

Day 1, Part 2: RStudio

Brennan & Matt

R Studio

- The most popular open-source IDE, developed by Posit
 - IDE = integrated **d**evelopment **e**nvironment
 - 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 😊

IAM3 2023 R / IAM3 2023 R RAM ⚙️ ⋮ BT Brennan Terhune-Cotter

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins R 4.3.0

12_1_begin.Rmd

Source Visual Outline

```
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
15 right.
16 Every R Markdown document has a `setup` chunk in the beginning. The `include = FALSE` parameter tells it not to
17 include this chunk in the final document.
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 -----
```

Environment History Connections Git Tutorial

Import Dataset 188 MiB

R 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		

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)

>

Posit RStudio (Desktop)

The screenshot displays the Posit RStudio Desktop interface. The main editor window contains R code for data cleaning and manipulation. The Environment pane on the right shows the current data objects. The bottom pane displays the documentation for the `mutate` function from the `dplyr` package.

```
1 library(readxl)
2 library(dplyr)
3
4 # assumes current workdir 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_2003_2022)
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(vear = as.numeric(vear)) %>%
6.1 (Top Level) - R Script
```

Environment

Data	Variables
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) R Documentation

Create, modify, and delete columns

Description

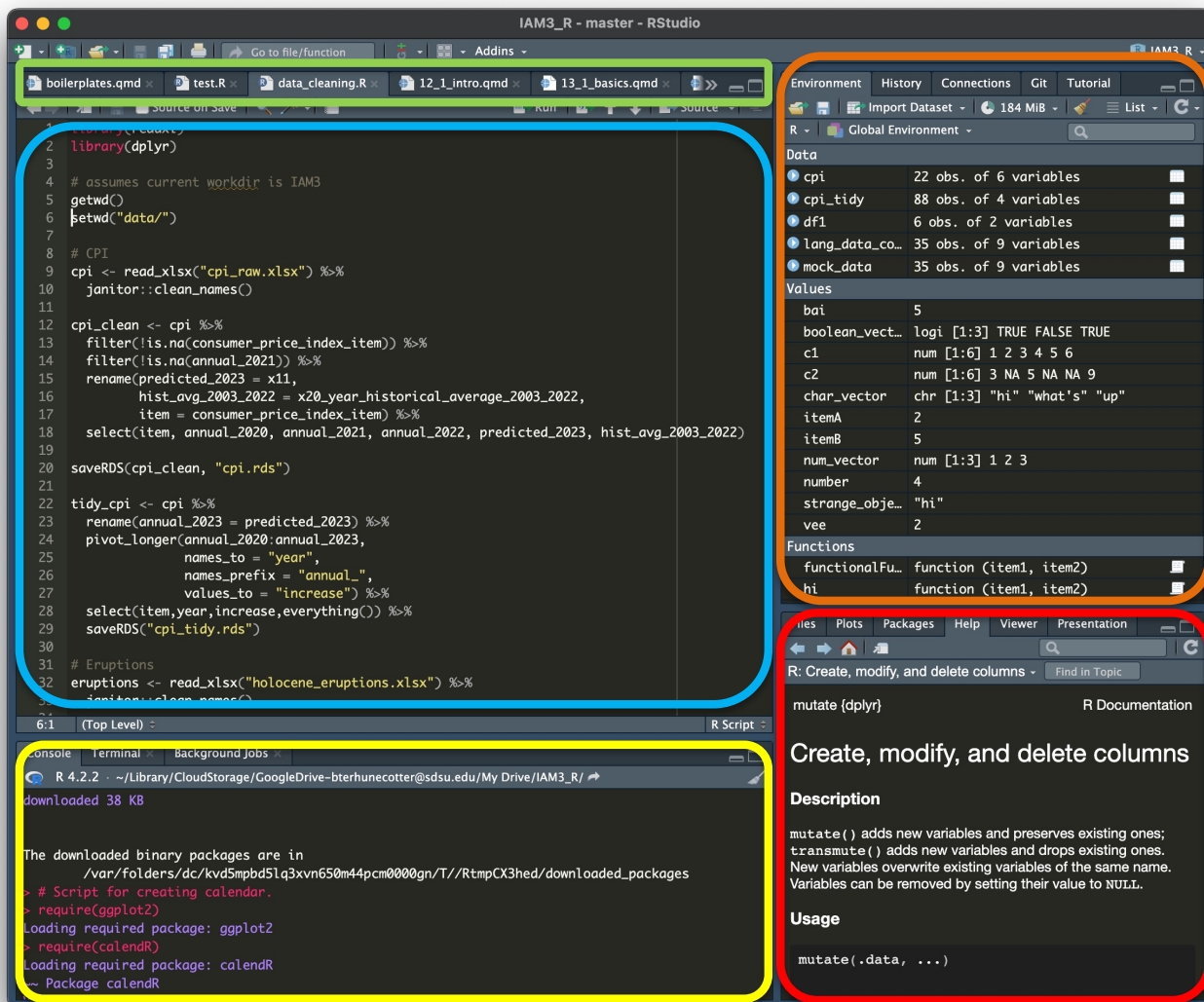
`mutate()` adds new variables and preserves existing ones; `transmute()` adds new variables and drops existing ones. New variables overwrite existing variables of the same name. Variables can be removed by setting their value to `NULL`.

Usage

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

S3 method for class 'data.frame'

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



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

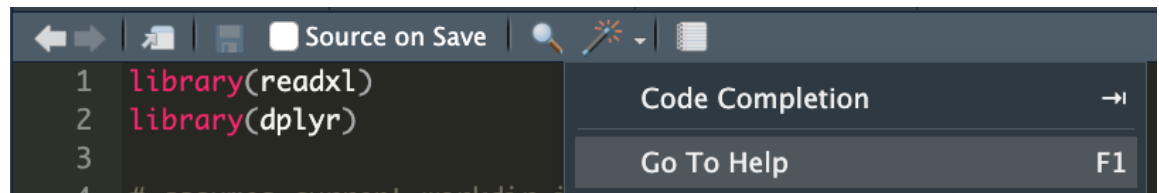
- *Console* shows you the *output* of all code you run, including source code
- When your code doesn't work, error messages show up in the console.
 - **Caution!** Error messages can be frustrating and do not tell you the *real* problem with the code. They just tell you what the computer noticed.
- You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



```
1 library(readxl)
2 library(dplyr)
3
4 # assumes current working directory is
```

Code Completion →
Go To Help F1

- **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 workdir 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, year)
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

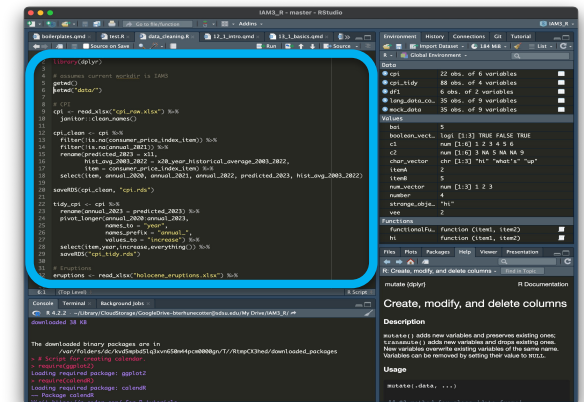
R Markdown

```
1 ---
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
12 R to put that data in a data frame called climate_data. If it finds any blank
13 cells, then those are NA data. NA values are treated specially in R; we will
14 cover how to work with NA values later.
15
16 After running this command, a new 'object' called climate_data appears in the
17 Environment pane in the top right corner of R Studio. Click on it and the
18 dataframe opens in a new window and can be inspected. This is equivalent to
19 typing View(climate_data) in the Console below.
20
21 Always be careful with dates. Looking at our CSV file, we can see that it is
22 in M/DD/YY HH:MM format. In R terminology, this is *%m%d%y %H%M*. It is best
23 to read dates into a data frame in ISO8601 format, which is *%Y-%m-%d
24 %H:%M:%S*. In tidyverse this is referred to as "datetime" format which has
25 the data type \<dtm\>. We use the col_types argument in read_csv to do this.
26
27 ```{r import_data}
28 climate_data <- read_csv("~/Documents/GitHub/2023_IAM3_R/data/climate-survey-
29 data.csv",
30                       na = "",
31                       col_types = cols(StartDate = col_datetime("%m/%d/%y
32 %H:%M"),
33                                       EndDate = col_datetime("%m/%d/%y
34 %H:%M"),
35                                       RecordedDate =
36 col_datetime("%m/%d/%y %H:%M"))
37 )
38 ```
39
40 Our next step is to rename the variables so that they conform to good R
41 practice. Note that by convention we put (what we will be creating / what we
42 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

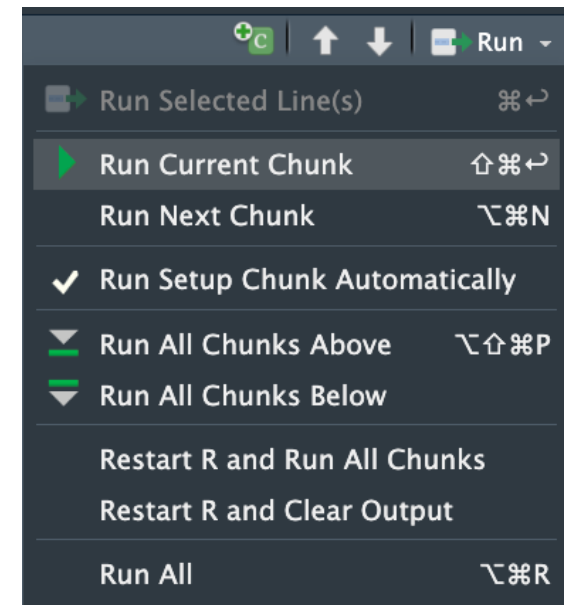
- You run source code in R scripts by:
 - Pressing CMD-ENTER (or CTRL-ENTER for PC)
 - Runs the line of code your text cursor is on
 - Highlighting code and pressing CMD-ENTER
 - Runs everything you highlighted
- If you want to run the entire script:
 - CMD-A then CMD-ENTER
 - If you accidentally press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



Running Source Code in R Markdown

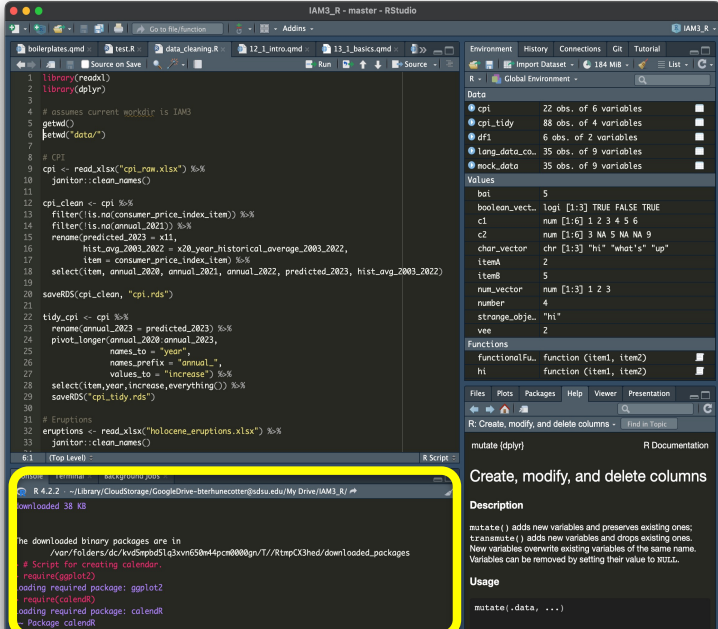
- .Rmd documents have code in *chunks*.
- Run code in a chunk by pressing 
- You can also run chunks in different ways ->

```
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
12 R to put that data in a data frame called climate_data. If it finds any blank
13 cells, then those are NA data. NA values are treated specially in R; we will
14 cover how to work with NA values later.
```



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 the RStudio interface. The main editor window contains R code for data cleaning and manipulation. The console window at the bottom shows the execution of the code, including the installation of the 'ggplot2' package. The Environment pane on the right shows the objects created in the workspace, such as 'cpi', 'tidy_cpi', and 'eruptions'.

```
1 library(readxl)
2 library(dplyr)
3
4 # assumes current working dir is IAM3
5 getwd()
6 #readxl::read_xlsx("cpi_raw.xlsx") %>%
7   janitor::clean_names()
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_2023)) %>%
15   rename(predicted_2023 = x11,
16           hist_ovg_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_ovg_2003_2022)
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("tidy.rds")
30
31 # Eruptions
32 eruptions <- read_xlsx("holocene_eruptions.xlsx") %>%
33   janitor::clean_names()
```

The console output shows the following commands and their results:

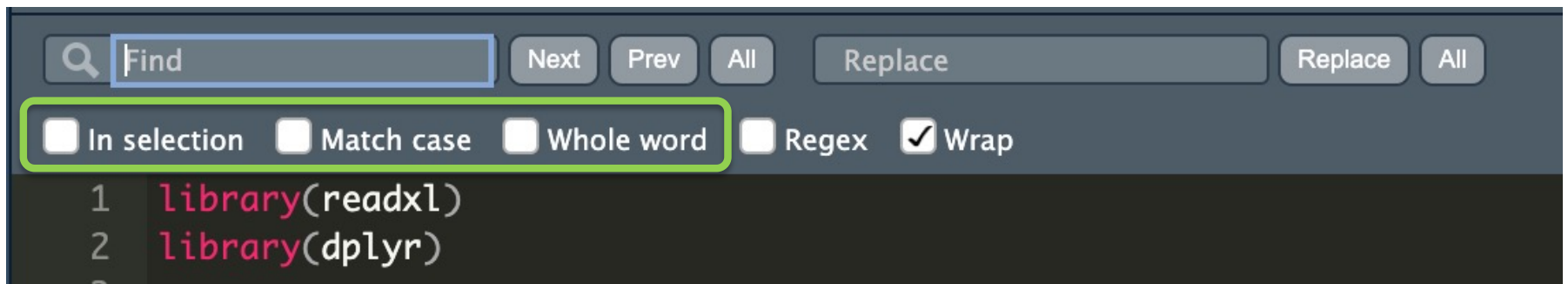
```
R 4.2.2 -> Library/CloudStorage/GoogleDrive-brerhuncotten@idsu.edu/My Drive/IAM3_R/
downloaded 38 KB
The downloaded binary packages are in
  /var/folders/dc/kvd5pbd5lq3xvri65h44pcn000gn/~/RtmpCX3hd/d/downloaded_packages
# Scripts for creating calendar.
# requires(ggplot2)
loading required package: ggplot2
# requires(calendar)
loading required package: calendar
# Package calendar
```

The Environment pane shows the following objects:

Object	Class	Attributes
cpi	tbl_df	22 obs. of 6 variables
cpi_tidy	tbl_df	88 obs. of 4 variables
df1	tbl_df	6 obs. of 2 variables
lang_data_co	tbl_df	35 obs. of 9 variables
noc_data	tbl_df	35 obs. of 9 variables
bool	logical	[1:3] TRUE FALSE TRUE
boolean_vect	logical	[1:6] 1 2 3 4 5 6
c1	numeric	[1:6] 3 NA 5 NA NA 9
c2	numeric	[1:6] 3 NA 5 NA NA 9
chr_vector	character	[1:3] "hi" "what's" "up"
item	character	2
item	numeric	[1:3] 1 2 3
num_vector	numeric	4
number	numeric	2
strange_obj	list	"hi"
vee	numeric	2

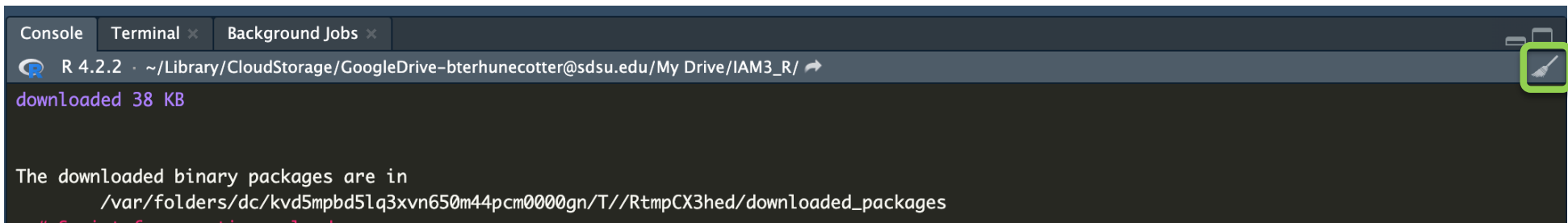
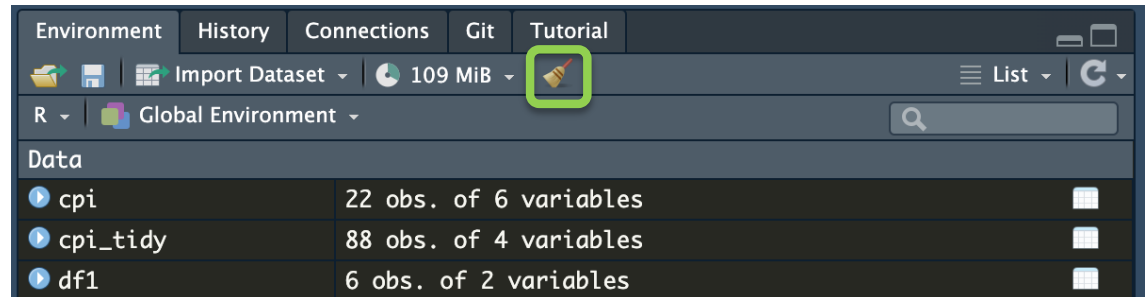
RStudio Tips

- 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 😊

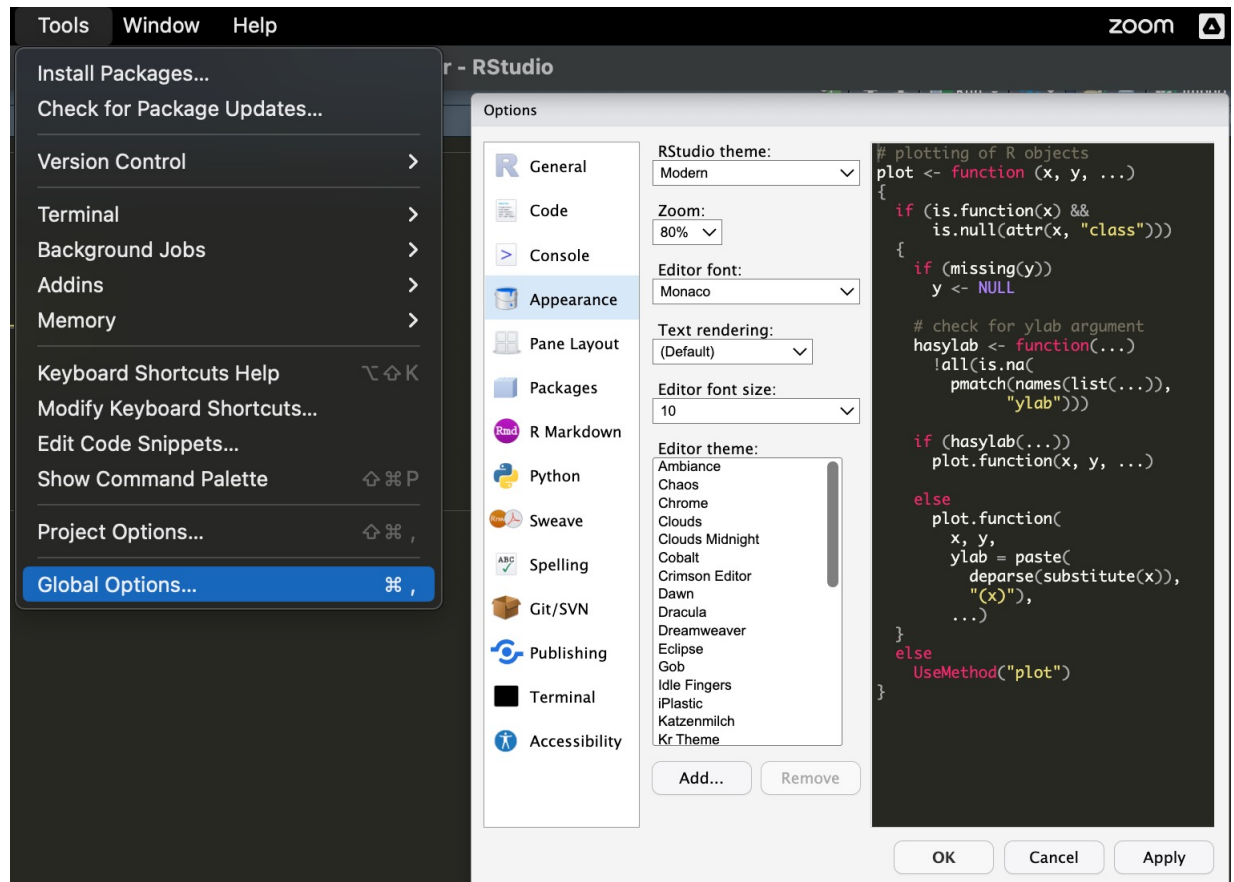


RStudio Tips

- 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 displays the Visual Studio Code interface with an R script named `data_cleaning.R` open. The Explorer sidebar on the left shows a project structure with folders like `14th_scripts` and `15th_scripts`. The main editor area contains the following R code:

```
1 library(readxl)
2 library(dplyr)
3
4 # assumes current workdir 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_2003_2022)
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, year)) %>%
```

The Problems panel at the bottom shows several warnings related to variable binding:

- no visible binding for global variable 'last_known_eruption' object_usage_linter [Ln 38, Col 31]
- no visible binding for global variable 'last_known_eruption' object_usage_linter [Ln 38, Col 31]
- no visible binding for global variable 'last_known_eruption' object_usage_linter [Ln 38, Col 31]
- no visible binding for global variable 'year' object_usage_linter [Ln 39, Col 30]
- no visible binding for global variable 'year' object_usage_linter [Ln 39, Col 30]
- no visible binding for global variable 'year' object_usage_linter [Ln 39, Col 30]
- no visible binding for global variable 'year' object_usage_linter [Ln 39, Col 30]
- no visible binding for global variable 'year' object_usage_linter [Ln 39, Col 30]
- no visible binding for global variable 'volcano_number' object_usage_linter [Ln 41, Col 12]
- no visible binding for global variable 'volcano_name' object_usage_linter [Ln 41, Col 27]

The status bar at the bottom indicates the current file is `R: (not attached)` at `Ln 1, Col 1` with `Spaces: 2`, `UTF-8` encoding, and `LF` line endings.

Next up...

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