

EPSY 5261 : Introductory Statistical Methods

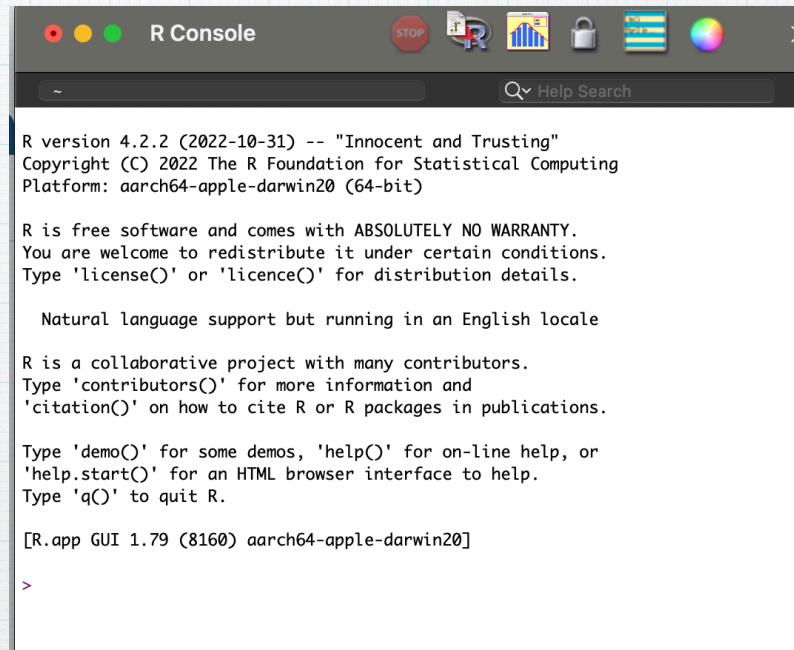
Day 3
Introduction to R and R Studio

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?

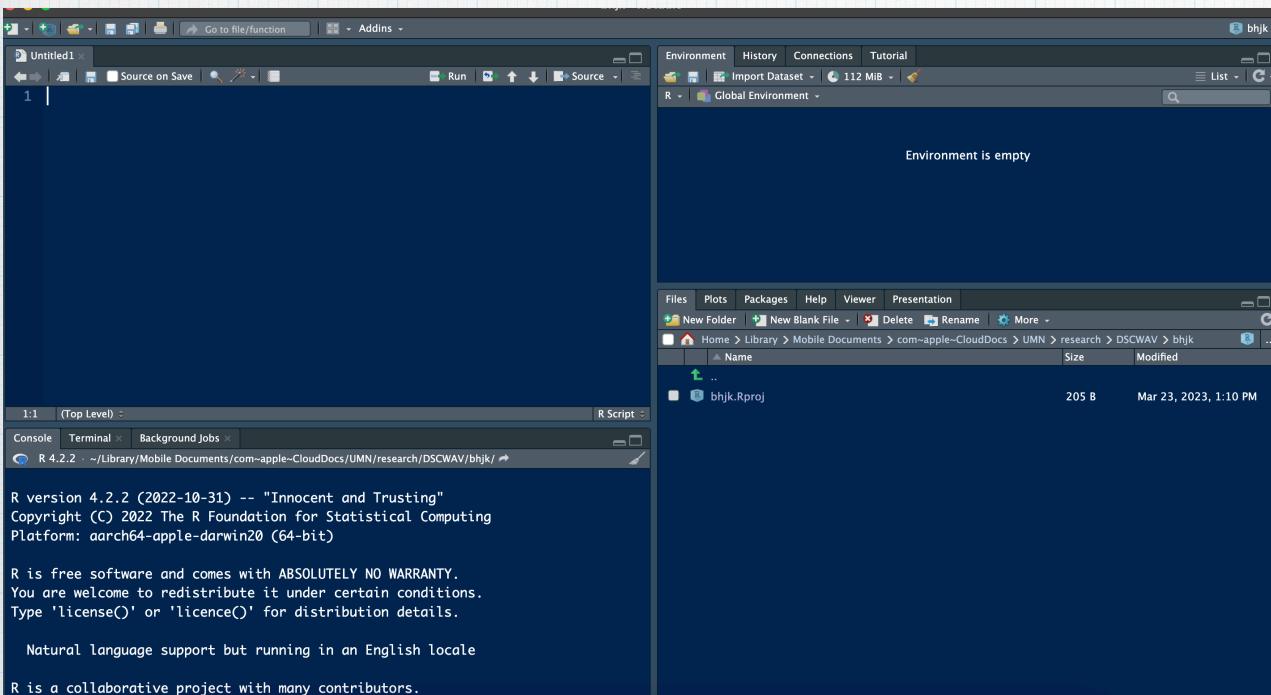
- R is a programming language used for statistical computing
- It is used to write code to communicate to the computer what you want to do with your data



Screenshot depicting R Application

What is R Studio?

- 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



Screenshot of R Studio Application

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

The screenshot shows the R Studio Application interface. On the left, a Quarto file named "quarto_file.qmd" is open. The code includes a YAML front matter section:

```
---
```

```
title: "test"
```

```
format: html
```

```
editor: visual
```

```
---
```

Below this, there is a section titled "Graph" containing the text "Here is a graph". A code chunk is shown:

```
{r}
```

```
1 + 1
```

Below the code, a note says: "You can add options to executable code like this". Another code chunk is shown:

```
{r}
```

```
#| echo: false
```

```
2 * 2
```

The note continues: "The echo: false option disables the printing of code (only output is displayed)".

At the bottom, the R console output is displayed:

```
R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"  
Copyright (C) 2022 The R Foundation for Statistical Computing  
Platform: arm64/v8 (64-bit)
```

The right side of the interface shows the "Environment" pane, which displays the message "Environment is empty".

R Studio Application

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

1. Download materials from the schedule page on the website.

- This is a zipped folder containing Quarto file (.QMD) and dataset (.csv).
- Unzip the folder.
- Double-click the QMD file to open it in R Studio.

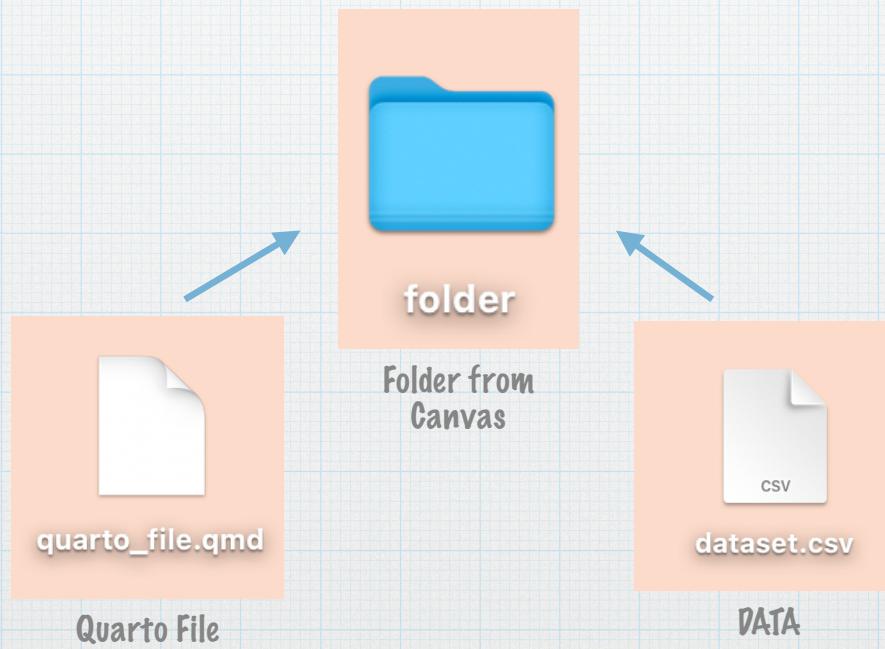
2. Load libraries (Be sure to install any libraries you do not have.)

3. Import data

4. Ready to perform analysis!

Download Folder from Canvas

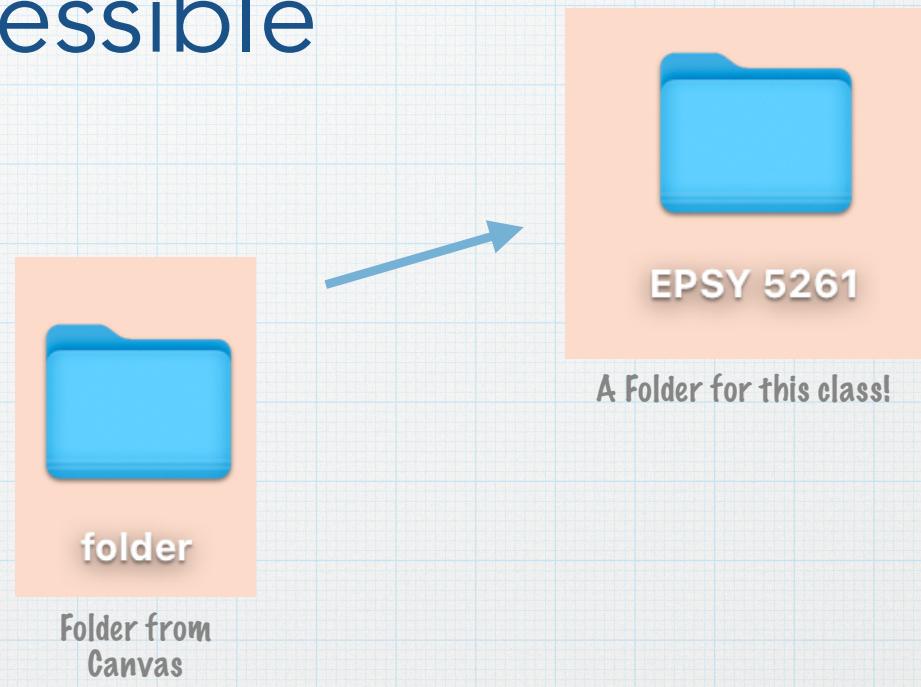
- * Note: Quarto Files and datasets must reside in the same folder to run properly



Put that folder somewhere easily accessible

Some suggestions:

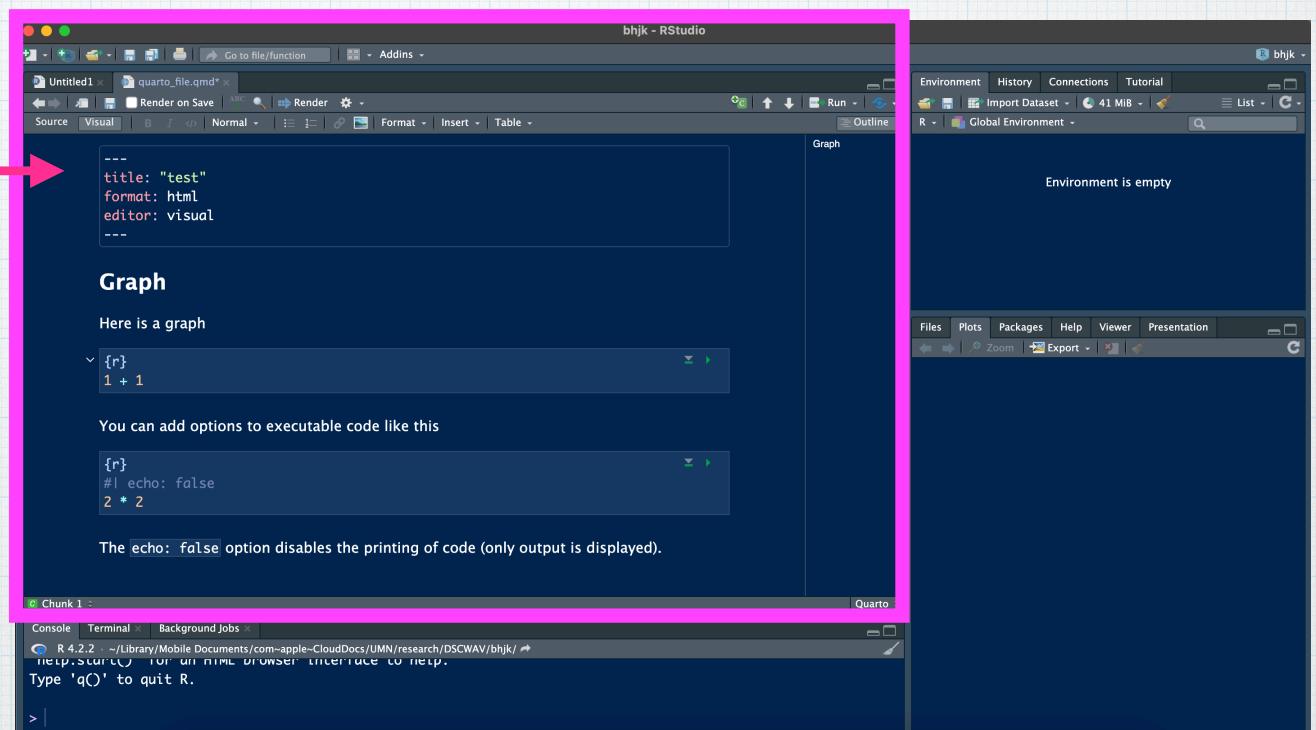
- Desktop
- “Grad School” Folder
- On your desktop?
- In your cloud?



Open .qmd file

Double-click the QMD file on your computer to open it in R Studio.

This area is where you will be working (adding code chunks, writing answers, etc.)



Functions

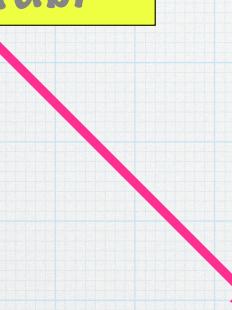
- To perform analyses on your data in R, you will need to use *functions*
- Functions tell R Studio what to do with your data
- Remember, **functions** have this structure:
 - `function_name()`
- They contain 1 or more **arguments**, which specify options for the function.

`function(argument1, argument2, argument3)`

Libraries/Packages

- Functions live in **libraries** (a.k.a. **packages**)
- Some packages come with your R and others need to be installed.
- Install the `{ggformula}` and `{tidyverse}` packages. (Reminder: Use the **Install** button in the **Packages tab** in RStudio.)

After installing the
a package, it will
show up in your
“Packages” tab.



Name	Description	Vers...	Actions
gdtools	Utilities for Graphical Rendering and Fonts Management	0.3.3	
generics	Common S3 Generics not Provided by Base R Methods Related to Model Fitting	0.1.3	
geosphere	Spherical Trigonometry	1.5–18	
gert	Simple Git Client for R	1.9.2	
gfonts	Offline 'Google' Fonts for 'Markdown' and 'Shiny'	0.2.0	
gganimate	A Grammar of Animated Graphics	1.0.8	
ggdag	Analyze and Create Elegant Directed Acyclic Graphs	0.2.10	
ggExtra	Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements	0.10.0	
ggforce	Accelerating 'ggplot2'	0.4.1	
ggformula	Formula Interface to the Grammar of Graphics	0.10.4	
ggmap	Spatial Visualization with ggplot2	3.0.2	
<input checked="" type="checkbox"/> ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	3.4.2	
ggraph	An Implementation of Grammar of Graphics for Graphs and Networks	2.1.0	

Load a Package

- To use the functions from a particular package, we will need to load the package (loading and installing are two different things!)
- To load the package and its functions we use the `library()` function

Load Packages (cntd.)

General Example: `library(libraryName)`

- `library()` is our function
- `libraryName` will specify the library we want to load
- To load the `{ggformula}` library we would use:

`library(ggformula)`

Once you load a package, the box next to the library (in the "Packages" tab) will be checked.

	Name	Description	Vers...	
<input type="checkbox"/>	gdtools	Utilities for Graphical Rendering and Fonts Management	0.3.3	
<input type="checkbox"/>	generics	Common S3 Generics not Provided by Base R Methods Related to Model Fitting	0.1.3	
<input type="checkbox"/>	geosphere	Spherical Trigonometry	1.5-18	
<input type="checkbox"/>	gert	Simple Git Client for R	1.9.2	
<input type="checkbox"/>	gffonts	Offline 'Google' Fonts for 'Markdown' and 'Shiny'	0.2.0	
<input type="checkbox"/>	gganimate	A Grammar of Animated Graphics	1.0.8	
<input type="checkbox"/>	ggdag	Analyze and Create Elegant Directed Acyclic Graphs	0.2.10	
<input type="checkbox"/>	ggExtra	Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements	0.10.0	
<input type="checkbox"/>	ggforce	Accelerating 'ggplot2'	0.4.1	
<input checked="" type="checkbox"/>	ggformula	Formula Interface to the Grammar of Graphics	0.10.4	
<input type="checkbox"/>	ggmap	Spatial Visualization with ggplot2	3.0.2	
<input checked="" type="checkbox"/>	ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	3.4.2	
<input type="checkbox"/>	ggraph	An Implementation of Grammar of Graphics for Graphs and Networks	2.1.0	

Install vs Loading

- You only need to install a package once
- But, you need to load that package every time you open a new session in RStudio. (After you load it, all the functions in the library can be used.)
- Phone App analogy

Importing Data

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

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

- `read_csv()` is our function
- “`Dataset.csv`” is our argument that specifies the name of the data file

In R Studio

Click the “play” button to run the syntax in the code chunk!

```
▼ {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")
```



Assignment

- The syntax to import the data was:

```
countries <- read_csv("World-Countries-Data-2019.csv")
```

- The `read_csv()` function literally only reads the data (and then forgets it).
- In order to keep the data from being forgotten, we need to **assign** it into an object. (An object is like an internal container in R.)
- To do this we use the assignment operator `<-`
- `countries` is the name of the object we want to use to assign the data into
- Now anytime we want to operate on the data we just use `countries`

Introduction to R Studio Activity

Summary: R Studio Workflow

1. Download materials from the schedule page on the website.
2. Load libraries (Be sure to install any libraries you do not have.)
3. Import data
4. Ready to perform analysis!