

# Day 1, Part 2: RStudio

Brennan & Matt

# R Studio

- The most popular open-source IDE, developed by Posit
  - IDE = integrated development environment
  - Basically a fancy word for software that makes your life easier when coding ☺
- Can be downloaded as [desktop software](#)
  - Recommended if you want to keep using R after the workshop!
- Can use on [the cloud](#)
  - Guided labs will be via RStudio Cloud

# RStudio Desktop vs. Cloud

- Posit has a desktop version of RStudio and a cloud version.
- For guided labs, we will be using the cloud to reduce time spent on troubleshooting while in class.
- We recommend familiarizing yourself with the desktop version if you want to keep using R after the workshop.
- We are available to help troubleshoot with the desktop version ☺

# Posit Cloud

IAM3 2023 R / IAM3 2023 R

File Edit Code View Plots Session Build Debug Profile Tools Help

RAM Go to file/function Addins

Source Visual

1 `---`  
2 `title: 'Lab: RStudio Setup and Data Import'`  
3 `author: "Brennan & Matt"`  
4 `date: "r Sys.Date()"`  
5 `output: html_document`  
6 `---`  
7  
8 `# R Markdown Files -----`  
9  
10 This is an R Markdown document, which is used to create documents with executable code.  
11  
12 See <http://rmarkdown.rstudio.com> for details on R Markdown.  
13  
14 This is a code chunk, in **gray**. You can run the code in this chunk by pressing the green Play button on the right.  
15  
16 Every R Markdown document has a `setup` chunk in the beginning. The `include = FALSE` parameter tells it not to include this chunk in the final document.  
17  
18 ````{r setup, include=FALSE}`  
19 `knitr::opts_chunk$set(echo = TRUE)`  
20 `---`  
21  
22 `# Make RStudio Yours! -----`  
23  
24 Unfortunately there are no themes for Posit Cloud, only R Studio Desktop.  
25  
26 If you want to personalize the desktop version, read on!  
27  
28 Personalize it to your heart's delight:  
29 1. Tools Menu > Global Options > Appearance...  
30 2. Change your Editor Theme and/or font size. There's some cool color themes in there!  
31 3. Apply & Save.

100:4 `# Make Sure Tidyverse Is Loaded`

Console Terminal Background Jobs

R 4.3.0 /cloud/project/ Error: object 'my\_data' not found

Restarting R session...  
Connected to your session in progress, last started 2023-Jun-11 09:50:06 UTC (4 minutes ago)

Environment History Connections Git Tutorial

Import Dataset 188 MIB RAM Global Environment

Data starwars\_data 87 obs. of 14 variables

Values

x	10
y	20
z	30

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project > 12th\_introduction

Name	Size	Modified
..		
12th_scripts		
12th_slides		

# Posit RStudio (Desktop)

The screenshot displays the Posit RStudio desktop application interface. On the left, the code editor shows an R script with several lines of code related to data cleaning and manipulation, including imports for `readxl` and `dplyr`, and operations like `filter`, `mutate`, and `select`. The code editor tabs include `boilerplates.qmd`, `test.R`, `data_cleaning.R`, `12_1_intro.qmd`, `13_1_basics.qmd`, `13_2_viz.qmd`, `13_3_viz.qmd`, `14_1_verse.qmd`, and `14_1_verse.qmd`. The status bar at the bottom indicates the R version is 4.2.2 and the current working directory is `~/Library/CloudStorage/GoogleDrive-bterhune@sdsl.edu/My Drive/IAM3_R`.

The right side of the interface features the Global Environment browser, which lists various objects such as `cpi`, `cpi_tidy`, `df1`, `lang_data_complex`, `mock_data`, `bai`, `boolean_vector`, `c1`, `c2`, `char_vector`, `itemA`, `itemB`, `num_vector`, `number`, `strange_object`, and `vee`. It also shows a list of functions: `functionalFunction` and `hi`.

Below the environment browser is a documentation viewer for the `mutate` function from the `dplyr` package. The documentation includes sections for **Description** and **Usage**, with examples of how to use the function.

```
library(readxl)
library(dplyr)

# assumes current workdir is IAM3
getwd()
setwd("data/")

# CPI
cpi <- read_xlsx("cpi_row.xlsx") %>%
  janitor::clean_names()

cpi_clean <- cpi %>%
  filter(!is.na(consumer_price_index_item)) %>%
  filter(!is.na(annual_2021)) %>%
  rename(predicted_2023 = x11,
         hist_avg_2003_2022 = x20_year_historical_average_2003_2022,
         item = consumer_price_index_item) %>%
  select(item, annual_2020, annual_2021, annual_2022, predicted_2023, hist_avg_2003_2022)

saveRDS(cpi_clean, "cpi.rds")

tidy_cpi <- cpi %>%
  rename(annual_2023 = predicted_2023) %>%
  pivot_longer(annual_2020:annual_2023,
               names_to = "year",
               names_prefix = "annual_",
               values_to = "increase") %>%
  select(item, year, increase, everything()) %>%
  saveRDS("cpi_tidy.rds")

# Eruptions
eruptions <- read_xlsx("holocene_eruptions.xlsx") %>%
  janitor::clean_names()

extract_year <- function(df, col_name) {
  library(stringr)
  df %>%
    mutate(year = str_extract(last_known_eruption, "\\\d+")) %>%
    mutate(year = as.numeric(year)) %>%
    mutate(year = as.numeric(year))

6:1 (Top Level) : R Script
```

Console Terminal Background Jobs

R 4.2.2 ~/Library/CloudStorage/GoogleDrive-bterhune@sdsl.edu/My Drive/IAM3\_R ↵

downloaded 38 KB

```
The downloaded binary packages are in
  /var/folders/dc/kv5mfpbd5lq3xvn650m44pcm000gn/T//RtmpCX3hed/downloaded_packages
> # Script for creating calendar.
> require(ggplot2)
Loading required package: ggplot2
> require(calendR)
Loading required package: calendR
~~ Package calendR
Visit https://r-coder.com/ for R tutorials ~~
> |
```

Environment History Connections Git Tutorial

Import Dataset 192 MB

Data

- cpi 22 obs. of 6 variables
- cpi\_tidy 88 obs. of 4 variables
- df1 6 obs. of 2 variables
- lang\_data\_complex 35 obs. of 9 variables
- mock\_data 35 obs. of 9 variables

Values

- bai 5
- boolean\_vector logi [1:3] TRUE FALSE TRUE
- c1 num [1:6] 1 2 3 4 5 6
- c2 num [1:6] 3 NA 5 NA NA 9
- char\_vector chr [1:3] "hi" "what's" "up"
- itemA 2
- itemB 5
- num\_vector num [1:3] 1 2 3
- number 4
- strange\_object "hi"
- vee 2

Functions

- functionalFunction function (item1, item2)
- hi function (item1, item2)

Files Plots Packages Help Viewer Presentation

R: Create, modify, and delete columns Find in Topic

mutate (dplyr)

Create, modify, and delete columns

Description

`mutate()` adds new variables and preserves existing ones; `transmute()` adds new variables and drops existing ones. New variables overwrite existing variables of the same name. Variables can be removed by setting their value to `NULL`.

Usage

```
mutate(.data, ...)

## S3 method for class 'data.frame'
mutate(
  .data,
  ...,
```

R Documentation

The screenshot shows the RStudio interface with several panes highlighted:

- Source:** The top-left pane contains R code for data cleaning and analysis, including reading Excel files, filtering data, and saving results to RDS files.
- Environment:** The top-right pane lists objects in the global environment, such as `cpi`, `cpi_tidy`, `df1`, `lang_data_co...`, and `mock_data`.
- Console:** The bottom-left pane shows the R command-line interface with history of commands run, including package installation and loading.
- Help:** The bottom-right pane displays the help documentation for the `mutate` function from the dplyr package.

Your open files (scripts, data frames, documents, etc.)

The source: where you'll do 99% of your coding

The console: displays errors and output; write one-time commands

Displays all objects in your environment

Displays help documentation for functions, or plots, or other things

## Source

- *Source* is where you add code to your script(s).
- It works like any document: you write code, Save As a file, and save as often as possible ☺
- You also will want to run code you write here pretty often

# Console

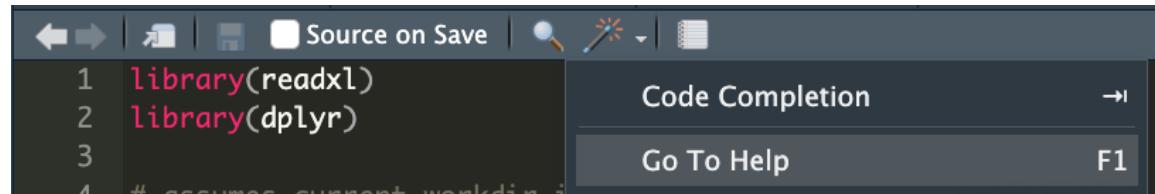
- *Console* shows you the *output* of all code you run, including source code
- When your code doesn't work, error messages show up in the console.
  - Caution! Error messages can be frustrating and do not tell you the *real* problem with the code. They just tell you what the computer noticed.
- You can also type and run “temporary” code in the console.

# Environment

- **Environment** shows all active objects and custom functions in your global environment
  - This is critical to understanding what your code has access to!
- **History** shows your command history (*I've never used this*)  
*(I've never used Packages, Viewer, or Presentations)*

# Files/Plots/Help

- **Files** displays the files in your working directory (current folder)
- **Plots** displays plots when you generate them
- **Help** displays documentation for functions
  - *Very useful!!*
  - Press **F1** or **magic wand** -> **Go To Help** when your text cursor (|) is on a function name:



- **OR** search for a function in the search box

*(I've never used Packages, Viewer, or Presentations)*

# R Scripts

```
1 library(readxl)
2 library(dplyr)
3
4 # assumes current working_dir is IAM3
5 getwd()
6 setwd("data/")
7
8 # CPI
9 cpi <- read_xlsx("cpi_raw.xlsx") %>%
10   janitor::clean_names()
11
12 cpi_clean <- cpi %>%
13   filter(!is.na(consumer_price_index_item)) %>%
14   filter(!is.na(annual_2021)) %>%
15   rename(predicted_2023 = x11,
16         hist_avg_2003_2022 = x20_year_historical_average_2003_2022,
17         item = consumer_price_index_item) %>%
18   select(item, annual_2020, annual_2021, annual_2022, predicted_2023, hist_avg_2
19
20 saveRDS(cpi_clean, "cpi.rds")
21
22 tidy_cpi <- cpi %>%
23   rename(annual_2023 = predicted_2023) %>%
24   pivot_longer(annual_2020:annual_2023,
25   |   names_to = "year",
26   |   names_prefix = "annual_",
27   |   values_to = "increase") %>%
28   select(item, year, increase, everything()) %>%
29   saveRDS("cpi_tidy.rds")
30
31 # Eruptions
32 eruptions <- read_xlsx("holocene_eruptions.xlsx") %>%
33   janitor::clean_names()
34
35 extract_year <- function(df, col_name) {
36   library(stringr)
37   df %>%
38     mutate(year = str_extract(last_known_eruption, "\\d+")) %>%
39     mutate(year = as.numeric(year)) %>%
40     mutate(year = if_else(str_detect(last_known_eruption, "BCE"), year * -1, yea
41   select(volcano_number, volcano_name, country, last_known_eruption, year, everythi
42 }
```

- *Scripts have code which can be executed*
- R scripts end with an ‘.R’ extension
- Everything in an R script will be executed unless it is *commented out* with # at the beginning of the line

# R Markdown

```
1 v ---
2 title: "Importing Data"
3 output: html_notebook
4 ^
5
6 ````{r setup}
7 library(tidyverse)
8 ^
9 |
10
11 Next, we use the read_csv() function to pull in data from our survey. We tell R to put that data in a data frame called climate_data. If it finds any blank cells, then those are NA data. NA values are treated specially in R; we will cover how to work with NA values later.
12
13 After running this command, a new 'object' called climate_data appears in the Environment pane in the top right corner of R Studio. Click on it and the dataframe opens in a new window and can be inspected. This is equivalent to typing View(climate_data) in the Console below.
14
15 Always be careful with dates. Looking at our CSV file, we can see that it is in M/DD/YY HH:MM format. In R terminology, this is %m/%d/%y %H:%M. It is best to read dates into a data frame in ISO8601 format, which is %Y-%m-%d %H:%M:%S. In tidyverse this is referred to as "datetime" format which has the data type \<dttm>. We use the col_types argument in read_csv to do this.
16 ````{r import_data}
17 climate_data <- read_csv("~/Documents/GitHub/2023_IAM3_R/data/climate-survey-data.csv",
18                           na = "",
19                           col_types = cols(StartDate = col_datetime("%m/%d/%y
19 %H:%M"),
20                           EndDate = col_datetime("%m/%d/%y
20 %H:%M"),
21                           RecordedDate =
22                           col_datetime("%m/%d/%y %H:%M"))
23 ^
24
25 Our next step is to rename the variables so that they conform to good R
practice. Note that by convention we put (what we will be creating / what we
want) to the left of the expression.
```

- RStudio also provides R Markdown/Notebook documents (“.Rmd”)
- Rmd files are *text files* that contain executable *code chunks*
- Better when you want to type a lot of non-code text
  - No need to comment out everything!

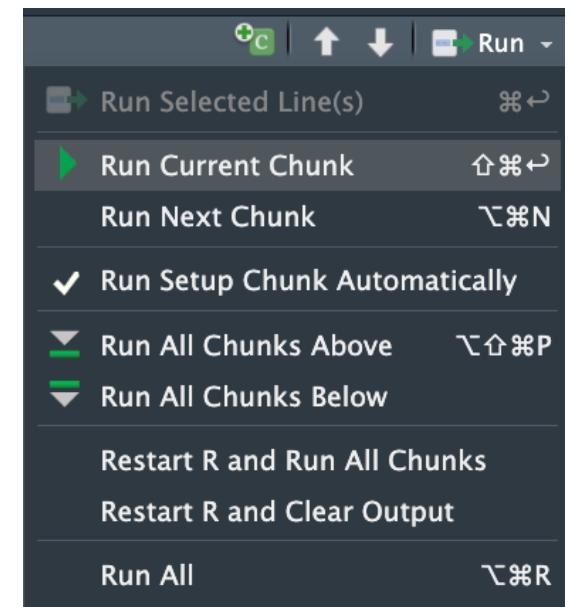
# Running Source Code in R Scripts

- You run source code in R scripts by:
    - Pressing CMD-ENTER (or CTRL-ENTER for PC)
      - Runs the line of code your text cursor is on
    - Highlighting code and pressing CMD-ENTER
      - Runs everything you highlighted
  - If you want to run the entire script:
    - CMD-A then CMD-ENTER
    - If you accidentally press ENTER and erase everything, CMD-Z!

# Running Source Code in R Markdown

- .Rmd documents have code in *chunks*.
- Run code in a chunk by pressing ➤
- You can also run chunks in different ways ->

```
1 ---  
2 title: "Importing Data"  
3 output: html_notebook  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ...  
9 |  
10  
11 Next, we use the read_csv() function to pull in data from our survey. We tell  
R to put that data in a data frame called climate_data. If it finds any blank  
cells, then those are NA data. NA values are treated specially in R; we will  
cover how to work with NA values later.  
12
```



# Running Console Code

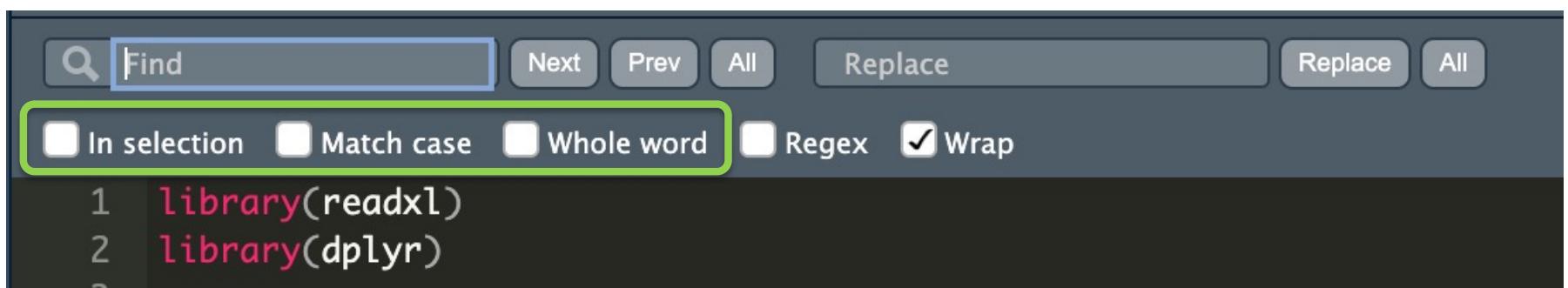
- Coding in the console works like in any console: you type out your command and press Enter 😊
  - You type in the line starting with >
- Meant to be for one-time code only
  - Viewing data
  - Doing calculations
  - Checking things before adding to source

The screenshot shows an RStudio interface with the following details:

- Code Area:** Contains R code for data cleaning and analysis, including reading Excel files, filtering data, and creating pivot tables.
- Environment Tab:** Shows the global environment with objects like `cpi`, `cpi_tidy`, `df1`, `long_data_co`, and `mock_data`.
- Data View:** Displays the structure of the `cpi` data frame, showing 22 observations and 6 variables.
- Output Area:** Shows the results of the R code execution, including the printed output of the pivot table and the final data frame.
- Help and Documentation:** A yellow box highlights the "Create, modify, and delete columns" section of the "Description" pane for the `mutate` function.

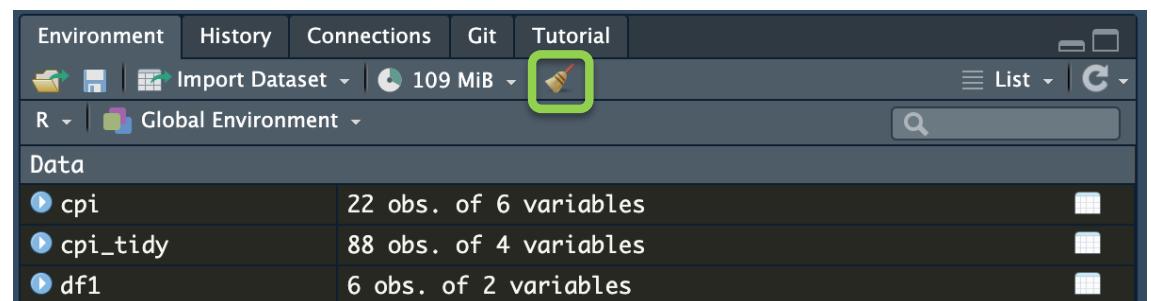
# RStudio Tips

- Their **find and replace** function is excellent
  - CMD-F on Mac | CTRL-F on PC
- Find specific text (names, etc.) and replace one or all
  - CMD-Z | CTRL-Z will undo replacements



# RStudio Tips

- The **broom icon** will wipe clean your environment or your console history.
- **For your environment:** this is important to make sure your script doesn't call "leftover" objects which should not exist at that point in the script.
- **For your console history:**  
it will just keep you sane ☺



A screenshot of the RStudio interface showing the Console tab. The console output shows R version 4.2.2 and a download of 38 KB. The broom icon in the top right corner of the Console tab is highlighted with a green box.

```
Console Terminal × Background Jobs ×
R 4.2.2 · ~/Library/CloudStorage/GoogleDrive-bterhuncotter@sdsu.edu/My Drive/IAM3_R/ ↵
downloaded 38 KB

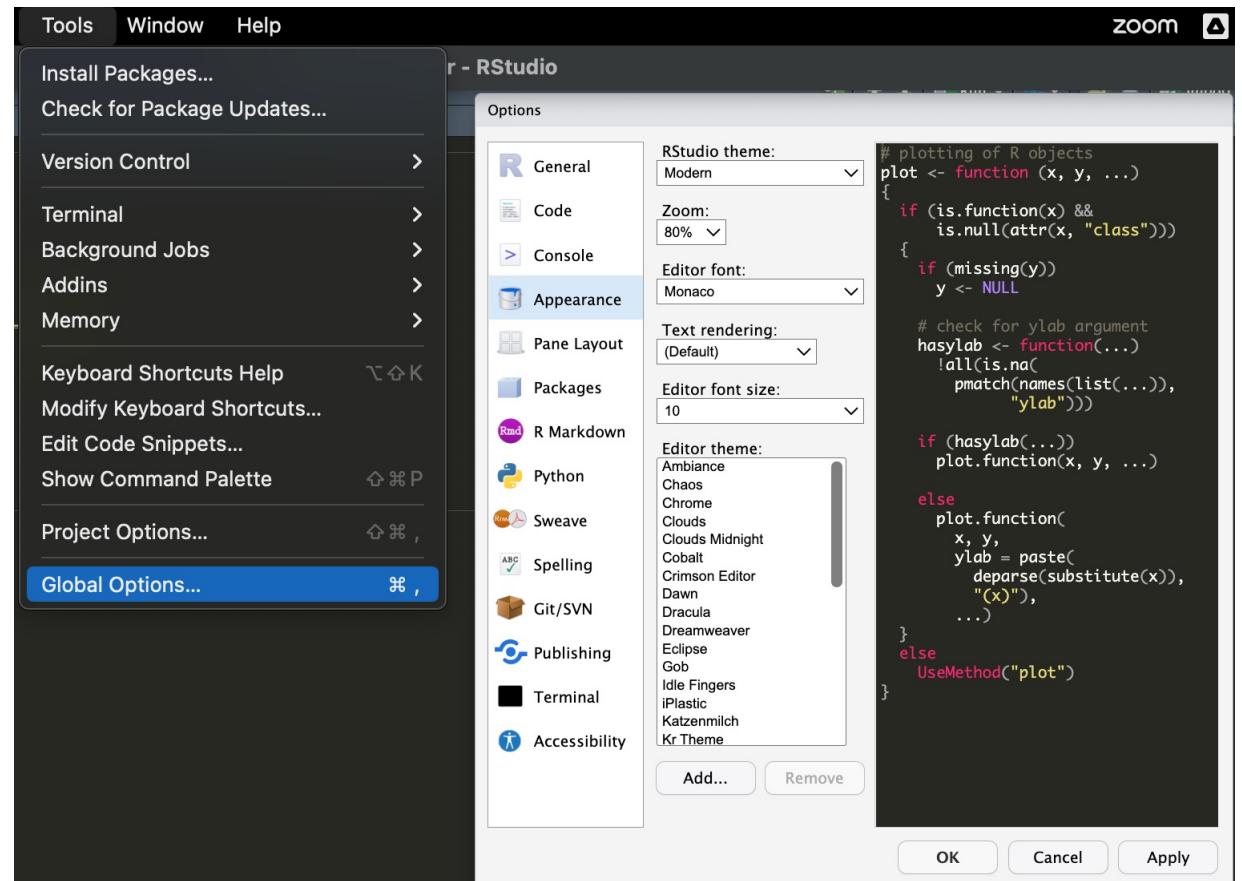
The downloaded binary packages are in
/var/folders/dc/kvd5mpbd5lq3xvn650m44pcm0000gn/T//RtmpCX3hed/downloaded_packages
# Script for creating calendar
```

# RStudio Tips

- To ensure you don't end up with a broken script, do this semi-frequently:
  1. Broom up everything in the Environment (and Console if you want, but don't need)
  2. Re-run your entire script:
    1. Press CMD-A (highlights everything)
    2. Press CMD-ENTER
  3. Fix the error that pops up and do steps 1-2 again
  4. You're done when everything runs correctly ☺
- This can be *super* annoying, but is critical to catching errors early
- The more often you do it, the less annoying it is because you know where the errors are

# Picking an RStudio & Posit Cloud color theme!

- You won't be a good coder if your screen doesn't look cool!
- Click on **Tools > Global Options > Appearance** and pick your favorite theme.
- There are light and dark themes
- I like Monokai and Tomorrow Night; choose whatever you think is prettiest! ☺



## Side Note...

- RStudio is not the only IDE for R
- Another good one is Visual Studio Code (VSC)
  - VSC is a great IDE for multiple programming languages
  - It is not specialized for R, but it is very polished and has cool features for programming.
    - GitHub Copilot (AI to help autocomplete code)
  - I don't recommend VSC for learning R but if you want to use other languages or explore more features, check it out!

# Visual Studio Code

The screenshot shows the Visual Studio Code interface with the following details:

- Explorer View:** Shows a file tree with a folder named "IAM3\_R" containing various R scripts, QMD files, and HTML files. A file named "data\_cleaning.R" is currently selected.
- Editor View:** Displays the content of the "data\_cleaning.R" script. The code uses dplyr and readxl packages to clean CPI and eruption data, saving results to RDS files.
- Terminal View:** Located at the bottom right, showing the command "R: (not attached) Ln 1, Col 1 Spaces: 2 UTF-8 LF R".
- Bottom Status Bar:** Shows icons for file operations like save, close, and refresh, along with the current file path "data\_cleaning.R" and line/character counts.

## Next up...

- We will tell you how to import files into RStudio
- Then you will try importing your (or our) data into RStudio!