EPSY 5261 : Introductory Statistical Methods

Day 3

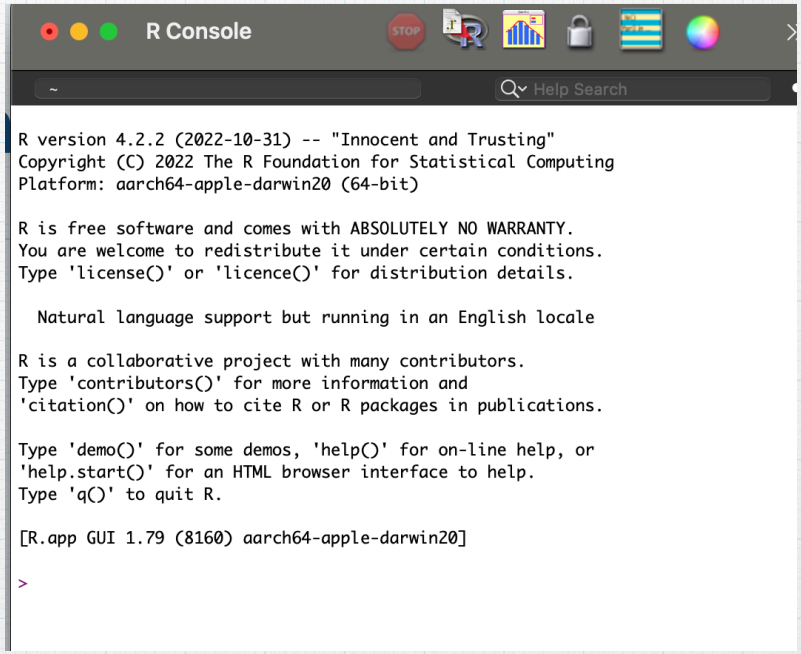
Introduction to R and R Studio

Learning Goals

- At the end of this lesson, you should be able to…
  - Explain what R Studio is
  - Explain why we use it
  - Carry out a basic workflow in R Studio for data analysis

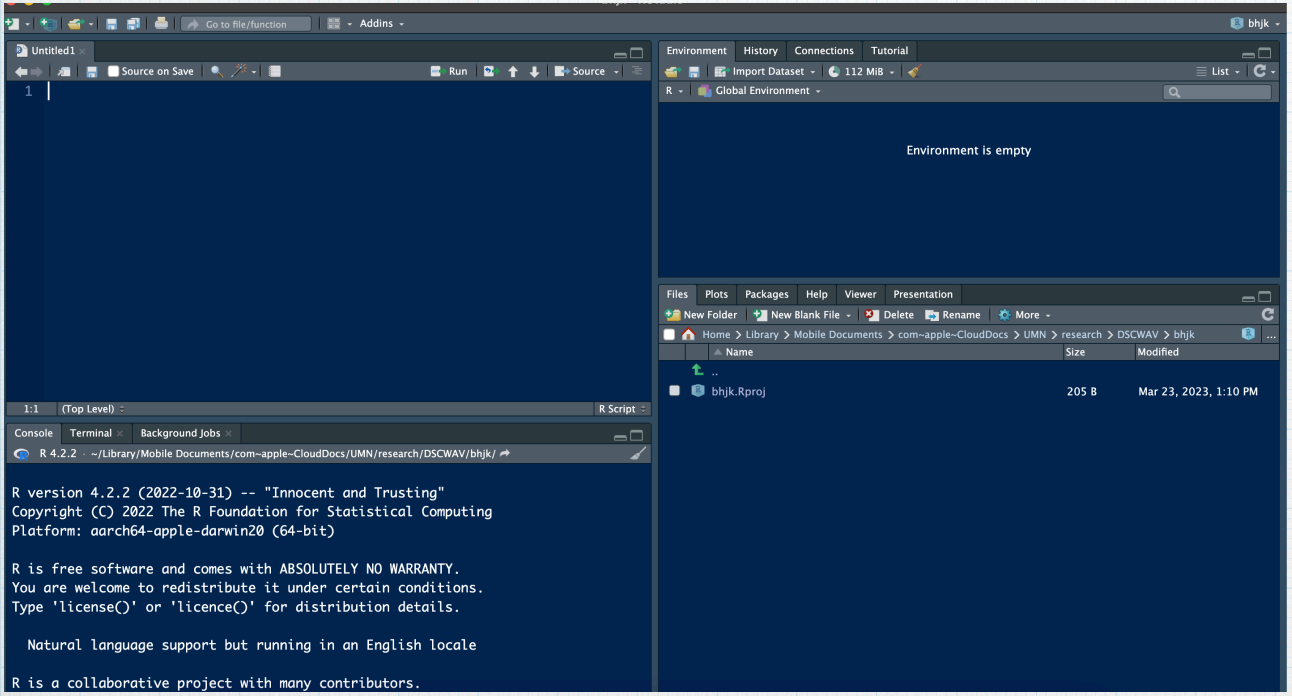
What is R?

- R is a programming language used for statistical computing
- It is used to write code to communicate to the computer what you want to do with your data



Screenshot depicting R Application

What is R Studio?



Screenshot of R Studio Application

- R Studio is an interface for programming in R
  - Point and click functionality makes it more user friendly than R alone
  - Allows integration of text and code to produce reports

Why R Studio?

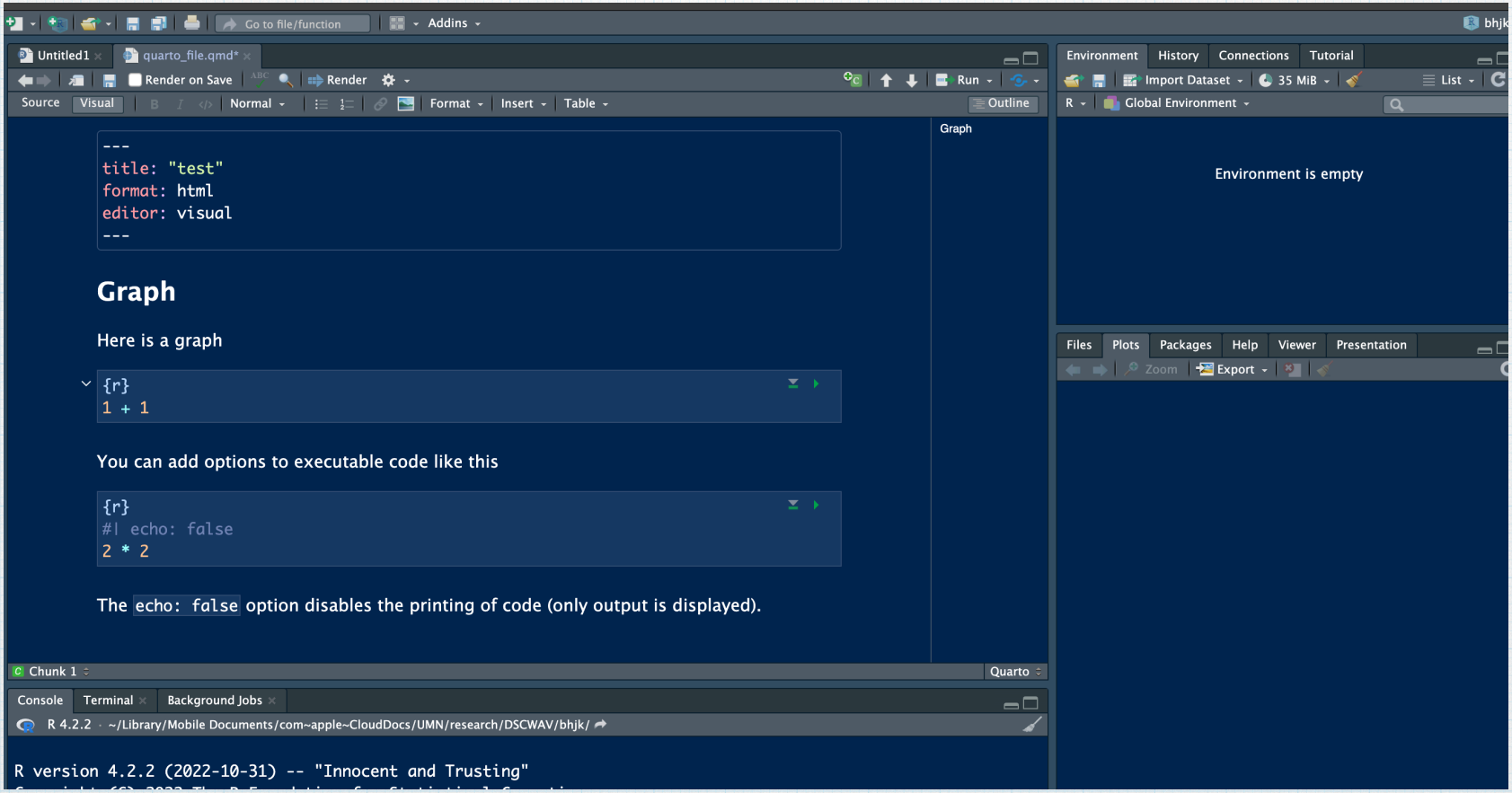
- It's free!
- Computing is a main component of statistics and science
- Open source
- Integration of code and text makes it easy to create reports
  - Supports reproducibility and open science goals

Reproducibility

- In general:
  - Experiments/studies repeated with the same methods and analysis should yield the same results
- In computing:
  - Code and data should be provided alongside detailed documentation of analysis so that they can be repeated with the same results

R Studio & Programming Introduction

Basics



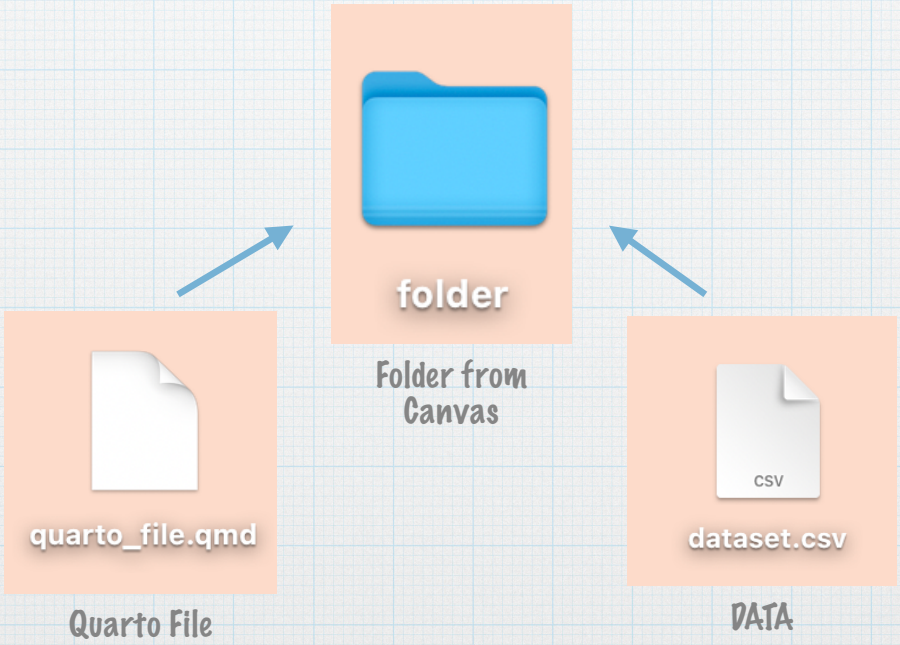
R Studio Application

Workflow (a.k.a. steps to using R Studio)

1. Download files from Canvas (folder containing Quarto file and dataset)
2. Load packages
3. Import data
4. Ready to perform analysis!

Download Folder from Canvas

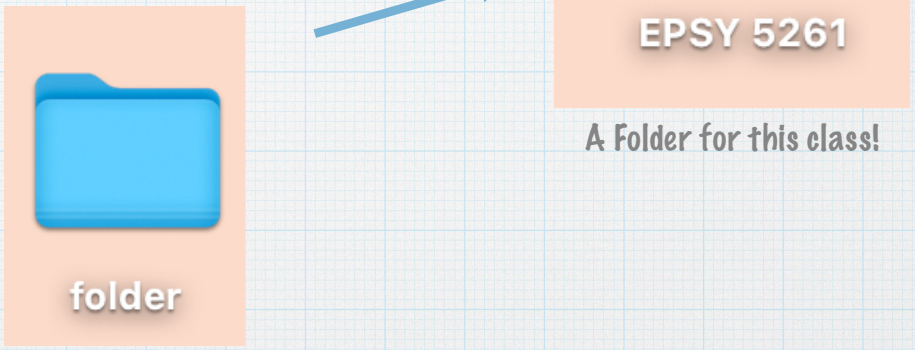
- Note: Quarto Files and datasets must reside in the same folder to run properly



Put that folder somewhere easily accessible

Some suggestions:

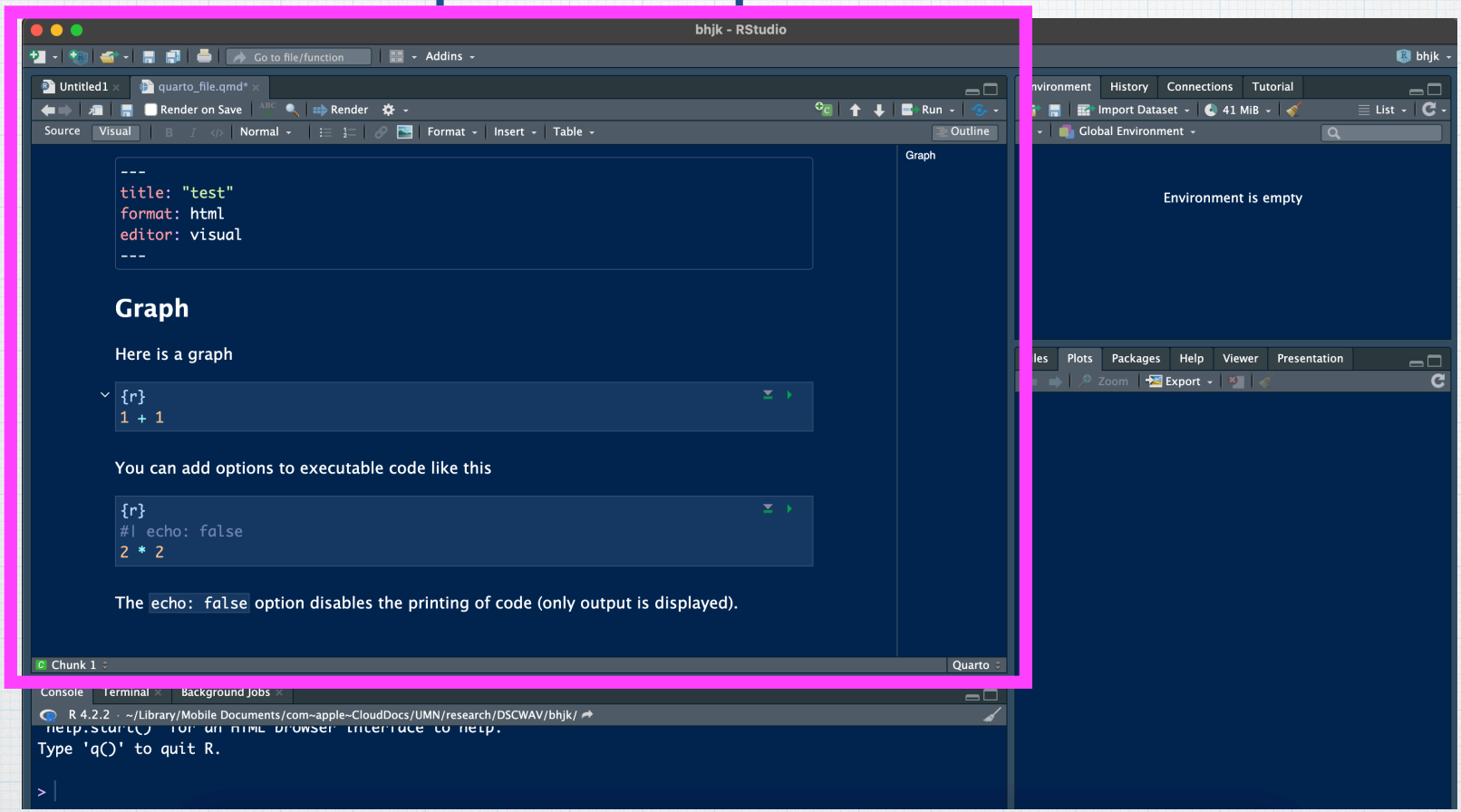
- Desktop
- "Grad School" Folder
- On your desktop?
- In your cloud?



Folder from Canvas

A Folder for this class!

Open .qmd file



Functions

- To perform analyses on your data in R, you will need to write *functions*
- Functions tell R Studio what to do with your data
- They look like
  - `function()`
- They contain 1 or more *arguments*
  - The arguments specify even further what you would like to do with your data

For example:

```
function(argument1, argument2, argument3)
```

Load Libraries

- To get all the functions we need in R Studio we will need to load a *package*
- *Packages* contain collections of functions
- To load the package and its functions we use the `library()` function

Load Libraries

General Example: `library(packageName)`

- `library()` is our function
- `package` will specify the package we want to load

Upload Data

- Using the function `read_csv()` we can load the data into R Studio

Example: `read_csv("dataset.csv")`

- `read_csv()` is our function
- `"Dataset.csv"` is our argument to specify the data file

In R Studio

```r
#This code chunk allows us to load the packages needed for our activity
library(tidyverse)
library(ggformula)
```

```r
#This code chunk allows us to read in data
countries <- read_csv("World-Countries-Data-2019.csv")
```

Introduction to R Studio Activity

Summary: R Studio Workflow

1. Download files from Canvas (folder containing Quarto file and dataset)

2. Load packages

3. Import data

4. Ready to perform analysis!