

Canadian Bioinformatics Workshops

# Introduction to R Programming for Bioinformatics

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

**Mohamed Helmy, PhD**

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

6-7 October 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 Azərbaycanca Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto  
Español Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)  
Euskara Suomi Galego ગુજરાતી hrvatski Magyar Italiano 日本語 한국어 Macedonian Malayu  
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски (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



**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

A screenshot of the R Console window. The title bar says 'R Console'. The window contains the following text:

```
R version 4.4.3 (2025-02-28) -- "Trophy Case"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20

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.

[R.app GUI 1.81 (8497) aarch64-apple-darwin20]

[History restored from /Users/cil291/.Rapp.history]

> |
```



# R Environment & Interface



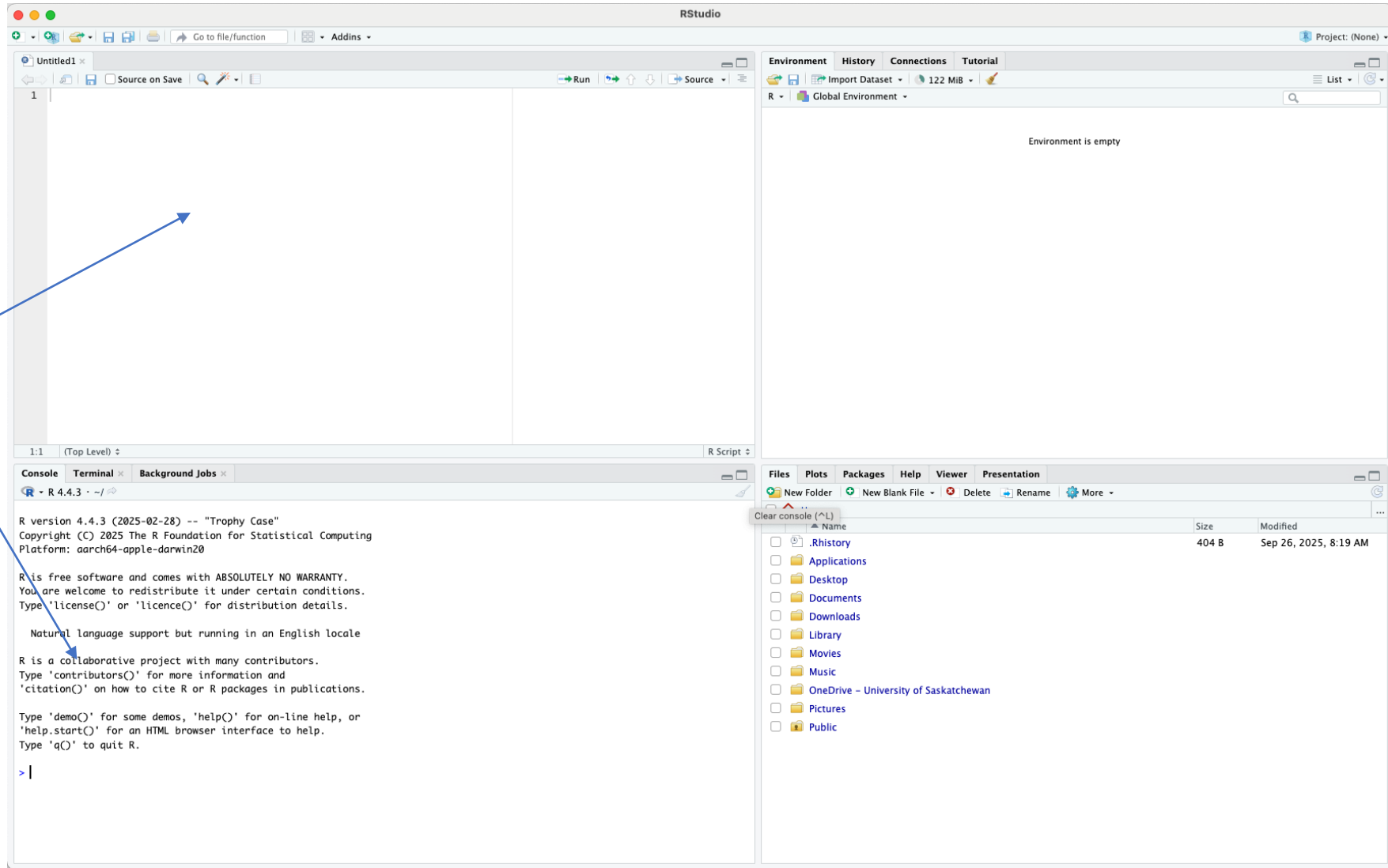
- Execute both R and Python code
  - Syntax highlighting and code completion
  - Manage multiple projects
- Another popular environment are Jupyter notebooks and Google CoLab:



# R Environment & Interface



Panes



# R Environment & Interface



The screenshot displays the RStudio IDE interface with four main panes:

- Source Pane:** Contains an R script named `ggplot2.R` with the following code:

```
1 library(ggplot2)
2 mpg_plot <- ggplot(mpg, aes(x = displ, y = hwy)) +
3   geom_point(aes(colour = class))
4
5 mpg_plot
6
```
- Environment Pane:** Shows the `Global Environment` with a table of objects:

Name	Type	Len...	Size	Value
mpg_plot	gg	9	29.1...	List of 9
- Console Pane:** Shows the execution of the script:

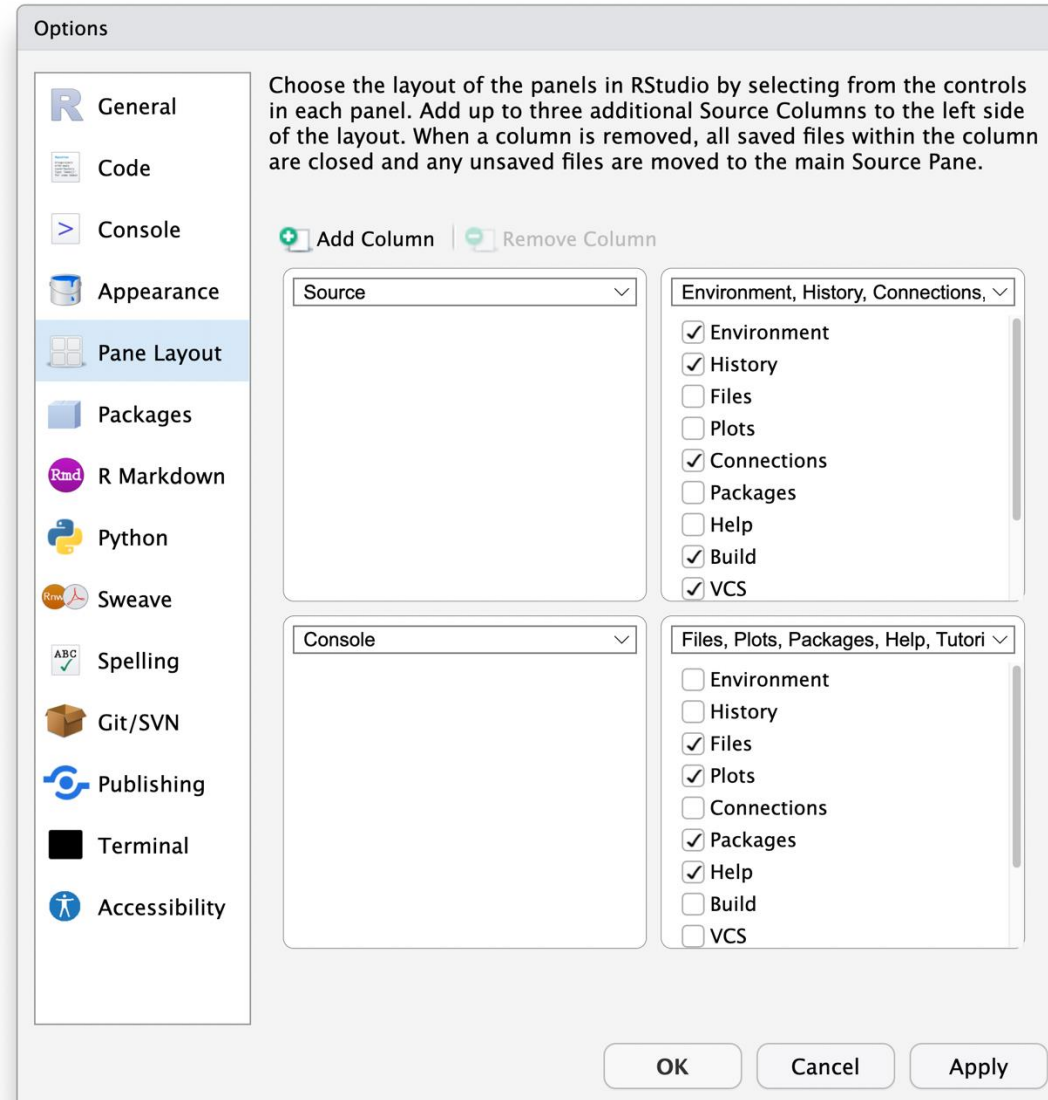
```
> library(ggplot2)
> mpg_plot <- ggplot(mpg, aes(x = displ, y = hwy)) +
+   geom_point(aes(colour = class))
>
> mpg_plot
>
```
- Plots Pane:** Displays a scatter plot of `hwy` vs `displ` colored by `class`. The legend shows the following categories:
  - 2seater
  - compact
  - midsize
  - minivan
  - pickup
  - subcompact
  - suv

# R Environment & Interface



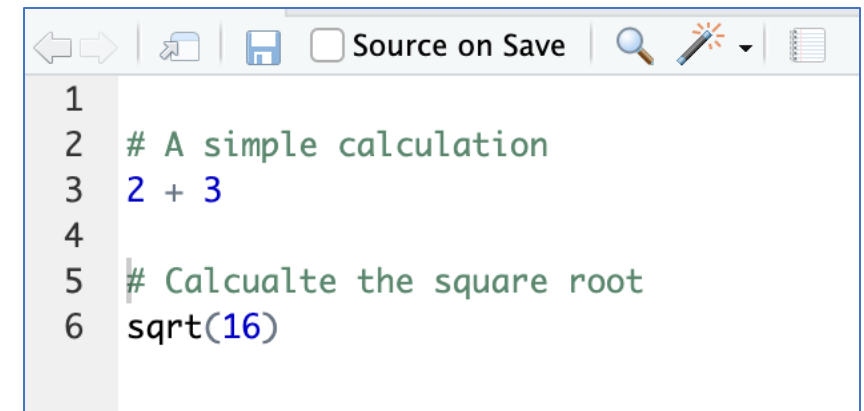
# R Environment & Interface

## Pane layout



# First Steps in R

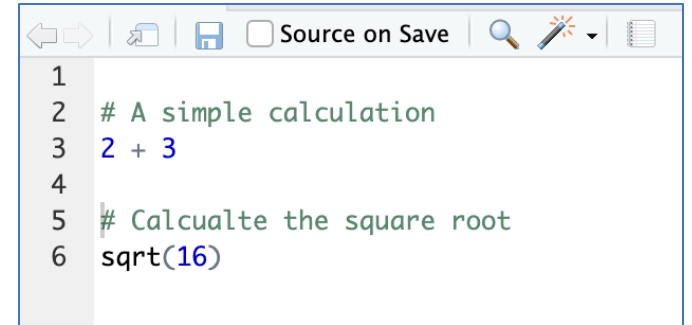
- We talk with R through a Syntax (grammar)
- It is as simple as using a Calculator
- And we add comments to make our code easy to understand



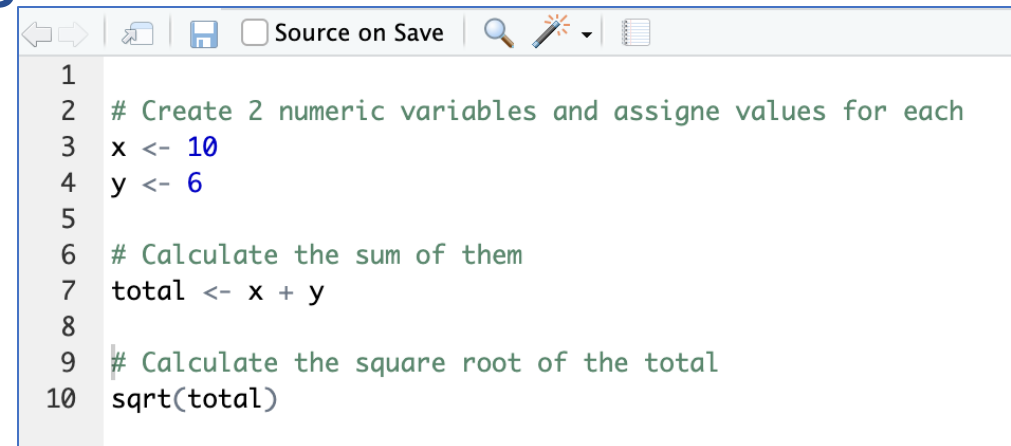
```
1  
2 # A simple calculation  
3 2 + 3  
4  
5 # Calcualte the square root  
6 sqrt(16)
```

# Variables and Assignments

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



```
1  
2 # A simple calculation  
3 2 + 3  
4  
5 # Calculate the square root  
6 sqrt(16)
```



```
1  
2 # Create 2 numeric variables and assign values for each  
3 x <- 10  
4 y <- 6  
5  
6 # Calculate the sum of them  
7 total <- x + y  
8  
9 # Calculate the square root of the total  
10 sqrt(total)
```

# 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



VACCINE AND INFECTIOUS DISEASE ORGANIZATION

**VIDO.ORG**

