Introduction to R and RStudio

Instituto Nacional de Medicina Genomica

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
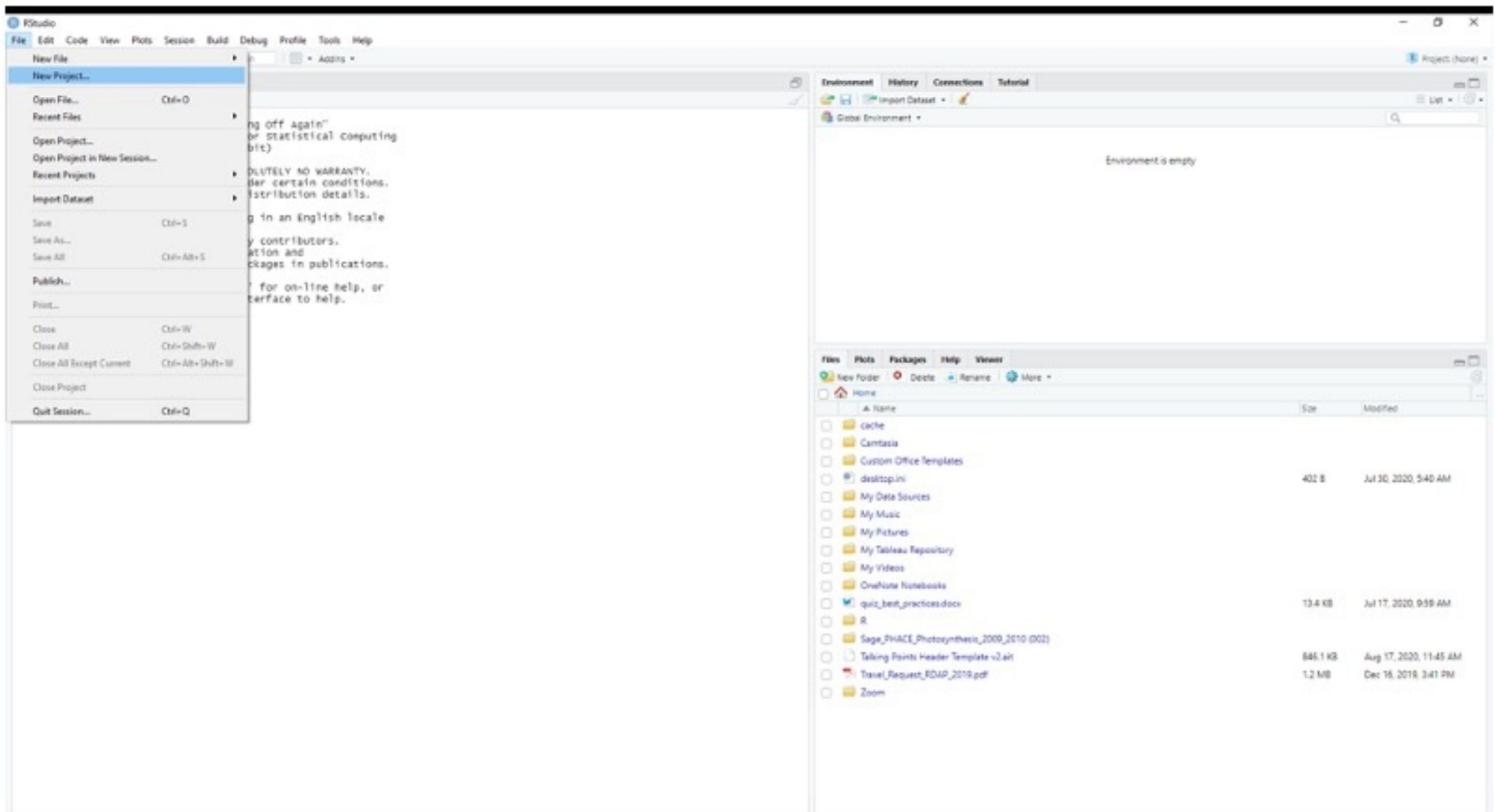   - Matrix based

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
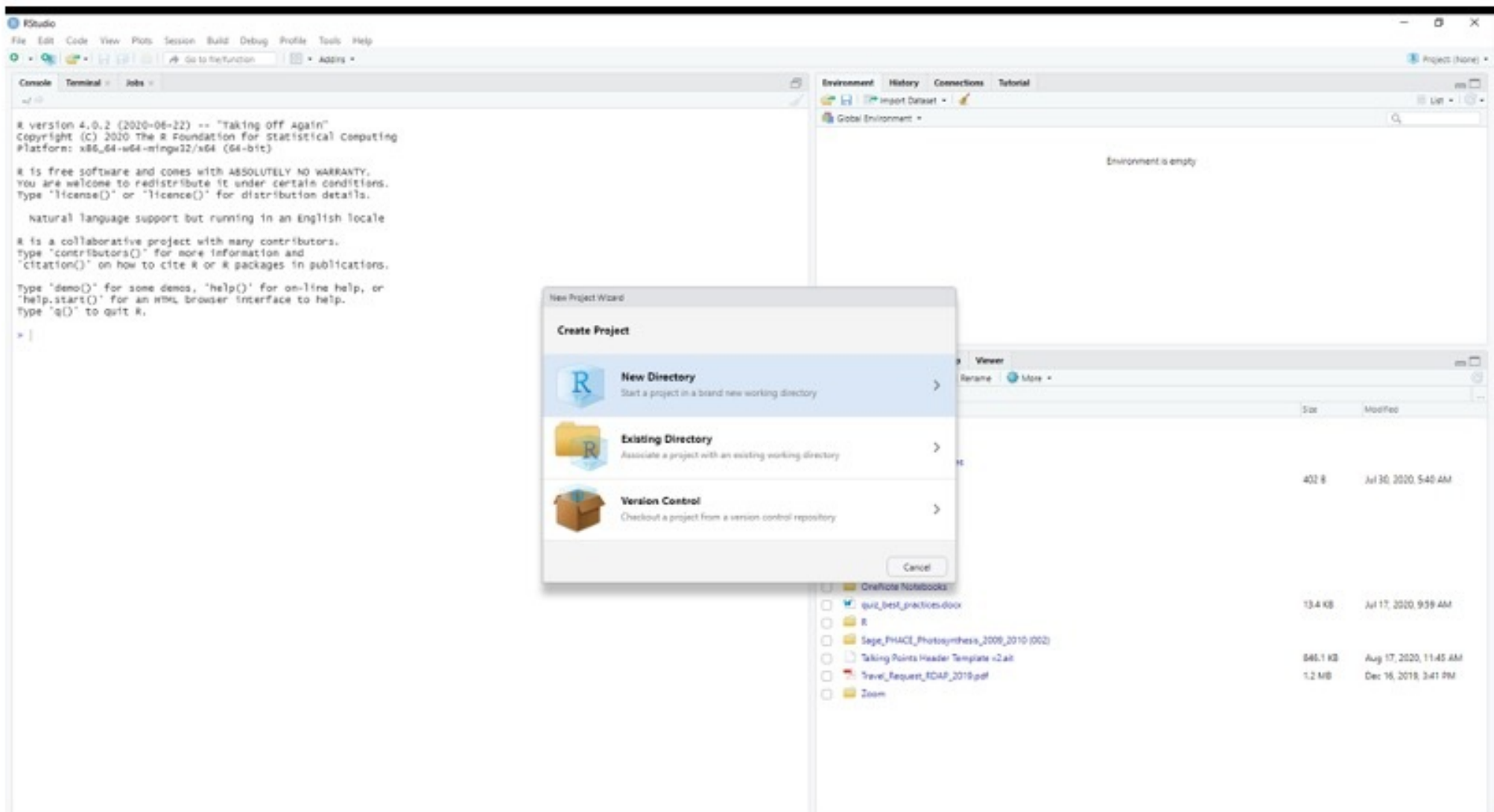   - Fast
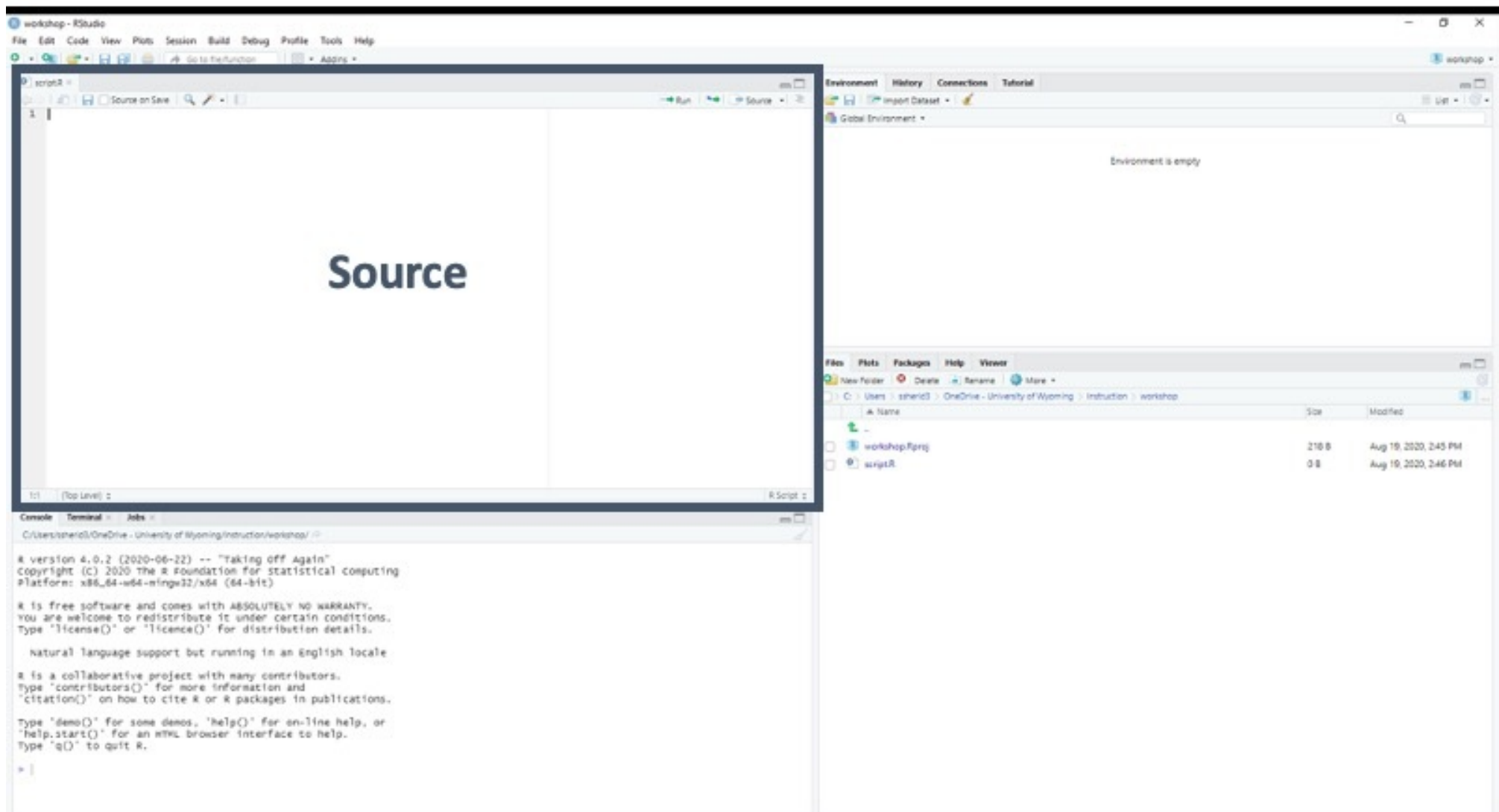   - Increasingly popular in macro / for solving complicated structural models
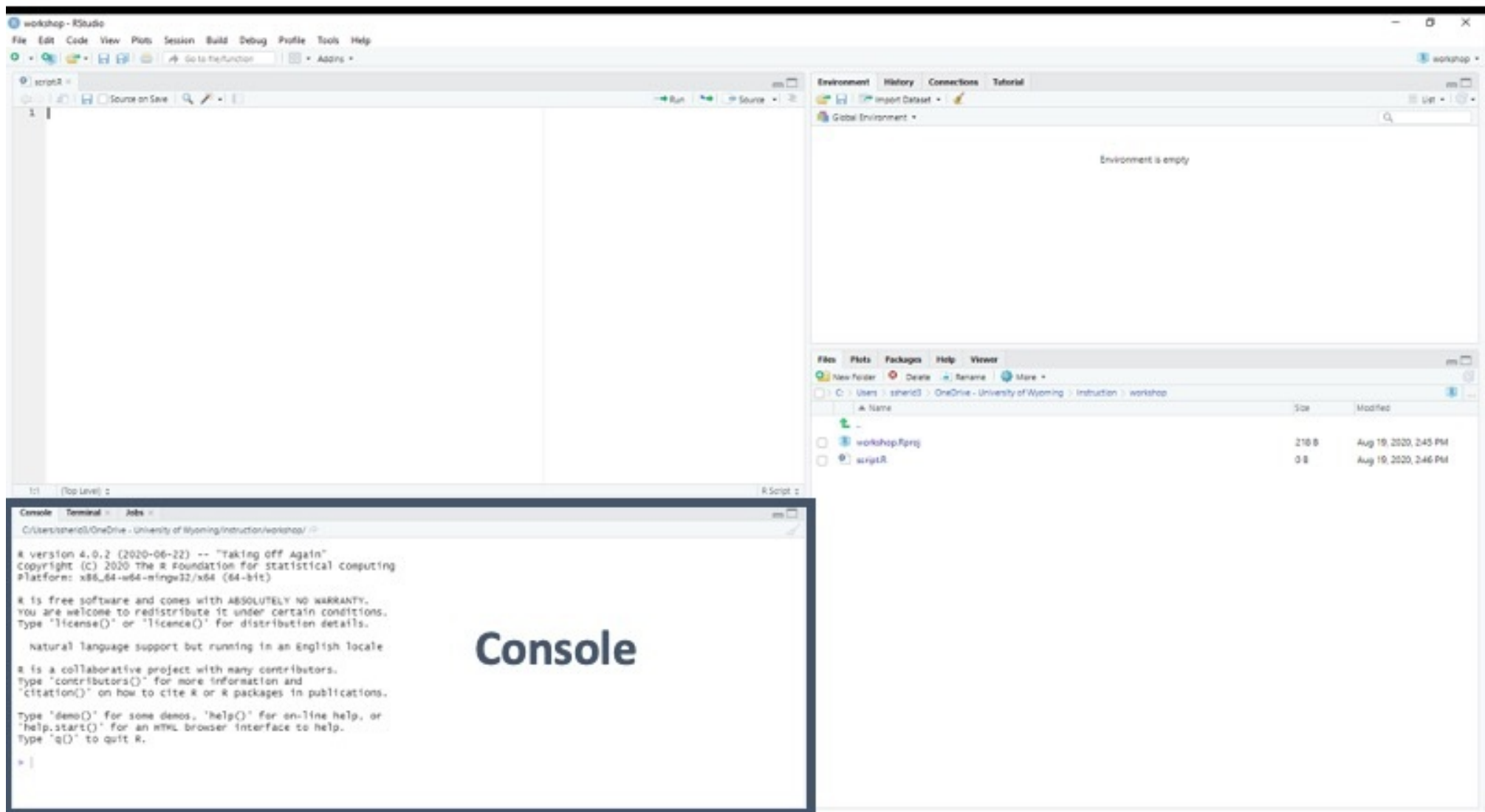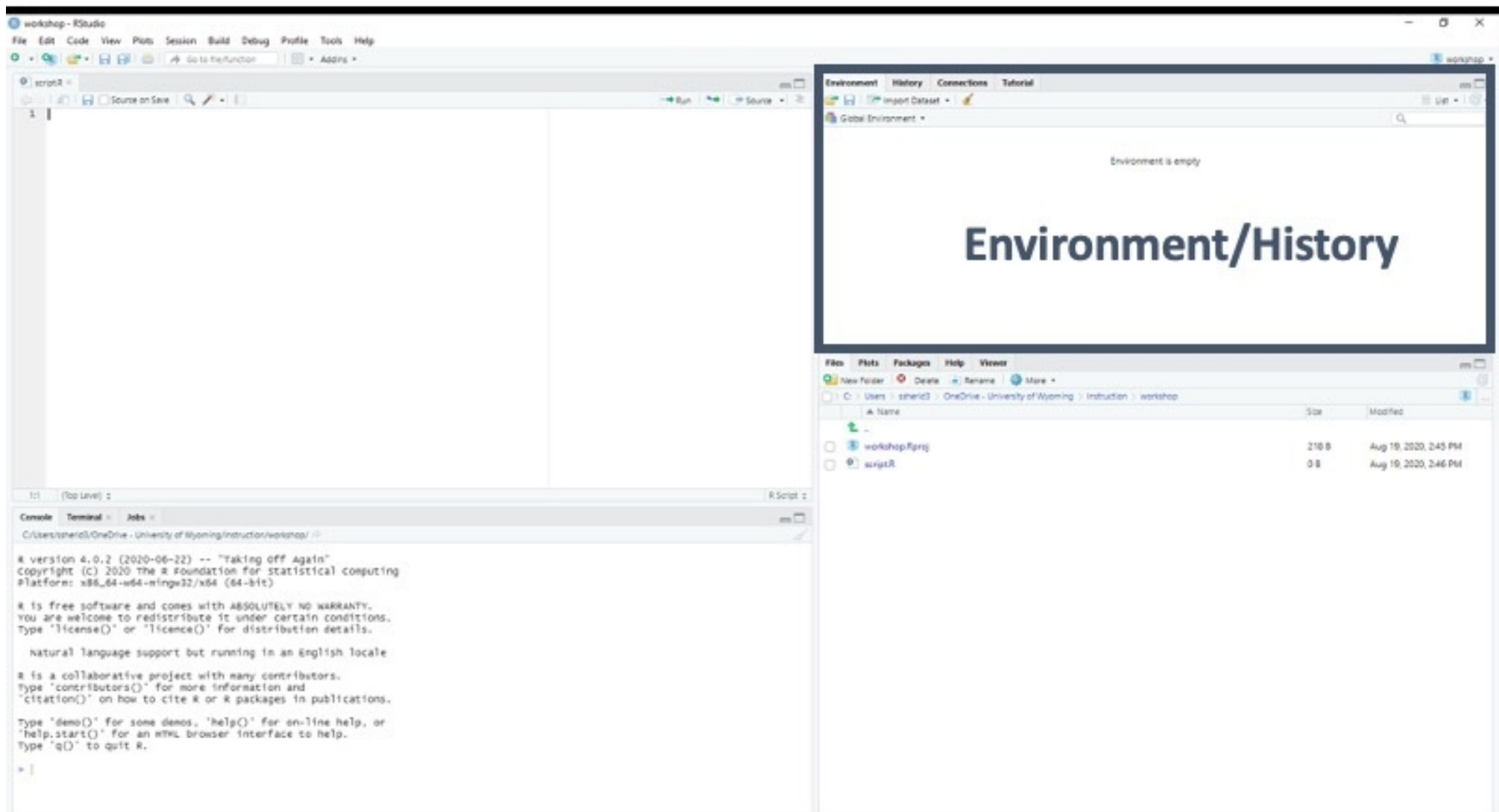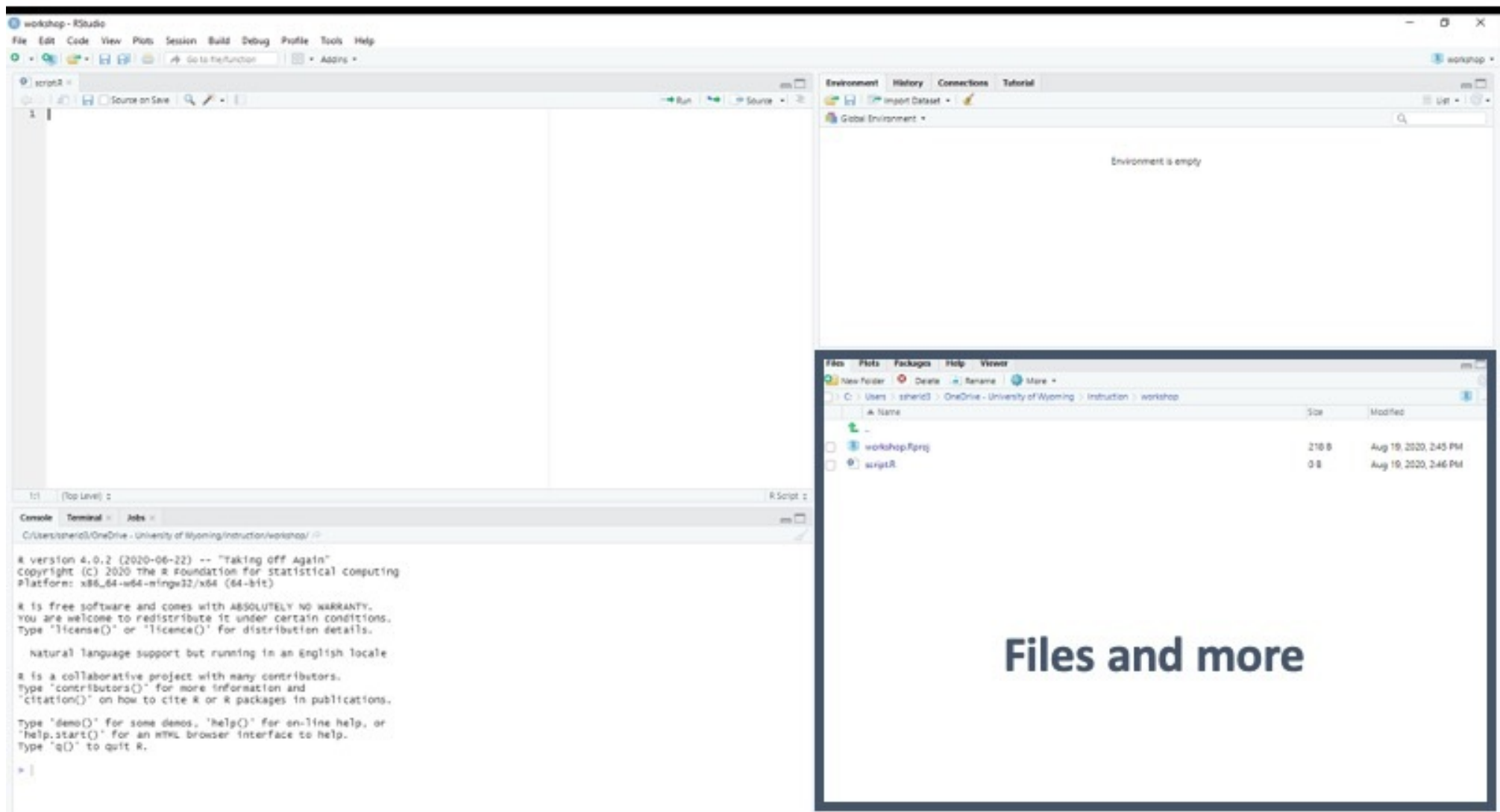
RStudio
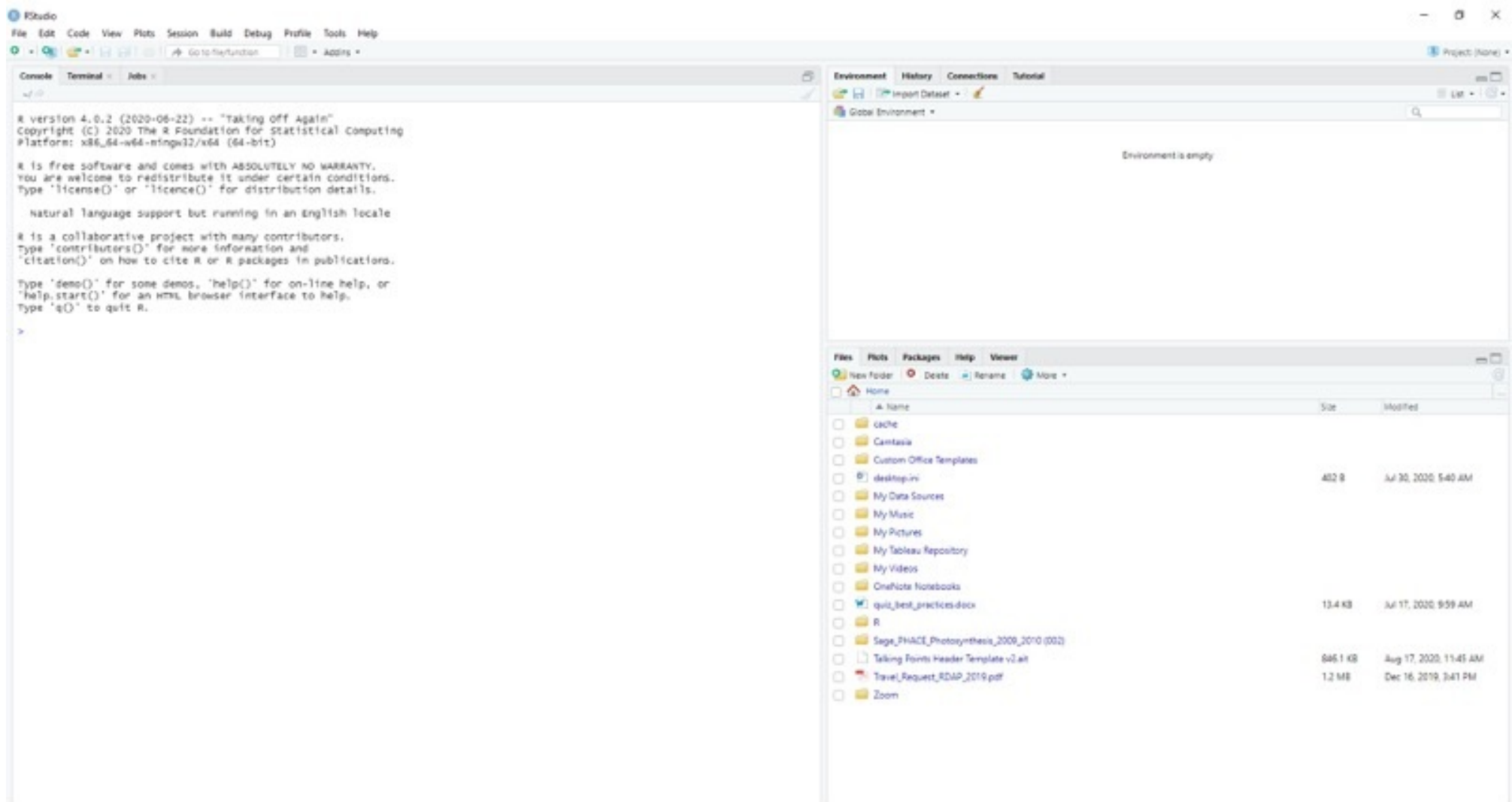
RStudio

RStudio

RStudio

RStudio

RStudio

RStudio

RStudio

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
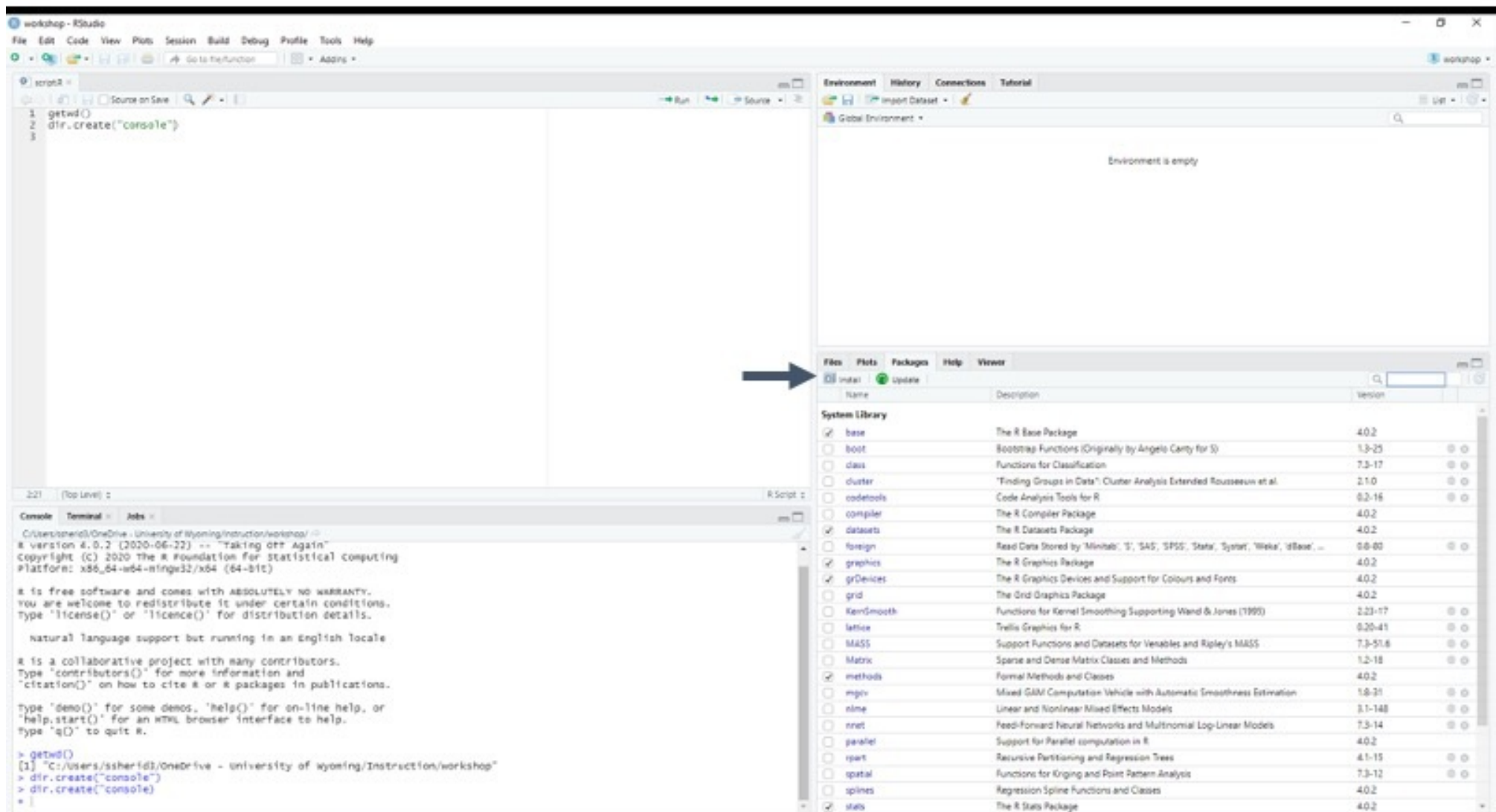- Third-party libraries also available

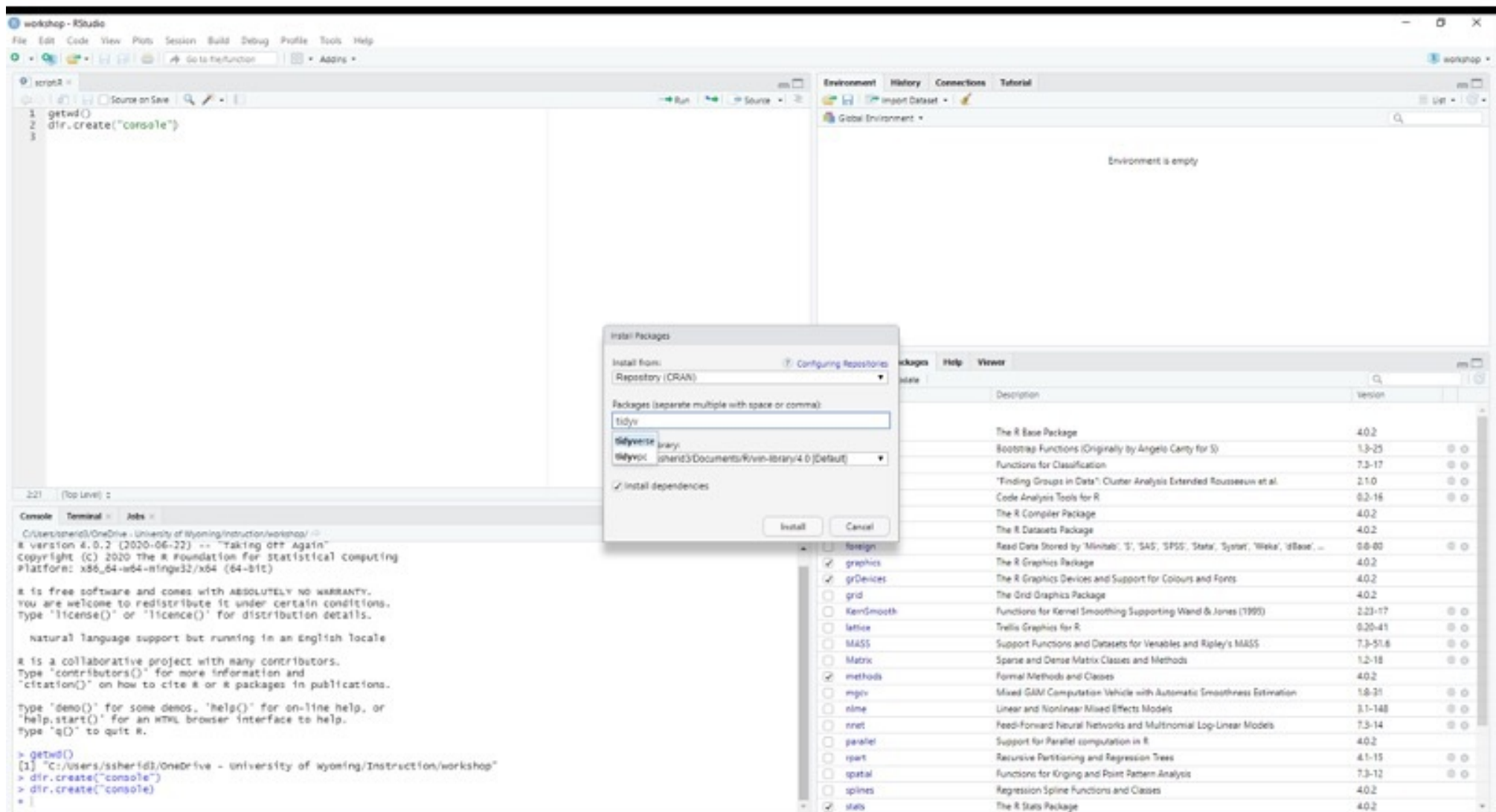# Third-party Libraries

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
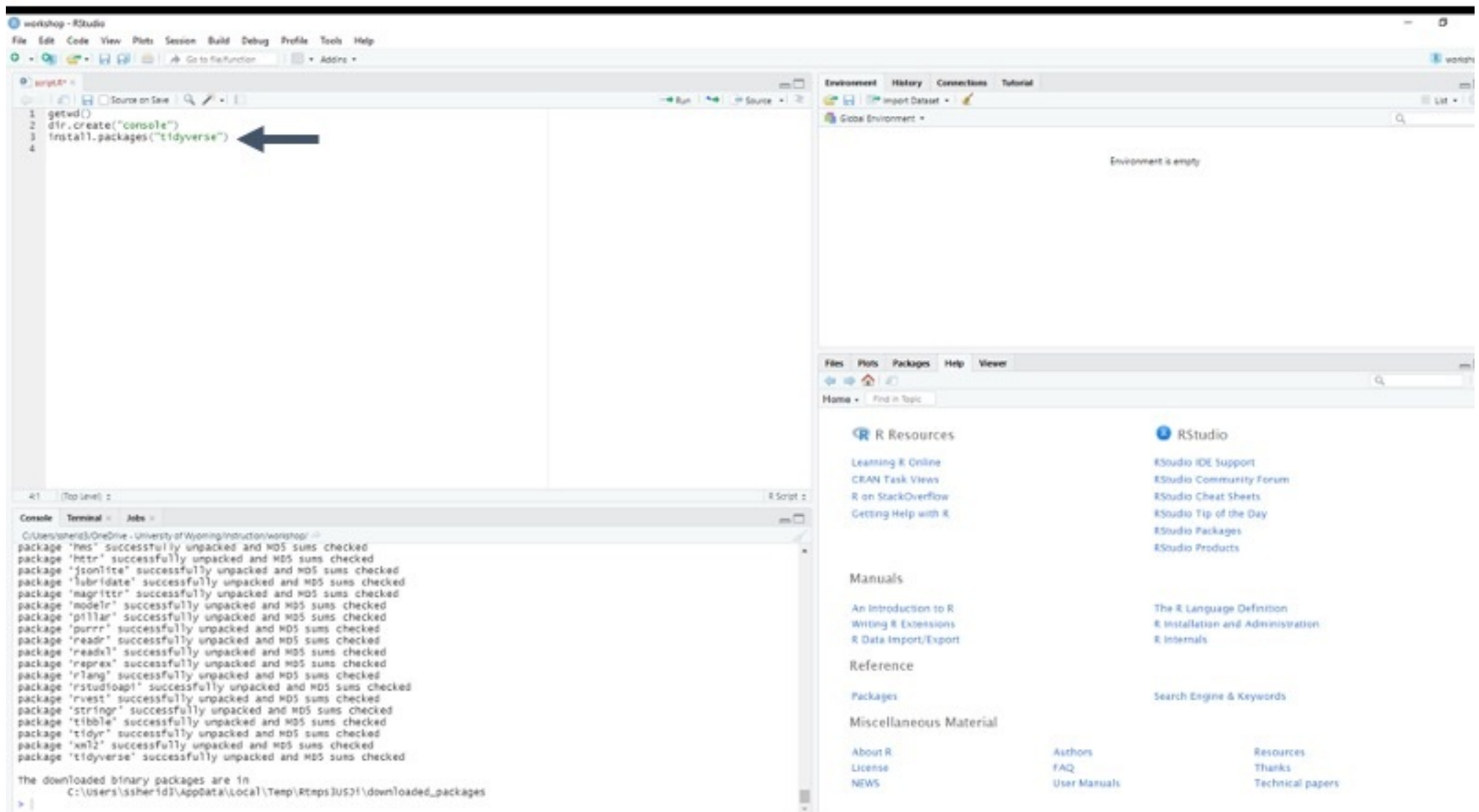- **CRAN** is the main repository for packages
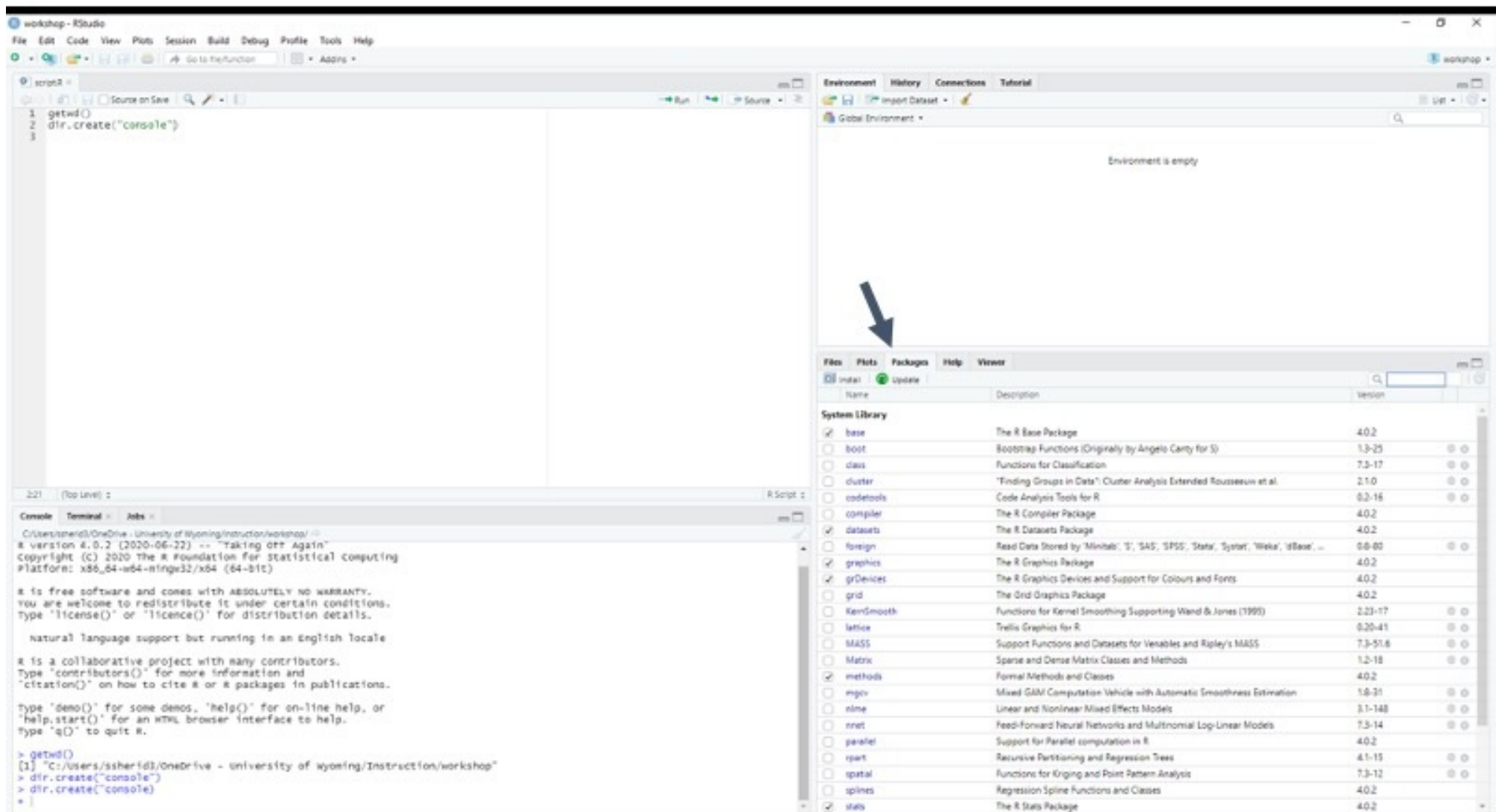
RStudio

RStudio

RStudio

RStudio

Workshop II

Daniel Rivas

danielrivasmd@gmail.com

Questions & Comments