VIDO BIOINFO Bioinformatics and Systems Biology Lab

Canadian Bioinformatics Hub

Canadian Bioinformatics Workshops

Introduction to R Programming for Bioinformatics

Day 1- Module 1A: Inturoduction to R programming Language

Mohamed Helmy, PhD
Principal Scientist and Adjunct Professor
Bioinformatics and Systems Biology Lab
VIDO, University of Saskatchewan

6-7October 2025, VIDO, Saskatoon

Land Acknowledgement

"As we gather here today, we acknowledge we are on Treaty 6 Territory and the Homeland of the Métis. We pay our respect to the First Nations and Métis ancestors of this place and reaffirm our relationship with one another."

- University Council, University of Saskatchewan

CBW and CBH Open Contents

Creative Commons

This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska 中文 華語 (台灣) isiZulu

creative commons

Attribution-Share Alike 2.5 Canada

You are free:

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

Free Cultural APPROVED FOR Works

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French

Learn how to distribute your work using this licence

House Keeping Rules

Bathrooms

Coffee breaks

Lunches

Internet access

Workshop Objectives

By the end of this workshop, participants will be able to:

- **Navigate the R environment:** Install, load, and manage R packages.
- **Apply core programming concepts:** Variables, data structures, loops, and functions.
- **Organize and manipulate data:** Import, subset, reshape, and prepare datasets.
- **Visualize biological data:** Generate and customize plots for exploration and communication.
- **Work with Bioconductor:** Access and explore omic datasets using specialized data structures.
- **Preprocess and analyze omics data:** Perform normalization, basic visualization (PCA, heatmaps), and annotation.
- **Build confidence in R for bioinformatics:** Apply learned skills to real-world biological datasets.

Module 1: Getting to Know R

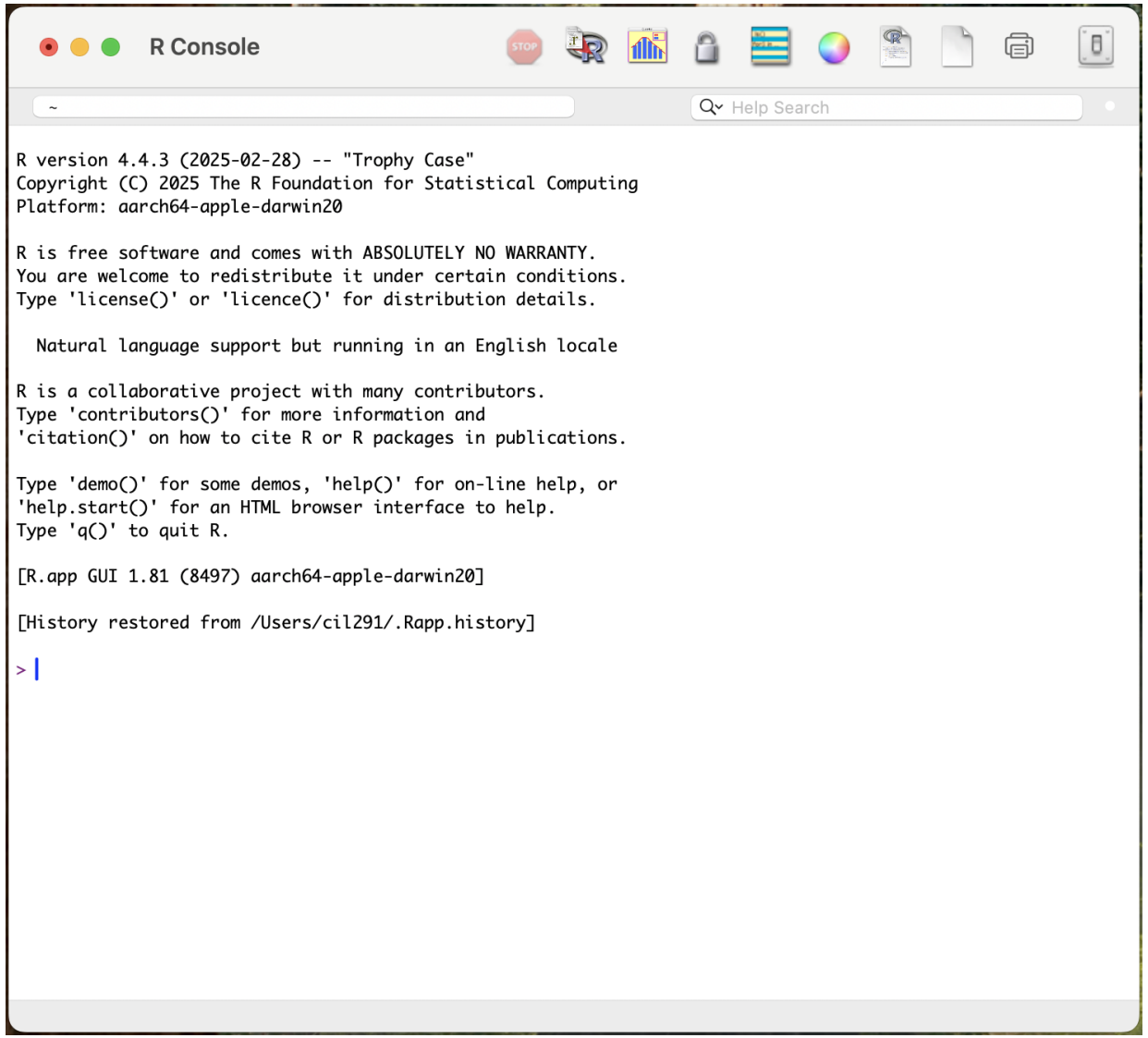
The environment, syntax, data types, and data handling

Coming next

What is R?

- **Open-source** programming language for statistics, visualization, and bioinformatics
- **Widely used** in biology, medicine, and data science
- **Supported** by a large community and thousands of packages
- **Free, cross-platform** (Windows, Mac, Linux)

R Environment & Interface



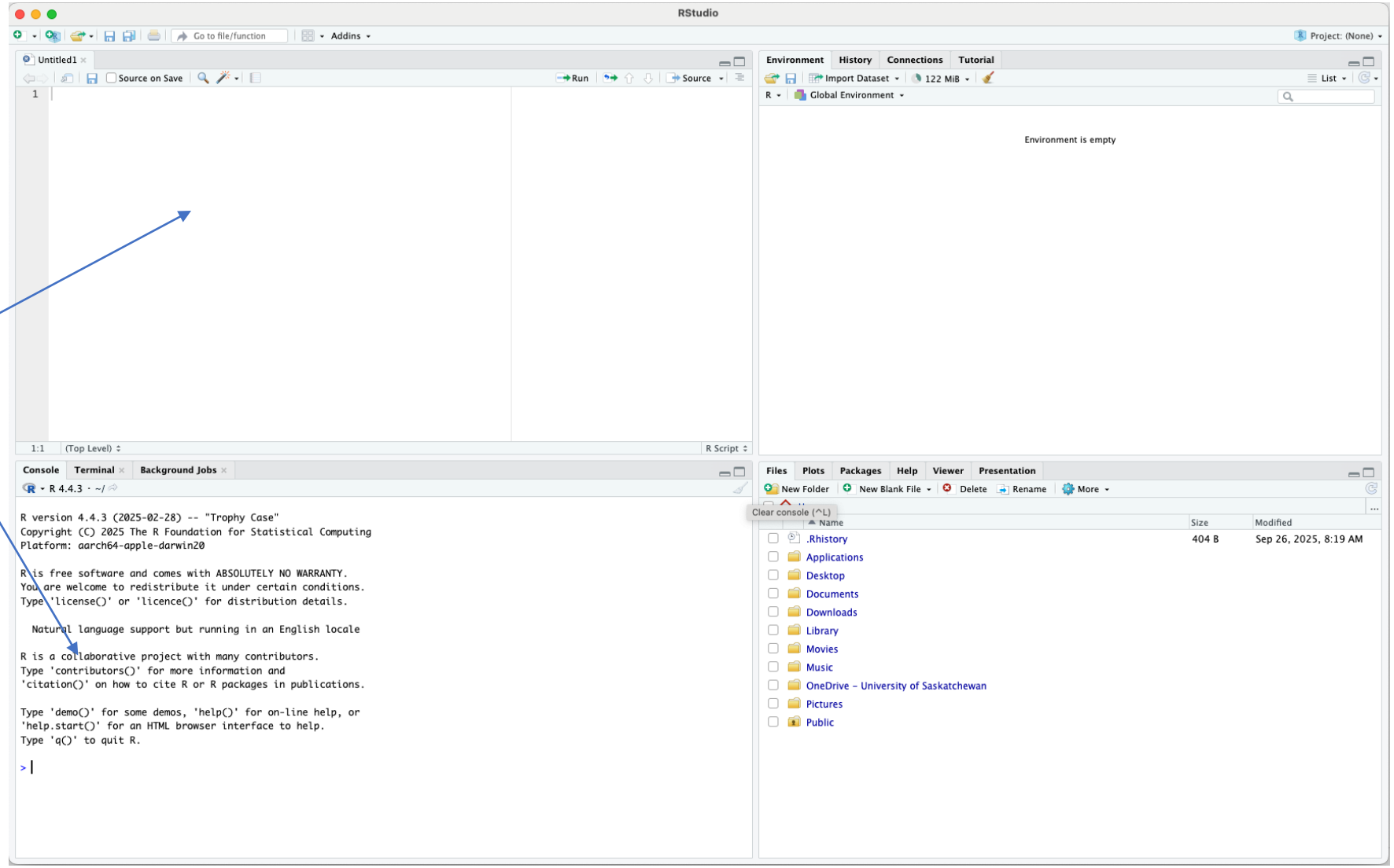
R Environment & Interface

Integrated development environment (IDE)

- Execute both R and Python code
- Syntax highlighting and code completion
- Manage multiple projects

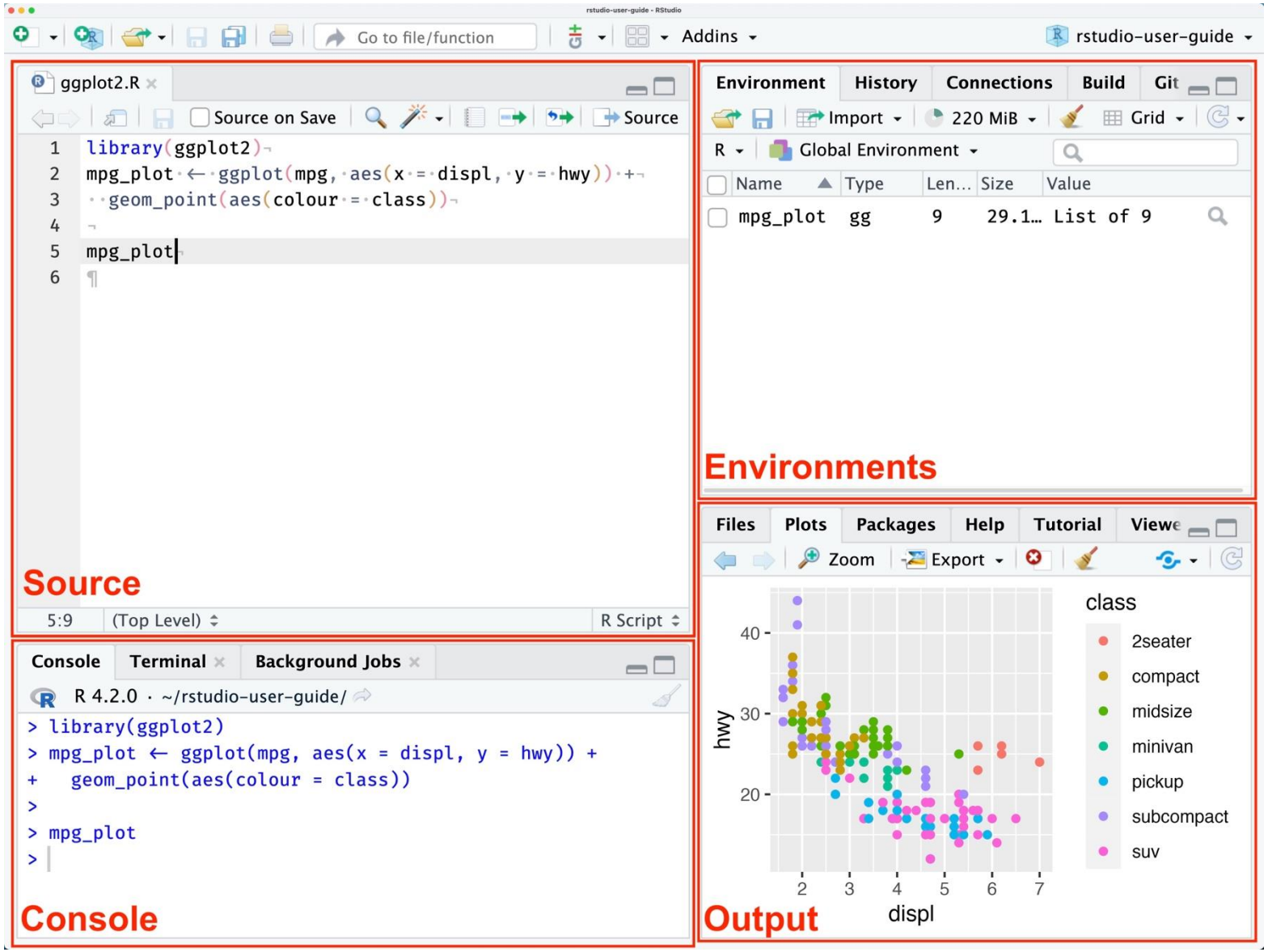
•Another popular environment are Jupyter notebooks and Google CoLab:

R Environment & Interface



Panes

R Environment & Interface

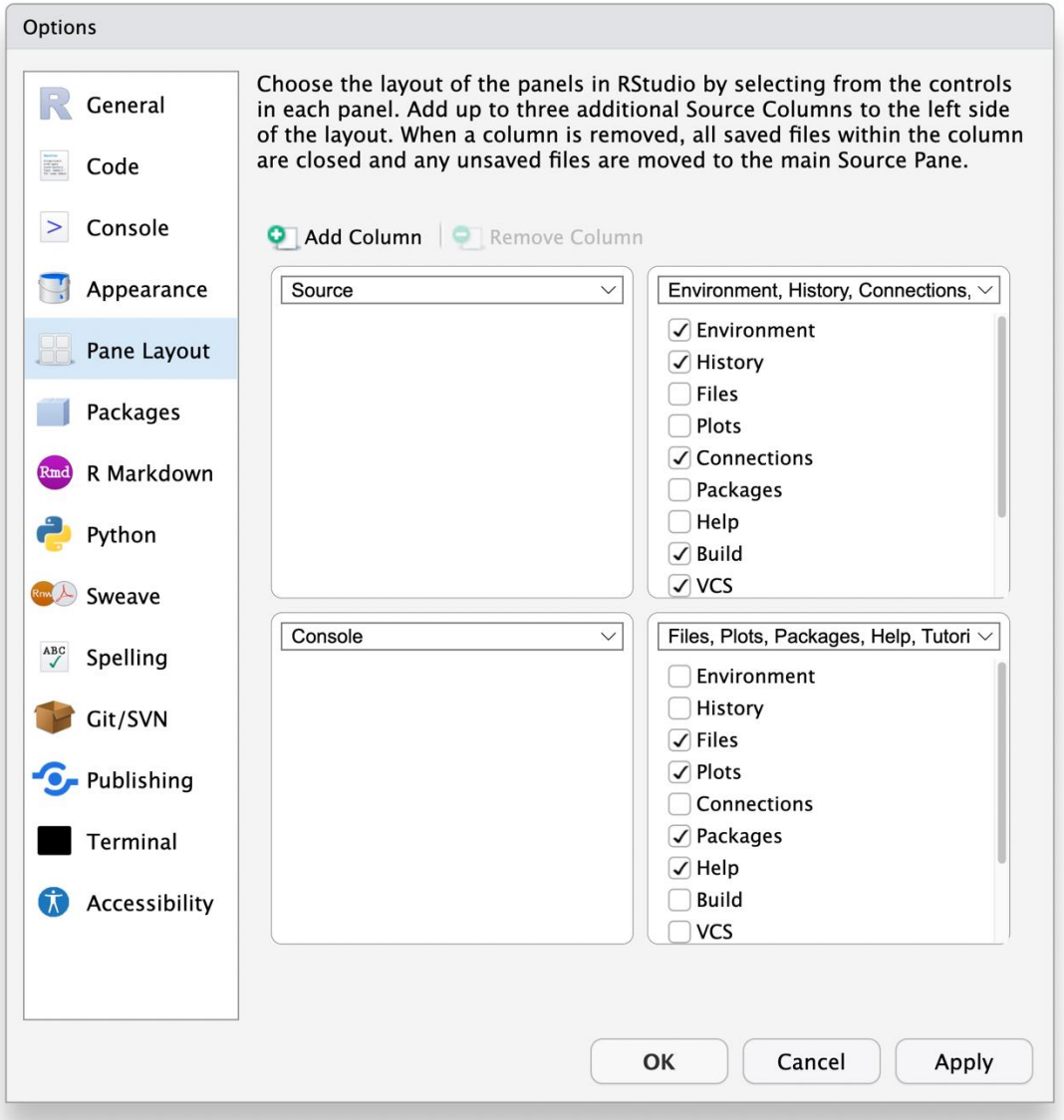


Source: RStudio IDE User Guide

R Environment & Interface

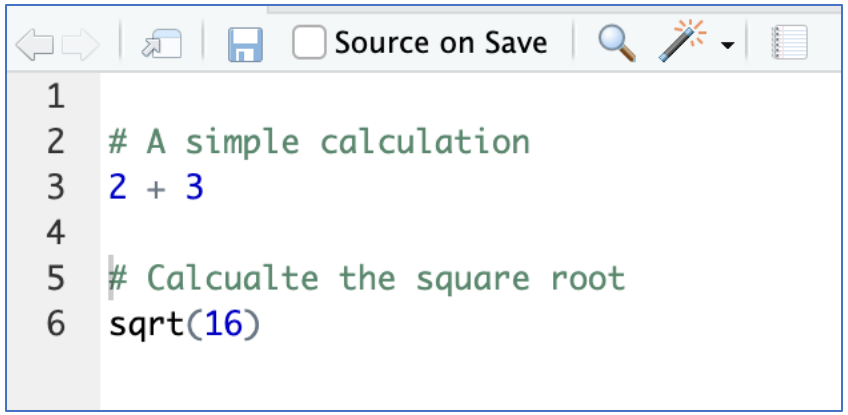
R Environment & Interface

Pane layout



First Steps in R

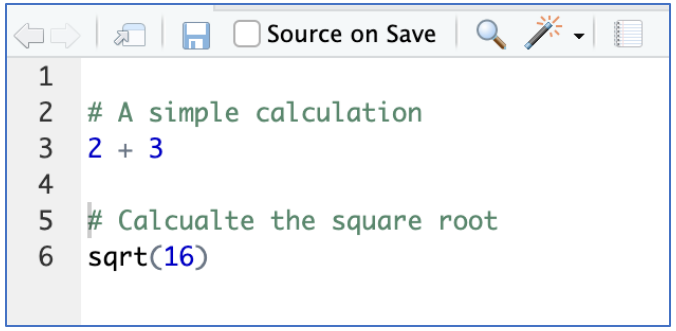
- We talk with R through a Syntax (grammar)
- It is as simple as using a Calculator

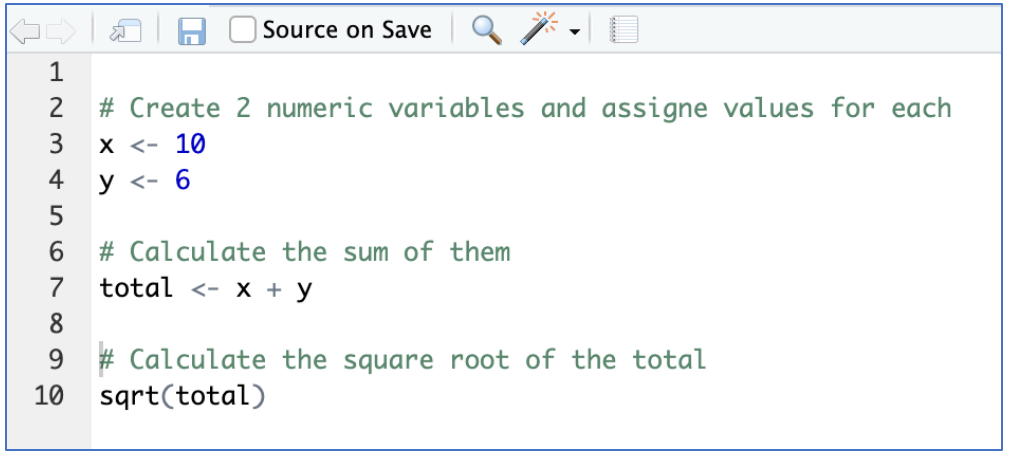


- And we add comments to make our code easy to understand

Variables and Assignments

- **Variables** are places to store **Values**
- **Values** are assigned to **Variable**
- The assignment in R is **=** or **<-** signs





Scripts and Projects in RStudio

- **Scripts**
  - Save your code in .R files → easy to re-run and share
  - Use comments (#) to explain steps
- **Projects**
  - Keep files, data, and scripts organized in one folder
  - Makes your work **reproducible**
- **Workspaces**
  - Save variables and reload sessions (.RData)

Basic Data Types in R

- **Numeric:** x <- 42
- **Character (string):** name <- "Alice", “Mike_123”
- **Logical (TRUE/FALSE):** flag <- TRUE, T
- **Factor (categorical):** group <- factor(c("control","treated"))
- **Date/Time:** Sys.Date()

Data Structures in R

- **Vector**: 1D collection of elements
- **Matrix:** 2D numeric data
- **Dataframe:** like a table with rows and columns
- **List:** collection of different objects

Data Structures in R

- **Vector**: v <- c(1, 2, 3)
- **Matrix:** m <- matrix(1:6, nrow=2)
- **Dataframe:** df <- data.frame(age=c(25,30), group=c("A","B"))
- **List:** lst <- list(numbers=v, info=df)

Saving & Loading Work

- **Save your script (.R file) to keep your code.**
- **Save your workspace (.RData) to reload objects later.**
- **Use save.image() and load() commands.**
- **Tip: Best practice is to save code, not just the workspace.**

THANK YOU

VIDO

VACCINE AND INFECTIOUS DISEASE ORGANIZATION

VIDO.ORG

UNIVERSITY OF SASKATCHEWAN