

# Introduction to R and RStudio

Instituto Nacional de Medicina Genómica

Daniel Rivas & Juan Manuel Mejia Arangure

August 11, 2023

# Outline - R basics

- Concepts
- Motivation
- RStudio
- Graphics
- Third-party Libraries

# Concepts

# Computer Programming

Computer programming is the process of performing particular computations, usually by designing and building executable computer programs.

# Programming language

A programming language is a system of notation for writing computer programs.

Most programming languages are text-based formal languages, but they may also be graphical.

# R

- Programming language
- For Statistics, Data Analysis & Graphics
- Free & Open-source
- Refers to both: the language & the interpreter

# RStudio

- User interface for working with R
- Wrapper around the R language
- Extends what R can do & facilitates writing R code
- Free & Open-source

# Motivation

# Why R?

- Rely on written commands instead of point & click
- Great for reproducibility
- Interdisciplinary & extensible
- Works for data of all shapes & sizes
- Amazing graphics!
- Large & welcoming community
- Lots of packages
- Open-source & cross-platform

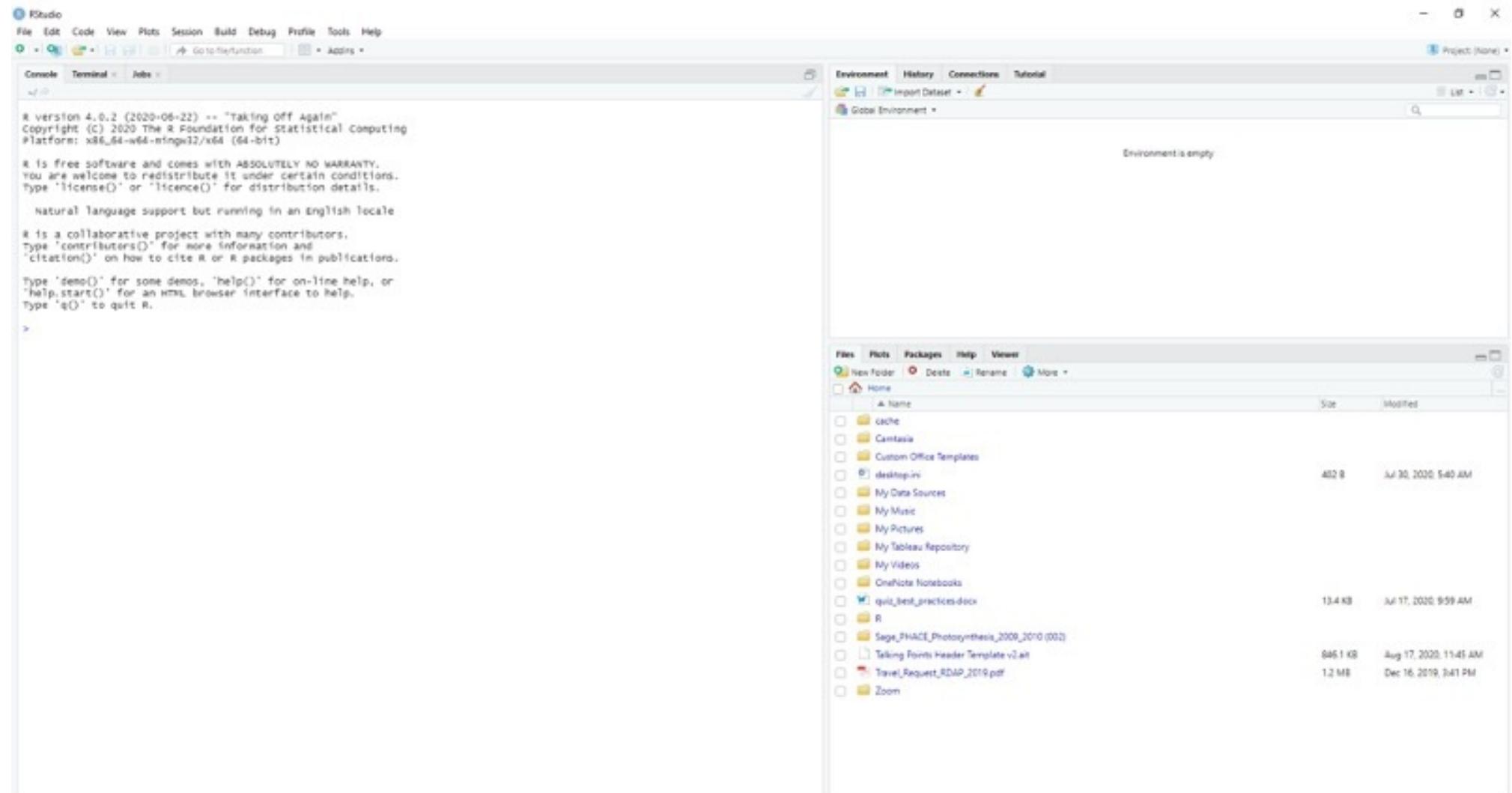
# Alternatives to R

- **Stata** (proprietary)
  - Most popular statistical software in economics
  - Easy to use for standard methods
  - Not a good programming language
- **Matlab** (proprietary)
  - Numerical programming environment
  - Matrix based

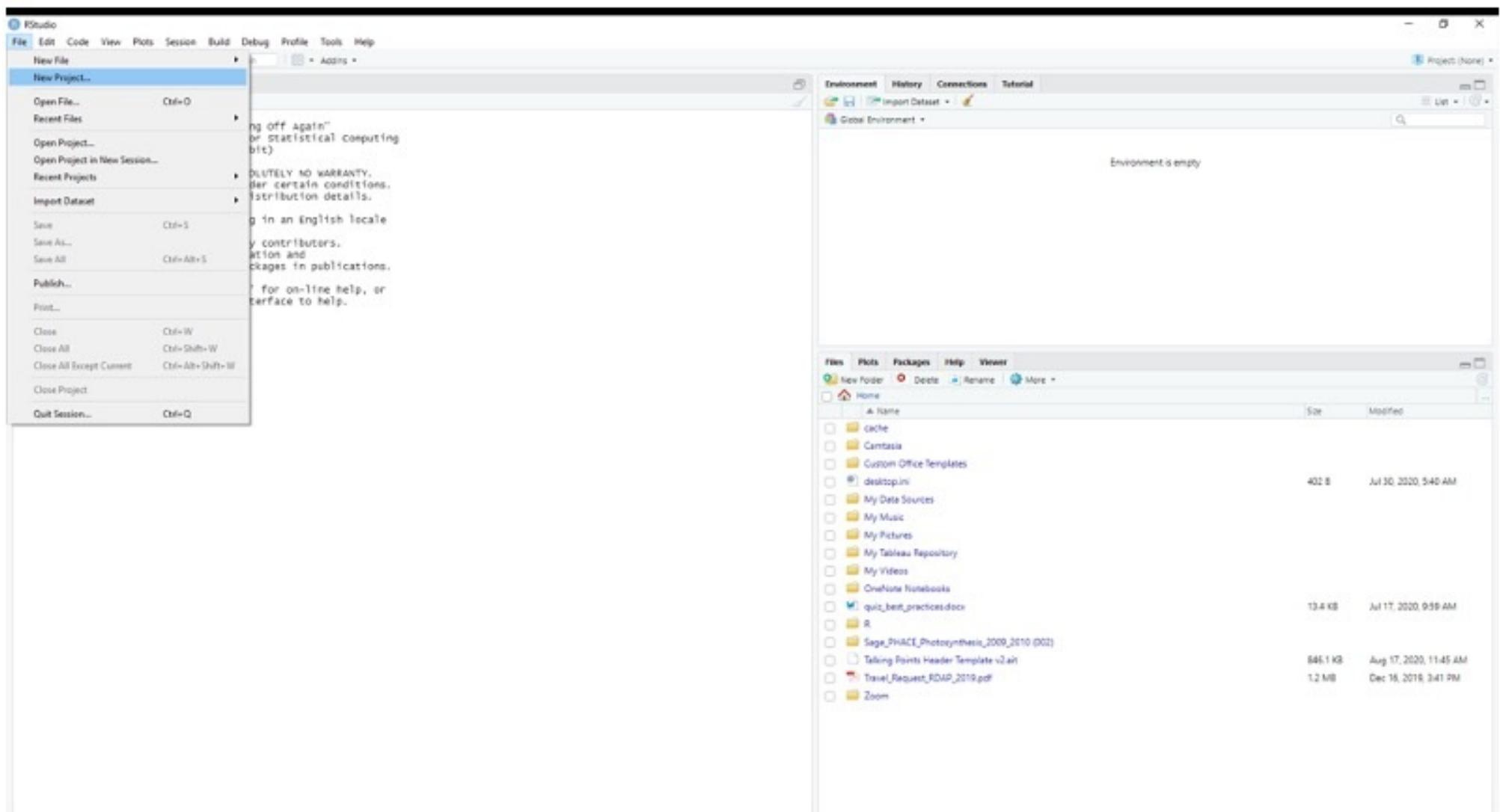
# Alternatives to R

- Python (open-source)
  - General purpose programming language
  - Standard in industry
  - Not targeted toward data analysis and statistics
- Julia (open-source)
  - New language for numerical programming
  - Fast
  - Increasingly popular in macro / for solving complicated structural models

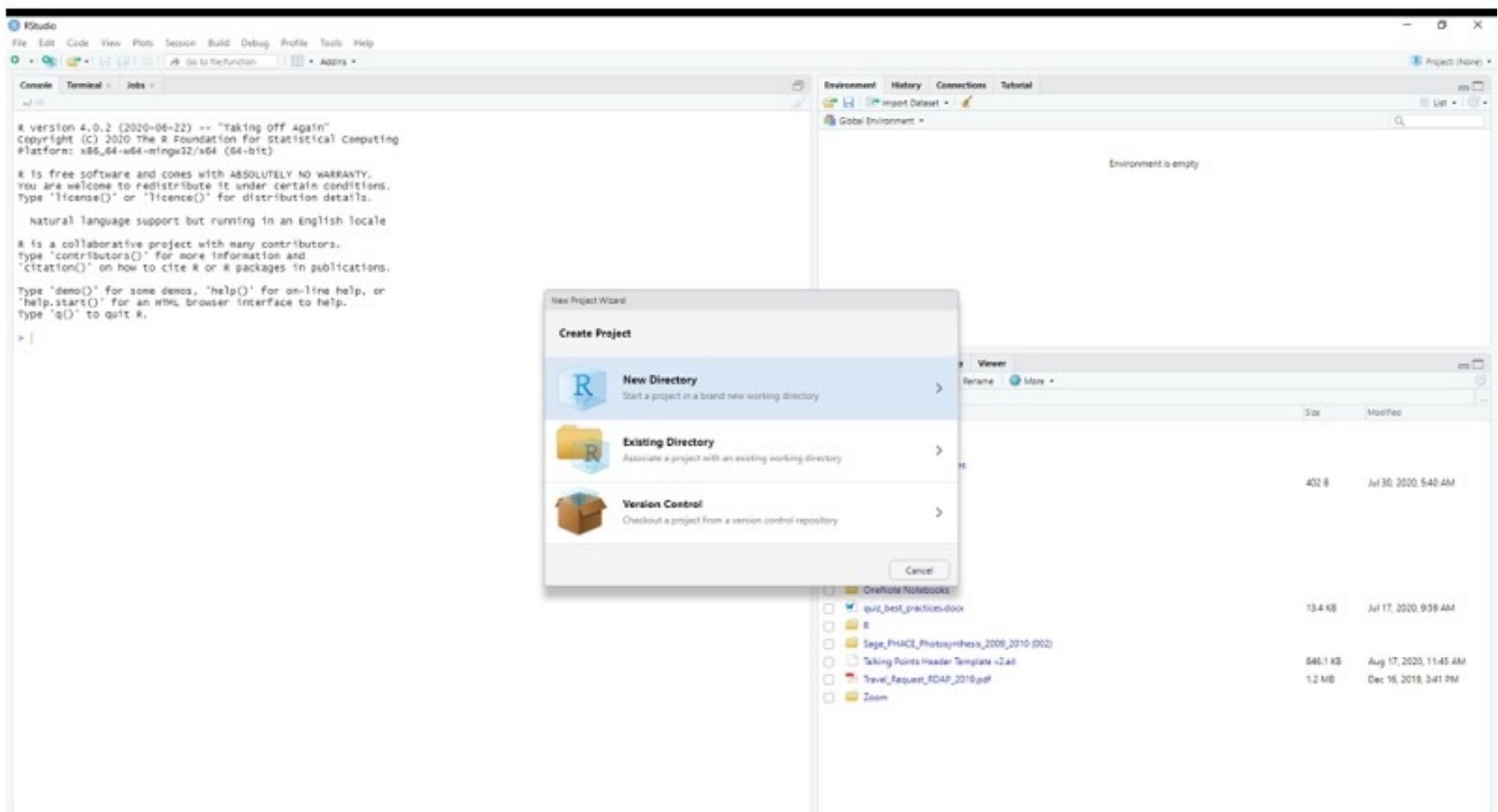
RStudio



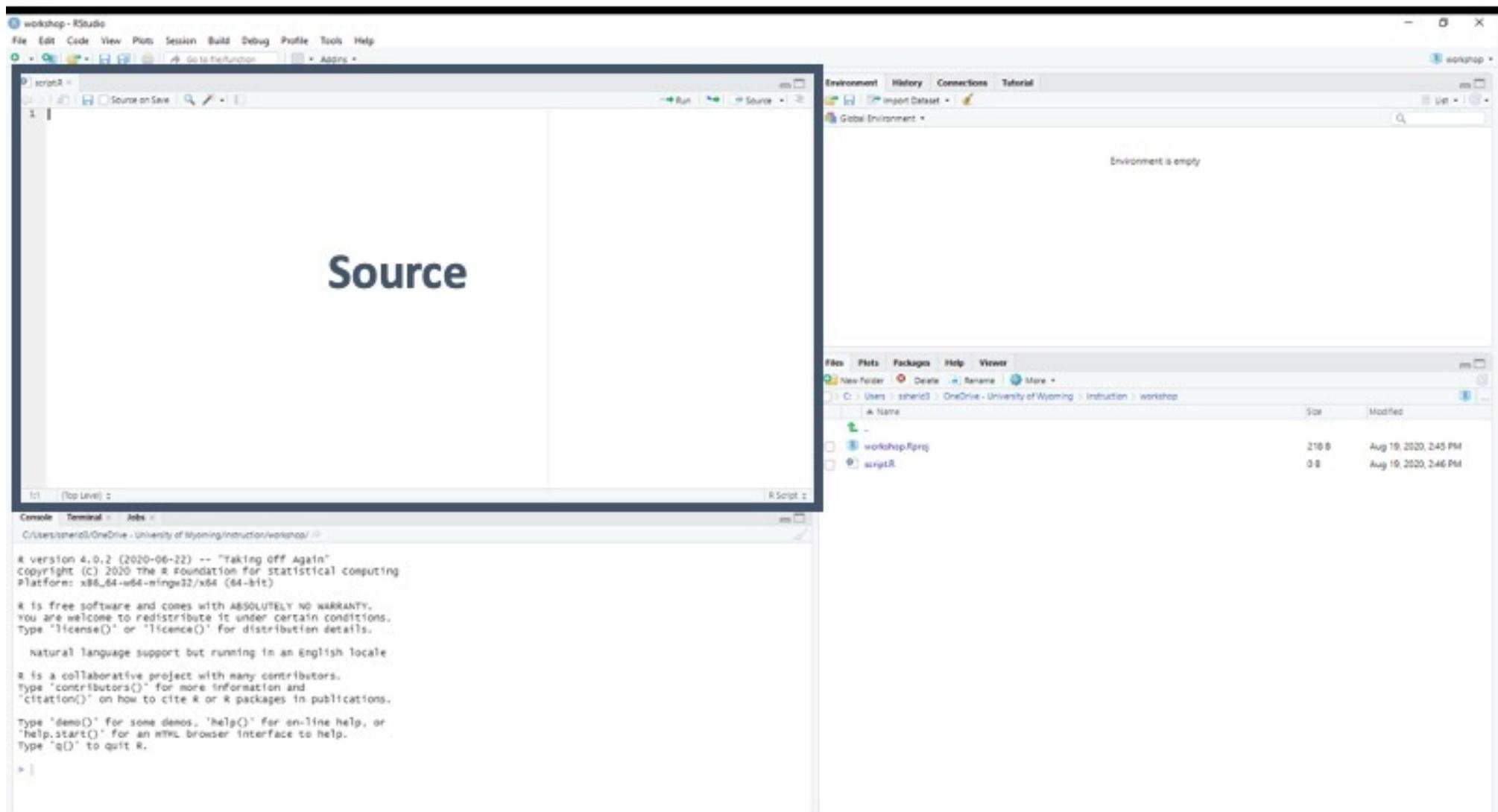
RStudio



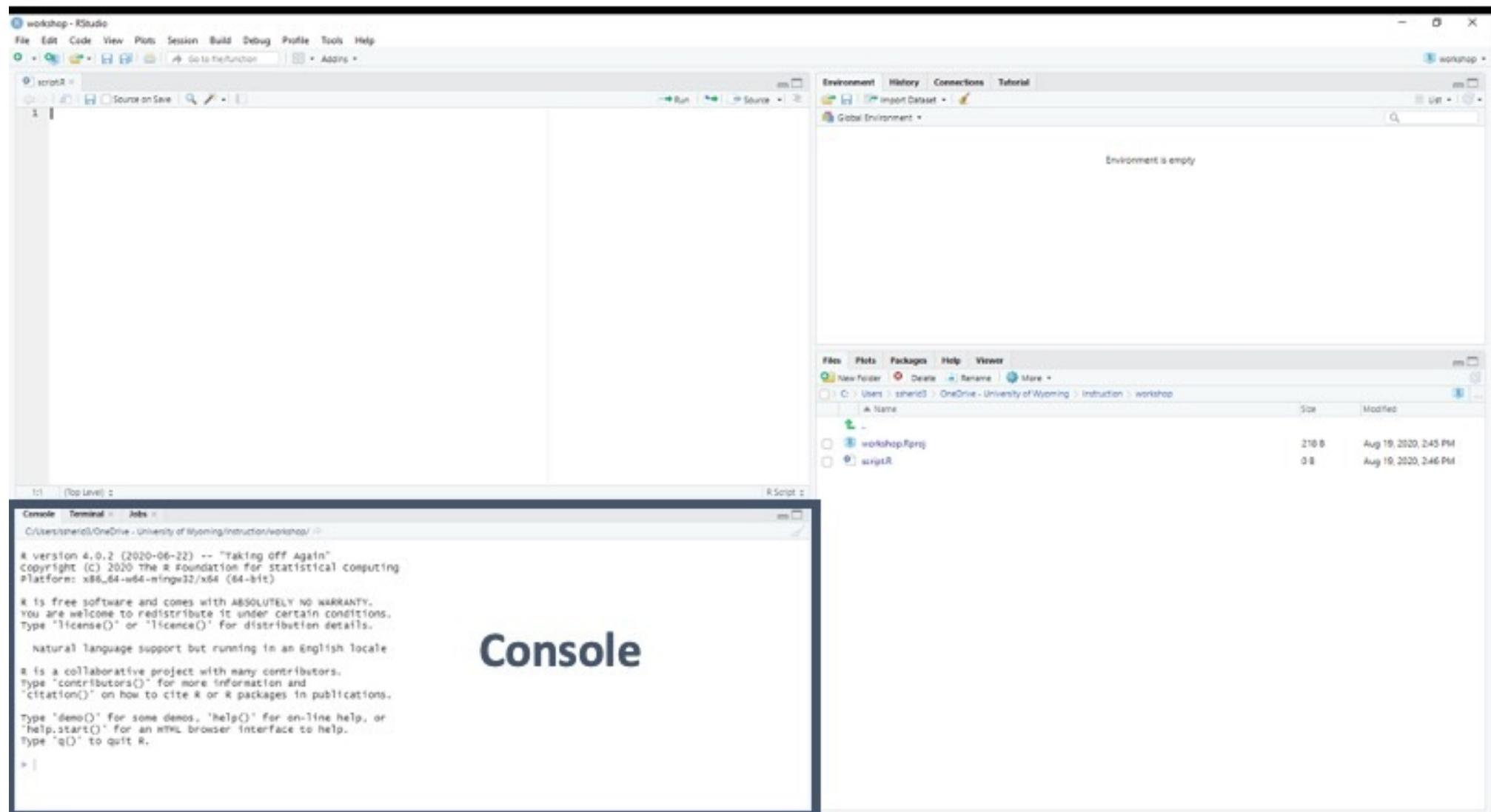
RStudio



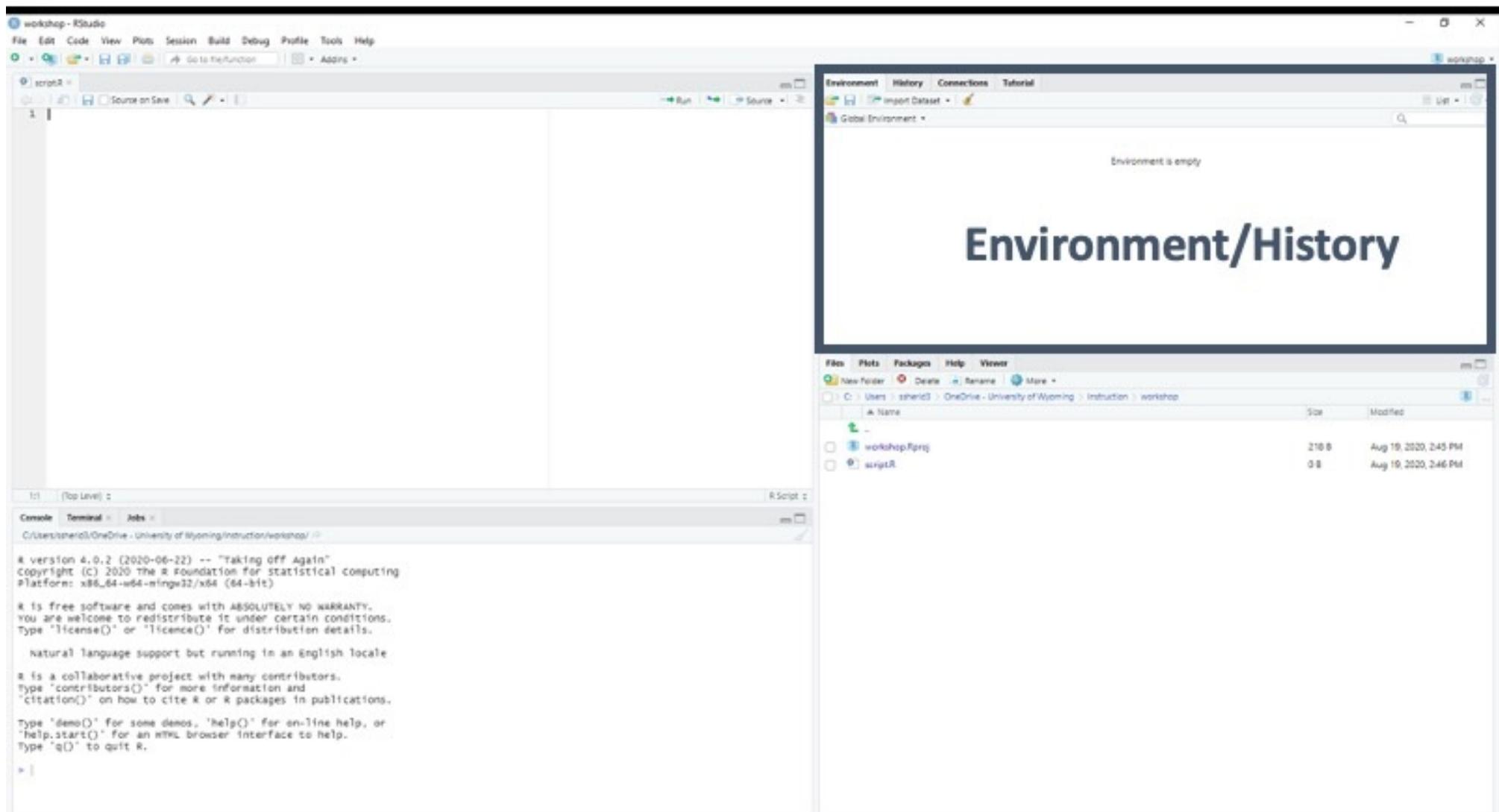
RStudio



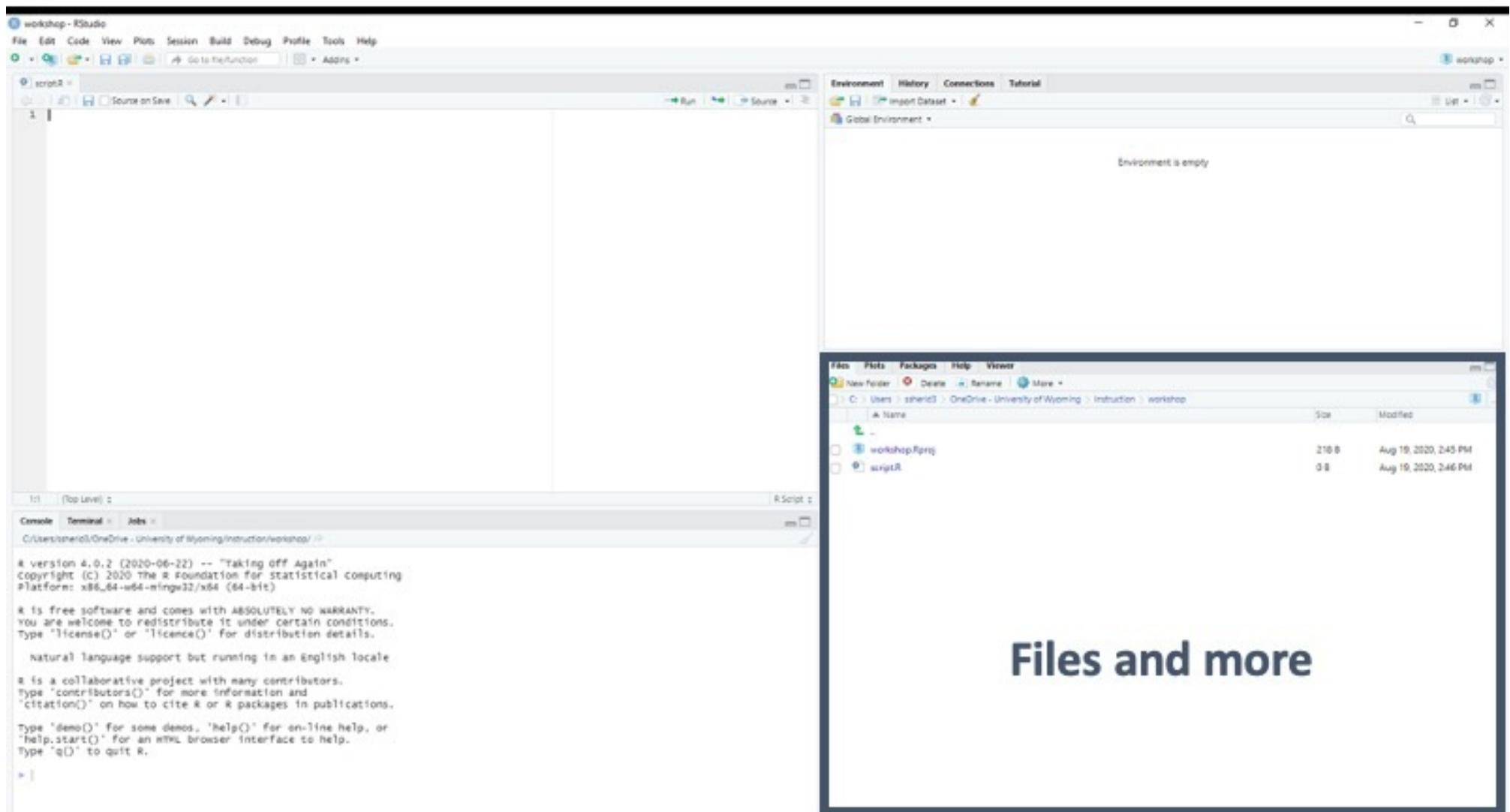
RStudio



RStudio



RStudio



RStudio

## Files and more

# The working directory

- Where R will look for & save files
- Scripts should refer to files in relation to root of working directory
- Only refer to files within this structure
- Check working directory with the `getwd()` function

# Suggested subdirectories

- `data/` => for raw data and intermediate datasets
- `data_output/` => modified versions of raw data
- `documents/` => outlines, drafts, other text
- `fig_output/` => graphics generated by scripts
- `scripts/` => R scripts for different analyses or plotting

# Working in the console or the script?

- Can type commands directly in console
- All console commands are *forgotten* after each session
- Script is key for reproducibility
- Send commands to R console from script using **Ctrl + Enter**

# Accepting commands

- When ready: >
- If waiting: +
- Cancel commands by pressing Esc
- Stop sign when working
- Anything after # will not be read

# Workshop I

# Graphics

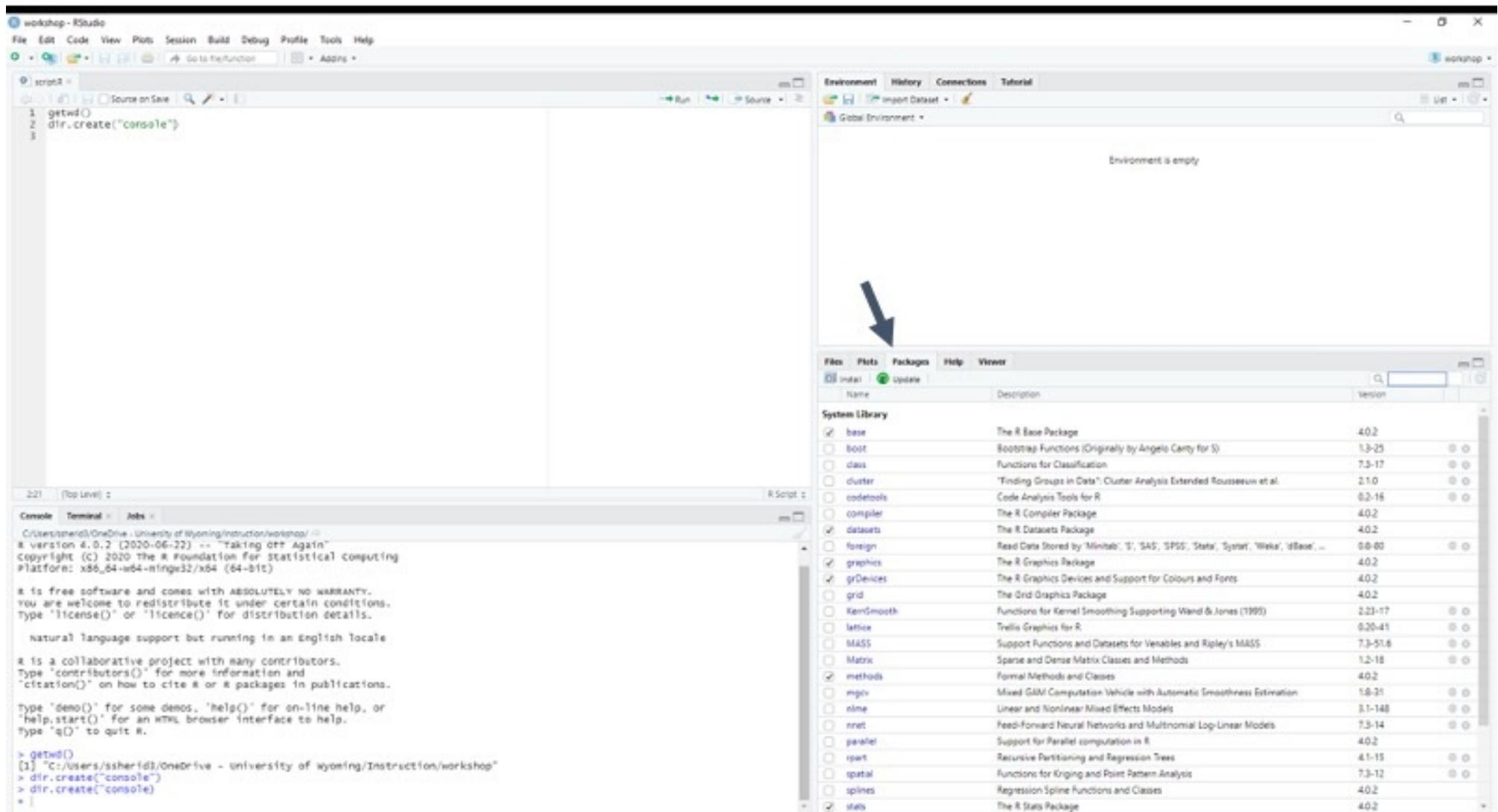
# Visualization

- R has one of the best graphics integrated
- Highly customizable
- Visualize plots on the spot or save to disk
- Third-party libraries also available

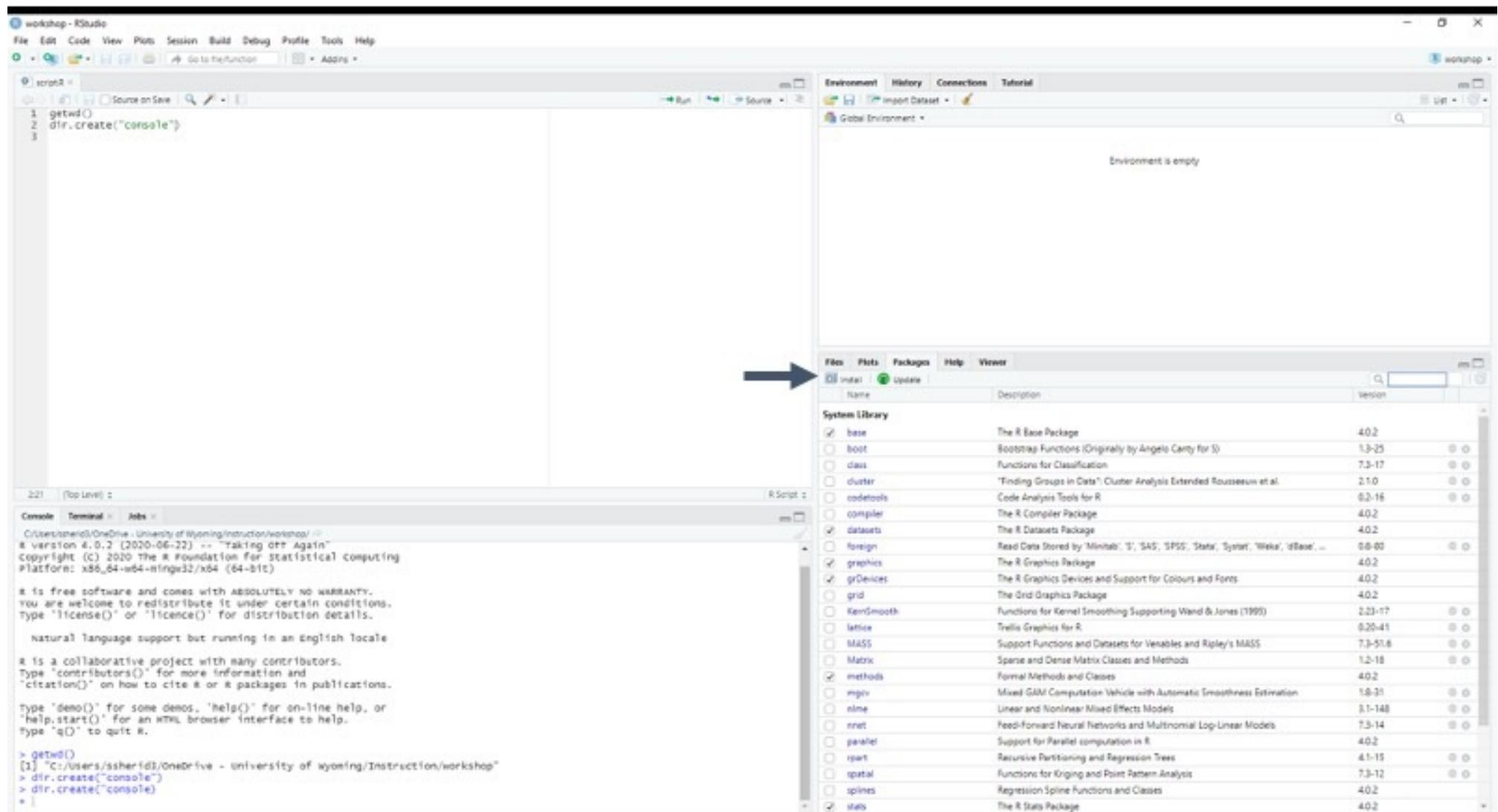
# Third-party Libraries

# Packages

- Created by other users
- Libraries of functions or data
- Provided to community at no charge
- Extends the capacity of R
- Often improve on Base R functions
- CRAN is the main repository for packages



RStudio



RStudio

The screenshot shows the RStudio interface with the following components visible:

- Top Bar:** workshop - RStudio. Includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help menus.
- Left Sidebar:** A tree view showing 'script.R' with code: `1 getwd(); 2 dfr.create("console");`
- Run Button:** Run button with dropdown menu.
- Source Button:** Source button with dropdown menu.
- Addons:** Addons button.
- Global Environment:** Environment pane showing 'Environment is empty'.
- Install Packages Dialog:** A modal window titled 'Install Packages'. It has a dropdown 'Install from:' set to 'Repository (CRAN)'. The 'Packages' input field contains 'tidyverse'. Below it, 'tidyverse' is listed under 'Available' with a status of 'Up-to-date'. A checked checkbox 'Install dependencies' is present. Buttons for 'Install' and 'Cancel' are at the bottom.
- Available Packages Table:** A table listing available R packages. The first few rows include:

Name	Description	Version
base	The R Base Package	4.0.2
boot	Bootstrap Functions (Originally by Angelo Carty for S)	1.3-25
ca	Functions for Classification	7.3-17
fcluster	'Finding Groups in Data': Cluster Analysis Extended Rousseeuw et al.	2.1.0
codetools	Code Analysis Tools for R	0.2-16
compiler	The R Compiler Package	4.0.2
datasets	The R Datasets Package	4.0.2
foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'idBase', ...	0.8-80
graphics	The R Graphics Package	4.0.2
grid	The R Graphics Devices and Support for Colours and Fonts	4.0.2
KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.3-17
lattice	Trellis Graphics for R	0.20-41
MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-51.8
Matrix	Sparse and Dense Matrix Classes and Methods	1.2-18
methods	Formal Methods and Classes	4.0.2
mgcv	Mixed GAM Computation Vehicle with Automatic Smoothness Estimation	1.8-31
nlme	Linear and Nonlinear Mixed Effects Models	3.1-148
nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models	7.3-14
parallel	Support for Parallel computation in R	4.0.2
rpart	Recursive Partitioning and Regression Trees	4.1-15
spatial	Functions for Kriging and Point Pattern Analysis	7.3-12
splines	Regression Spline Functions and Classes	4.0.2
stats	The R Stats Package	4.0.2
- Console:** Shows the R startup message, current working directory, and the command `dfr.create("console")` being run.

RStudio

The screenshot shows the RStudio interface with the following components:

- Top Bar:** workshop - RStudio
- File, Edit, View, Plots, Session, Build, Debug, Profile, Tools, Help**
- Toolbar:** Go to Function, Addins
- Script Editor:** A script named "script.R" containing the following code:

```
1 getwd()
2 dir.create("console")
3 install.packages("tidyverse") ← Arrow points here
4
```
- Console:** Output of the R session showing the installation of the tidyverse package and its dependencies. The output includes messages like "package 'magrittr' successfully unpacked and MD5 sums checked" and "The downloaded binary packages are in C:/Users/ssherrid/AppData/Local/Temp/Rtmp3USZf/downloaded\_packages".
- Environment:** Global Environment tab, showing "Environment is empty".
- Help:** Files, Plots, Packages, Help, Viewer
- RStudio Home:** Includes links to R Resources (Learning R Online, CRAN Task Views, R on StackOverflow, Getting Help with R), Manuals (An Introduction to R, Writing R Extensions, R Data Import/Export, The R Language Definition, R Installation and Administration, R Internals), Reference (Packages, Miscellaneous Material), and Search Engine & Keywords.
- Bottom Status Bar:** Home, Find in Topic

RStudio

# Workshop II

Daniel Rivas

[danielrivasmd@gmail.com](mailto:danielrivasmd@gmail.com)

# Questions & Comments