sequence of numbers from 1 to 100 and assign it to the variable name x. Save the variable in a .Rdata file to your desktop.

Feel free to work together but make sure you understand how to do these things. Use any of the resources we have discussed today.

Next Class: Reproducible Research

- I will be posting resources on Blackboard before the next class. Be sure to take a look at them.
- Check out the following reproducible research tools:
  - R Markdown
  - GitHub
  - GitHub's Wikipedia page
- As always, feel free to share interesting resources on Blackboard!