

# Introduction to Statistical Inference (QTM 100 Lab)

## Lecture 0

---

Justin Eloriaga — Emory University

Fall 2024

Laboratory Course Structure and Lab Policies

Lab 0: Introduction to R and RStudio

# **Laboratory Course Structure and Lab Policies**

---

Apply the statistical concepts covered in the lecture using R

- Work with real data and perform statistical techniques learned in class using R.

Apply the statistical concepts covered in the lecture using R

- Work with real data and perform statistical techniques learned in class using R.
- 25% of your lecture grade will be based on the Lab Course!

# Lab Grading

- Lab Attendance - 3%
  - Lab Homework - 6%
  - Lab Exams - 8%
  - Lab Project - 8%
  - **Total - 25%**
- Lab Attendance is **mandatory** every Friday.
  - Lab Homework is graded on *completeness*.
  - Lab Exams and the Lab Project are announced ahead of time.

# Lab Attendance

- Attendance will be checked every meeting.
- You must attend all classes.
- If you miss one class, it is okay! You get one freebie.
- Each absence will correspond to a demerit of 1% (up to a cap of 3%)

- Posted on **Wednesdays**
- You are responsible for reviewing lab manual every week!
- Manuals will contain instructions, tips, and examples
- **Everything you need to know to complete the lab homework**



# Lab Homeworks

- Due on **Wednesdays at 7:00pm**
- Completion grade! (i.e. for as long as you answered everything regardless of whether the answer is correct or not, you will get credit)
- Do NOT include your name in the submission. This will be *peer reviewed!* By your classmates! You must complete the peer review (otherwise, there is a -10% demerit)
- Feel free to collaborate with other people! I encourage you to do this, in fact!
- Still an individual submission!
- Submit your R script (.R file)
- Individual submission

# Lab Mini-Exams

- There will be two, in person, lab exams.
  - First Mini-Exam on **October 25** worth 3%
  - Second Mini-Exam on **November 15** worth 5%
- You must physically attend your lab section to participate in these exams, and you must bring your laptop
- Several multiple-choice questions, followed by a blank space where you submit the code you used to answer the questions
- Closely modeled after the lab homework
  - No extra time is allowed to download and import the dataset because the same dataset(s) used in your homework will also be used for the exam.
  - You must complete the exam independently; No collaboration of any kind is allowed during the exam.
  - Make sure to arrive a few minutes early. If you arrive late, you will not be given the full 30 minutes to complete the exam.

- Counts for 8%
- completed in small groups, and it is designed to give you hands-on experiences on developing/answering research questions using real-world data
- Divided into *two* parts
  - Preliminary - 2% due by **October 23, Wednesday**
  - Final - 6% due by **December 10, Tuesday**

Details on this project will be distributed later!

- I will have *one* weekly office hour.
  - Tuesday, 6:15 pm - 7:15 pm online via Zoom.
  - We can also meet by Zoom or you visit my office at the R.Randall Rollins office and consult with me on an appointment basis. Send an email to [jeloria@emory.edu](mailto:jeloria@emory.edu) at least 24 hours in advance with your availability and I will reply with either Zoom invitation or a calendar invite if I am available.

- Darya Dahi (email: [darya.dahi@emory.edu](mailto:darya.dahi@emory.edu))
- She will check attendance every class.
- Keep communication with all of you in case you have questions (through the discussion board)
- Hold office hours

# Typical Week

- Mondays, Tuesdays, and Thursdays are mostly for your *lecture requirements*
- On Wednesday
  - New **Lab Manual** posted (in Lab Canvas Page)
  - DUE: **Lab Homework**
- On Friday: Your lab peer review is due!

- **ALWAYS** contact me first about any and all lab matters! I will try to resolve any problems you may have to the best that I can.
- If I can't, I will contact Dr. Kim on your behalf

# Accommodations

- Please contact me immediately if you have any DAS accommodations (or if you need help obtaining them)
- Do NOT leave this to the last minute!
- This is in addition to telling Dr. Kim, not in lieu of



# **Lab 0: Introduction to R and RStudio**

---

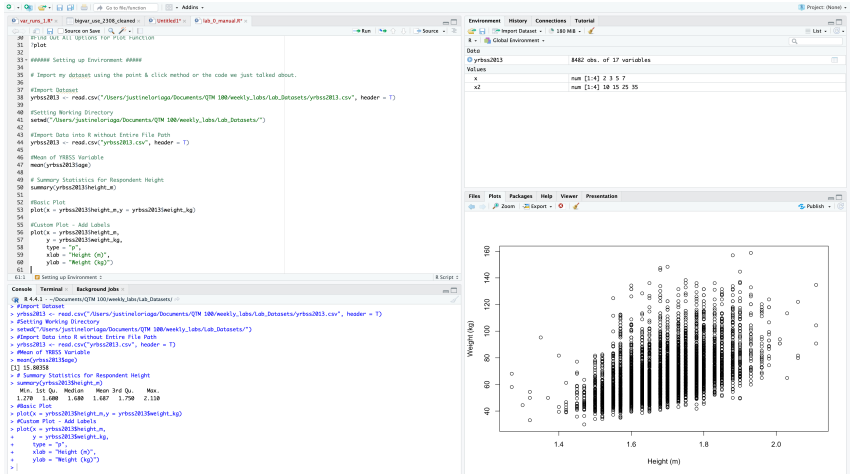
# Installing R and RStudio

You will need both R and RStudio



Does anyone have any problems installing these?

# Sample Workspace in RStudio



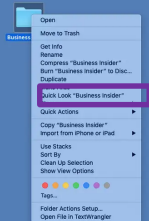
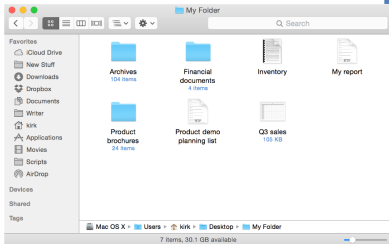
Usually the first step in any coding work is to import data!

- There are two ways to do this in R
  - Point and Click (you'll probably choose this).
  - Using a File Path/Setting up a Working Directory.

# Finding a File Path on Mac

- Click on file of interest
- Then right click on file of interest
- Scroll down to and click on “Get Info”

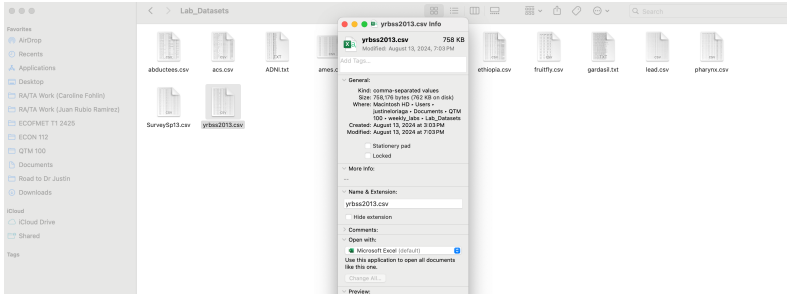
# Finding a File Path on Mac



# Finding a File Path on Mac

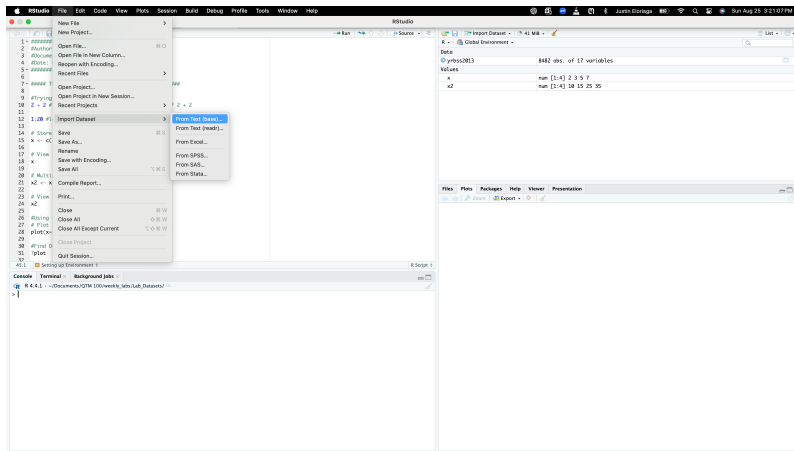
- Look for “Where” and copy and paste the information found on that line
- **That will be your file path**

# Finding a File Path on Mac



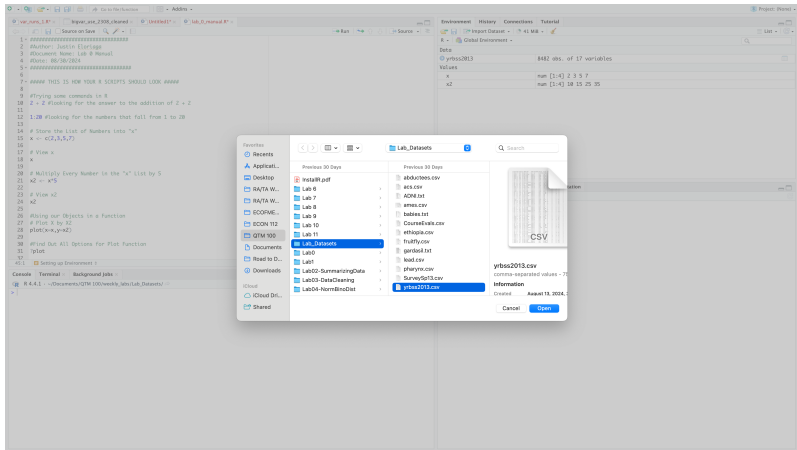


# Point and Click Importing a Dataset



Start by going to **File** then **Import Dataset**. We use **From Text (base)** since we have a .csv file.

# Point and Click Importing a Dataset



Find the file you want to import

# Point and Click Importing a Dataset

The screenshot shows the RStudio interface with a script editor on the left, a console at the bottom, and a 'Import Dataset' dialog box in the center. The dialog box has the following settings:

- Name: yrbcs2013
- Input File: yrbcs2013.csv
- Encoding: Automatic
- Heading: ☒ Yes ☐ No
- Row names: Automatic
- Separator: Comma
- Decimal: Period
- Quote: Double (")
- Comment: None
- NA strings: NA
- Strings as factors: ☐

The 'Data Frame' preview shows the following data:

age	gender	height_cm	weight_kg	bed	BHPTCT	sustbelt	saustbelt2_r
15	female	1.73	64.37	28.2	04.63	never	
15	female	1.68	55.79	21.8	68.15	sometimes	
15	female	1.58	49.72	29.8	57.56	always	
15	female	1.52	67.13	27.2	59.29	always	
15	female	1.68	69.85	24.7	80.67	always	
15	female	1.65	66.68	24.5	85.65	rarely	
16	male	1.85	76.39	21.7	61.13	always	
16	male	1.79	79.21	22.2	73.88	always	
14	male	1.77	72.44	26.0	81.84	most of the	
15	male	1.83	67.59	20.2	58.85	always	
14	male	1.68	46.27	16.4	6.18	most of the	
16	male	1.83	73.48	21.9	63.87	most of the	
16	female	1.54	52.14	21.7	82.52	always	

You will find a preview of the file. Select **heading** to be **yes**. Datasets usually have the first row as the names of the **variables**

# Point and Click Importing a Dataset

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Shows the R script used to import the dataset:

```
R 4.4.1 -> Documents/QTM 100/weekly_lab/Lab_Datasets/
> yrbss2013 <- read.csv("~/Documents/QTM 100/weekly_lab/Lab_Datasets/yrbss2013.csv")
> View(yrbss2013)
```
- Environment Pane:** Shows the loaded dataset 'yrbss2013' with 8482 observations and 17 variables. The 'Values' section shows the first few rows of the 'x' and 'x2' variables.
- Data View:** Displays a table of the first 22 rows of the dataset, showing variables like 'age', 'gender', 'height\_in', 'weight\_lb', 'level', 'BMSPEC', 'searbleth', 'searblet2', 'ride\_drunkdriver', 'drive\_drunk', 'drive\_test', and 'carried\_weap'.

Done! You will also see the dataset in the variable window!

# Importing a Dataset using the File Path

```
#Import Dataset  
yrbss2013 <- read.csv("/Users/justineloriaga/Documents/QT100/weekly_labs/Lab_Datasets/yrbss2013.csv", header = T)
```

- `yrbss2013` is the name of the *object* (in this case, our dataset). You can name it anything, this just makes it clear
- `< -` is used to assign an object something
- **`read.csv()`** is the command for importing a csv file
- `header = T` is an option to make the first row as variable names

# Format of R Scripts

Include a header at the top of each R script that includes the following information:

1. Name of person who wrote the code for the R script
2. The name of the R script that tells some information of what it includes
3. The date that the R script was written or completed or the working version.

## **Comment, comment, comment!**

There should be a comment for every line of code included in your R script

```
1 #####
2 #Author: Justin Elorriaga
3 #Document Name: Lab 0 Manual
4 #Date: 08/30/2024
5 #####
6
7 ##### THIS IS HOW YOUR R SCRIPTS SHOULD LOOK #####
8
9 #Trying some commands in R
10 2 + 2 #looking for the answer to the addition of 2 + 2
11
12 1:20 #looking for the numbers that fall from 1 to 20
13
14 # Store the List of Numbers into "x"
15 x <- c(2,3,5,7)
16
17 # View x
18 x
19
```

# Key Things to Remember

- Typos are the #1 cause of errors in R!
- Comment **EVERYTHING**. The more you comment, the easier coding becomes!
- Save **EVERYTHING**. Type everything into the upper left-hand section of your screen and save your file!