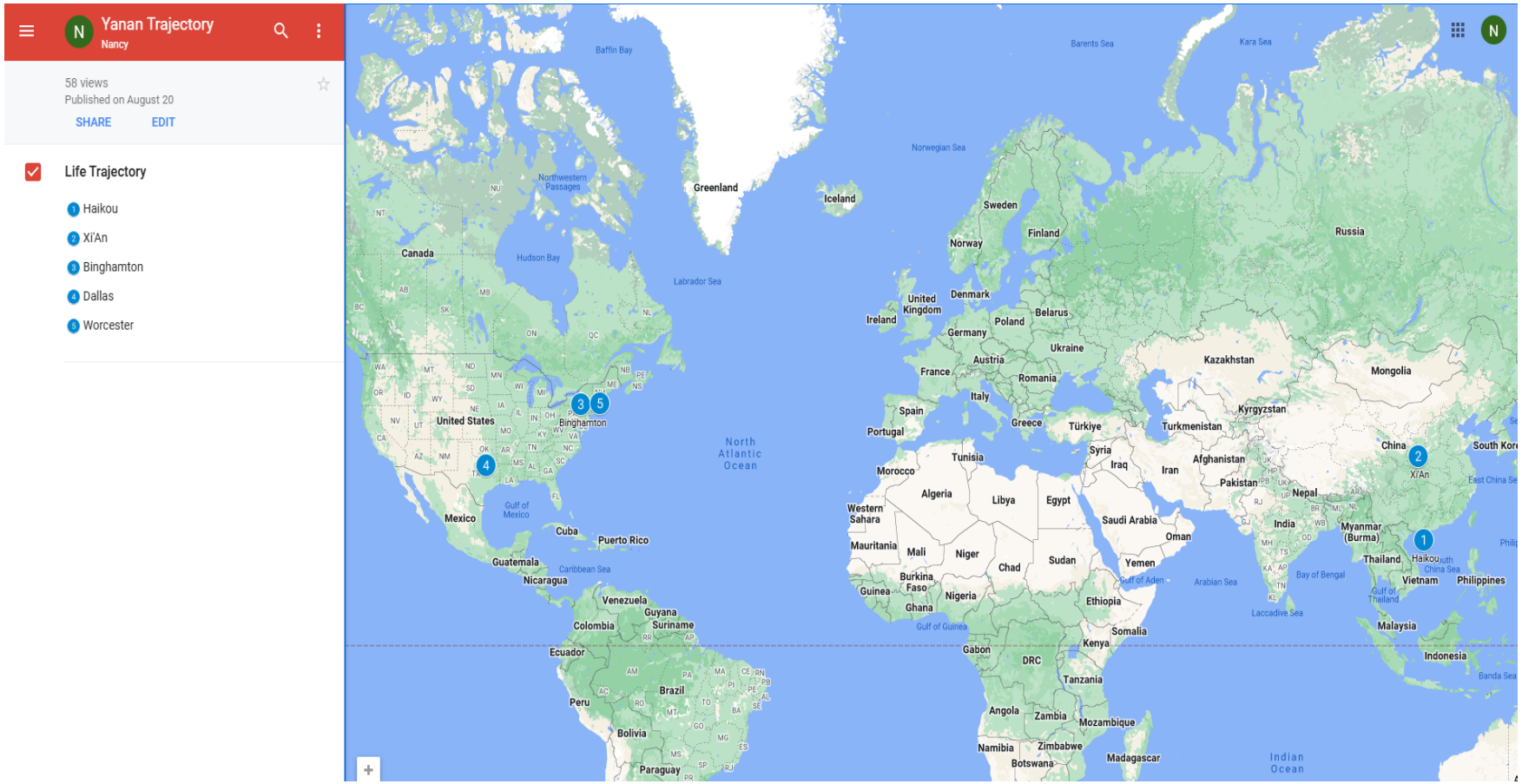
WEEK 01

Instructor: Yanan Wu
TA: Nisar Khadija

Spring 2025

YANAN WU – VISITING ASSISTANT PROFESSOR

- Education & Experience



TEACHING

- Intermediate Statistics
- Python Programming in GIS
  - Manipulating spatial data
  - Web mapping
  - Processing Raster
  - Data Analysis
  - Creating Custom Tool
  - Data visualization
- Spatial Database

INSTRUCTOR OFFICE HOUR

- Instructor: Yanan Wu
- Email: yanawu@clarku.edu
- Hours: Tuesday & Thursday 1:30 pm – 2:30 pm
- Location

TA INTRODUCTION

- Office Hours:
- Office Location

HOW ABOUT YOU?

- Your background (e.g., name, major, where you come from)
- What is your funniest thing that happened during your winter break?
- What relevant experience do you have with statistics?
- What are your expectations for this course?

RESOURCES

- Introductory text is available for free via https://www.openintro.org/book/os/
- Intermediate book is Joseph Hair, William Black, Barry Babin and Rolph Anderson. Multivariate Data Analysis. Edition 7 or 8. Upper Saddle River NJ: Prentice Hall.

Amazon.com usually offers used copies for less than $30.

https://www.amazon.com/Multivariate-Analysis-Joseph-Anderson-William/dp/9353501350/ref=pd_cp_14_2/144-5828787-2724822?_encoding=UTF8&pd_rd_i=9353501350&pd_rd_r=397e3554-2af1-476b-8336-5c1018af6453&pd_rd_w=Q3cmn&pd_rd_wg=XzEv7&pf_rd_p=0e5324e1-c848-4872-bbd5-5be6baedf80e&pf_rd_r=FYB3ZG6A42Z1ANQMRH1V&psc=1&refRID=FYB3ZG6A42Z1ANQMRH1V

COURSE REQUIREMENTS

- Assignments: 9 in total

  For any graded assignment, if the you do not agree with the grade received, the instructor and TA must be notified within one week after the assignment is graded.

- Late policy for lab (excluding midterm and final project)
- One final project (oral presentation and paper report)

EVALUATION

- Assignments 80% = nine assignments
  - Each assignment need to be completed using R
  - You can ask for help with assignments from the instructor and TA, No plagiarism is allowed

- Final Project 20% = 15% oral + 5% written
  - To apply & to interpret statistical procedures
  - To make an oral presentation of a statistical analysis
  - To write a report

A	**93.0 - 100.0**	**B+**	**88.0 - 89.9**	**C+**	**77.0 - 79.9**	**D+**	**67.0 - 69.9**
		B	83.0 - 87.9	C	73.0 - 76.9	D	60.0 - 66.9
A-	90.0 - 92.9	B-	80.0 - 82.9	C-	70.0 - 72.9	F	0.0 - 59.9

IMPORTANT DATES

- **Jan 22.** Add/Drop ends (& last day to request audit) Full Semester
- No class
  - Jan 20. University holiday
  - Feb 17. Wellness day
  - March 3-7 Spring Break
  - March 24-27 AAG Conference
- Final project
  - April 14-24 Working on final project
  - April 28 – May 1 Final project presentation
  - May 5 Final report due

COURSE WEBSITE

- Clark Canvas
- Course Website on Github

SOFTWARE

- R

- R-Studio
  - Open-source IDE (integrated development environment)

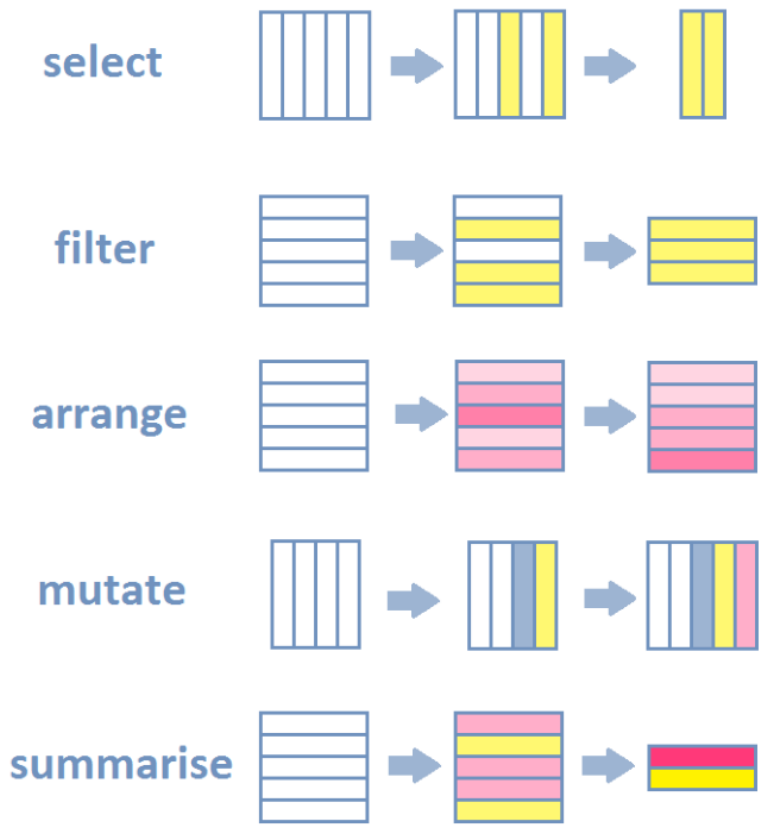
PREAMBLE

- A good online textbook, Hands-on Programming with R, for R beginner.
- Explore the R project website: https://www.r-project.org/
- Explore R Studio: https://posit.co/

R-INTRODUCTION

- An online free learning source: An Introduction to R

Data Manipulation



Data Visualization

- Data Visualization Section in R for Data Science
- Modern Data Visualization with R

Interactive Applications

- Shiny Gallery in R

R - INTRODUCTION

Statistical Analysis

- Descriptive analysis (mean, median., etc)
- Regression analyses (linear, logistic, ect)
- Time series analysis (ARIMA, etc)
- Multivariate analysis (PCA, factor analysis)
- A handbook of statistical analysis in R

Geospatial Data Analysis

- Handle raster and vector data
- Analyze spatial data with sf, sp or raster

R - INTRODUCTION

Machine Learning

- Implement supervised learning (classification, regression).
- Apply unsupervised learning (clustering, dimensionality reduction).
- Perform deep learning with packages like keras or torch.
- Evaluate models using cross-validation and other metrics.

COURSE TOPICS

Week 2

- Foundations of Inferential Statistics
  - Point estimates and sampling variability
  - Confidence intervals and hypothesis testing
  - …
- Inference for numerical data
  - t-distribution
  - …

WEEK 01

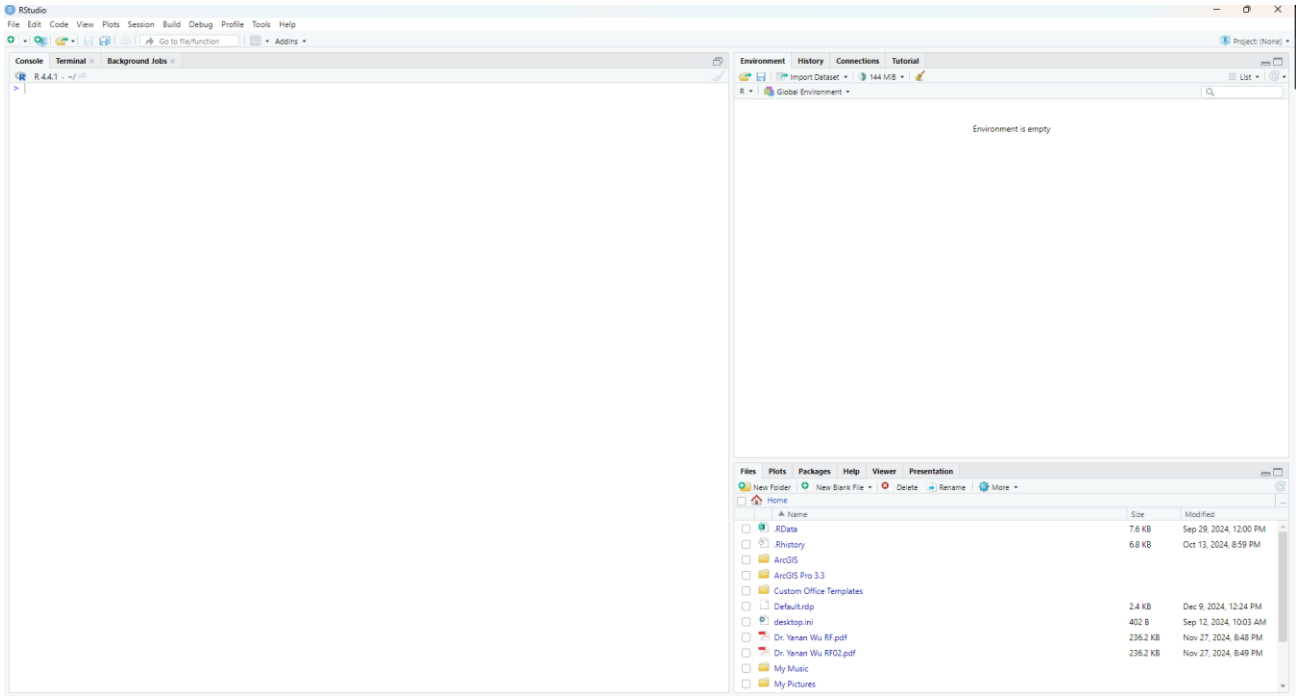
LAB SESSION

Instructor: Yanan Wu
TA: Nisar Khadija

Spring 2025

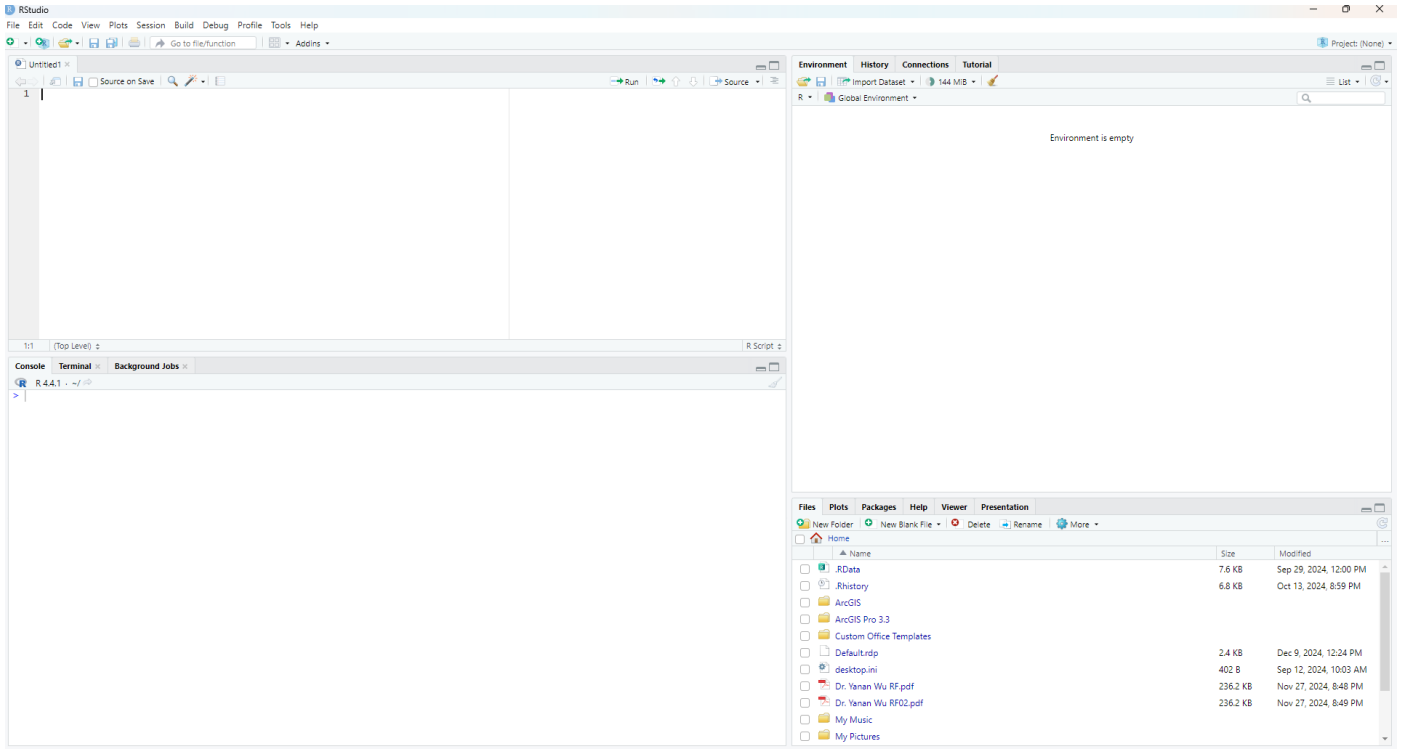
OVERVIEW OF RSTUDIO INTERFACE

- The panes
  - Left pane: R console
  - Right top pane: includes tabs such as *Environment* and *History*
  - *Right bottom pane: File, Plots, Packages, Help and Viewer*



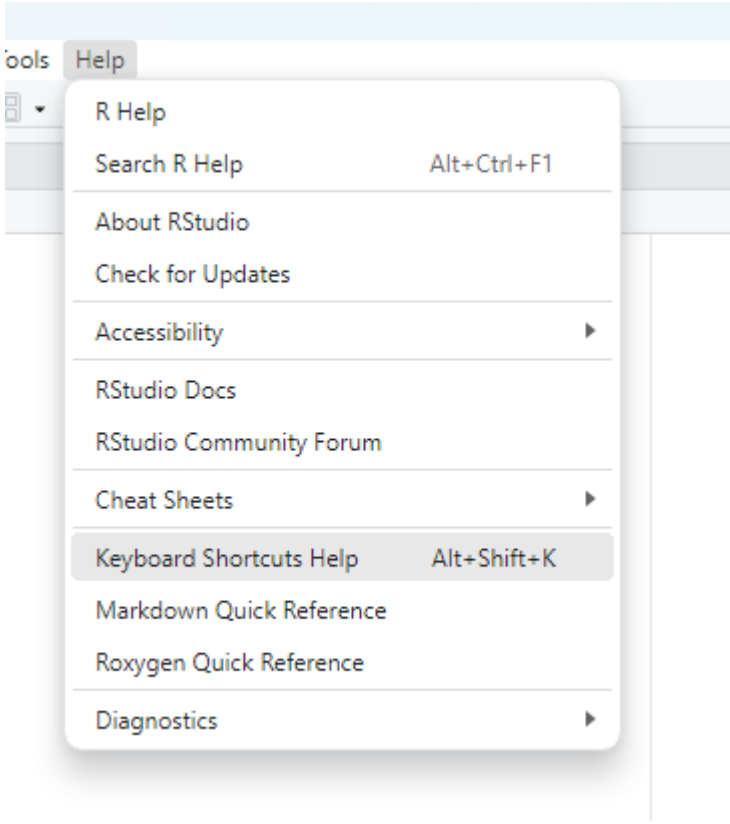
OVERVIEW OF RSTUDIO INTERFACE

- Starts a new pane on the left
  - File – New File – R Script



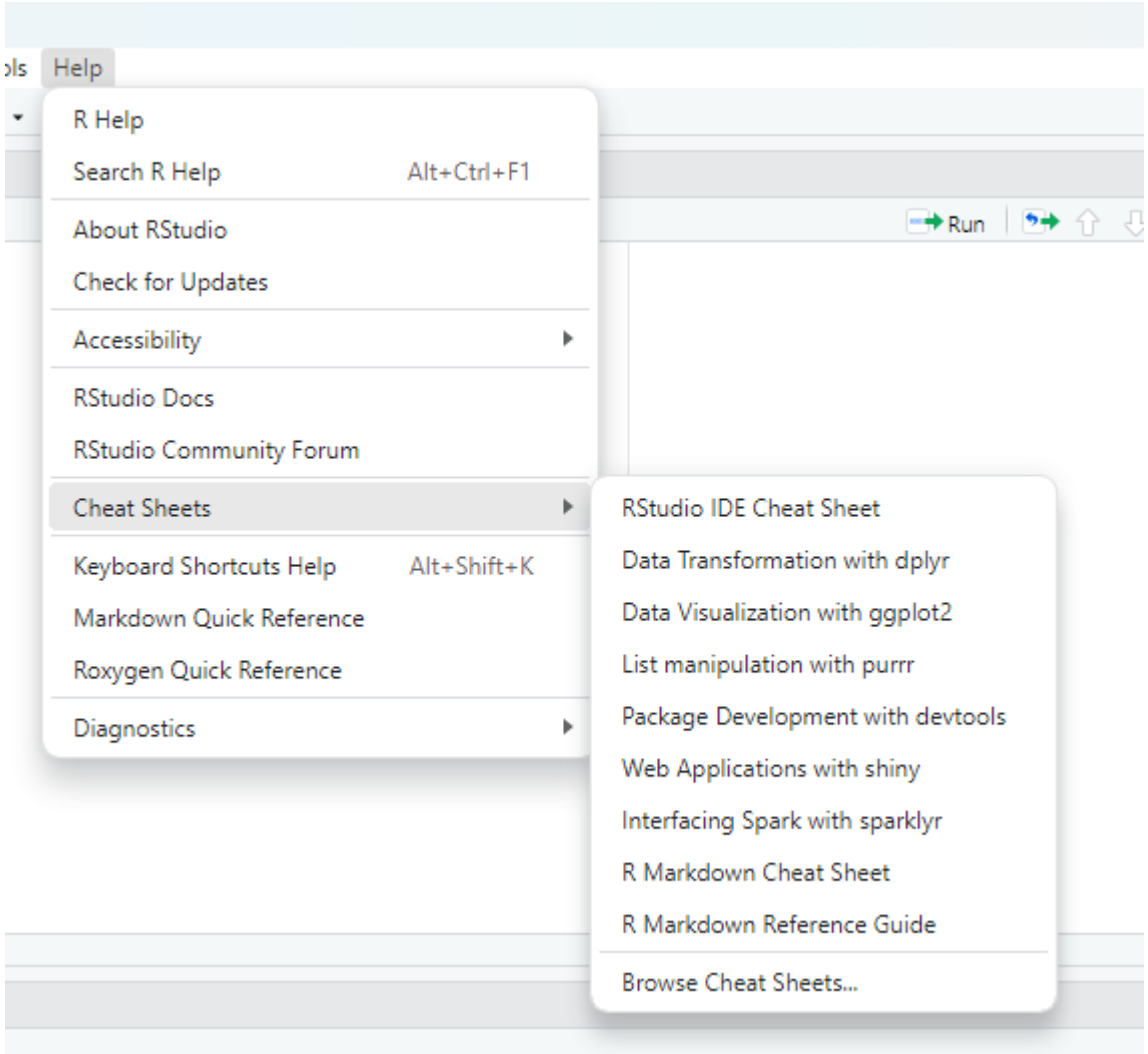
OVERVIEW OF KEYBOARD SHORTCUTS

- Keyboard shortcuts
  - Help – Keyboard Shortcuts Help



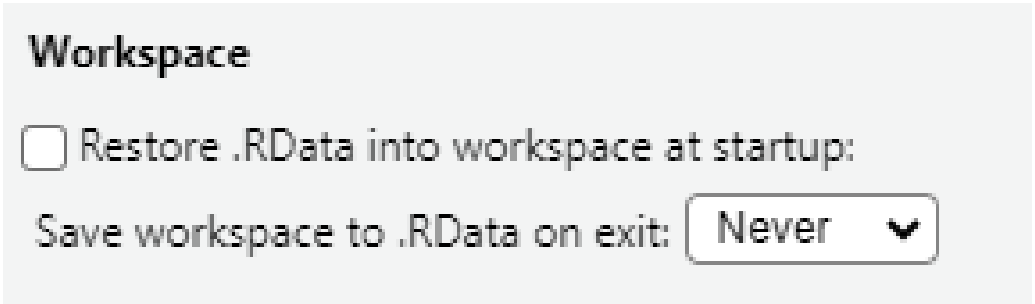
OVERVIEW OF CHEATSHEETS

- Cheatsheets in Rstudio
  - Help - Cheatsheets



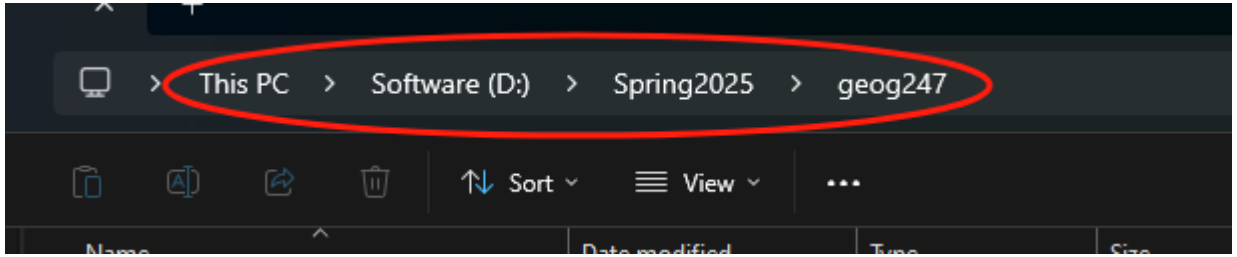
GLOBAL SETTING

- .RData
  - Save your workspace, including variables, data frames, lists, and other objects
- Cons
  - Causing confusions especially when we share code with others and assume they have this .Rdata file
- Tools – Global Options
  - Change the setting as below



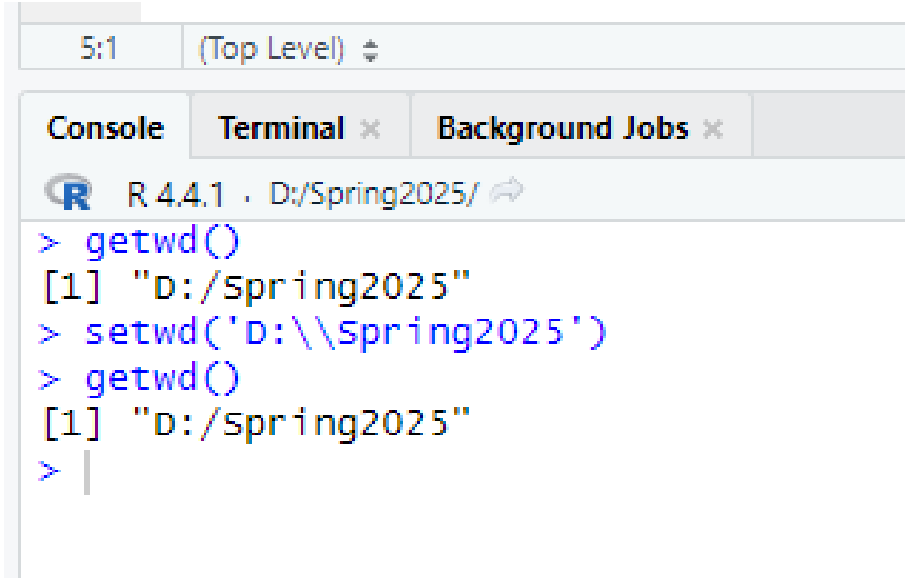
WORKING DIRECTORY

- Get working directory where your scripts and workspaces are stored
  - **getwd()**
  - Run this command:
    - Ctrl + Enter
    - Or Run in R
  - The returning strings, e.g.,"C:/Users/yy00021/Documents“ is the path to the working directory
  - The windows convention uses slash \ to separate sub-directories
  - However, R uses forward slash / or a double backward slash \\
- Change working directory
  - I suggest you to setup a specific directory for this course
  - **setwd('D:\\Spring2025\\geog247’)**
  - Now check your working directory again



CONSOLE WINDOW

- The character > in CONSOLE window indicates that R is ready to receive new commands
- It show up when R completed executing a script



TERMINATE SCRIPT

- The Esc Key or pressing  in the CONSOLE window to terminate the script

```
## Terminate script
i <- 1
while (i>0) {
  print('good')
}
```

GET HELP

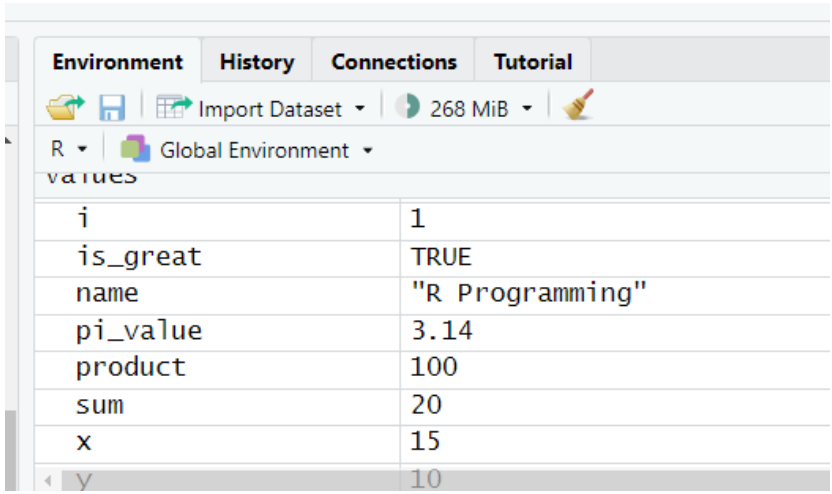
- Get help for activate libraries
  - **help('dplyr')**
  - **?dplyr**
- Get help for all installed libraries
  - **??dplyr**
  - **help.search('dplyr')**

INTERACTING WITH THE R-CONSOLE

- All commands (or programs) can be stored in external *.R script-files
- Single command or a set of highlighted commands can be run using shortcut (shift+enter) or Run button in R
- All commands can be run use the Source button in R
- Scroll through the history of previously commands in R
- Using shortcut key (Ctrl + L) or broom icon to clean the  Console window

VARIABLES IN R

- Variable names
  - Variables are created using the assignment operator <
  - Variables can store different types of data (numeric, character, logical, etc.).
  - Variables can be reassigned new values anytime.
  - The document shows professional naming for your code
- Object in the ENVIRONMENT
  - Any data structure or function that is defined using commands becomes an object in the ENVIRONMENT



- Remove objects
  - The objects can be removed from the ENVIRONMENT
  - rm(x)
- Clean ENVIRONMENT
  - Broom icon in the ENVIRONMENT mean bar
  - or **rm(list=ls())**

LIST IN R

- Creating a list
  - A list in R is a flexible data structure that can contain elements of different types: numbers, characters, vectors, matrices, data frames, or even other lists.
  - It's like a container for multiple objects.

- Accessing elements in a list
  - Use [[ ]] to access elements by position or name.
  - Use $ to access elements by name.

DATA SETS

- Read csv
  - **read.csv()** for reading CSV files.
- Check columns
  - Accessing column names using **colnames()**
- Add new columns
  - Adding columns based on calculations or conditions

WEEK 01

PRACTICES

Instructor: Yanan Wu
TA: Nisar Khadija

Spring 2025

PRACTICES

- Explore Tools and Help in RStudio
- Explore the different tables in RStudio