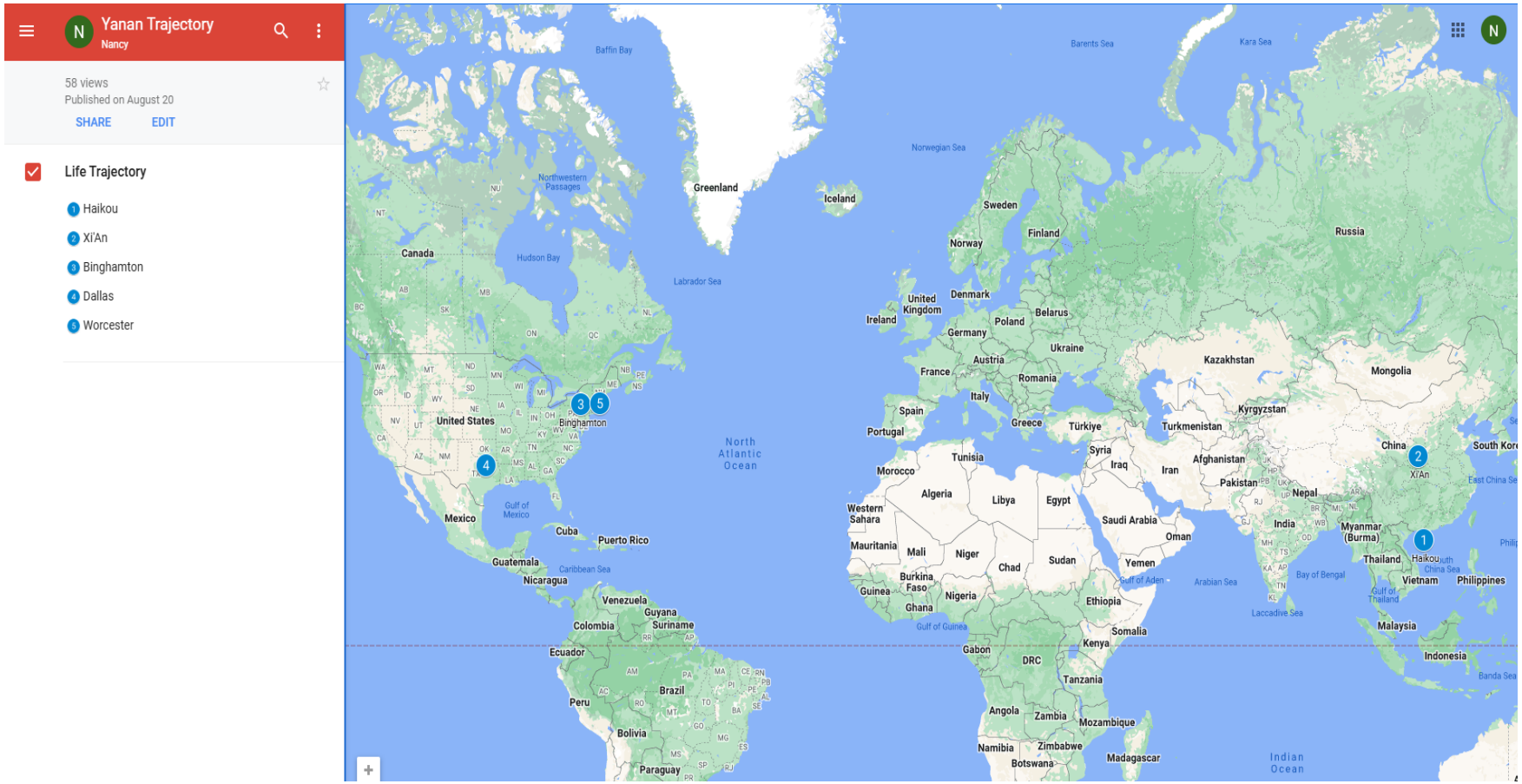
WEEK 01

Instructor: Yanan Wu
TA: Khadija Nisar

Spring 2025

YANAN WU – VISITING ASSISTANT PROFESSOR

- Education & Experience



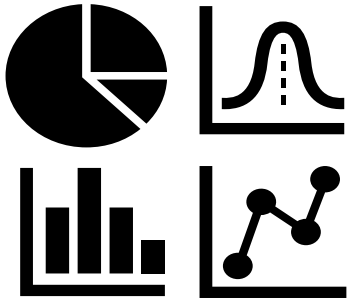
- **Python Programming**

1. Manipulating Spatial Data
2. Web Mapping
3. Processing Raster
4. Data Analysis
5. Creating Custom Tool
6. Data Visualization
7. …

- **Spatial Database**

1. Geodatabase
2. SQL
3. Proximity Analysis
4. Geometry processing
5. Raster processing
6. PostSQL with python
7. …

- **Intermediate Statistics**



1. Bivariate regression
2. Logistics regression
3. PCA
4. GWR
5. Spatial Autocorrelation
6. …

INSTRUCTOR OFFICE HOUR

- Instructor: Yanan Wu
- Email: yanawu@clarku.edu
- Office Hours: Tuesday & Thursday 1:30 pm – 2:30 pm or by appointment
- Location: Jeff 220

TA INTRODUCTION

- Office Hours: 3: 00 – 5: 00 PM on Tuesday
- Office Location: Room 102 B, Main Geography Building

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
- 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 WEBSITES ON GITHUB

- Spatial Database

- Intermediate Statistics

- Python Programming

- Web Mapping

GIS HELP DESK

NEW SPRING 2025 HOURS:

Need help with GIS? Can't make it to visit your professor or TA for office hours? Stop by our GIS Help Desk in the Jefferson Tower (6th floor), or in Jefferson 220A (2nd floor Jefferson building, Geog Main office) on Wednesdays, and visit one of our Help Desk Assistants!

MONDAYS	SASHA 9 AM - 11 AM
TUESDAYS	SASHA 9 AM - 12 PM
WEDNESDAYS @ JF220A	SASHA 9 AM - 12 PM WYNNIE 2 PM - 5 PM
THURSDAYS	WYNNIE 2 PM - 5 PM
FRIDAYS	WYNNIE 3 PM - 5 PM

GIS Help Desk hours follow a University schedule. If the University is closed from remote/online operations for any reason, the GIS Help Desk may also be unavailable. GIS Help Desk Assistants operate on a first-come, first-serve drop-in basis. Contact Marjorie Miller (marmiller@clarku.edu) with any questions, or call the Geography Main Office at 508-793-7336 for more information. The GIS Help Desk is sponsored by the Graduate School of Geography at Clark University.

SPRING 2025 HOURS

GIS HELP DESK

Please contact either of our GIS Help Desk Assistants during their specified hours for more information.

SASHA GANNON | GEOG '24 MS-GIS '25

MONDAYS 9AM – 11AM

TUESDAYS 9AM – 12PM

WEDNESDAYS* 9AM –12PM

WYNNIE GROSS | GEOG '24 MS-GIS '25

WEDNESDAYS*
3PM – 5PM
THURSDAYS
2PM – 5PM
FRIDAYS
3PM – 5PM

Help Desk hours follow a University schedule. If the University is closed (or if the Geography office is closed), the GIS Help Desk will be unavailable. Hours may be limited or extended during midterms/final exams. Any changes will be announced or posted on the Clark University Geography Facebook page. Appointments operate on a first-come, first-serve drop-in basis unless otherwise scheduled.

508.793.7336 | JEFFERSON TOWER, 6TH FLOOR; *WEDNESDAYS ARE IN JEFFERSON 220A, GEOG MAIN OFFICE, JEFFERSON BUILDING, 2ND FLOOR*

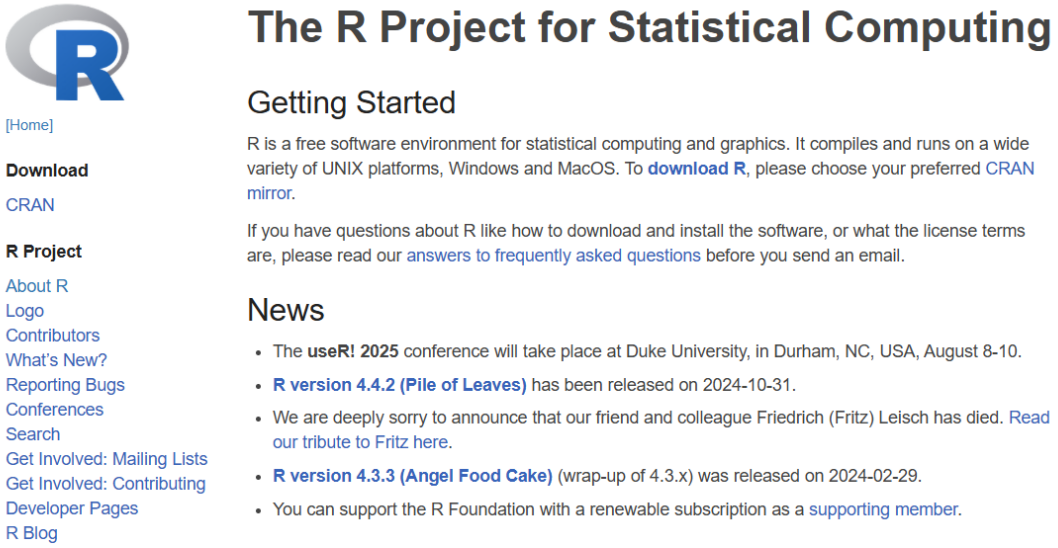
WEEK 01

WEEKLY SCHEDULE

Instructor: Yanan Wu
TA: Khadija Nisar

Spring 2025

WEEK 01



[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Get Involved: Contributing
Developer Pages
R Blog

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

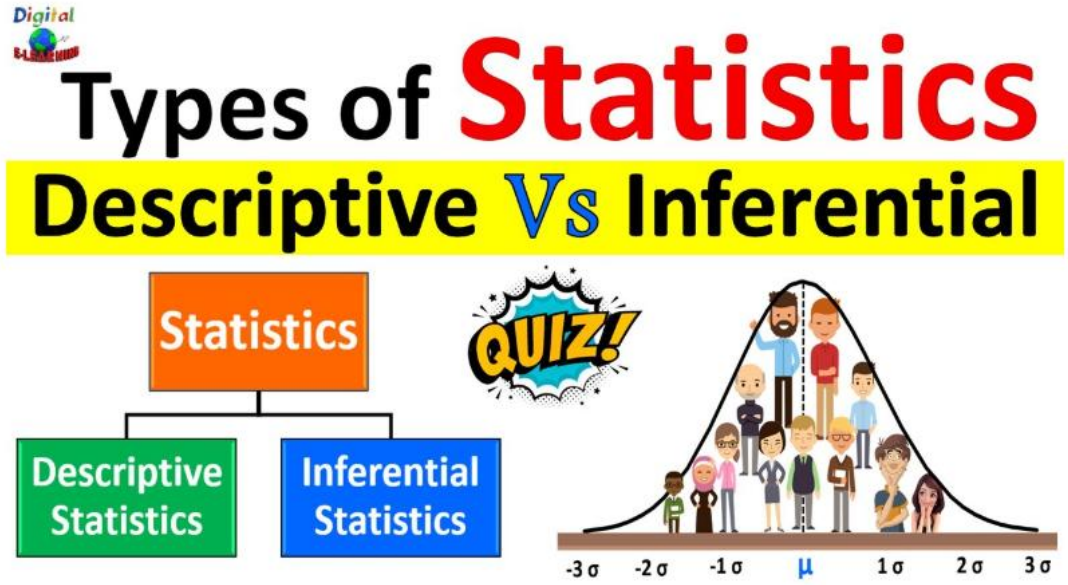
If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

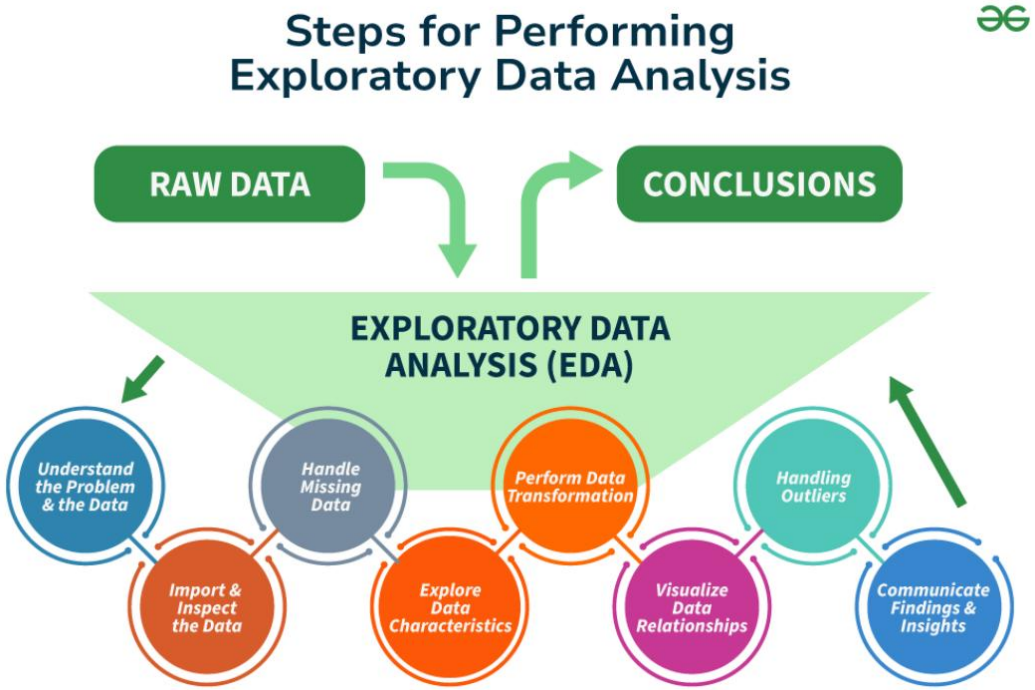
News

- The **useR! 2025** conference will take place at Duke University, in Durham, NC, USA, August 8-10.
- **R version 4.4.2 (Pile of Leaves)** has been released on 2024-10-31.
- We are deeply sorry to announce that our friend and colleague Friedrich (Fritz) Leisch has died. Read our tribute to Fritz here.
- **R version 4.3.3 (Angel Food Cake)** (wrap-up of 4.3.x) was released on 2024-02-29.
- You can support the R Foundation with a renewable subscription as a supporting member.

- COURSE INTRODUCTION
  SOFTWARE INSTALLATION
- R programming
- R-Studio
  - Open-source IDE (integrated development environment)

WEEK 02 & 03



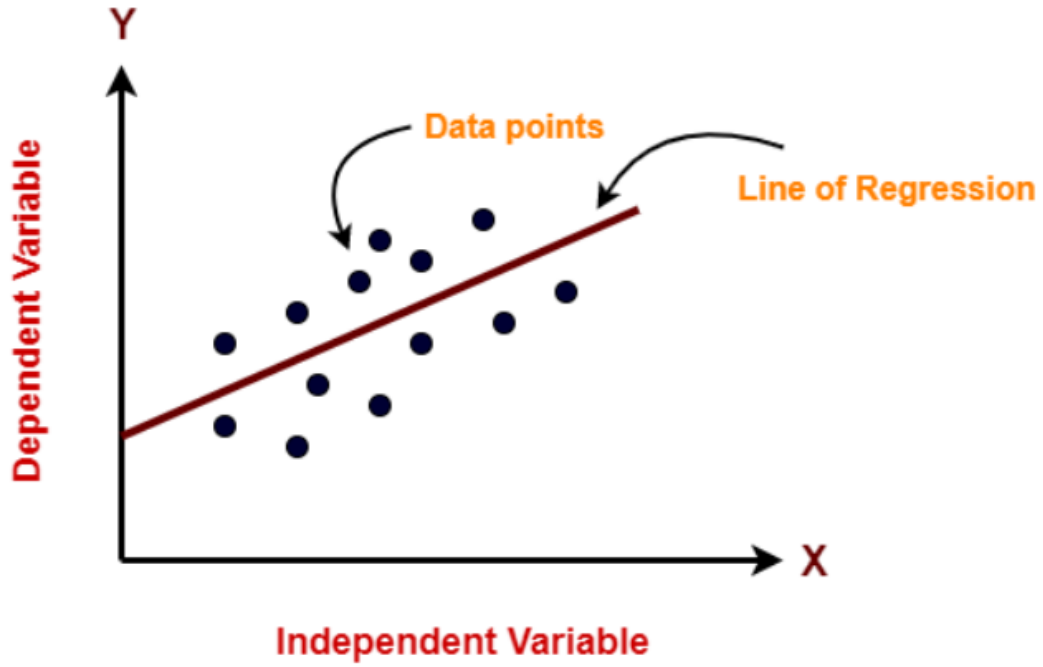


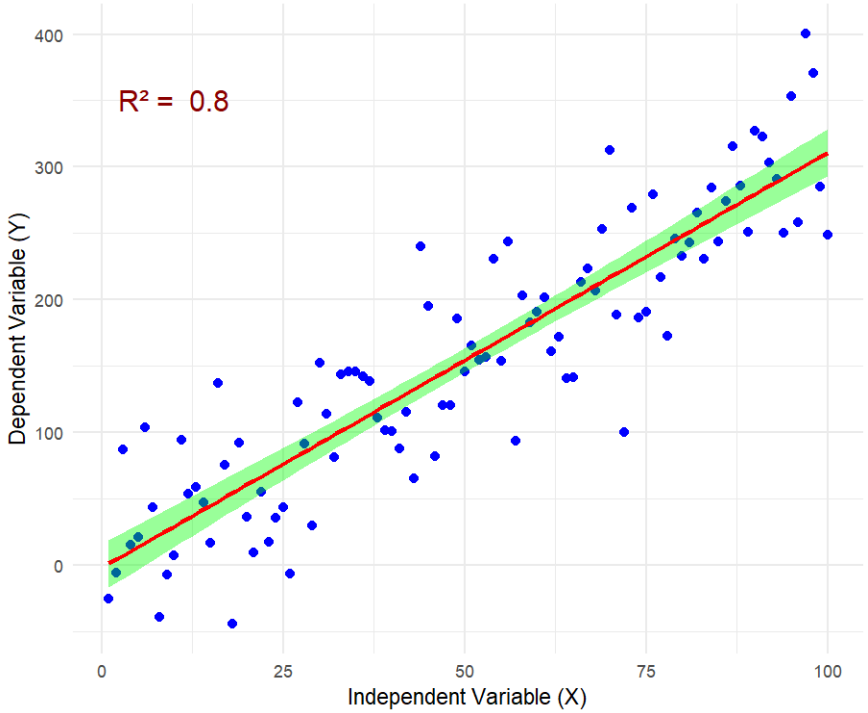
WEEK 02&3

- Inferential statistics
- Data Exploration

IMAGE SOURCE: HTTPS://WWW.LINKEDIN.COM/PULSE/DESCRIPTIVE-STATISTICS-VS-INFERENTIAL-MEASURE-CENTRAL-NHTOC/

HTTPS://WWW.GEEKSFORGEEKS.ORG/WHAT-IS-EXPLORATORY-DATA-ANALYSIS/



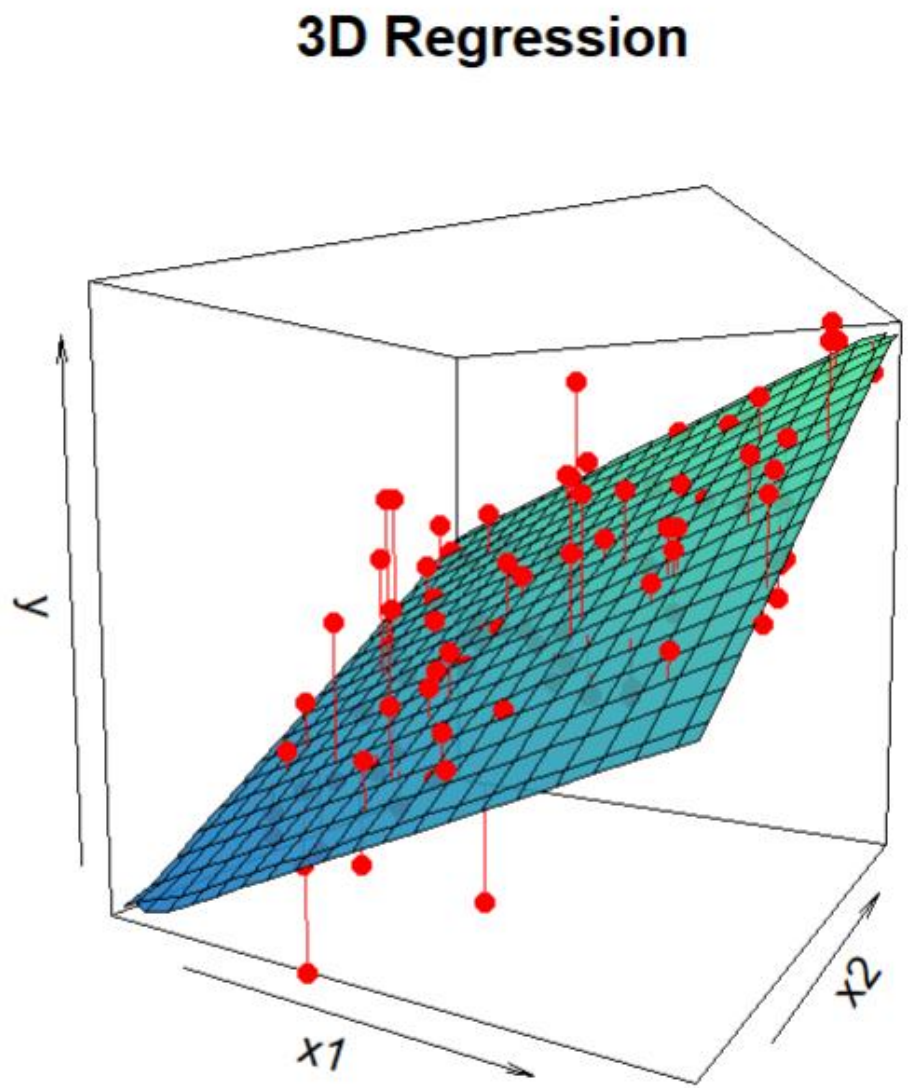


WEEK 04

- Bivariate Regression
  - Key Assumption
  - Slope and Intercept
  - R square
  - Confidence Interval

WEEK 05 & 06

MULTIPLE REGRESSION



WEEK 07

REGRESSION CRITICISM

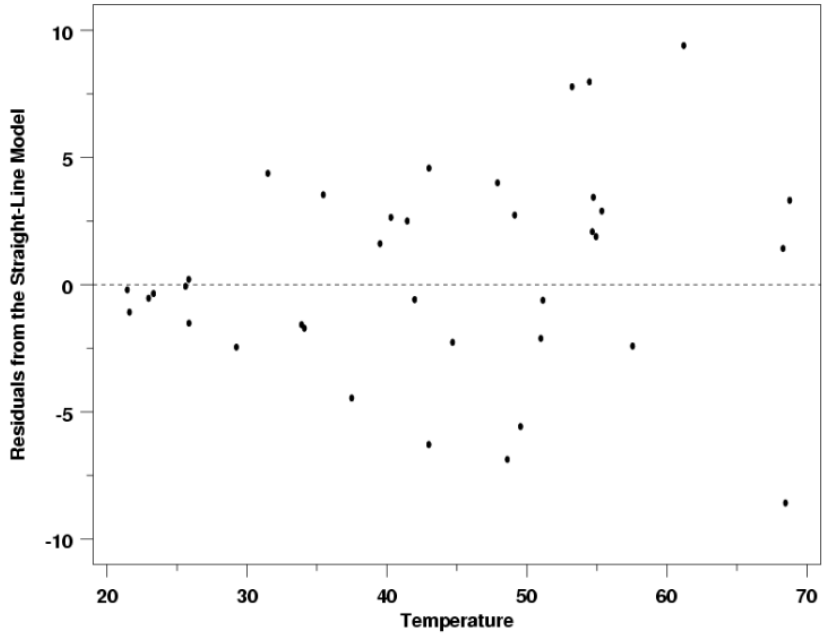
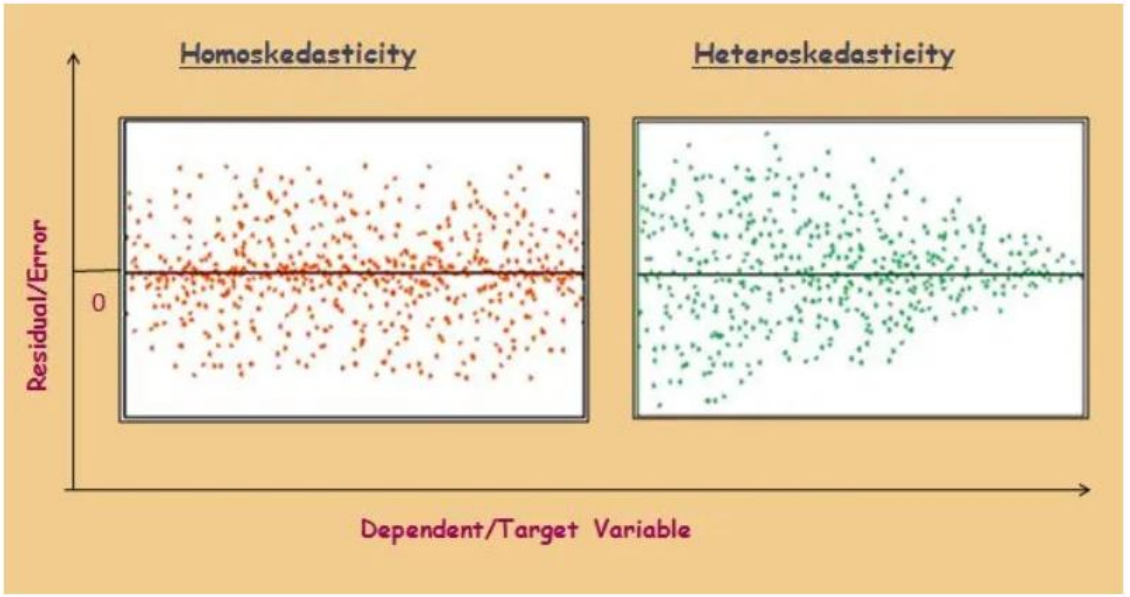
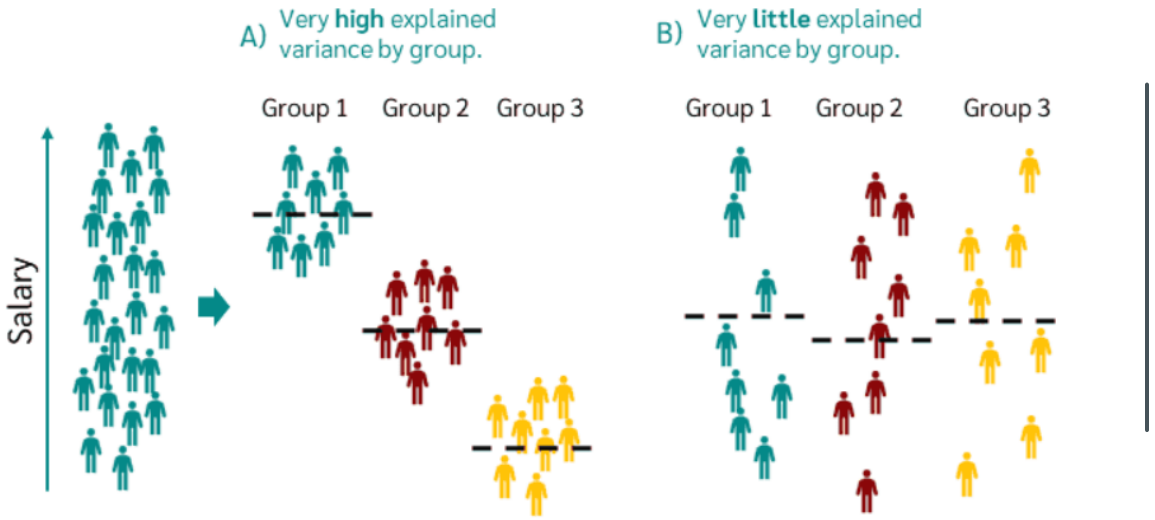


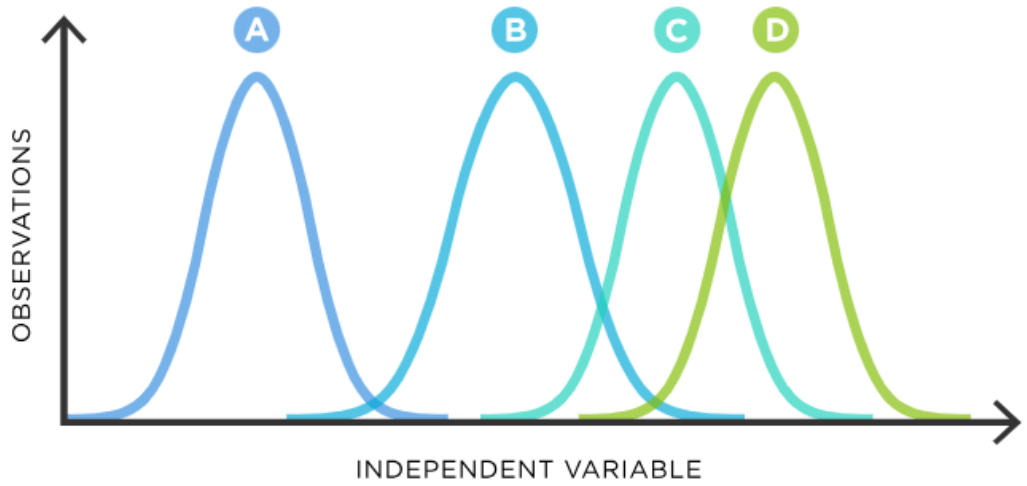
Image: itl.nist.gov



WEEK 08

- Happy Spring Break!





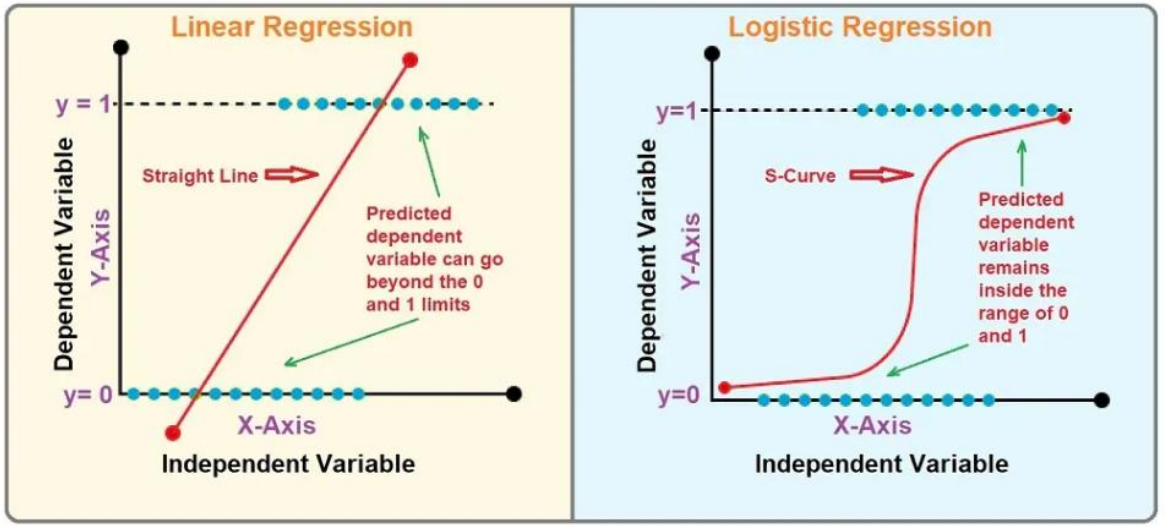
WEEK 09 ANOVA

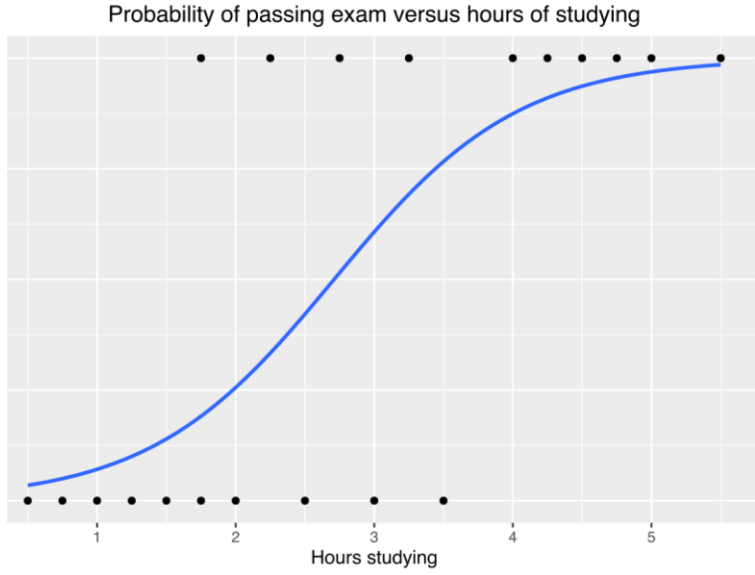
- Assumption of ANOVA
- Type of ANOVA
- Diagnostic in ANOVA

IMAGE SOURCE: HTTPS://DATATAB.NET/TUTORIAL/ANOVA

WEEK 10

LOGISTIC REGRESSION





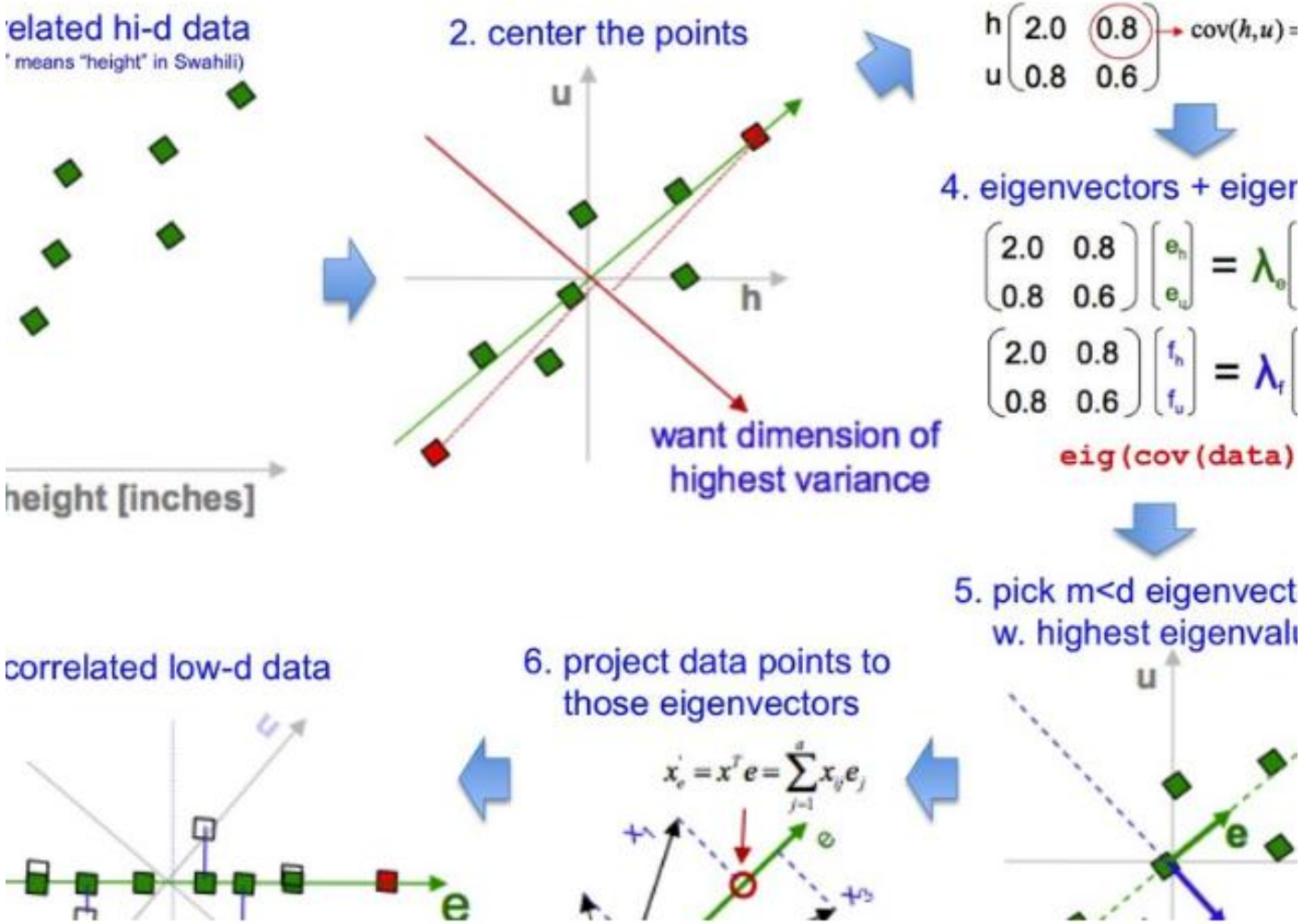
WEEK 11

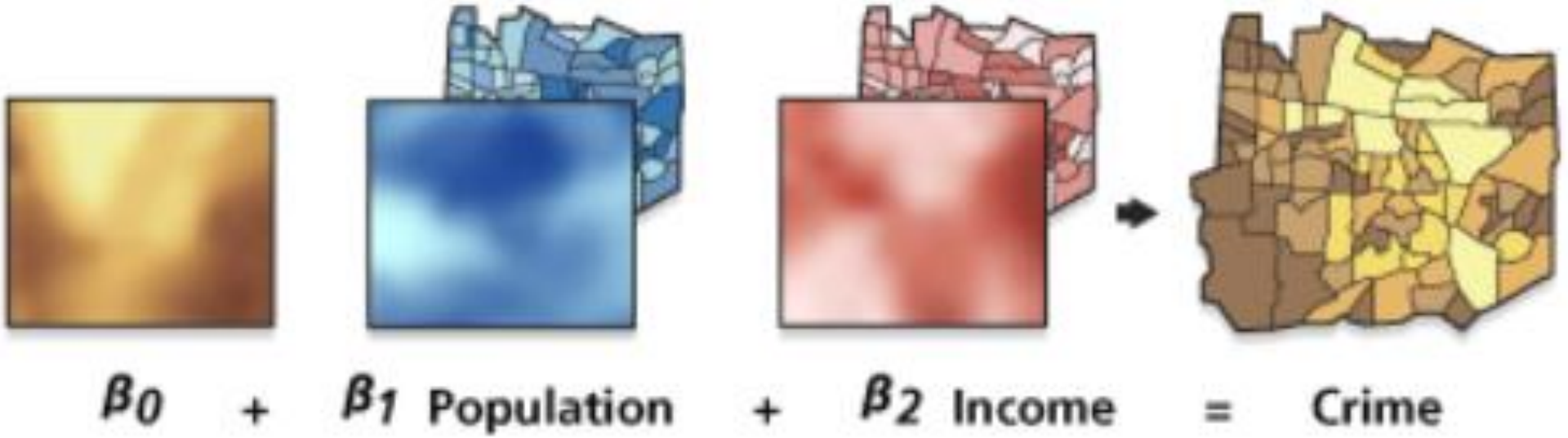
- 2025 AAG Conference

WEEK12

- Principal Component Analysis

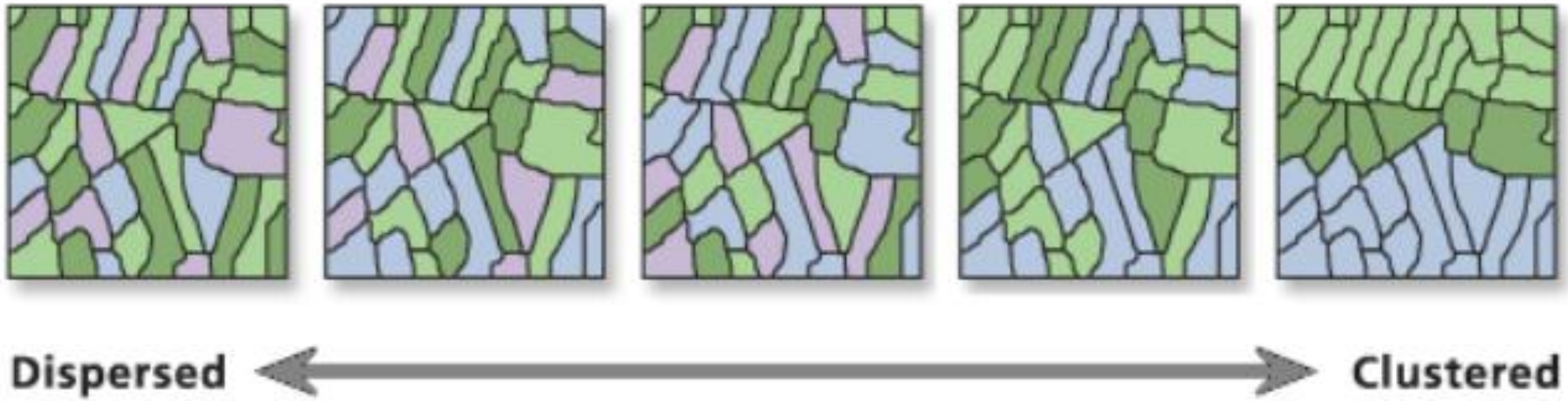
A powerful dimensionality reduction technique used in statistics and machine learning to simplify complex datasets.





WEEK 13

GEOGRAPHICALLY WEIGHTED REGRESSION



WEEK 14

- Spatial Autocorrelation

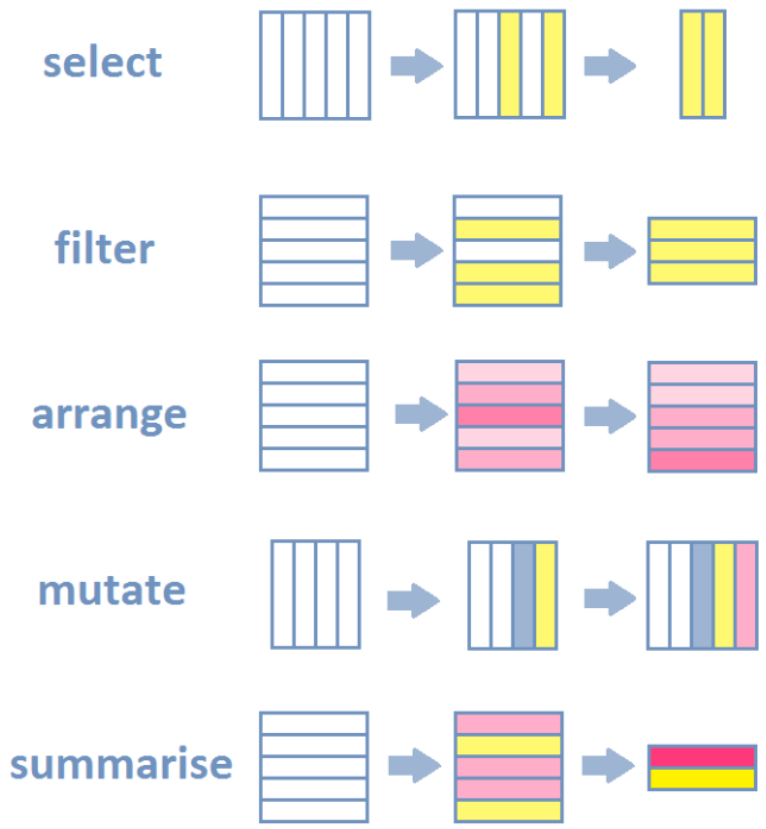
PREAMBLE

- 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.

WEEK 01

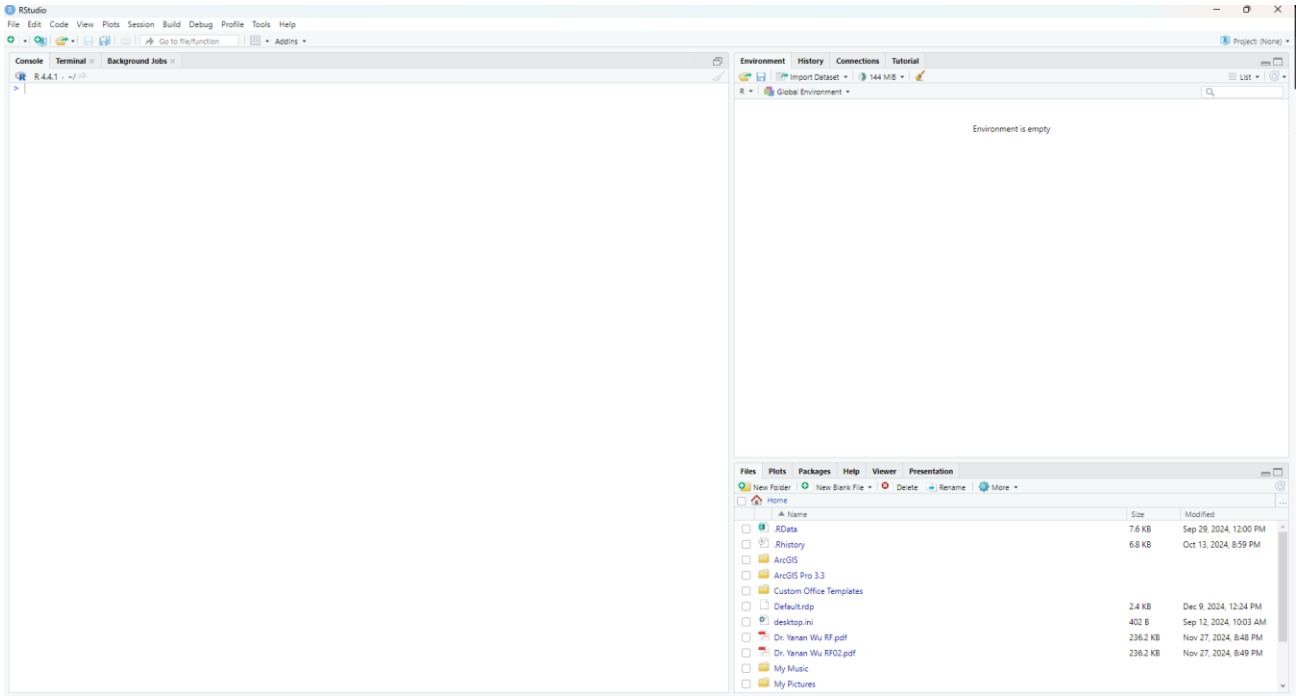
LAB SESSION

Instructor: Yanan Wu
TA: Khadija Nisar

Spring 2025

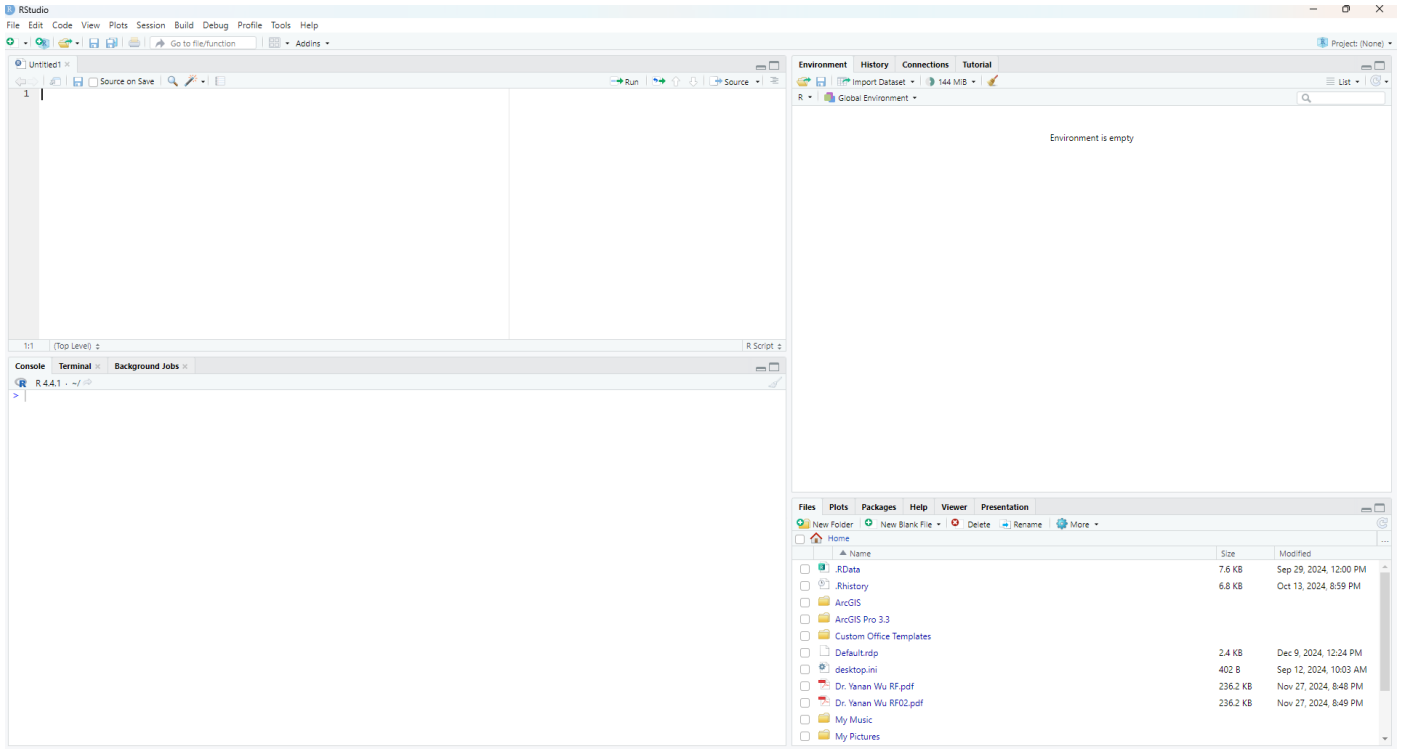
OVERVIEW OF RSTUDIO INTERFACE

- 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*



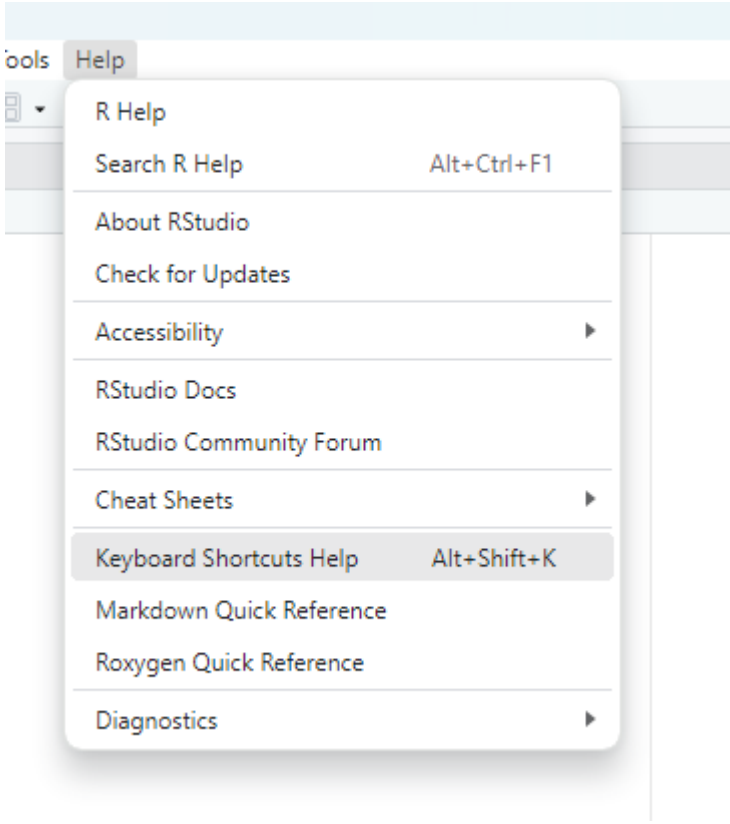
OVERVIEW OF RSTUDIO INTERFACE

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



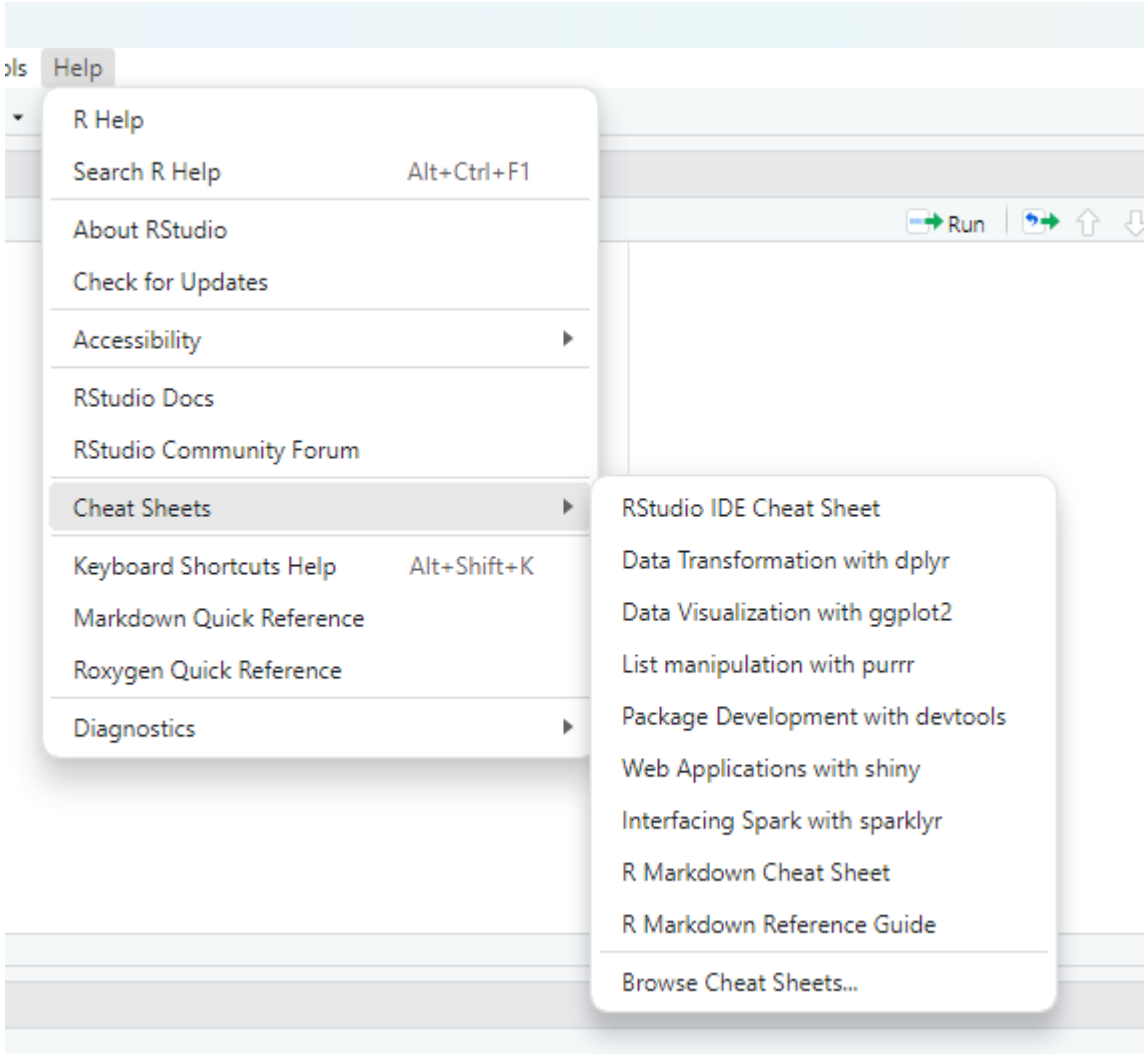
OVERVIEW OF KEYBOARD SHORTCUTS

- Keyboard shortcuts
  - Help – Keyboard Shortcuts Help



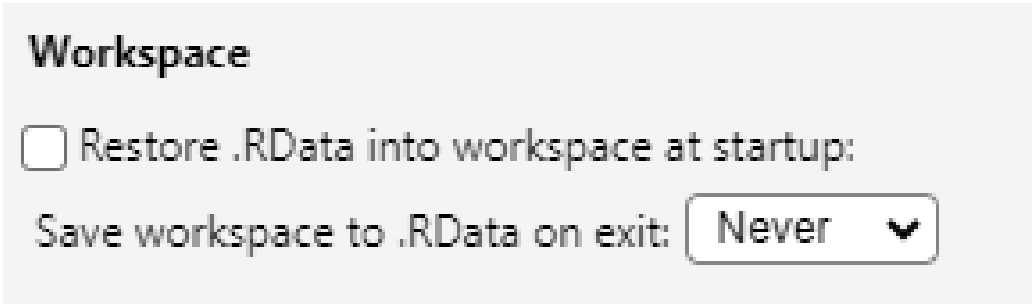
OVERVIEW OF CHEATSHEETS

- Cheatsheets in Rstudio
  - Help - Cheatsheets



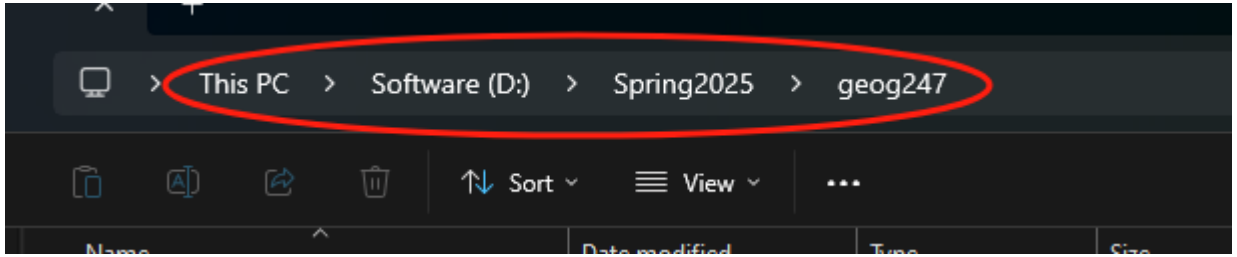
GLOBAL SETTING

- .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



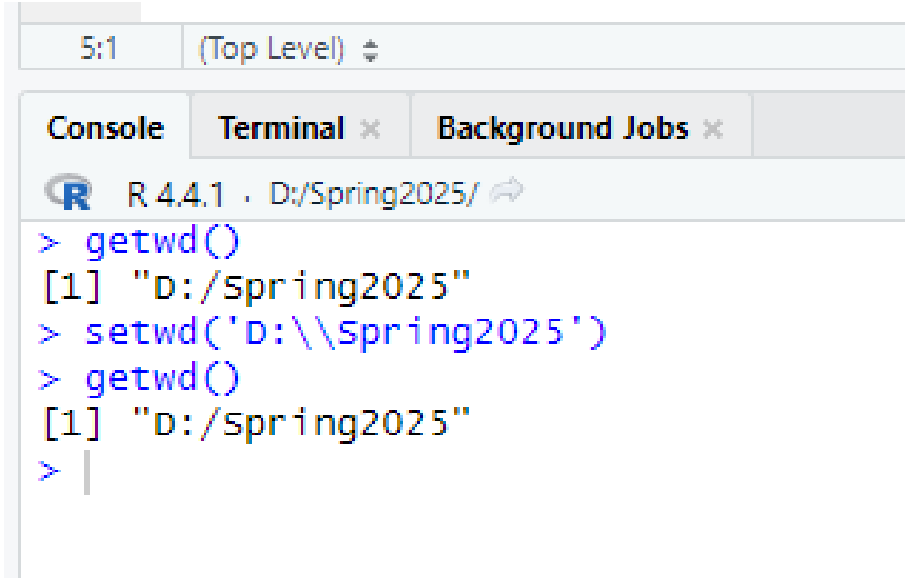
WORKING DIRECTORY

- 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

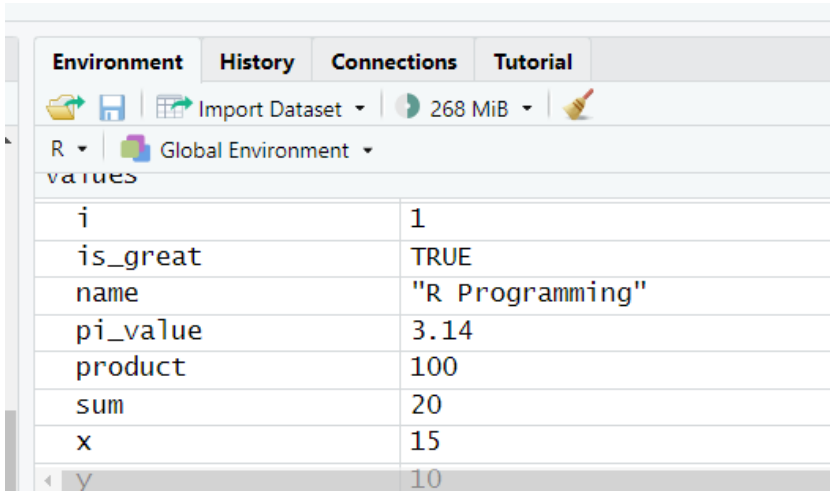
- Get help for activate libraries
  - **help('dplyr')**
  - **?dplyr**
- Get help for all installed libraries
  - **??dplyr**
  - **help.search('dplyr')**

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: Khadija Nisar

Spring 2025

PRACTICES

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