

# EARN Talks

Preparing for Analysis with R: A guided tutorial for installing R and RStudio

July 29<sup>th</sup>, 2025, 2-3pm EST

The webinar/discussion will begin soon.



# Meet today's facilitators!



Emma Cohn (she/her)  
Research Assistant



Daniel Perez (he/him)  
State Economic Analyst

# Announcements

- Upcoming EARN Data Workshops:
  - Foraging for Data in the Wild: Finding, loading, and wrangling economic data with R - August 19th, 2025, 2:30-3:45pm ET
  - Using EPI's CPS Microdata Extracts in R - September 4th, 2025, 2:00-3:15pm ET
- EARNCon 2025
  - Lowes New Orleans Hotel, Nov. 12 – 14<sup>th</sup>
  - State of Working X Bootcamp - November 12th, 2025

# Today's Agenda

- Distinguishing R vs. RStudio
- Installation walk-through
  - Windows and Mac
- Navigating the RStudio user interface (UI)
- Basics and key topics for beginners
  - Projects-oriented workflow
  - Basic commands / R is a calculator
  - Installing and using packages
- Basic workflow example
- Helpful resources
- Technical assistance following webinar conclusion



# R vs. RStudio

```
R version 4.5.1 (2025-06-13) -- "Great Square Root"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu
```

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

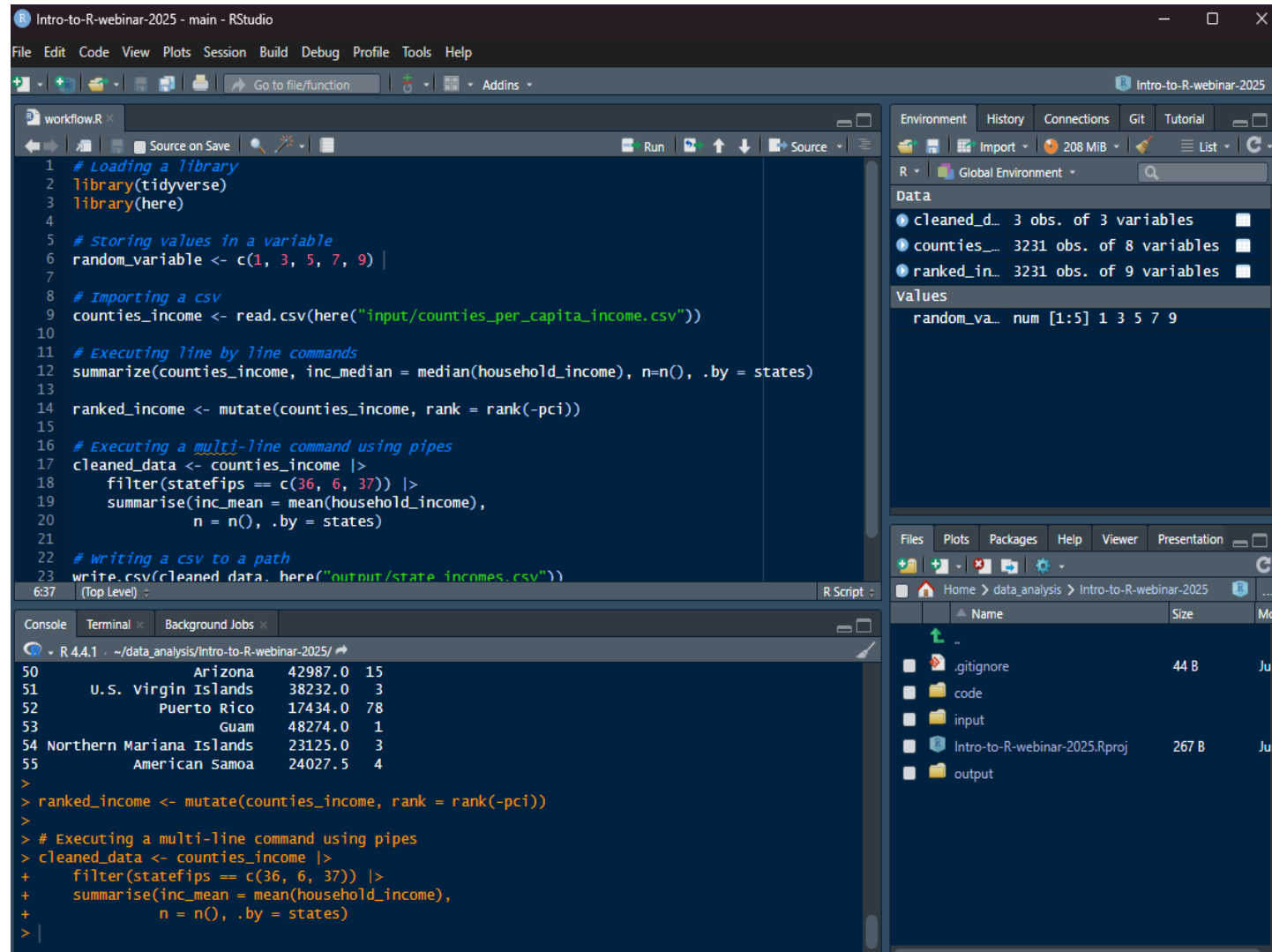
Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

```
[Previously saved workspace restored]
```

```
> 
```



The screenshot displays the RStudio IDE interface. The main editor window shows a script titled 'workflow.R' with the following R code:

```
1 # Loading a library
2 library(tidyverse)
3 library(here)
4
5 # Storing values in a variable
6 random_variable <- c(1, 3, 5, 7, 9)
7
8 # Importing a csv
9 counties_income <- read_csv(here("input/counties_per_capita_income.csv"))
10
11 # Executing line by line commands
12 summarize(counties_income, inc_median = median(household_income), n=n(), .by = states)
13
14 ranked_income <- mutate(counties_income, rank = rank(-pci))
15
16 # Executing a multi-line command using pipes
17 cleaned_data <- counties_income |>
18   filter(statefips == c(36, 6, 37)) |>
19   summarise(inc_mean = mean(household_income),
20             n = n(), .by = states)
21
22 # Writing a csv to a path
23 write_csv(cleaned_data, here("output/state_incomes.csv"))
```

The Environment pane on the right shows the following objects:

- cleaned\_data: 3 obs. of 3 variables
- counties\_income: 3231 obs. of 8 variables
- ranked\_income: 3231 obs. of 9 variables

The Console pane at the bottom shows the output of the R script, displaying a table of income data for various states and territories:

```
50      Arizona      42987.0    15
51  U.S. Virgin Islands 38232.0     3
52      Puerto Rico    17434.0    78
53          Guam     48274.0     1
54 Northern Mariana Islands 23125.0     3
55      American Samoa   24027.5     4
```

The file explorer on the right shows the project structure, including files like '.gitignore', 'code', 'input', 'Intro-to-R-webinar-2025.Rproj', and 'output'.

# Installing R and RStudio

- Download link:  
<https://posit.co/downloads/>

## Which version to download

1. Download R for Windows/Mac
  - a. If on Windows, download Rtools & Base R installer
  - b. If on Mac, install R
2. Download R Studio
  - a. If on Windows, download Rstudio
  - b. If on Mac, download Rstudio
    - i. In a terminal, run  
`/usr/sbin/softwareupdate --install-rosetta`

	Windows users	Mac users
	Intel vs ARM?	Intel vs M ?
<b>Step 1:</b>	Press windows key	Click Apple icon on top left corner
<b>Step 2:</b>	Type “about” and select “About your pc”	Select “About This Mac
<b>Step 3:</b>	Intel vs ARM?	Intel vs M chip?
<b>Step 4 (Mac):</b>		If M chip, install Rosetta

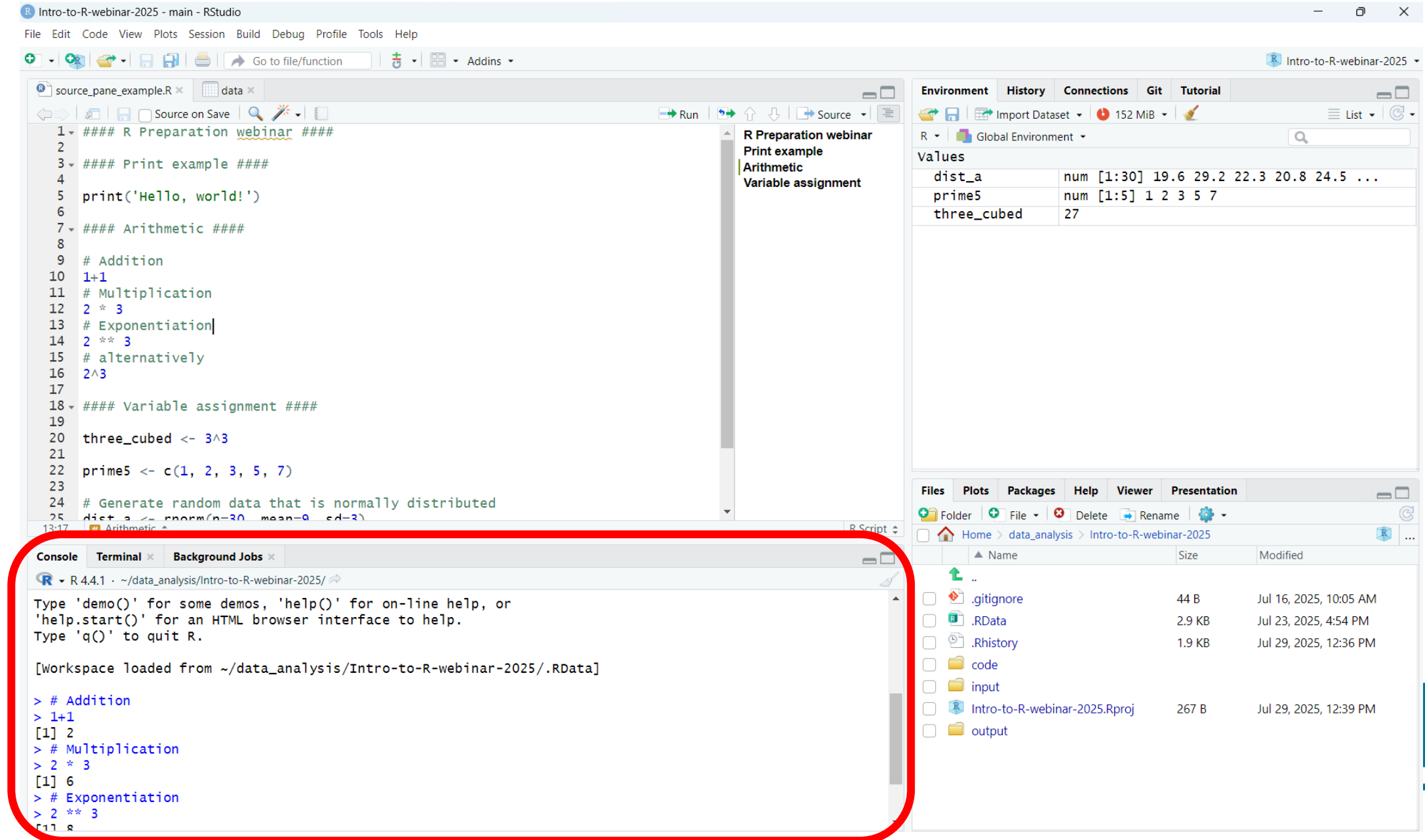


# Navigating the user interface (UI)

- Welcome R, users! 🎉
- Four primary panes:
  - Console
  - Source
  - Environment
  - Output



# Navigating the user interface (UI) – Console pane



The screenshot displays the RStudio user interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for file operations and running code. The source editor on the left shows a script named 'source\_pane\_example.R' with R code for preparation, printing, arithmetic, and variable assignment. The environment pane on the right shows the 'Global Environment' with variables 'dist\_a', 'prime5', and 'three\_cubed'. The console pane at the bottom is highlighted with a red border and shows the R prompt and the output of the code execution.

**Source Editor (source\_pane\_example.R):**

```
1 ##### R Preparation webinar #####
2
3 ##### Print example #####
4
5 print('Hello, world!')
6
7 ##### Arithmetic #####
8
9 # Addition
10 1+1
11 # Multiplication
12 2 * 3
13 # Exponentiation
14 2 ** 3
15 # alternatively
16 2^3
17
18 ##### Variable assignment #####
19
20 three_cubed <- 3^3
21
22 prime5 <- c(1, 2, 3, 5, 7)
23
24 # Generate random data that is normally distributed
25 dist_a <- rnorm(n=30, mean=9, sd=3)
```

**Environment Pane:**

Variable	Value
dist_a	num [1:30] 19.6 29.2 22.3 20.8 24.5 ...
prime5	num [1:5] 1 2 3 5 7
three_cubed	27

**Console Pane:**

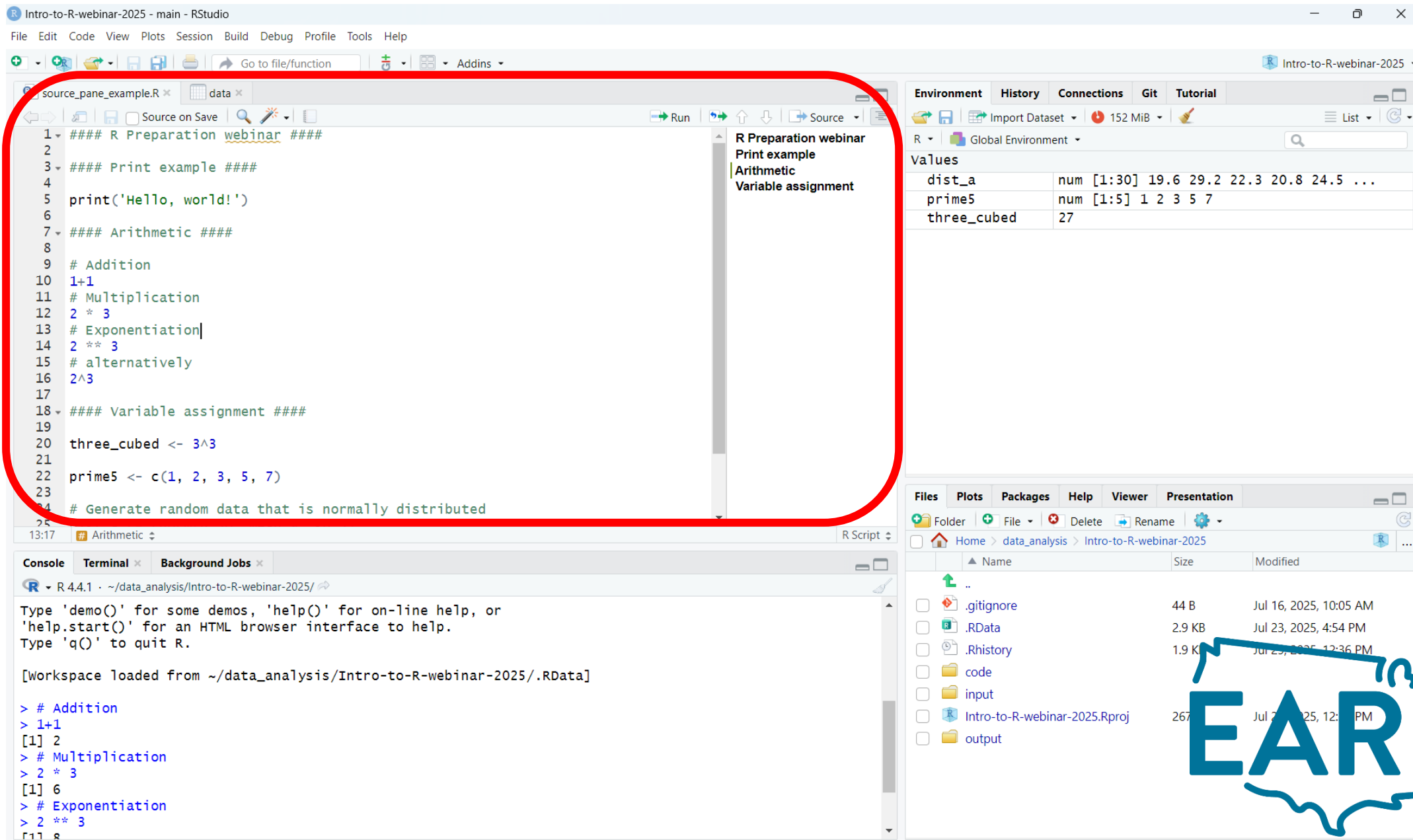
```
R 4.4.1 ~ /data_analysis/Intro-to-R-webinar-2025/
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/data_analysis/Intro-to-R-webinar-2025/.RData]

> # Addition
> 1+1
[1] 2
> # Multiplication
> 2 * 3
[1] 6
> # Exponentiation
> 2 ** 3
[1] 8
```



# Navigating the user interface (UI) – Source pane



The screenshot displays the RStudio interface with the Source pane highlighted by a red rounded rectangle. The Source pane contains the following R code:

```
1 ##### R Preparation webinar #####
2
3 ##### Print example #####
4
5 print('Hello, world!')
6
7 ##### Arithmetic #####
8
9 # Addition
10 1+1
11 # Multiplication
12 2 * 3
13 # Exponentiation
14 2 ** 3
15 # alternatively
16 2^3
17
18 ##### Variable assignment #####
19
20 three_cubed <- 3^3
21
22 prime5 <- c(1, 2, 3, 5, 7)
23
24 # Generate random data that is normally distributed
```

The Environment pane on the right shows the following values:

Variable	Value
dist_a	num [1:30] 19.6 29.2 22.3 20.8 24.5 ...
prime5	num [1:5] 1 2 3 5 7
three_cubed	27

The Console pane at the bottom shows the output of the code:

```
R 4.4.1 ~ /data_analysis/Intro-to-R-webinar-2025/
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/data_analysis/Intro-to-R-webinar-2025/.RData]

> # Addition
> 1+1
[1] 2
> # Multiplication
> 2 * 3
[1] 6
> # Exponentiation
> 2 ** 3
[1] 8
```

EARN

# Navigating the user interface (UI) – Environment pane

The screenshot displays the RStudio user interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with icons for saving, running, and other functions. The main editor window on the left shows a script named 'source\_pane\_example.R' with R code for a preparation webinar, including comments, a print statement, arithmetic operations, and variable assignments. The Environment pane on the right, highlighted with a red box, shows the current environment with variables 'dist\_a', 'prime5', and 'three\_cubed'. The Console at the bottom shows the output of the R script, including the workspace loading message and the results of the arithmetic operations.

**Environment pane (highlighted):**

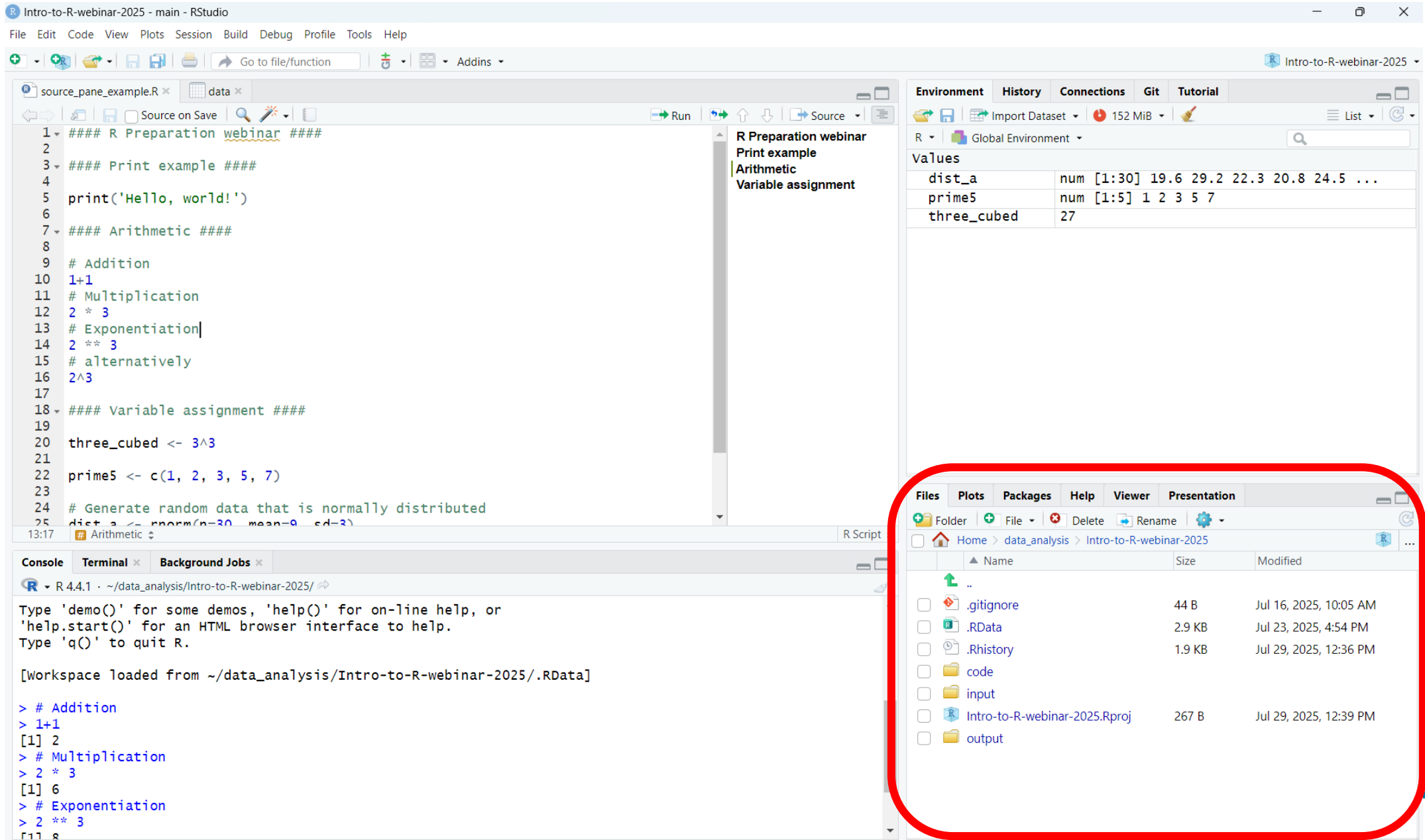
Variable	Value
dist_a	num [1:30] 19.6 29.2 22.3 20.8 24.5 ...
prime5	num [1:5] 1 2 3 5 7
three_cubed	27

**Console output:**

```
R 4.4.1 ~ /data_analysis/Intro-to-R-webinar-2025/  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
[Workspace loaded from ~/data_analysis/Intro-to-R-webinar-2025/.RData]  
  
> # Addition  
> 1+1  
[1] 2  
> # Multiplication  
> 2 * 3  
[1] 6  
> # Exponentiation  
> 2 ** 3  
[1] 8
```

EARN

# Navigating the user interface (UI) – Output pane



The screenshot displays the RStudio user interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for file operations and a search bar. The main editor window shows a script named 'source\_pane\_example.R' with R code for preparation, printing, arithmetic, and variable assignment. The Environment pane on the right shows the Global Environment with variables 'dist\_a', 'prime5', and 'three\_cubed'. The Console at the bottom shows the R prompt and the output of the executed code. A red box highlights the Files pane in the bottom right corner, which shows the file structure of the project.

**Source Editor:**

```
1 ##### R Preparation webinar #####
2
3 ##### Print example #####
4
5 print('Hello, world!')
6
7 ##### Arithmetic #####
8
9 # Addition
10 1+1
11 # Multiplication
12 2 * 3
13 # Exponentiation
14 2 ** 3
15 # alternatively
16 2^3
17
18 ##### Variable assignment #####
19
20 three_cubed <- 3^3
21
22 prime5 <- c(1, 2, 3, 5, 7)
23
24 # Generate random data that is normally distributed
25 dist_a <- rnorm(n=30, mean=9, sd=3)
```

**Environment Pane:**

Variable	Value
dist_a	num [1:30] 19.6 29.2 22.3 20.8 24.5 ...
prime5	num [1:5] 1 2 3 5 7
three_cubed	27

**Files Pane (highlighted):**

Name	Size	Modified
..		
.gitignore	44 B	Jul 16, 2025, 10:05 AM
.RData	2.9 KB	Jul 23, 2025, 4:54 PM
.Rhistory	1.9 KB	Jul 29, 2025, 12:36 PM
code		
input		
Intro-to-R-webinar-2025.Rproj	267 B	Jul 29, 2025, 12:39 PM
output		

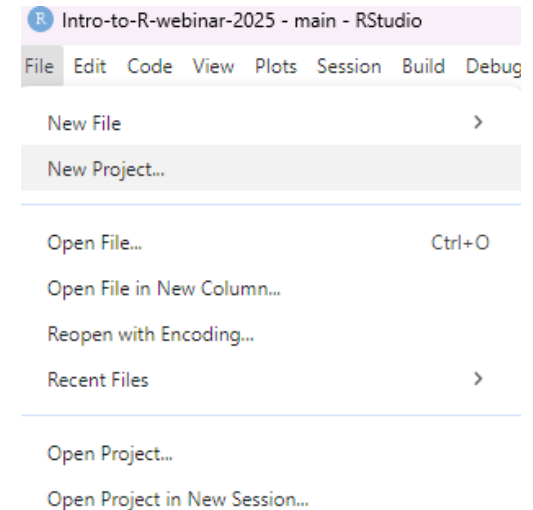
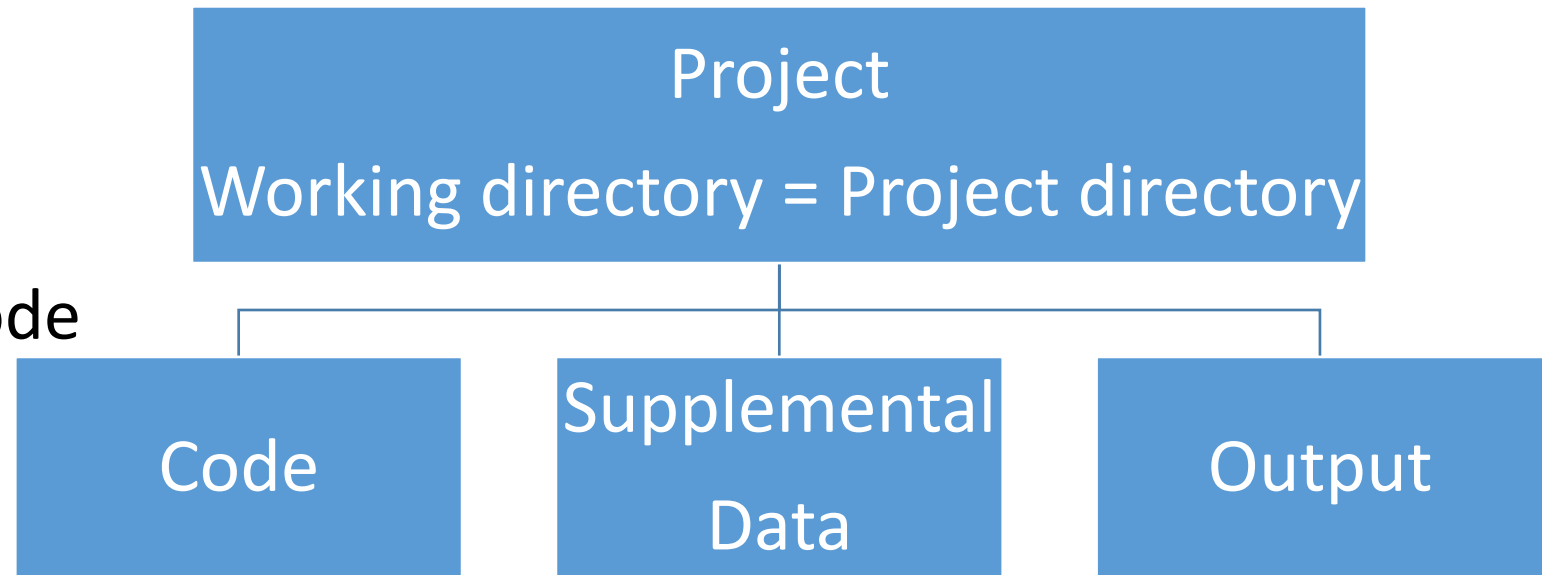
**Console:**

```
> # Addition
> 1+1
[1] 2
> # Multiplication
> 2 * 3
[1] 6
> # Exponentiation
> 2 ** 3
[1] 8
```

# Project-oriented Workflow

- File system discipline: Project root directory
- Working directory intentionality
- File path discipline: files relative to working directory

- Self-contained & portable
  - Reliable and precise code
- Organization
  - Accurate code



"Any resident R script is written assuming that it will be run from a fresh R process with working directory set to the project directory. It creates everything it needs, in its own workspace or folder, and it touches nothing it did not create. This convention guarantees that the project can be moved around on your computer or onto other computers and will still "just work". "

# Basic commands: R is a calculator

- Simple arithmetic calculations

```
# run in console
```

```
1+1
```

```
## [1] 2
```

```
4*9
```

```
## [1] 36
```

```
3^(1/2)
```

```
## [1] 1.732051
```

```
1+3+5+7+9
```

```
## [1] 25
```

- Use **<-** to save objects in the workspace

```
# define x as 5
```

```
x <- 5
```

```
# x + 1 ~ 5 + 1
```

```
x + 1
```

```
## [1] 6
```

# Basics and key topics: Packages

- Installing a package:
  - `install.packages("tidyverse")`
  - `install.packages("here")`
- Loading a package
  - `library(tidyverse)`
  - `library(here)`



- Tidy philosophy:
1. Human readable
  2. Consistent
  3. Composable



# Chained operations using “pipes”

- Piping: passing the value of the left-hand-side (LHS) to the to the first argument on the right-hand-side (RHS)

```
• x |> f(y) ~ f(x, y)
# use pipe to filter org_sample
org_sample |> select(wage)
## # A tibble: 242,582 × 1
##   wage
##   <dbl>
## 1  30.6
## 2  15.6
## 3    9
## # i 242,572 more rows
```

|> or %>%

- Chaining multiple operations:

```
org_sample |>
# TX wage and salary workers
filter(statefips == 48,
       age >= 16, cow1 <= 5, wage > 0) |>
# annualize weight
mutate(adj_wgt = orgwgt/12)
## # A tibble: 7,336 × 139
##   year month minsamp hrhhid          hrhhid2 hrsample hrsersuf huhhnum pulineno
##   <int> <int>   <int> <chr>          <chr>      <chr>      <chr>      <int>    <int>
## 1  2023     1       8 6074173588100... 13011      ""         ""         NA        1
## 2  2023     1       8 2100074149414... 13011      ""         ""         NA        2
## 3  2023     1       4 0408103688873... 15011      ""         ""         NA        1
## # i 7,326 more rows
```

# Basic workflow example

1. Create project
2. Create file structure as needed (code, data, output folders)
3. Create and save script
4. Install and load packages
5. Data analysis/manipulation code
6. Export to .xlsx or .csv file





# Helpful resources

- EARN Code Library
  - [https://economic.github.io/earn\\_code\\_library/](https://economic.github.io/earn_code_library/)
- R for Data Science book
  - <https://r4ds.had.co.nz/>
- Free R tutorial
  - <https://pll.harvard.edu/course/data-science-r-basics>
- Troubleshooting tips
  - Stack overflow, Reddit, other Q&A forums
  - AI assistance?
  - <https://statisticsglobe.com/errors-warnings-r> (needs to be reviewed)



# More helpful resources

- [Quick list of useful R packages \(posit support\)](#)
- [What They Forgot to Teach You About R](#)
- [Beginner's tutorial to R packages \(DataCamp\)](#)
- Stay in touch! Reach out to the EARN team with coding questions!

