

Introduction to R and RStudio



Session outline

1. What is R?
2. Why should we use R?
3. R and RStudio
4. Packages
5. Using the console
6. Scripts

What is R?

- R is a free, **open source** language for statistical computing.

1992 Development began as research project in Auckland, by Robert Gentleman and Ross Ihaka, based on the S (1976) and S-PLUS (1988) languages.

1993 First release of R

2000 Version 1.0 released

2004 First UseR! conference in Vienna, Austria

2013 Version 3.0 released

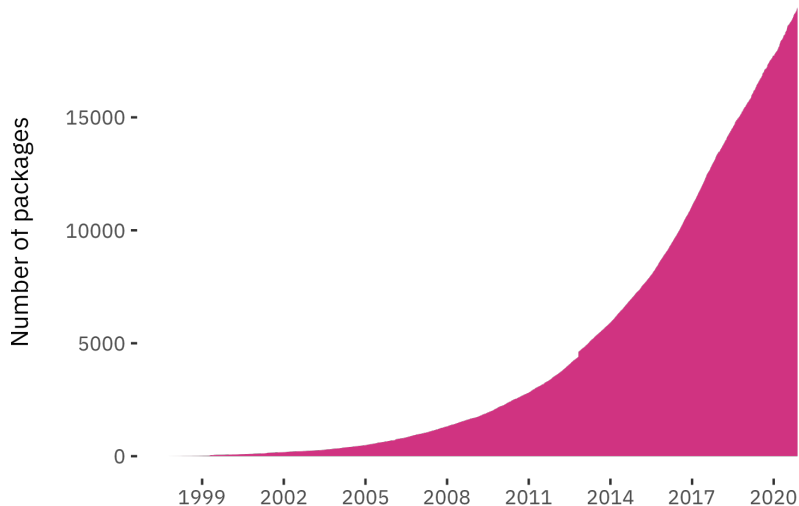
2015 R Consortium founded

2017 10,000 packages published on CRAN

2020 Version 4.0 released

- R is one of the **fastest growing** programming languages, especially for statistics and data science.
- A big, friendly ecosystem.

Over 23,000 packages (as of 2026)



Why use R?

Strengths

- Free and open source
- Big user community
- Excellent resources
- Incredibly flexible
- >23k packages
- Rapid development
- Industry adoption
- It's a programming language

Why use R?

Strengths

- Free and open source
- Big user community
- Excellent resources
- Incredibly flexible
- >23k packages
- Rapid development
- Industry adoption
- It's a programming language

Weaknesses

- >23k packages.
Which should I use?
- Rapid development
- Package management can be a pain
- ~~R is slow~~
(Less true today; many functions written in faster, low-level languages like C++ or Fortran.)
- It's a programming language



Licence

Open

Open

Open

Closed

Open

Version 1.0

2000

1994

2018

1985

1985

Easy of use*

Moderate

Hard

Moderate

Easy

Very hard

Extensibility

Excellent

Good

Good

Limited

Very limited

Industry support

Excellent

Excellent

Limited

Limited

Good

Performance

OK

Good

Very good

Good

Excellent

Execution

Interpreted

Interpreted

JIT

Interpreted

Compiled

*For statistical computing.

AS SEEN BY USERS OF ...

STATA



sas



STATA



sas



Should I learn R or Python?



RStudio

Installing R and RStudio

R is a programming language.

RStudio is a software environment for working with R.

RStudio is developed by Posit, a data science company that builds tools for R, Python, and publishing.

You need both R and an editor (like RStudio) to get started.

Installing R and RStudio

R is a programming language.

RStudio is a software environment for working with R.

RStudio is developed by Posit, a data science company that builds tools for R, Python, and publishing.

You need both R and an editor (like RStudio) to get started.

Other options you might consider:


- Positron – a “data science code editor”

<https://posit.co/products/ide/positron/>

- VS Code: a popular general purpose editor

<https://code.visualstudio.com/>

Install R from CRAN



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

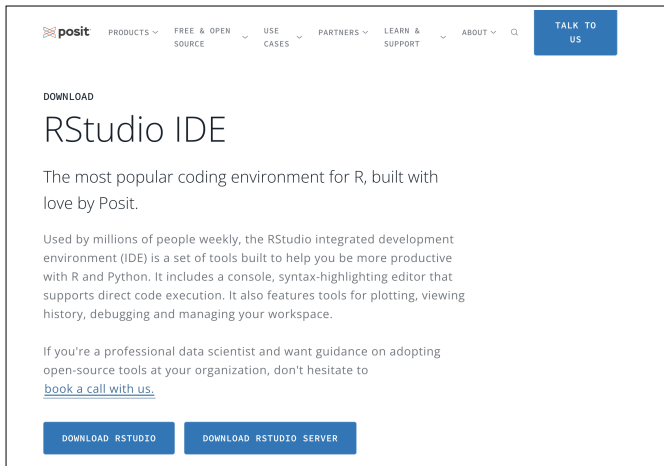
Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-07-02, Feather Spray) [R-3.5.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing

cran.r-project.org

Install RStudio from Posit



The screenshot shows the Posit website's page for downloading RStudio IDE. At the top, there is a navigation bar with the Posit logo and several menu items: PRODUCTS, FREE & OPEN SOURCE, USE CASES, PARTNERS, LEARN & SUPPORT, and ABOUT. A search icon is also present. A blue button labeled "TALK TO US" is in the top right corner. Below the navigation bar, the word "DOWNLOAD" is displayed in a smaller font. The main heading is "RStudio IDE". Below this, a paragraph states: "The most popular coding environment for R, built with love by Posit." Another paragraph describes the IDE: "Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution. It also features tools for plotting, viewing history, debugging and managing your workspace." A third paragraph says: "If you're a professional data scientist and want guidance on adopting open-source tools at your organization, don't hesitate to [book a call with us.](#)" At the bottom, there are two blue buttons: "DOWNLOAD RSTUDIO" and "DOWNLOAD RSTUDIO SERVER".

posit PRODUCTS FREE & OPEN SOURCE USE CASES PARTNERS LEARN & SUPPORT ABOUT TALK TO US

DOWNLOAD

RStudio IDE

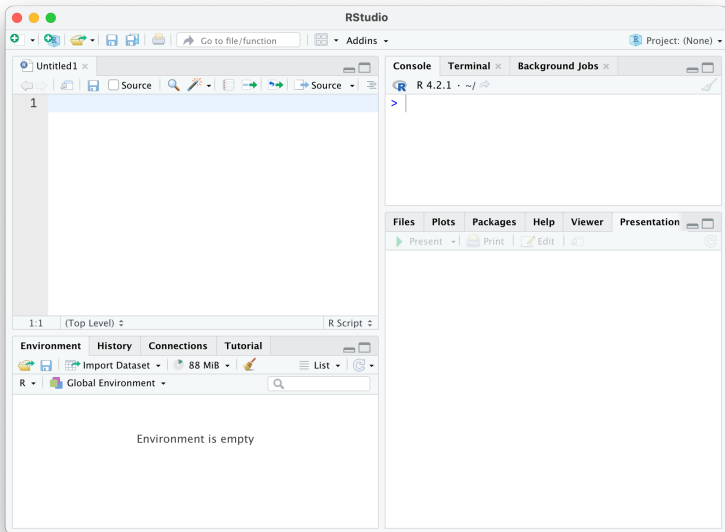
The most popular coding environment for R, built with love by Posit.

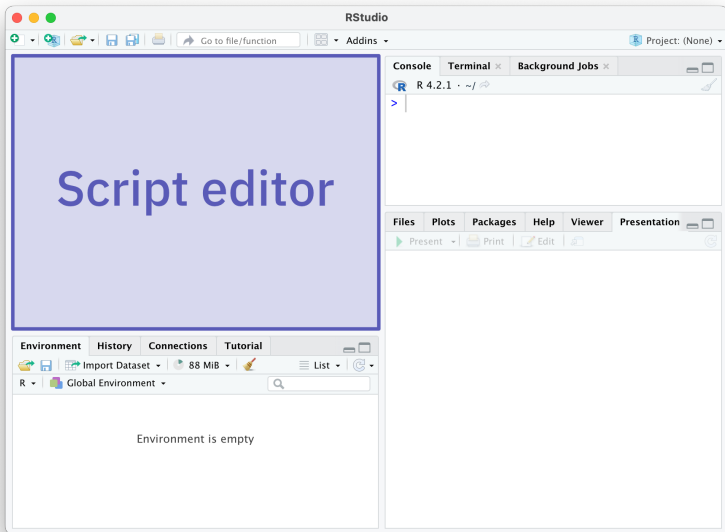
Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution. It also features tools for plotting, viewing history, debugging and managing your workspace.

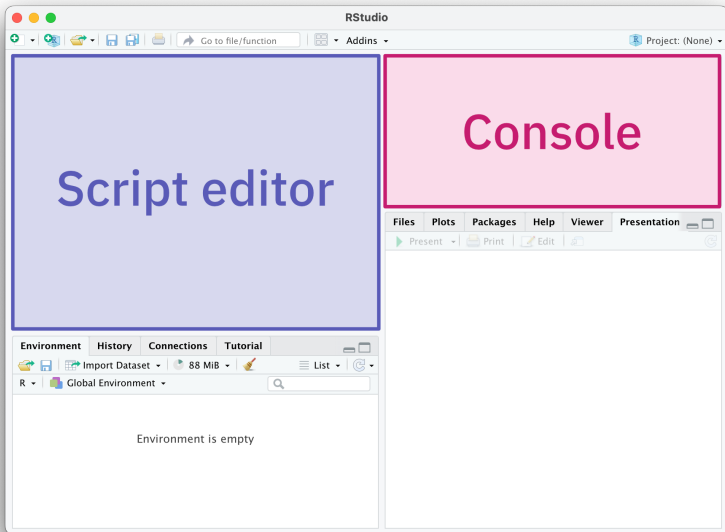
If you're a professional data scientist and want guidance on adopting open-source tools at your organization, don't hesitate to [book a call with us.](#)

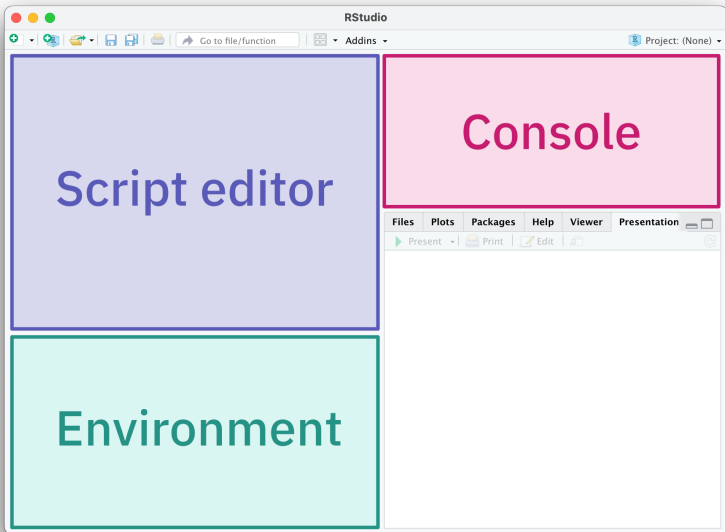
DOWNLOAD RSTUDIO DOWNLOAD RSTUDIO SERVER

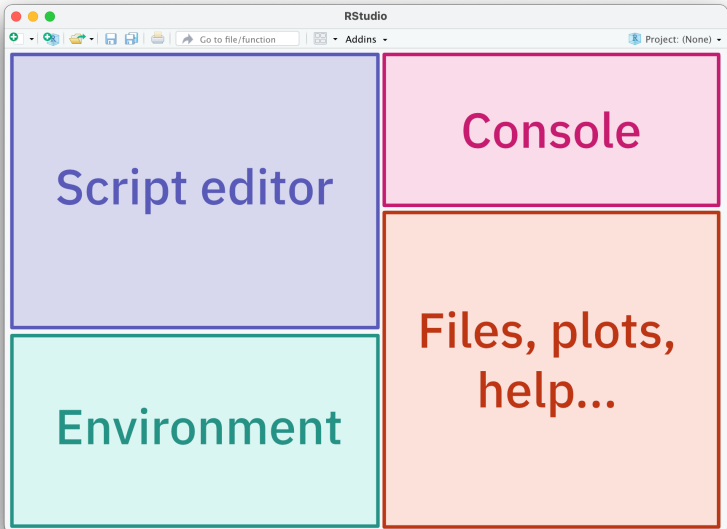
posit.co/download/rstudio-desktop



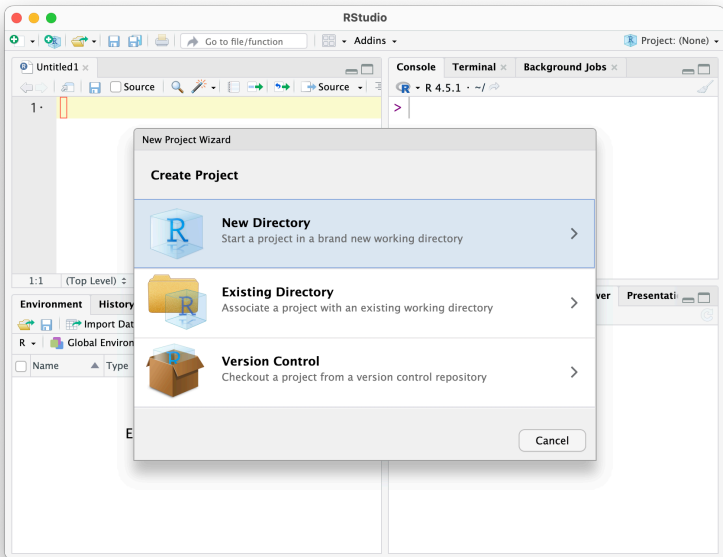








Projects in RStudio



R packages

R packages

We use packages to **add functionality** to R.

A collection of scripts designed to implement a particular function or method.

They represent one of the principle strengths of R: if a method exists, someone has probably written a package.

R comes with many packages out-of-the-box, such as **datasets**, **graphics**, and **stats**. These are referred to as **'base'** packages.

Installing packages

We can install packages by typing:

```
install.packages("lme4")
```

To load this package, we'd then type:

```
library(lme4)
```

(Note the use of quotation marks.)



You should **install** packages once. But you need to **load** them every time you use them.

How to choose which packages to use?

My preferences:

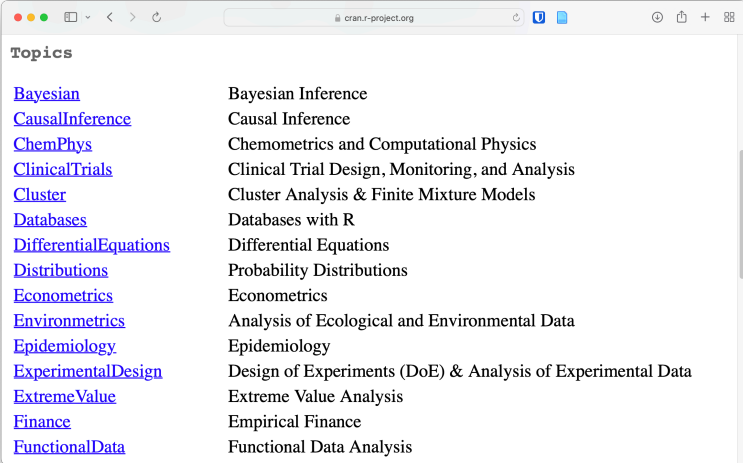
1. Use as **few** packages as possible (but as many as needed).
2. Use big **general** packages (rather than small specialised ones).
3. Use packages that have been **around for a long time** and are **regularly updated**.
4. Use **stable, reliable** packages.
5. At times, ignore all these rules.

Your preferences may differ!

For example: `ggplot2`, `dplyr`, `tidyr`, `data.table`, `lme4`

Use the CRAN Task Views

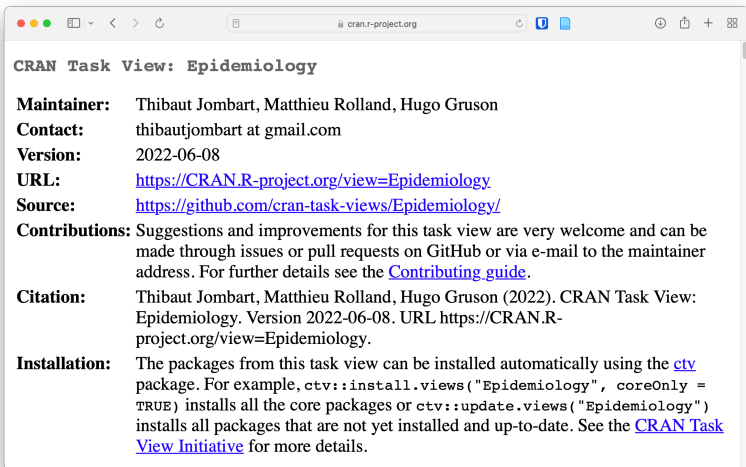
cran.r-project.org/web/views



The screenshot shows a web browser window with the address bar displaying `cran.r-project.org`. The page content is titled "Topics" and lists various statistical and computational topics, each with a blue underlined link on the left and a descriptive text on the right.

Topic Link	Description
Bayesian	Bayesian Inference
CausalInference	Causal Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Databases	Databases with R
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
Epidemiology	Epidemiology
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
ExtremeValue	Extreme Value Analysis
Finance	Empirical Finance
FunctionalData	Functional Data Analysis

cran.r-project.org/web/views/Epidemiology



The image shows a browser window with the address bar containing "cran.r-project.org". The page title is "CRAN Task View: Epidemiology". The content is organized into several sections:

- Maintainer:** Thibaut Jombart, Matthieu Rolland, Hugo Gruson
- Contact:** thibautjombart at gmail.com
- Version:** 2022-06-08
- URL:** <https://CRAN.R-project.org/view=Epidemiology>
- Source:** <https://github.com/cran-task-views/Epidemiology/>
- Contributions:** Suggestions and improvements for this task view are very welcome and can be made through issues or pull requests on GitHub or via e-mail to the maintainer address. For further details see the [Contributing guide](#).
- Citation:** Thibaut Jombart, Matthieu Rolland, Hugo Gruson (2022). CRAN Task View: Epidemiology. Version 2022-06-08. URL <https://CRAN.R-project.org/view=Epidemiology>.
- Installation:** The packages from this task view can be installed automatically using the [ctv](#) package. For example, `ctv::install.views("Epidemiology", coreOnly = TRUE)` installs all the core packages or `ctv::update.views("Epidemiology")` installs all packages that are not yet installed and up-to-date. See the [CRAN Task View Initiative](#) for more details.

Tip: Install packages with pak

pak

A Fresh Approach to R Package Installation

pak installs R packages from CRAN, Bioconductor, GitHub, URLs, git repositories, local files and directories. It is an alternative to `install.packages()` and `devtools::install_github()`. pak is fast, safe and convenient.

- [📖 Short tour](#)
- [🔗 Quick links \(start here if in doubt!\)](#)
- [🌟 Features](#)
- [📦 Installation](#)
- [📄 License](#)

LINKS

[View on CRAN](#)

[Browse source code](#)

[Report a bug](#)

LICENSE

[GPL-3](#)

COMMUNITY

[Code of conduct](#)

CITATION

[Citing pak](#)

DEVELOPERS

[Gábor Csárdi](#)

Author, maintainer

```
install.packages("pak")  
pak::pak("dplyr")
```

Scripts

Scripts are text files containing R code

Similar to Stata do-files or SPSS syntax files.

```
# Example of a script  
# Ewan Carr      Header explaining what  
# 2026-03-02    the script does  
  
library(tidyverse)  Load packages at the top  
library(lme4)  
  
# Import raw data Comments  
raw <- read_csv("raw.csv")  
  
# Fit a linear model  
model <- lm(y ~ x1 + x2, data = raw)
```

Scripts (cont.)

- You should annotate your scripts with **comments**.

```
# This is a comment.
```

What makes a good comment?

- Save your scripts with the **.R** file extension
e.g., 1-Data-prep.R.
- Install packages once, load them **every time you use them**. Place at the top of your script.

```
library(tidyverse)  
library(here)
```

Use the console for exploration; use scripts for anything you want to keep.

Organising your scripts

You can organise your scripts with [sections](#) and [headers](#).

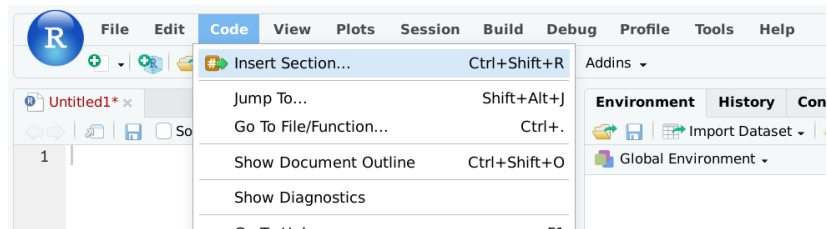
```
# Section One -----  
  
# Section Two =====  
  
### Section Three #####
```

Organising your scripts

You can organise your scripts with [sections](#) and [headers](#).

```
# Section One -----  
# Section Two =====  
### Section Three #####
```

These can be inserted in RStudio with **Ctrl** + **↑** + **R** or
Code > **Insert Section...**

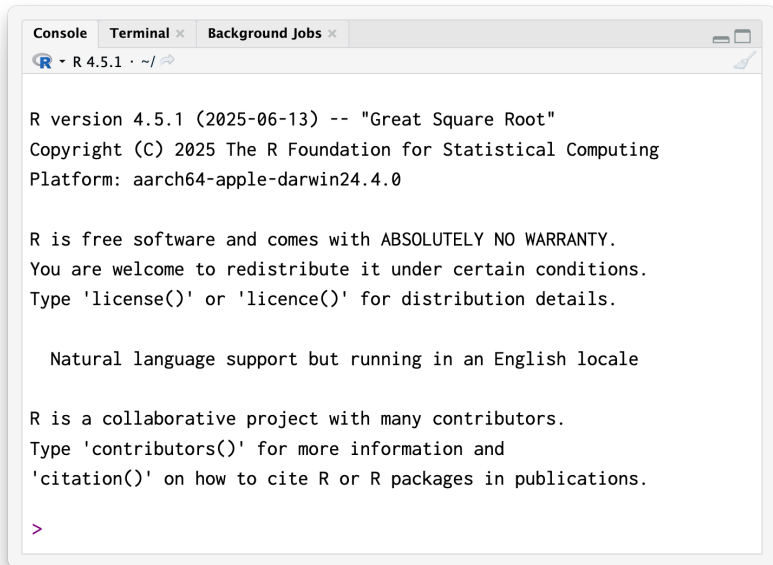




Combining several scripts

- Most analyses won't fit in a single script/file.
 - ↪ You might have separate scripts for data import, cleaning, analysis, plotting...
- We can use the `source` command to execute code contained in another file.







```
source("1-Import-raw-data.R")  
source("2-Data-cleaning.R")
```

The R console



```
Console Terminal x Background Jobs x  
R - R 4.5.1 - ~/    
  
R version 4.5.1 (2025-06-13) -- "Great Square Root"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin24.4.0  
  
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.  
  
>
```

Tips for the R console

- Access previous commands with  / .
- Search previous commands with  + .
- Clear the screen with  + .

Operators

Symbols that tell the compiler to perform specific mathematical or logical manipulations.

Arithmetic

- + addition
- subtraction
- * multiplication
- / division

\wedge or $**$ raising to a power

Relational

- $x < y$ less than
- $x > y$ greater than
- $x \leq y$ less than or equal to
- $x == y$ equal
- $x != y$ not equal

For more, see <https://www.statmethods.net/management/operators.html>

Getting help

You can get help for any function with `?`.

```
?rnorm    # Help for the 'rnorm' function  
?install.packages
```

All help files have a consistent structure:

Description → Usage → Arguments → Details → See Also → Examples

- I find the 'Arguments' and 'Examples' sections the most useful.
- It's worth familiarising yourself with this structure.

Resources

1. **A Succinct Introduction to R**

Steve Haroz

<http://r-guide.steveharoz.com>

2. **R for Data Science**

Hadley Wickham and Garrett Golemund

<https://r4ds.hadley.nz>

3. **Advanced R**

Hadley Wickham

<https://adv-r.hadley.nz>



Should I use generative AI to learn R?

Generative AI for Learning R

Good uses

- Quick help debugging
- Explain why code fails
- Speed up routine work

Be careful

- When it matters
- When you don't yet understand the problem
- Out-of-date or overconfident answers
- Data science has minimal boilerplate
- We need more thinking, not less