R You Ready for Some R & R?:

An Intro to R Programming & Data Analysis

Laura Walker

Northeast Association for Institutional Research
Tools to See Workshop Series
July 28, 2020

University System of Maryland

Why R?

- Started working with R in 2016
  - The work I need to do can **not** be done in Excel.
  - Getting data from various sources
  - Cleaning the data
  - Transforming the data
  - Communicate inferences
- Plus—it's free!

Learning Outcomes & Outline

- Become familiar with the R environment and customize it
- Load data into the R environment from the computer and internet.
- Perform a variety of analytic functions on IPEDS data
  - Renaming
  - Rearranging
  - Sorting
  - Filtering
  - Creating new variables
  - Summarizing

Software

- You'll need BOTH of these software:
  - **R for Statistical Computing**
  - **RStudio Desktop**
- For this session, we'll primarily use **RStudio.**

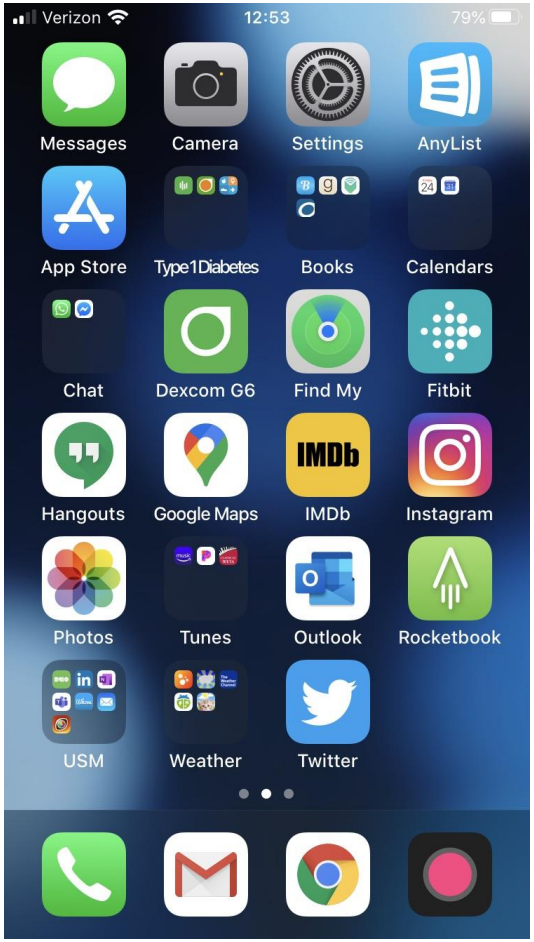
Why both?

- R for Statistical Computing allows RStudio to run.
- RStudio is an **integrated development environment (IDE)** that provides a more user-friendly way to code in R.

https://moderndive.com/1-getting-started.html

R Packages

- Think of **packages** like we may think of apps on our cell phones.
  - RStudio is the cell phone.
  - Packages are its apps.
- Packages have a similar focus:
  - Manipulating Data
    - *dplyr*– Allows us to subset, summarize, rearrange and join datasets.
  - Visualizing Data
    - *ggplot2*—Allows us to create impressive graphics after analysis is conducted
  - Reporting Results
    - *RMarkdown*—Allows you to store reproducible code and can produce Word, PDF



R Files

- Two types of files beginners often create:
  - R scripts
  - **Rmarkdowns**

Why?

- Personal preference—chunks
- Reproducible
- “Knitting” the file together to get HTML, MS Word, PDF, etc.

Ann Laura Walker
July 17, 2020

Example of R Markdown Output

Bachelor Degree Counts, 2015-2019

USM Institution	2015	2016	2017	2018	2019	Pct. Diff. from 2015 to 2019
BSU	801	832	713	781	826	↑ 3.1%
CSU	416	464	421	399	378	↓ -9.1%
FSU	1,032	964	1,060	1,027	1,078	↑ 4.5%
SU	1,935	1,982	2,026	1,872	1,805	↓ -6.7%
TU	4,422	4,428	4,628	4,609	4,619	↑ 4.5%
UB	694	721	755	711	615	↓ -11.4%
UMB	333	399	416	433	438	↑ 31.5%
UMBC	2,432	2,521	2,572	2,578	2,658	↑ 9.3%
UMCP	7,166	7,253	7,292	7,559	7,768	↑ 8.4%
UMES	577	574	514	482	508	↓ -12.0%
UMGC	5,240	5,638	5,883	6,206	6,346	↑ 21.1%
Total	25,048	25,776	26,280	26,657	27,039	↑ 7.9%

Master Degree Counts, 2015-2019

USM Institution	2015	2016	2017	2018	2019	Pct. Diff. from 2015 to 2019
BSU	304	337	313	245	252	↓ -17.1%
CSU	75	77	78	74	66	↓ -12.0%
FSU	235	217	268	221	208	↓ -11.5%
SU	262	272	308	323	323	↑ 23.3%
TU	1,097	974	920	906	887	↓ -19.1%
UB	504	523	561	533	436	↓ -13.5%
UMB	845	887	827	921	894	↑ 5.8%
UMBC	694	666	631	665	624	↓ -10.1%
UMCP	2,562	2,836	2,829	2,929	2,769	↑ 8.1%
UMES	46	89	50	52	53	↑ 15.2%

Data were extracted from:
https://www.usmd.edu/IRIS/DataJournal/Degrees/?report=Degrees-Awarded

BuzzR

RStudio anatomy

https://buzzrbeeline.blog/

Emma Rand

Script file

Write code here

To run code put your cursor on the line and click the run button

Edit to correct errors

⇒ record of commands that worked

Save scripts with the **.R** extension

⇒ syntax will be highlighted

⇒ good practice

<- is the assignment operator

⇒ puts what is on the right in to the object on the left

⇒ Assign results if you want to use them again

Console

When you click run, code is sent to the console and executed

> is the prompt

⇒ do not type it

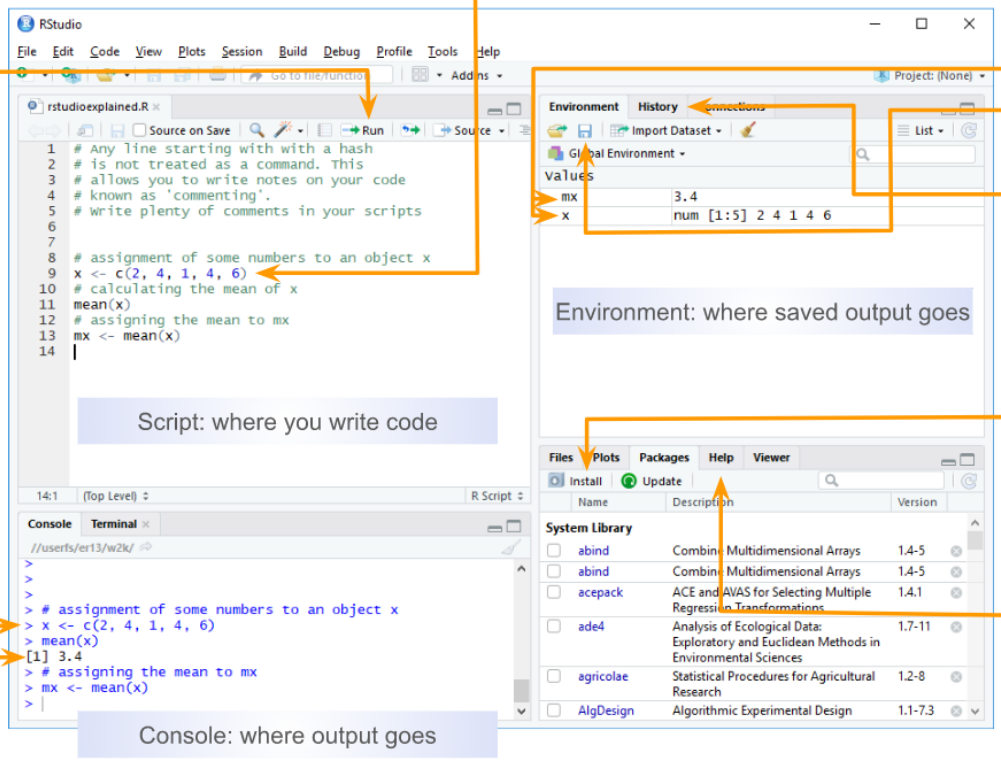
⇒ appears when R is ready for next command

Command output goes here by default

⇒ output is in a different colour

⇒ [1] indicates 3.4 is the first element of the output

⇒ many commands will not have output, the prompt just reappears



Environment

Name objects by assignment to use them again

All the objects you created in your session

Saving the environment saves all the objects, but not the code with a **.RData** extension

History

A history of every command you sent to the console, mistakes included.

File can be saved but usually you just need the script

Packages

Many functions come with R

A huge amount of extra functionality is available in packages

Packages can be installed by clicking the Install button

Help

Access to manual pages for all installed packages

Plots

Figure output appears here

Image Credit: https://buzzrbeeline.blog/2018/07/04/rstudio-anatomy/

Laura Walker
Senior Institutional Research Data Analyst
University System of Maryland
alwalker@usmd.edu