Intro to R

Lecture 3
BIOS 6660, Spring 2019
Instructor: Pam Russell

https://rforcats.net/

Homework 1 grading

- Grades and comments will be in Canvas
- Mostly just verifying tool setup
- Will check accuracy of commands for Problem 2
  - Solution available on Canvas later today

Overall homework grading

First 5 assignments are frequent and worth 6.7% each

Later assignments less frequent

- Homework 6-9: 8.4% each
- Homework 10-12: 11% each

Yampa/GitHub issue

- Fix is in email from 1/28
- To make the fix permanent, issue this command:
  - `echo “unset SSH_ASKPASS” >> ~/.bashrc`
- Otherwise you will have to type `unset SSH_ASKPASS` every time you want to push to GitHub

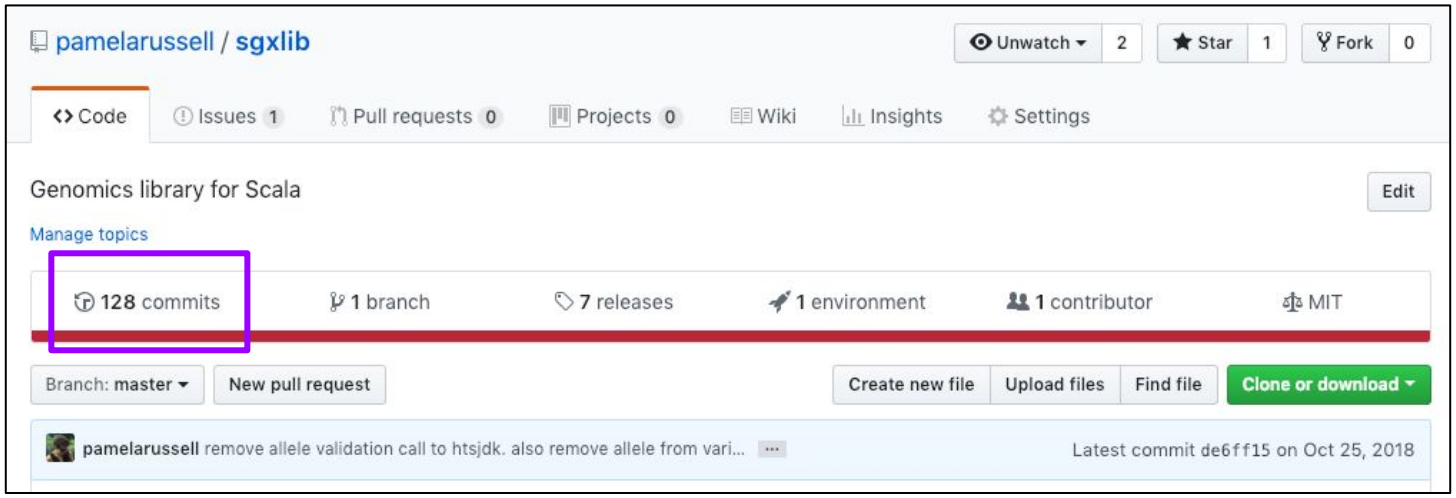
`~/.bashrc`

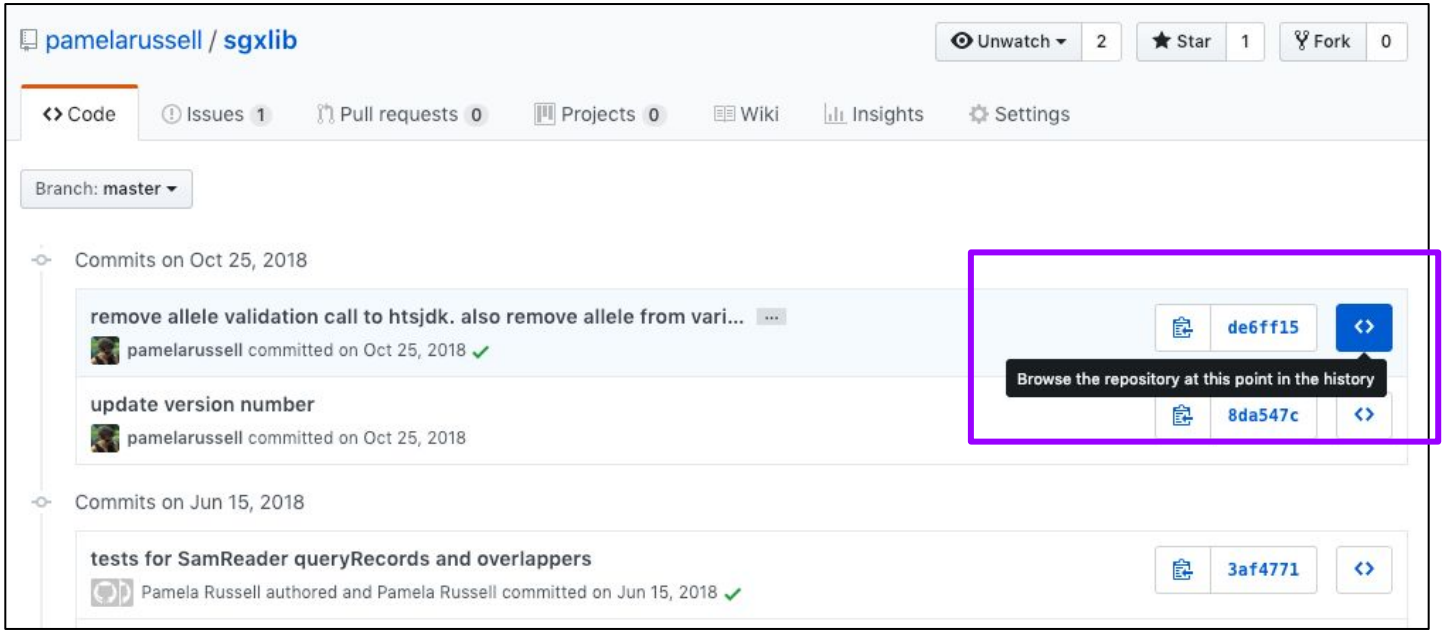
- The shell is a programming environment
- Bash is a language
- Your commands are Bash commands that modify the shell environment
- `~/.bashrc` is a script that is run every time you start the shell

Homework 2

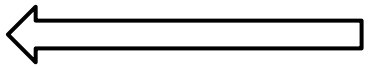
- Set up R coding environment
- Write an R script
- Add files to Git and GitHub
- Submit a GitHub URL through Canvas

Homework 2: Submit GitHub URL

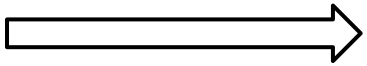




Misc recommendation: moving cursor on command line



ctrl + a
Jump to beginning of line



ctrl + e
Jump to end of line

ctrl + u
Clear command line

Misc recommendation: preprints

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

Search

Advanced Search

Subject Areas

All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics
Cancer Biology
Cell Biology
Clinical Trials
Developmental Biology
Ecology
Epidemiology
Evolutionary Biology
Genetics
Genomics
Immunology
Microbiology
Molecular Biology
Neuroscience
Paleontology
Pathology
Pharmacology and Toxicology
Physiology
Plant Biology
Scientific Communication and Education
Synthetic Biology
Systems Biology
Zoology

What is R?

- Free and open source
- Programming language and environment
- Statistics and graphics
- Thousands of packages
- Reproducibility with R Markdown
- RStudio: powerful development environment

R (S) philosophy

"We wanted users to be able to begin in an interactive environment, where they did not consciously think of themselves as programming. Then as their needs became clearer and their sophistication increased, they should be able to slide gradually into programming, when the language and system aspects would become more important."
- R. John Chambers, one of the creators of S

For more history and background on R

An excellent article and video on the history and many sides of R

(by Roger Peng, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health)

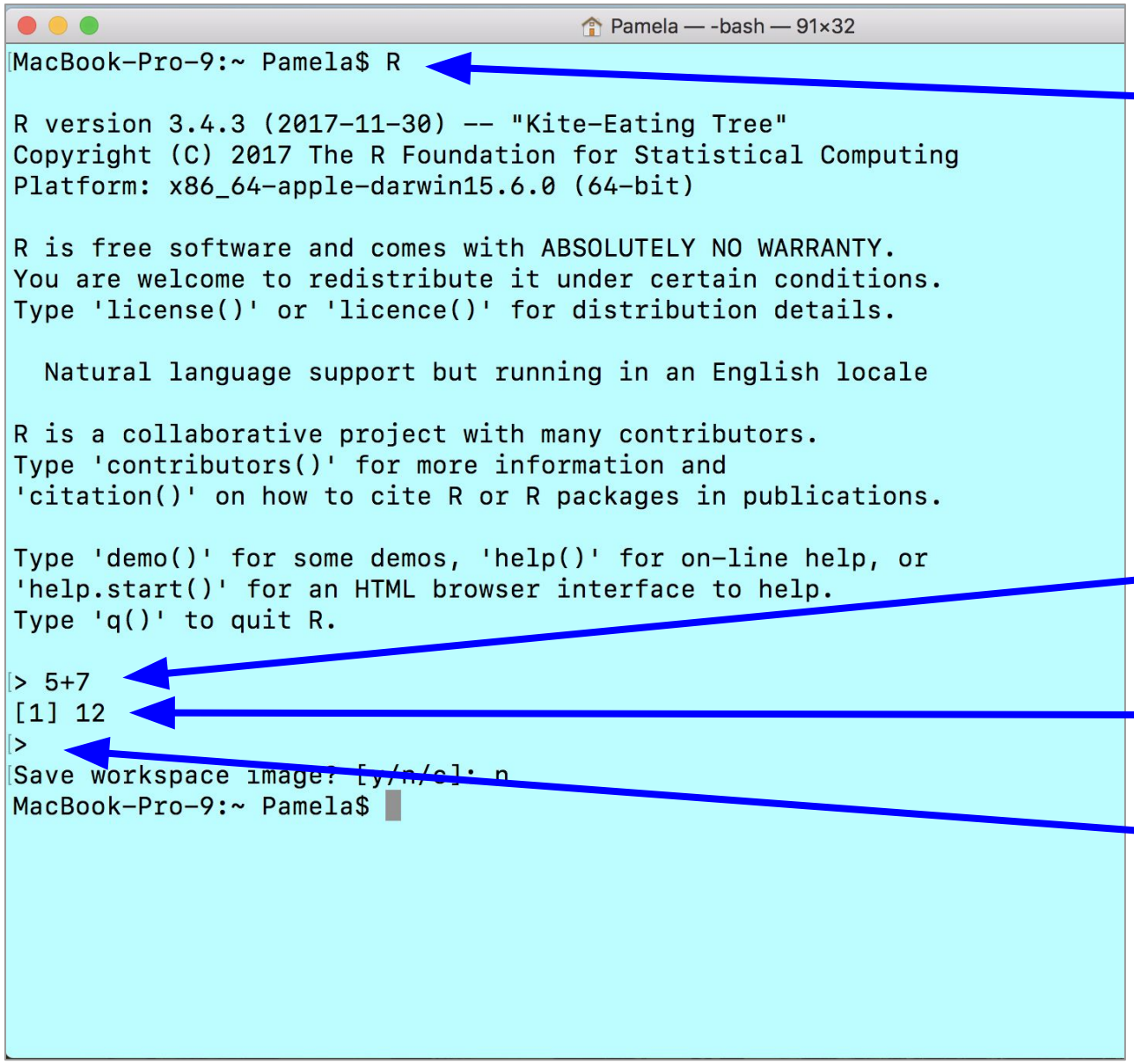
Ways to use R

Interactive console: type lines of code at a command prompt and see results in the console

Scripts: save code in a file and run it all at once

R Markdown: generate reports that mix code, results, graphics, and text explanations

R console



Start R console

Type a command and press enter

Command output

`Ctrl+D` to exit console

RStudio

A powerful free integrated development environment (IDE) designed specifically for R

Standard IDE features

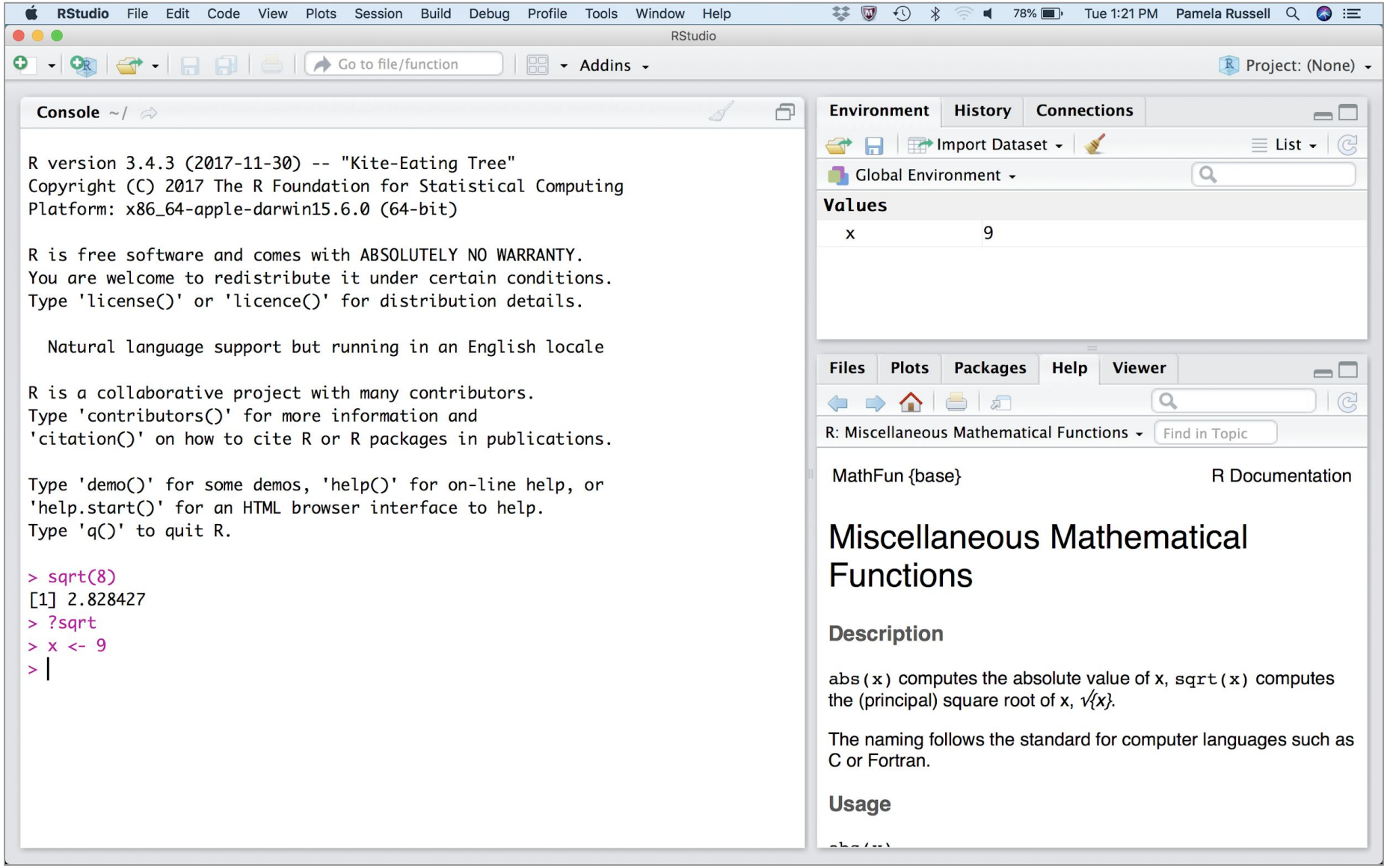
- Syntax highlighting
- Code completion
- Smart indentation
- Jump to function definitions

Special RStudio features

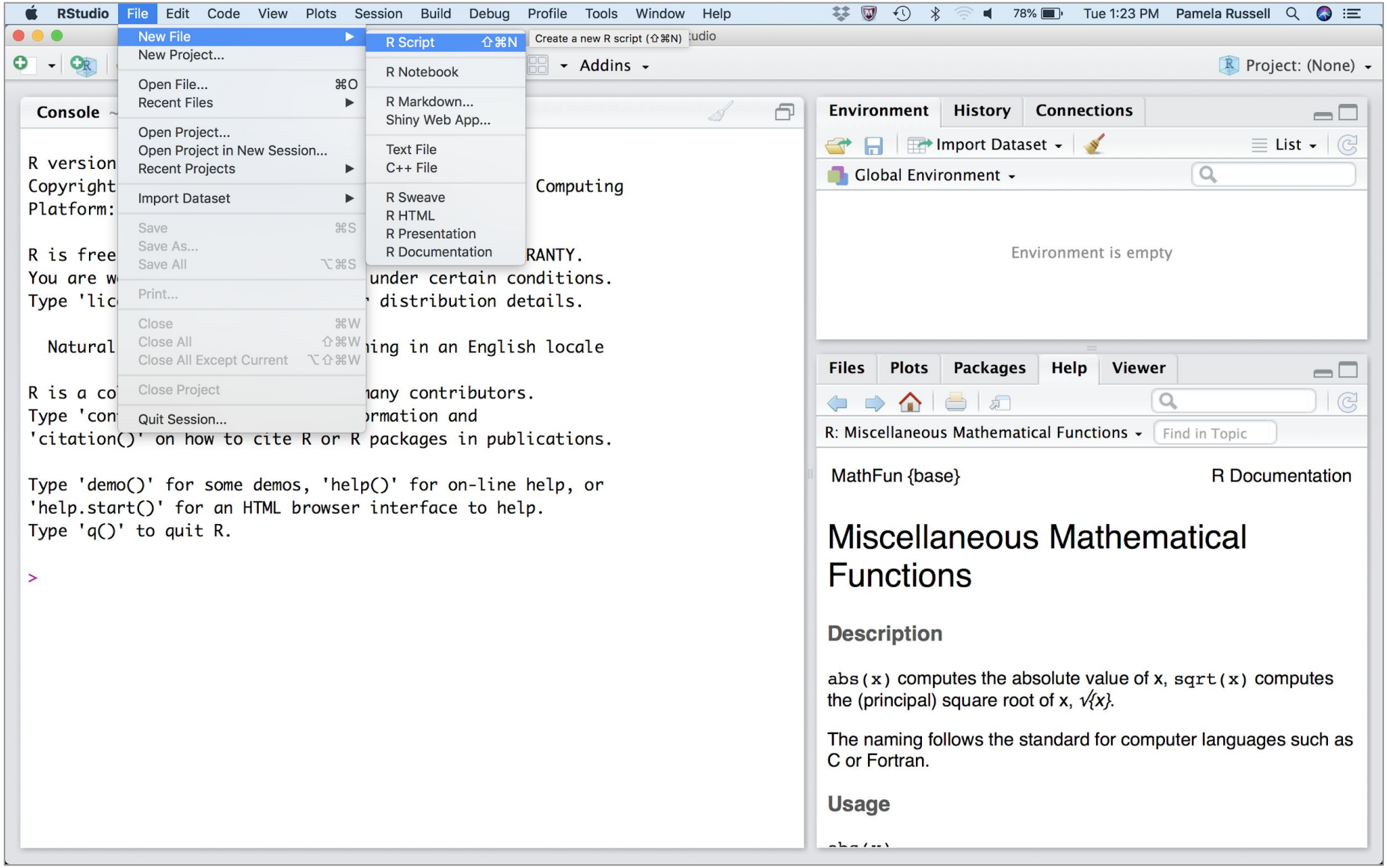
- Integrated console
- Data viewer
- Integrated plotting
- Execute code directly
- Integrated documentation

https://www.rstudio.com/products/rstudio/features/

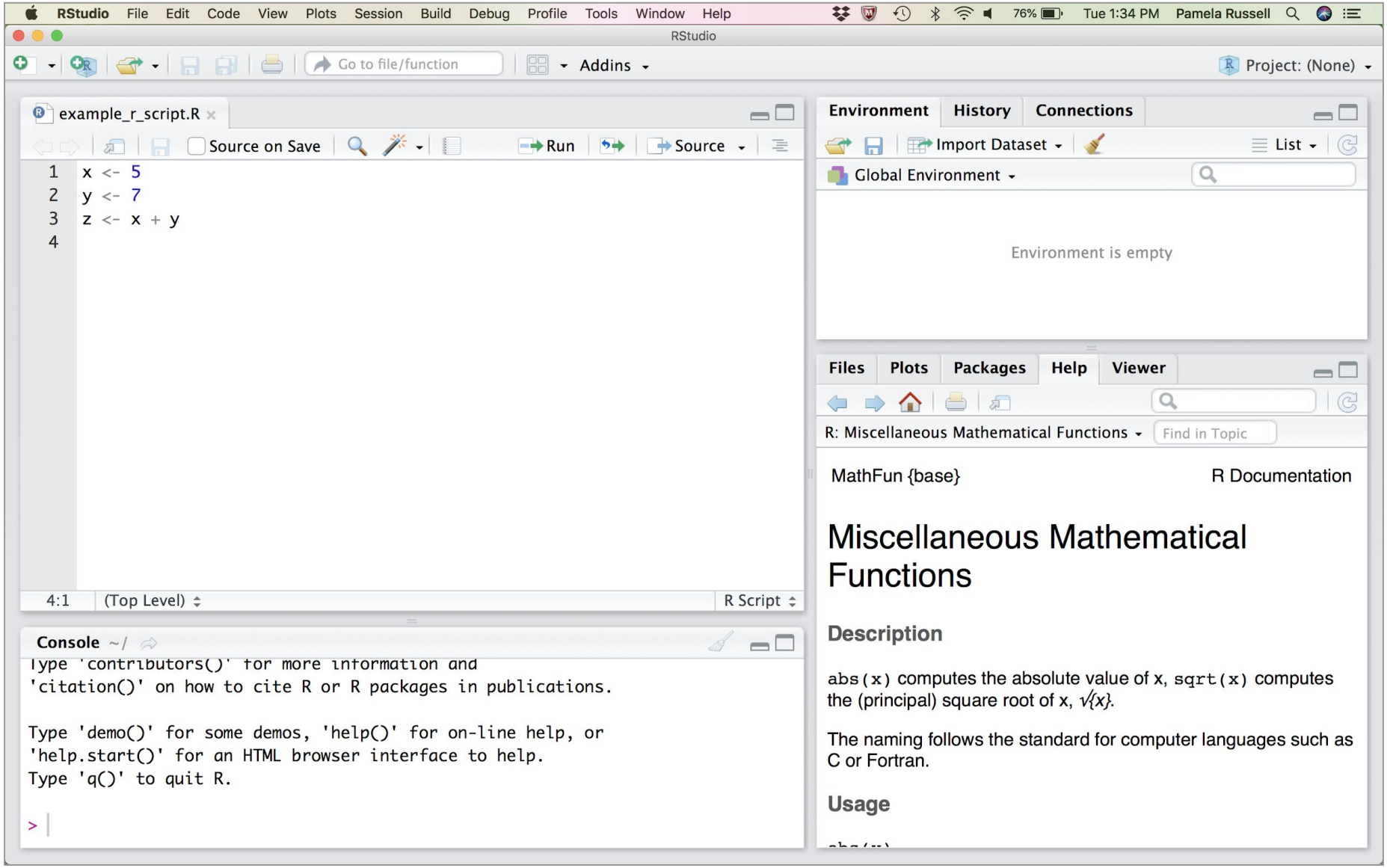
R console in Rstudio



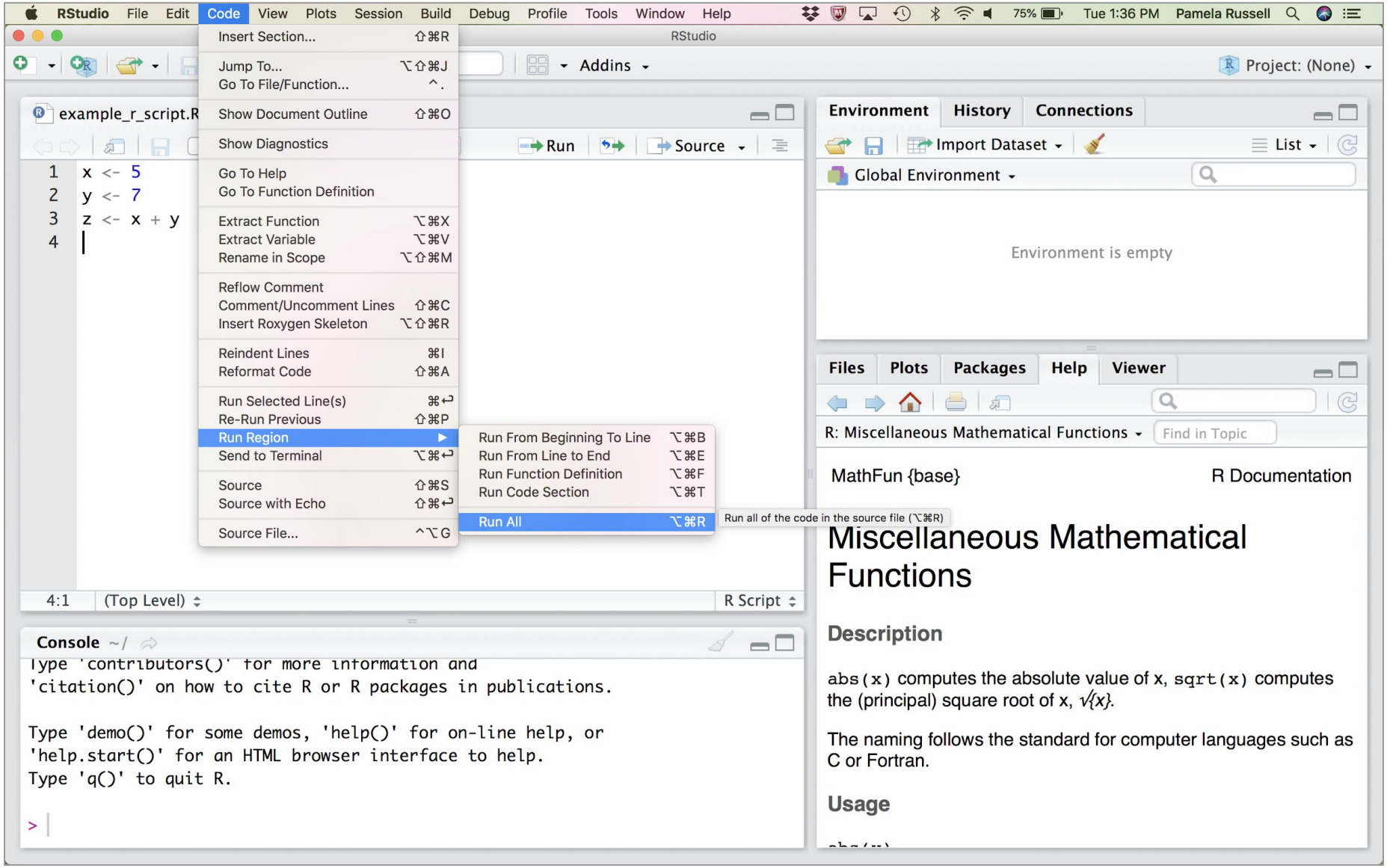
Saving commands as a script



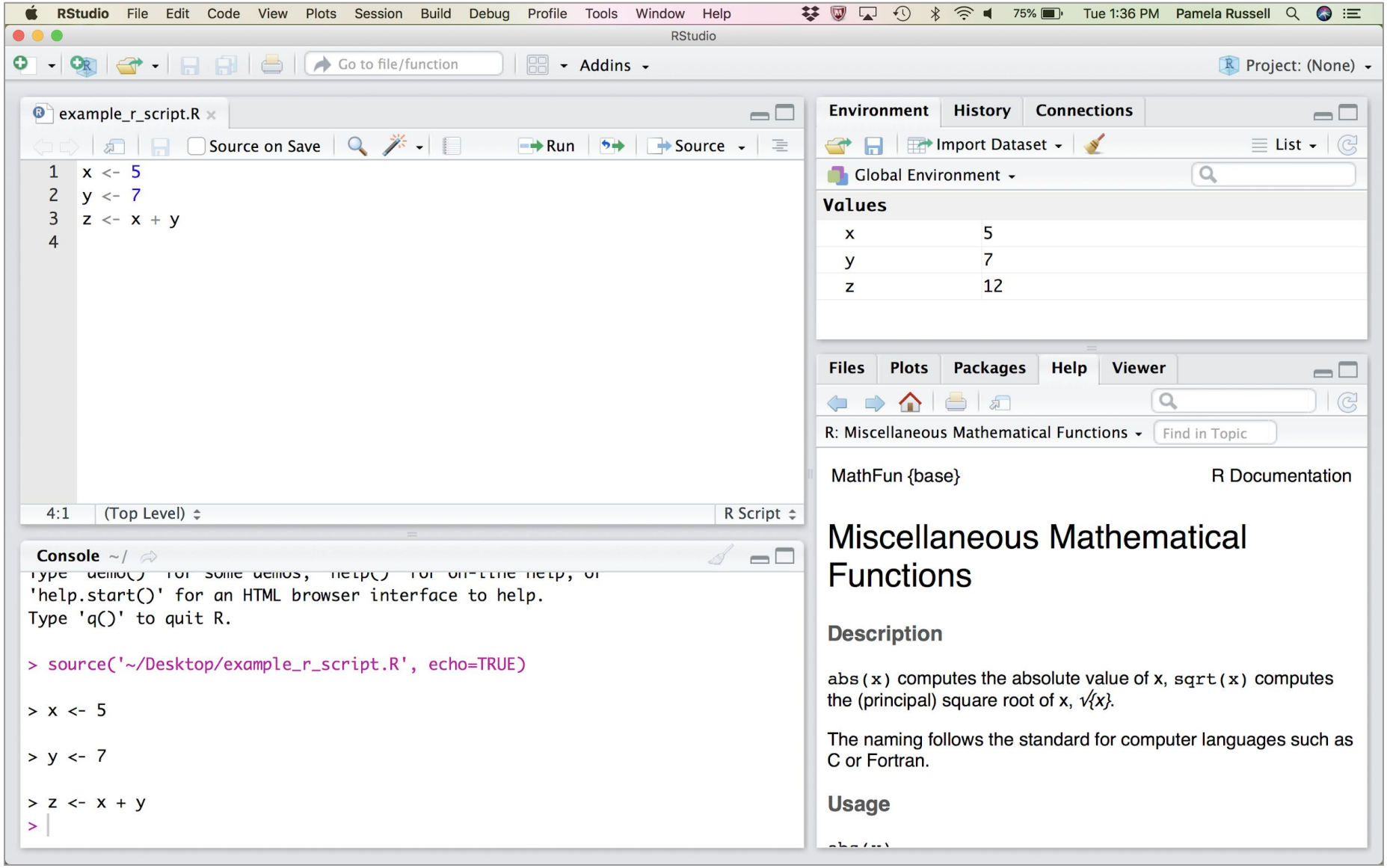
Saving commands as a script



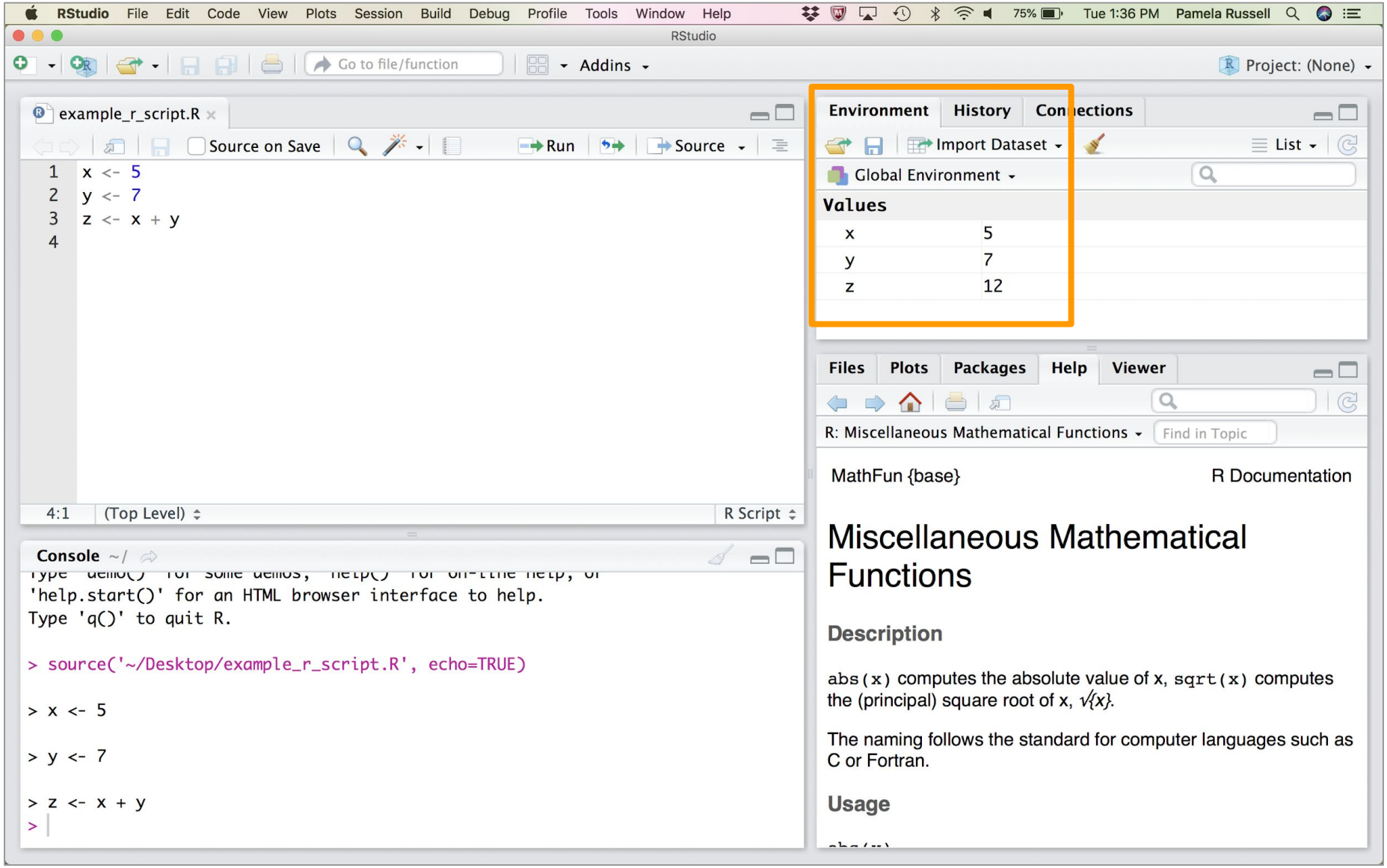
Running a script from within Rstudio



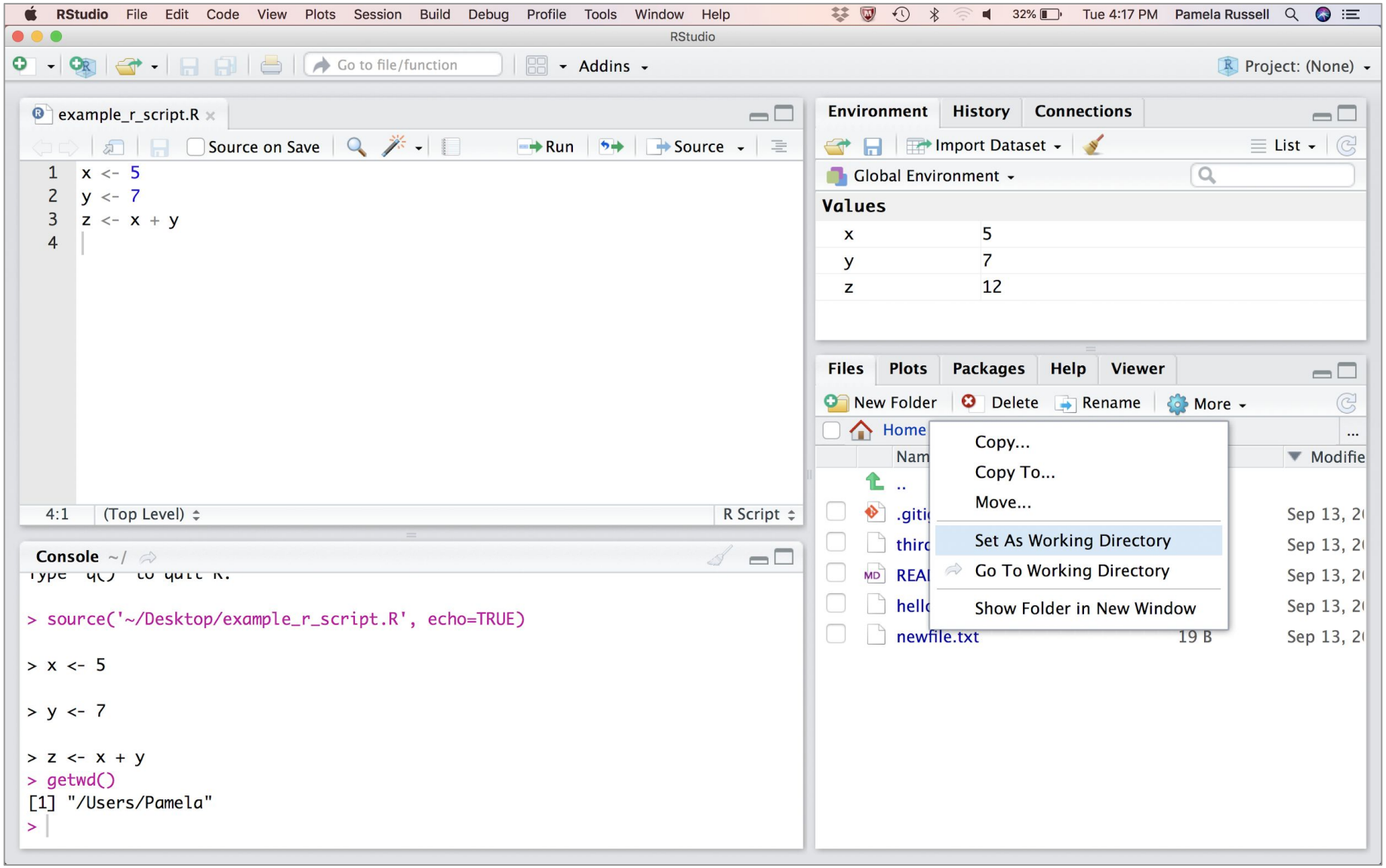
Running a script from within Rstudio



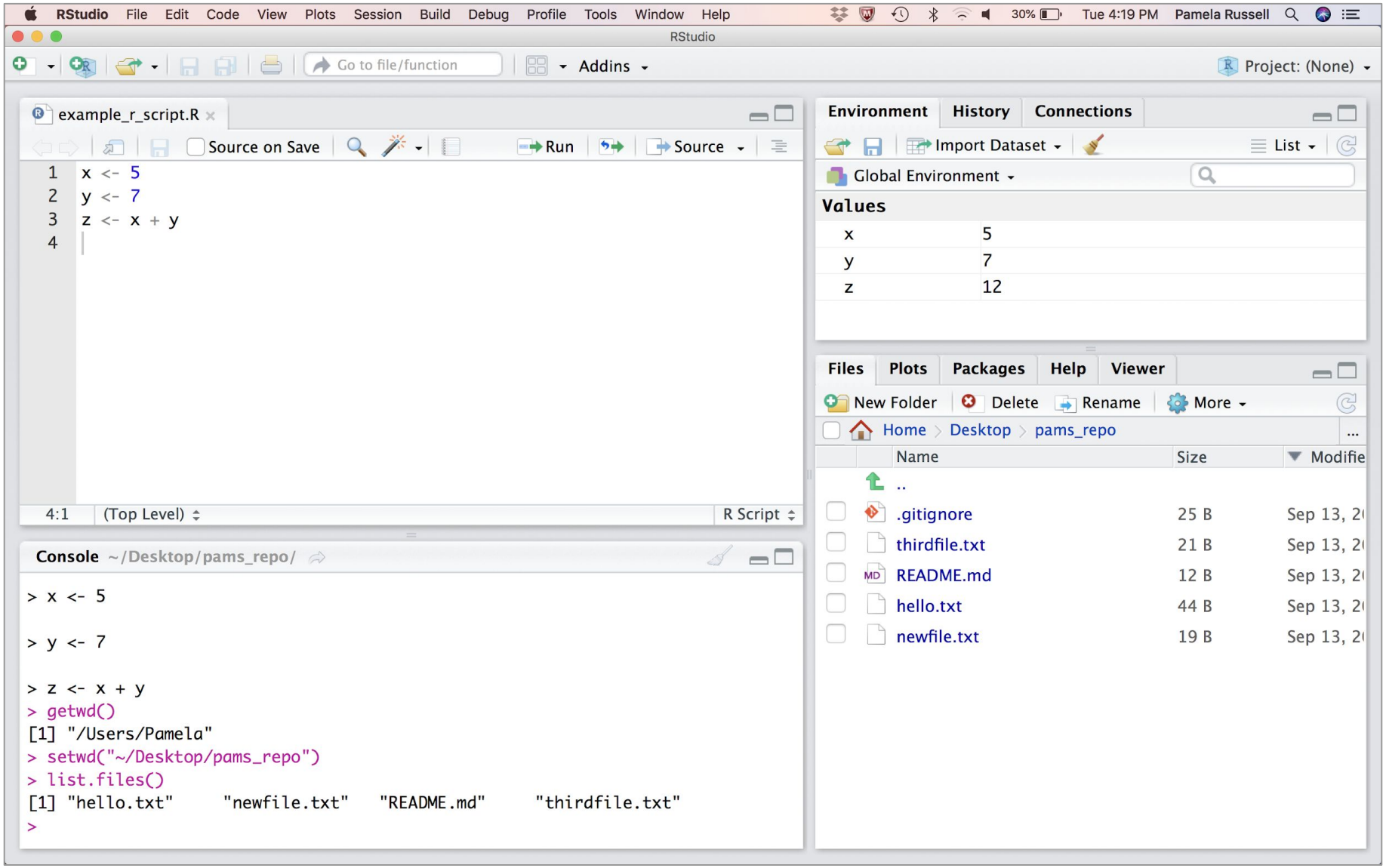
R session environment



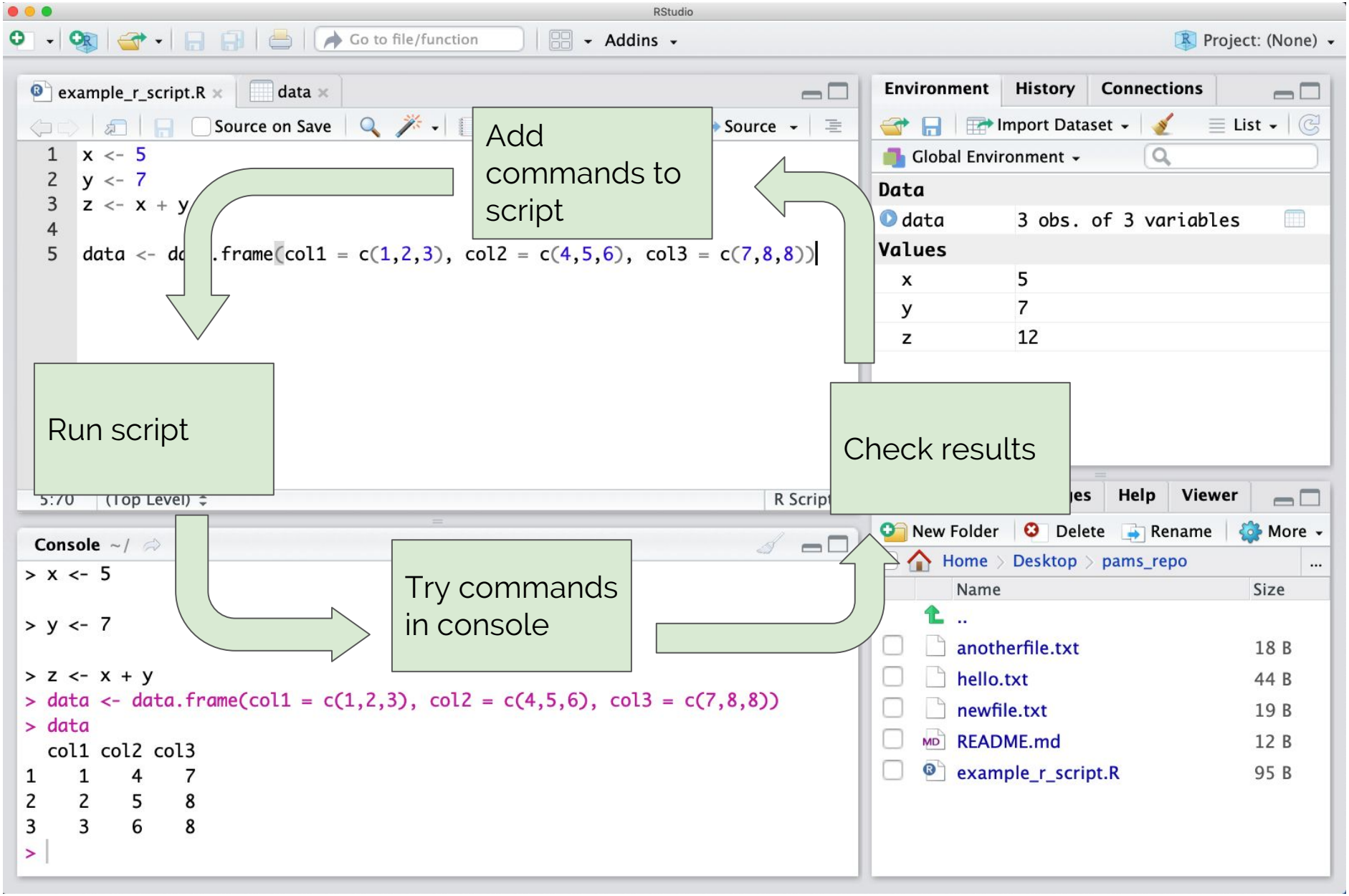
Working directory



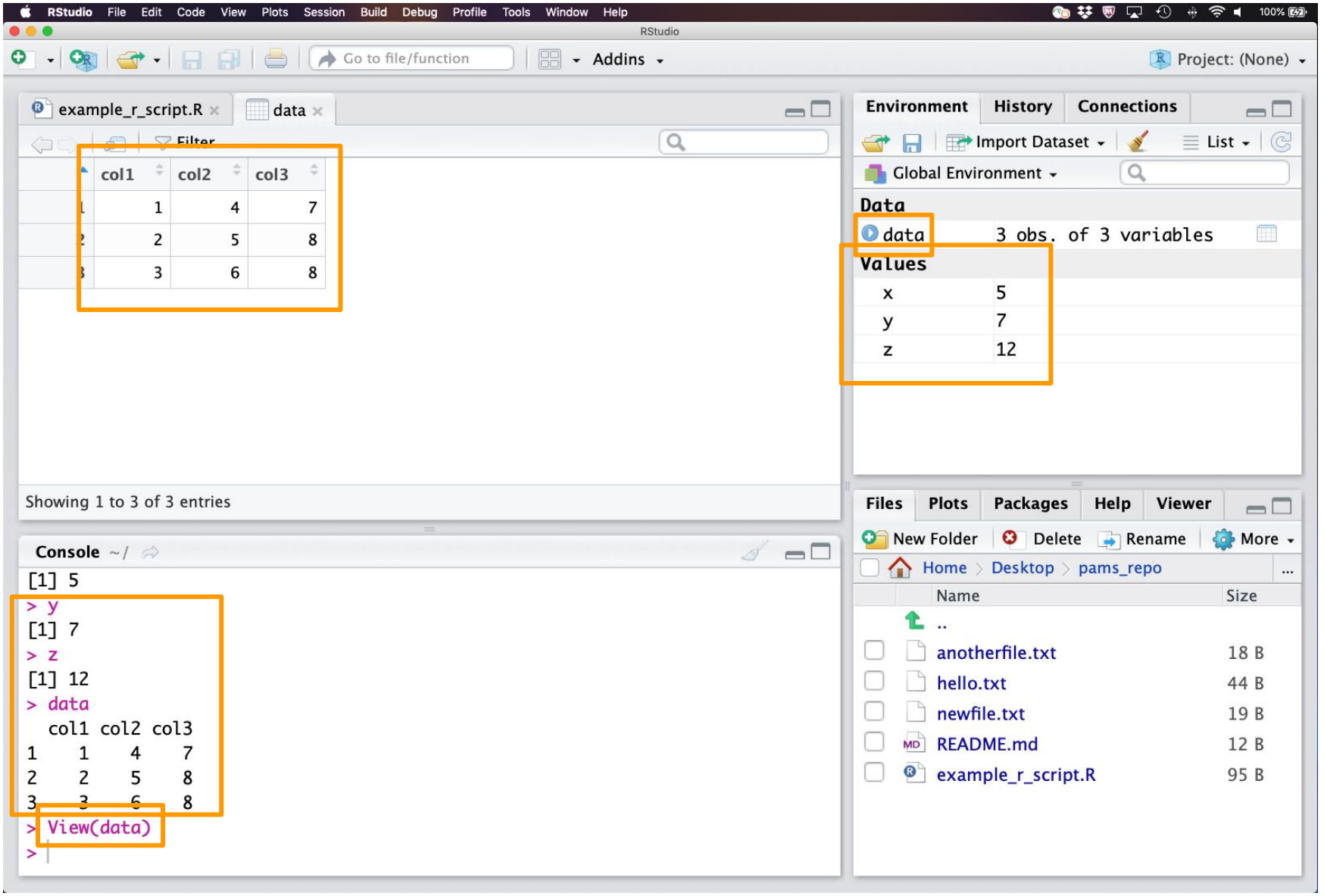
Working directory



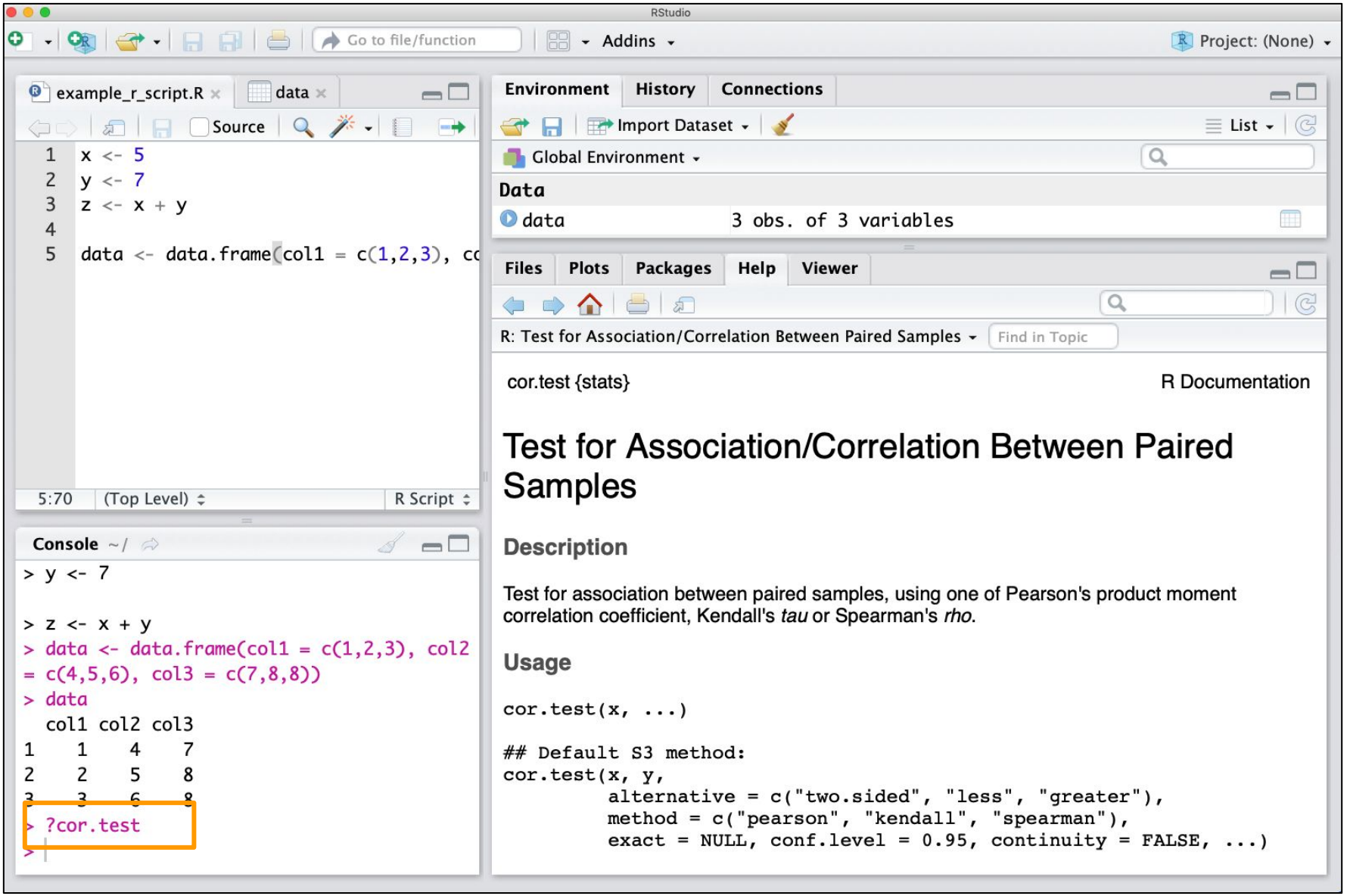
Development cycle



Viewing variable values



Function documentation

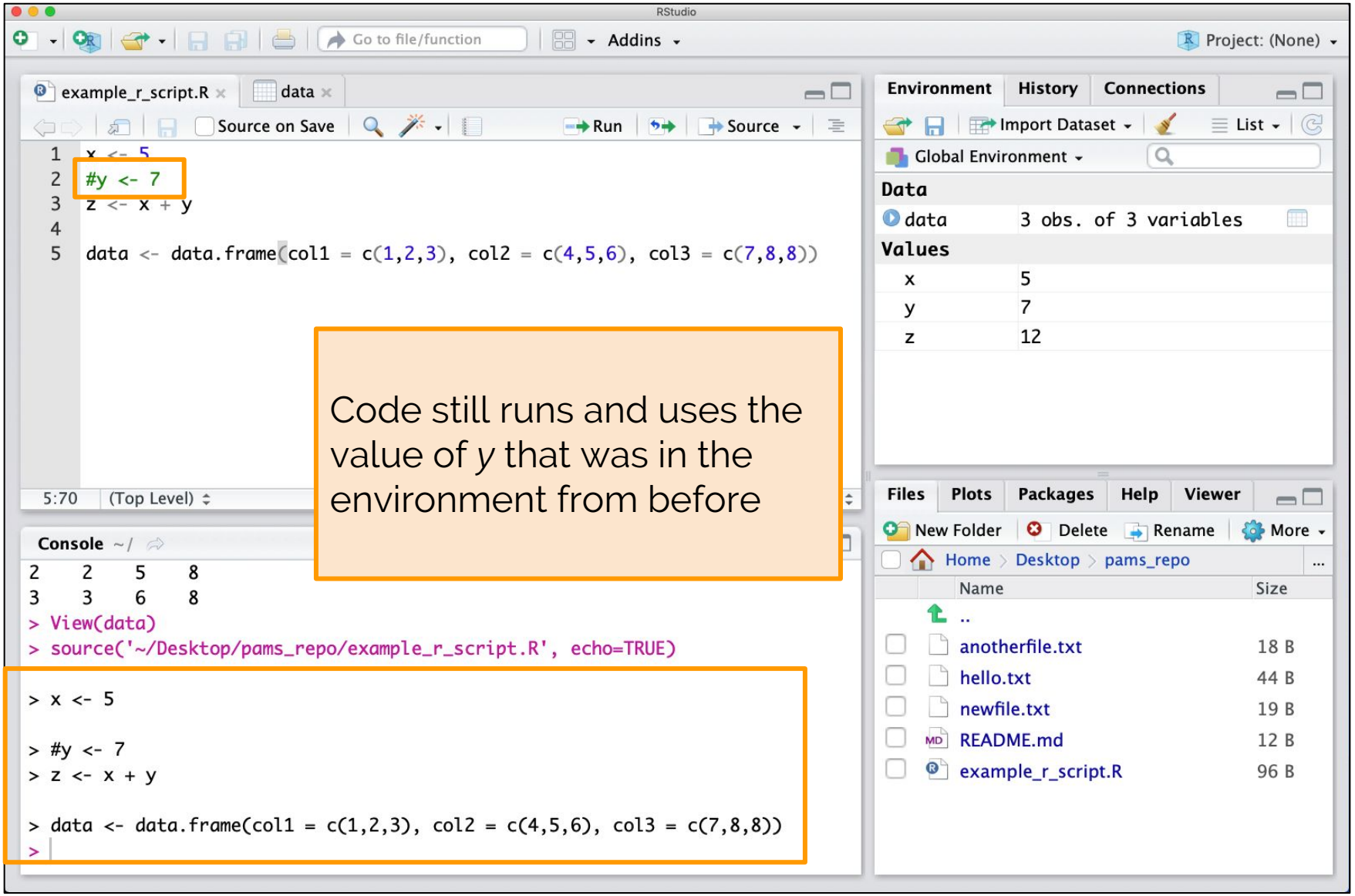


Global environment persists throughout R session

RStudio always has an active R session in progress with its own global environment (variable values)

When running script within RStudio, watch out for variables left over in the environment from previous runs/commands.

Global environment persists throughout R session



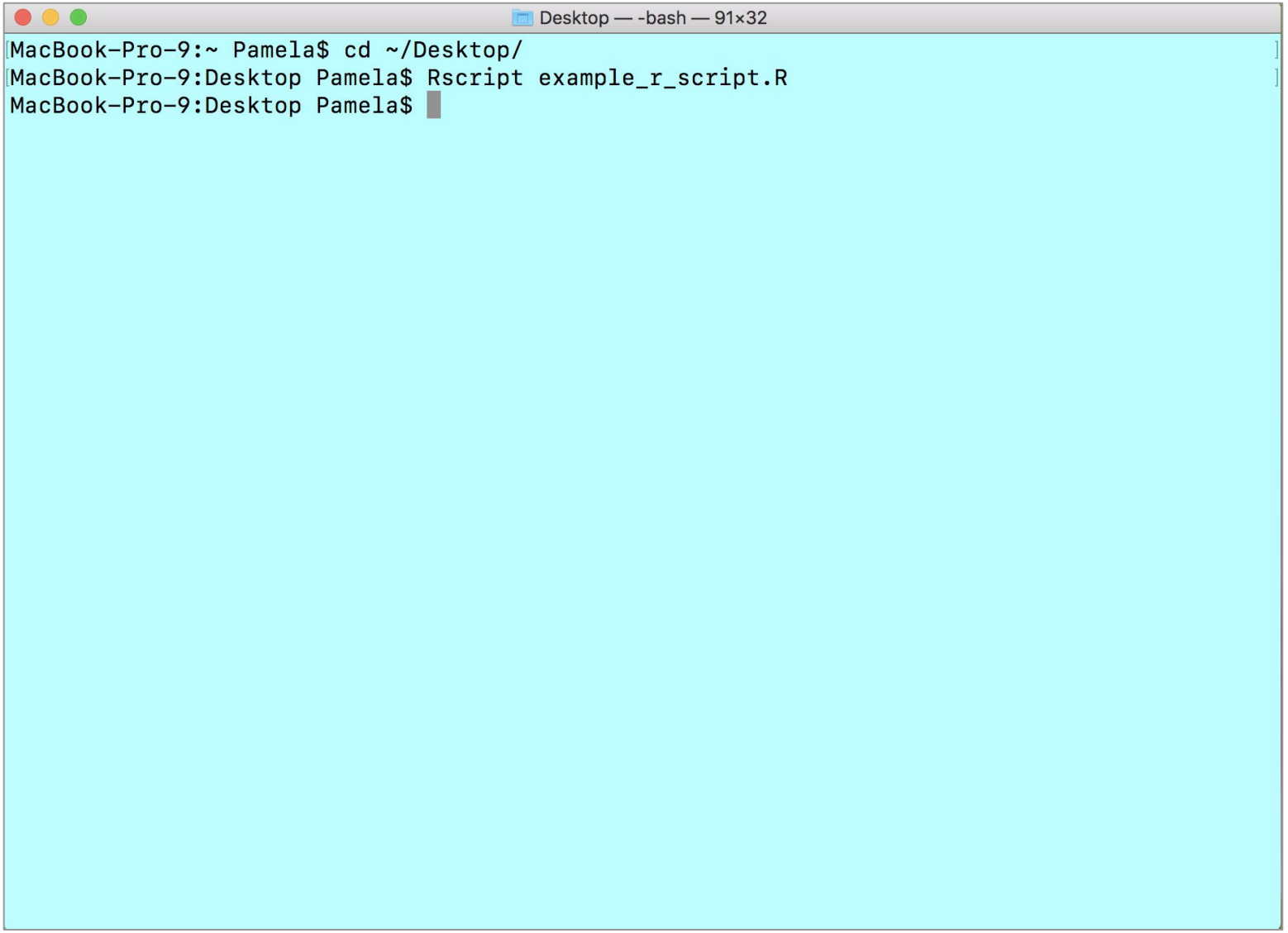
Use Rstudio locally; use Git/GitHub to move scripts back and forth to Yampa

Last lecture: how to use Git and GitHub to work on the same files from multiple computers

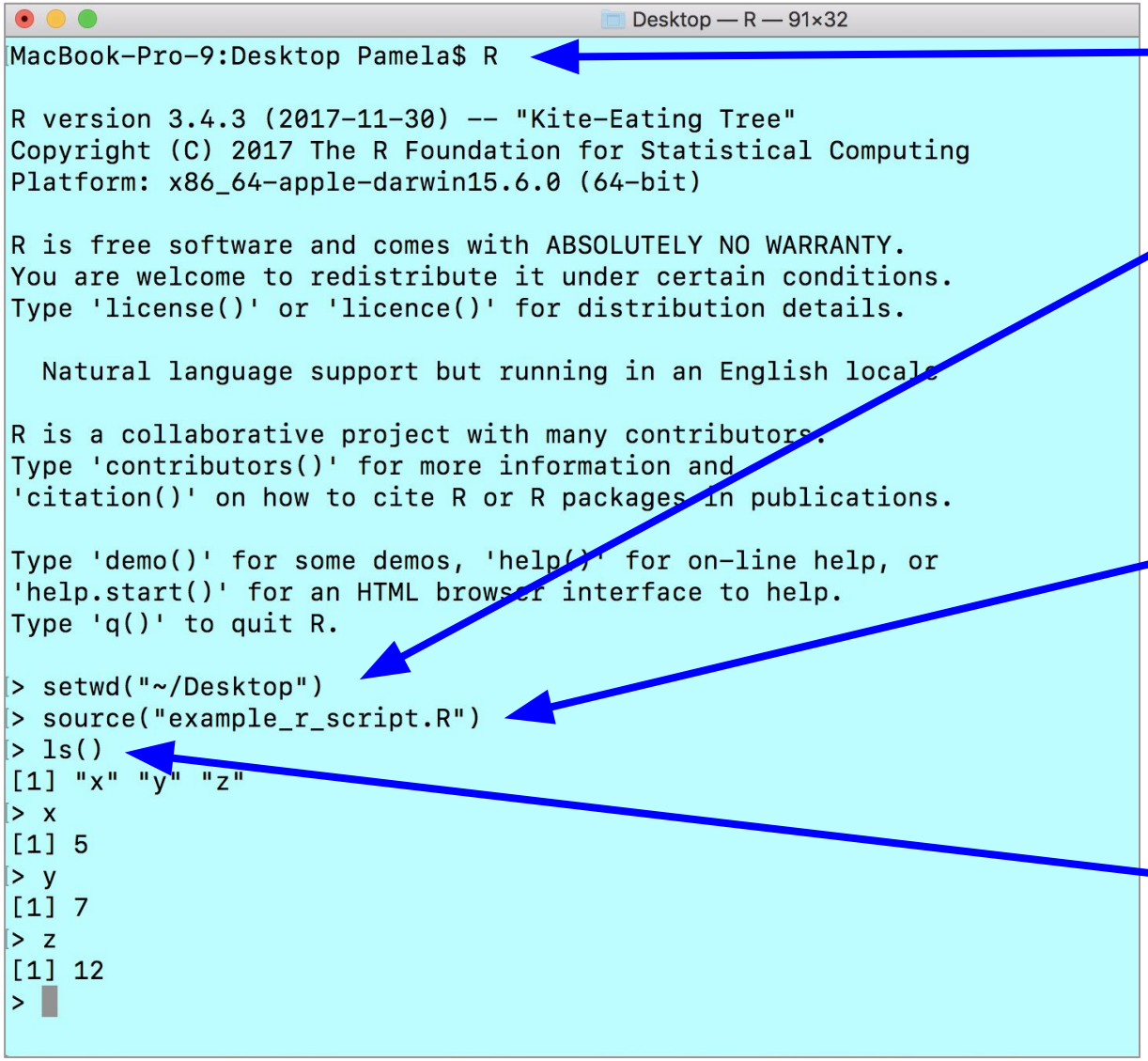
For R scripts that you want to run on Yampa: you should still write them on your computer with Rstudio.

Add scripts to GitHub from your computer, then pull them down from Yampa.

Running a script from the command line



Running a script from the console



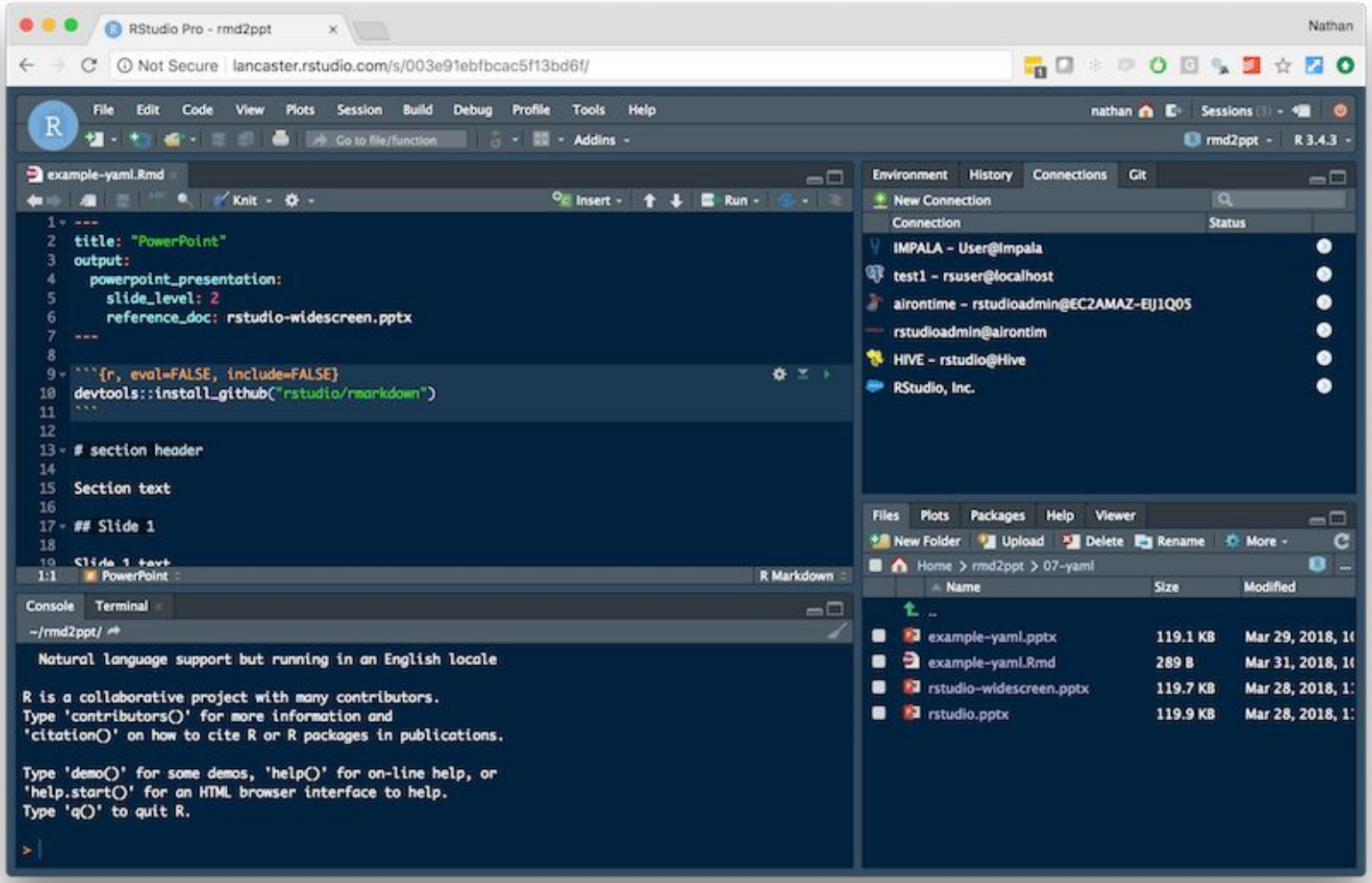
Start R console

Set working directory to the directory containing the script

The `source` command loads a script and runs all of the code

The `ls` command lists variables in the environment. The code in the script has added variables to our environment.

RStudio Server: RStudio on Linux server accessible by web browser



https://docs.rstudio.com/products.html

R variables and handling data

Slides:

L3_intro_to_R.html