Introduction to R

Lesson 1

3rd September 2024

In this course (6.5 hours) you will learn:

- How to use R studio to write R code
- R syntax, data types, operators and functions
- Dara wrangling*
- Visualizations

Introduction to R language

- R is a **language** and **environment** written and released by Ross Ihaka and Robert Gentleman in the 90s for statistical computing and graphics-facilitated by packages (21,203)

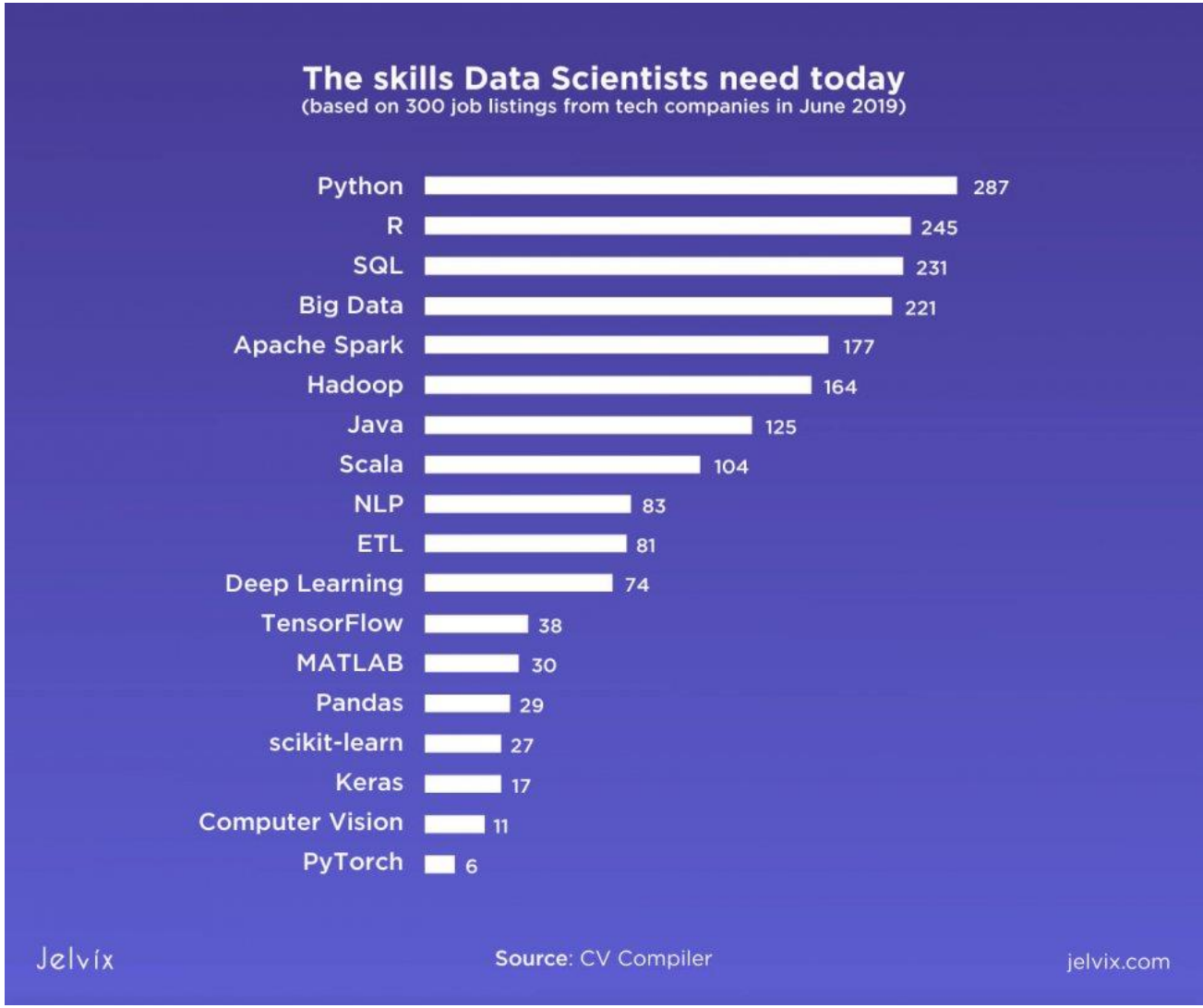
- R: effective data handling, a suite of operators, collection of packages for data analysis, graphical facilities for data visualization

- R: open source software. Not only is the software free but packages are also free.

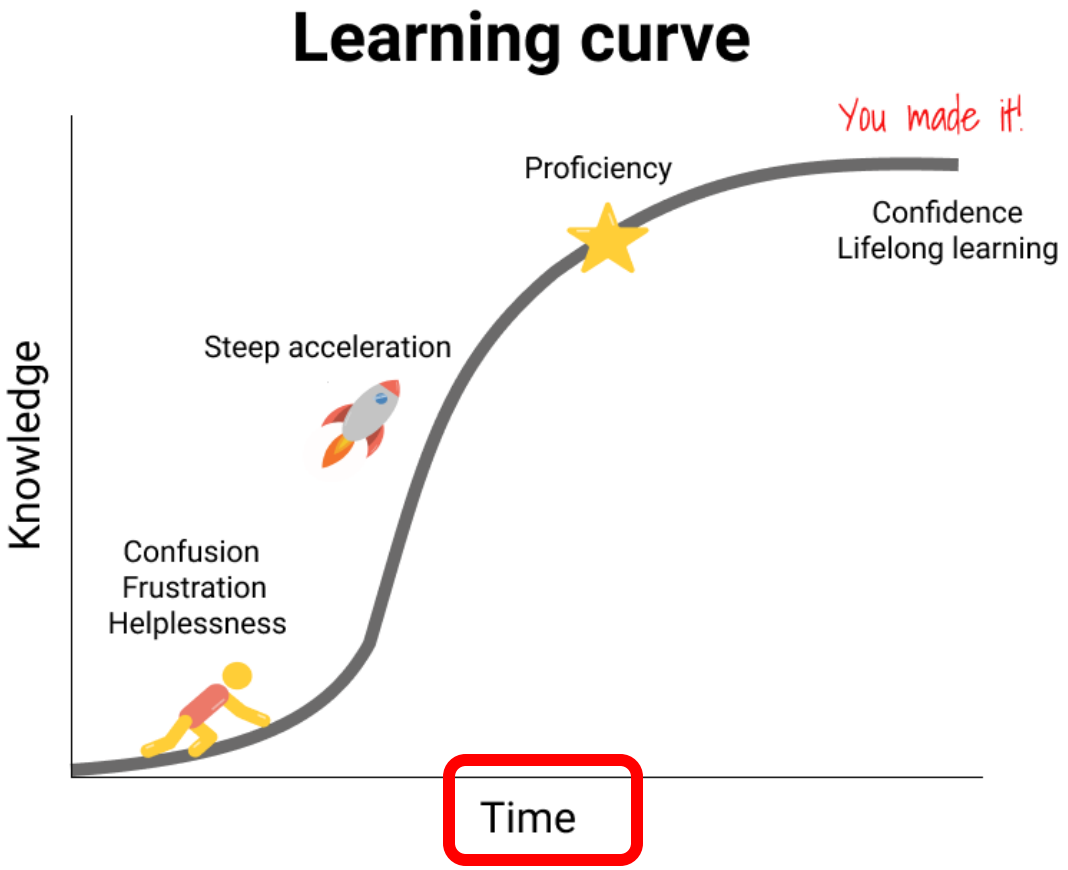
R: Among many other Programming Languages

IntelliPaat

Programming Language	Data Science Tasks Each Performs
python	• Conducts data mining • Carries out ML algorithms
R	• Performs data visualization • Conducts Data Analytics
Scala	• Eases the performance on high datasets • Sculpts data in any given form
julia	• Solves mathematical complications at high speed • Performs Data Analytics
Java	• Wise option for IoT and Big Data • Secure enough to work with sensitive data
SQL	• Manages large databases • Compliant toward Data Science workflow
MATLAB	• Performs profound mathematical operations • Highly specialized in working with Big Data
JavaScript	• Sets up data visualizations perfectly • Good fit for projects based on web and Big Data technologies
SAS	• Manipulates and manages data • Administers data analysis through statistical models
C++	• Used in Big Data in collaboration with Java • Computes large datasets quickly



Patience | Time | Practice | Learning from Mistakes

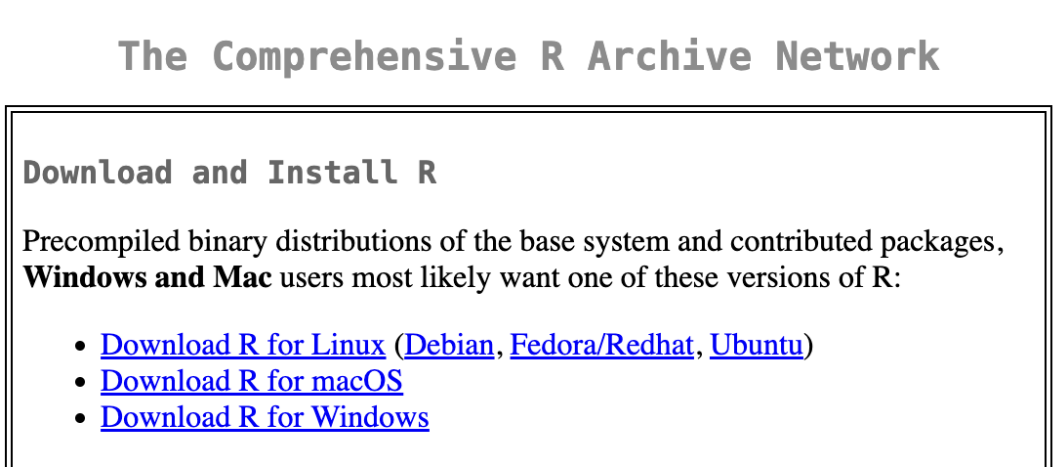


Common applications of R

- I have two datasets, and I want to merge them
- I have a very large file (Excel can’t open) how do I select a few columns to work with?
- Data transformations
- Data cleaning
- 16S analysis: *DADA2*
- RNA-transcriptomics analysis: *DESeq2*
- Statistical analysis
- Visualizations: how to I generate publication ready plots?
- Spatial analysis: creating maps
- Dashboards and interactive interfaces
- **Other examples?**

Where do I start

Step 1: Downloading & Install R (https://cran.r-project.org/)



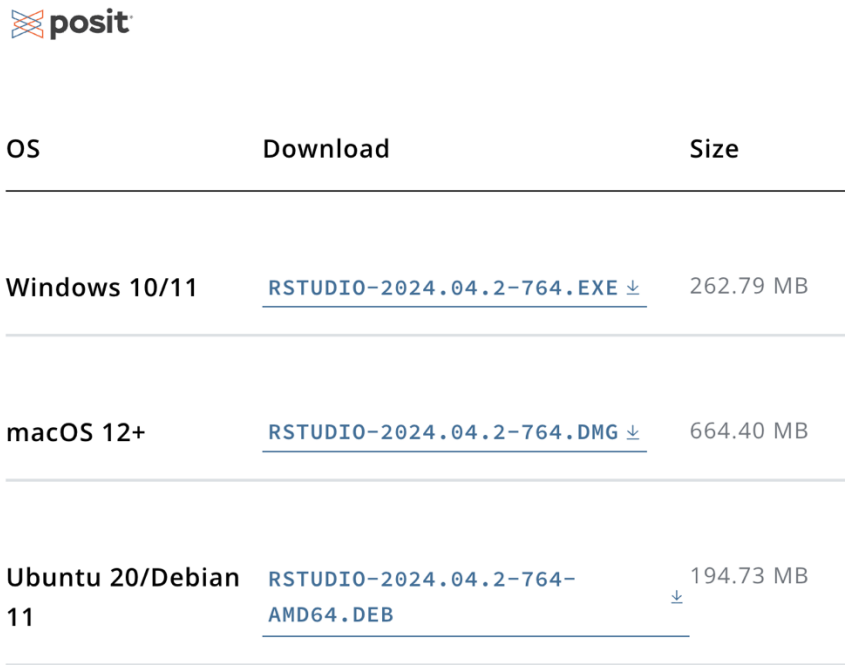
The Comprehensive R Archive Network

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 (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

Step 2: Download and install R studio (user interface) (https://posit.co/download/rstudio-desktop/)



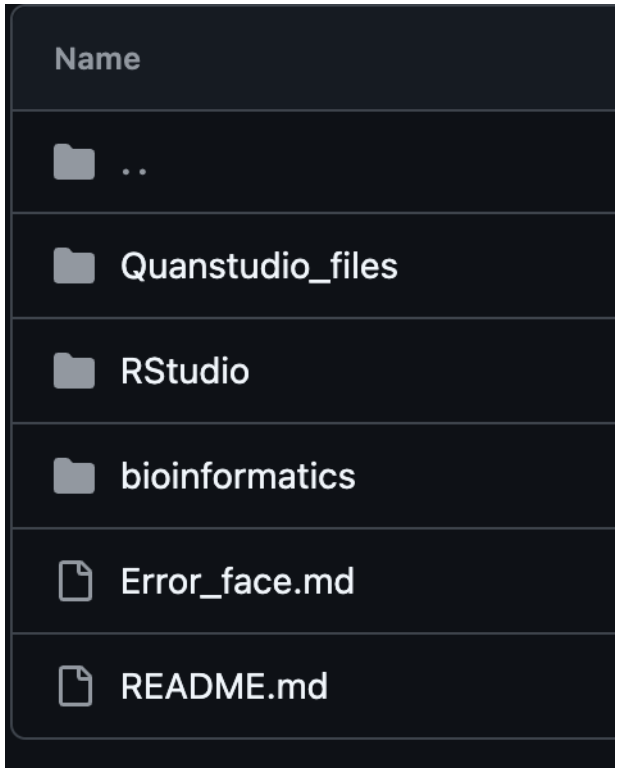
posit

OS	Download	Size
Windows 10/11	RSTUDIO-2024.04.2-764.EXE	262.79 MB
macOS 12+	RSTUDIO-2024.04.2-764.DMG	664.40 MB
Ubuntu 20/Debian 11	RSTUDIO-2024.04.2-764-AMD64.DEB	194.73 MB

Now we open R studio

1

2

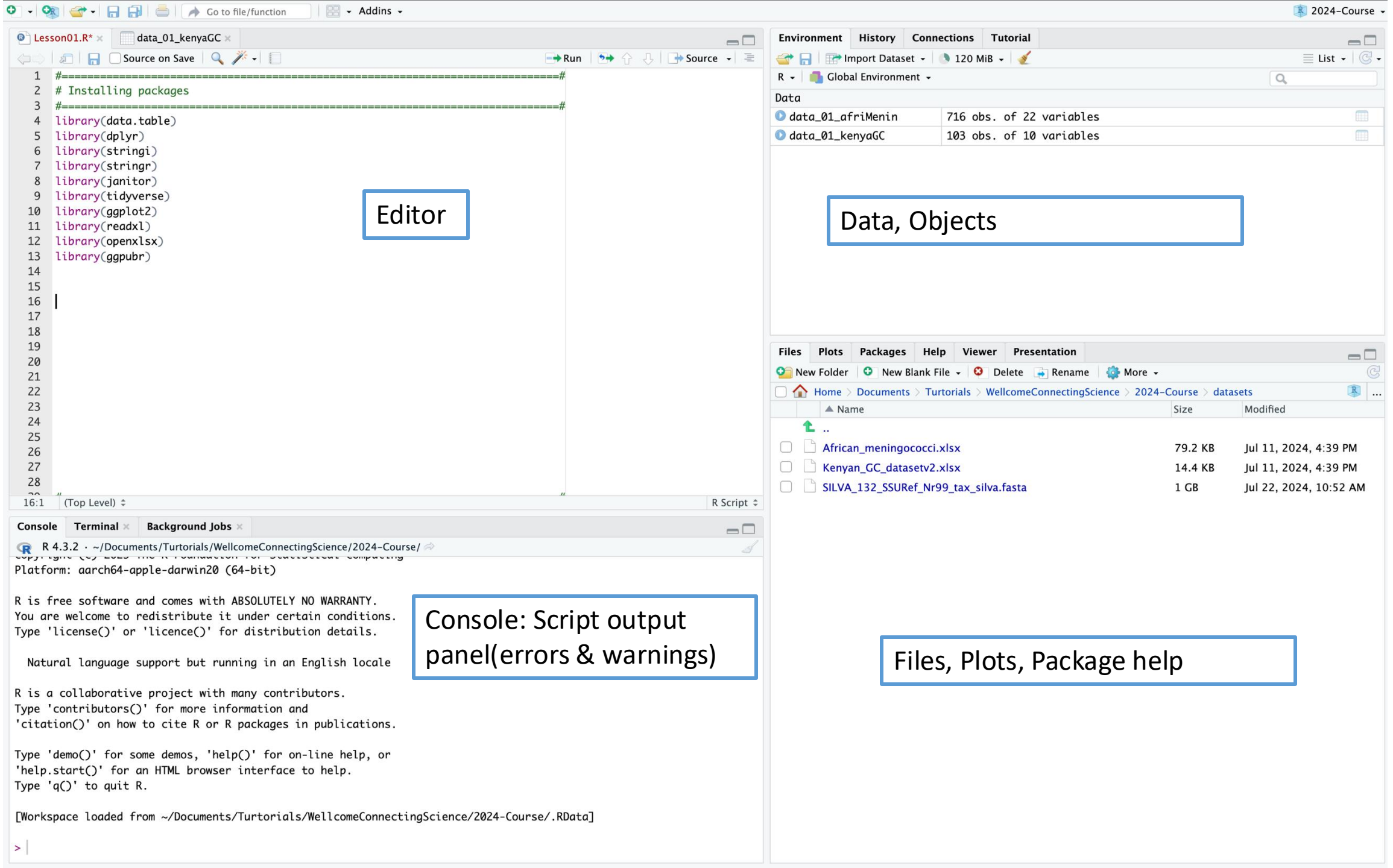


3

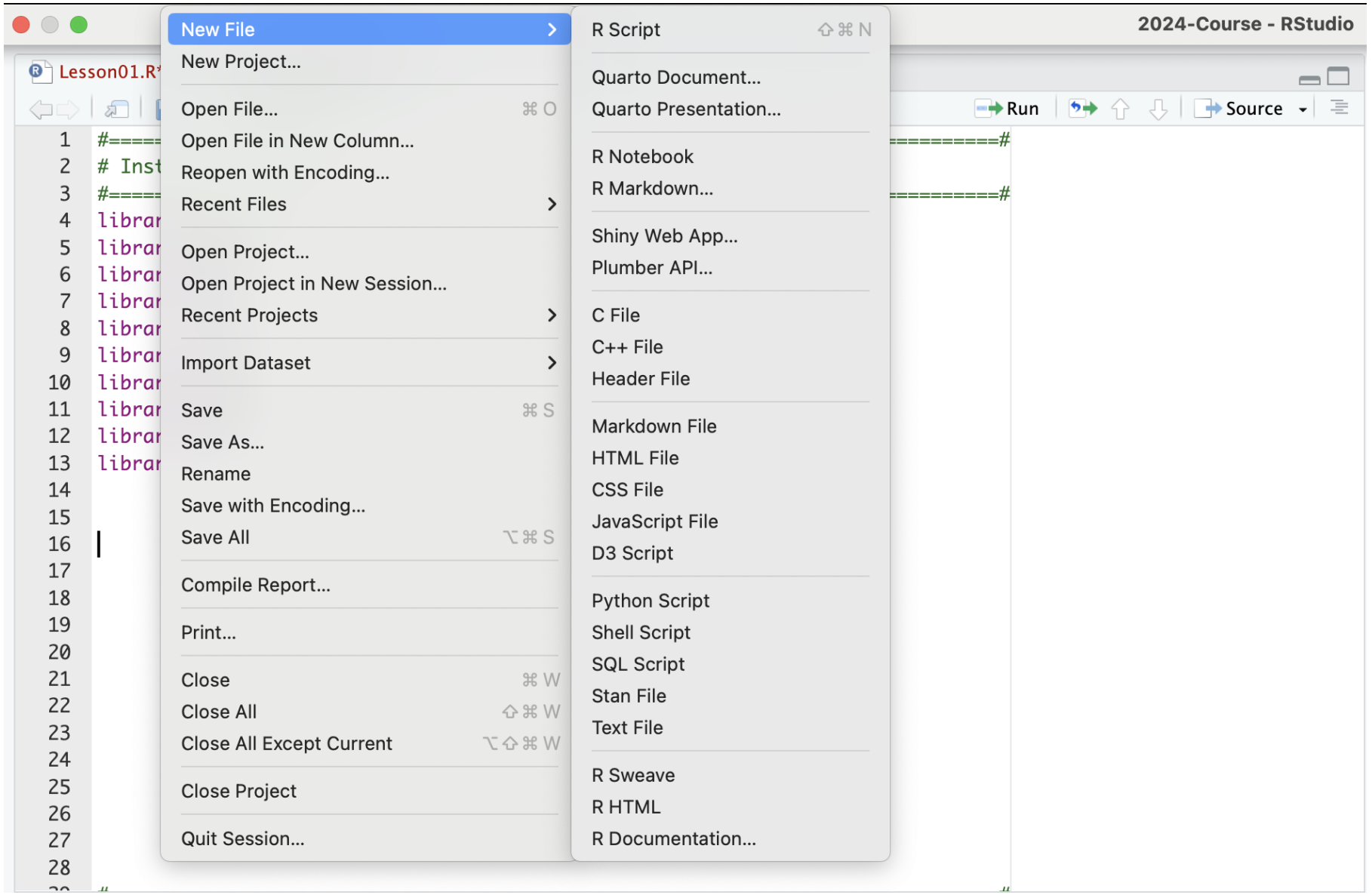
Errors Faced and changes made:

1. Changing file permission `sudo chmod 777 filename` another alternative is `sudo chmod a+rwx filename`
2. Rstudio not opening and crashing `sudo sysctl -w kernel.apparmor_restrict_unprivileged_userns=0`
3. Rstudio package installation line `install.packages(c("data.table", "janitor", "dplyr", "stringr", "stringi","tidyverse","ggplot2","tidyr", "ggpubr","plotly","arsenal","cowplot", "openxlsx"))`

Paste it on your terminal and press Enter. Now you can open R studio


Lesson01.R*
data_01_kenyaGC
Source on Save
Run
Source
#==========#
Installing packages
#==========#
library(data.table)
library(dplyr)
library(stringi)
library(stringr)
library(janitor)
library(tidyverse)
library(ggplot2)
library(readxl)
library(openxlsx)
library(ggpubr)
16:1
(Top Level)
R Script
Editor
Environment
History
Connections
Tutorial
Import Dataset
120 MiB
List
R
Global Environment
Data
data_01_afriMenin
716 obs. of 22 variables
data_01_kenyaGC
103 obs. of 10 variables
Data, Objects
Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Delete
Rename
More
Home
Documents
Turtorials
WellcomeConnectingScience
2024-Course
datasets
Name
Size
Modified
..
African_meningococci.xlsx
79.2 KB
Jul 11, 2024, 4:39 PM
Kenyan_GC_datasetv2.xlsx
14.4 KB
Jul 11, 2024, 4:39 PM
SILVA_132_SSURef_Nr99_tax_silva.fasta
1 GB
Jul 22, 2024, 10:52 AM
Files, Plots, Package help
Console
Terminal
Background Jobs
R 4.3.2 · ~/Documents/Turtorials/WellcomeConnectingScience/2024-Course/
Platform: aarch64-apple-darwin20 (64-bit)
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.
[Workspace loaded from ~/Documents/Turtorials/WellcomeConnectingScience/2024-Course/.RData]
>
Console: Script output panel(errors & warnings)

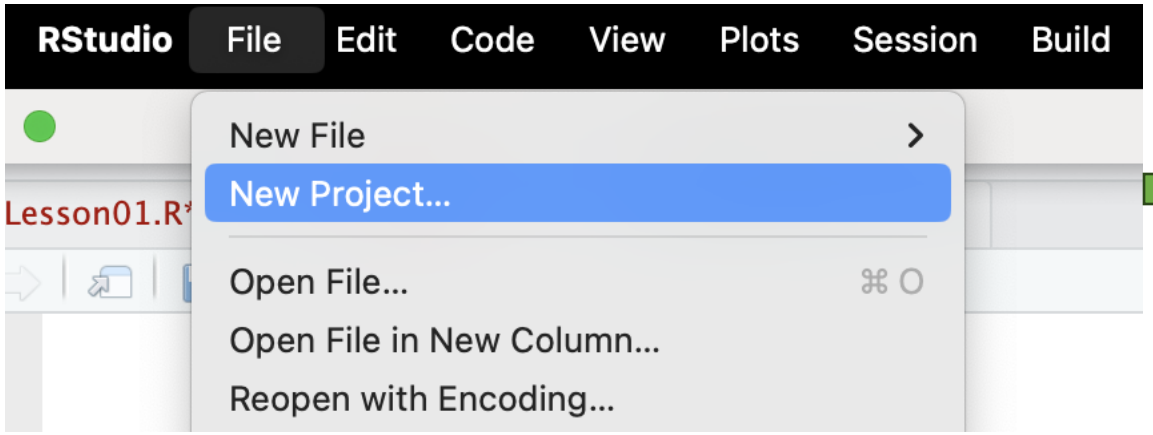
R studio



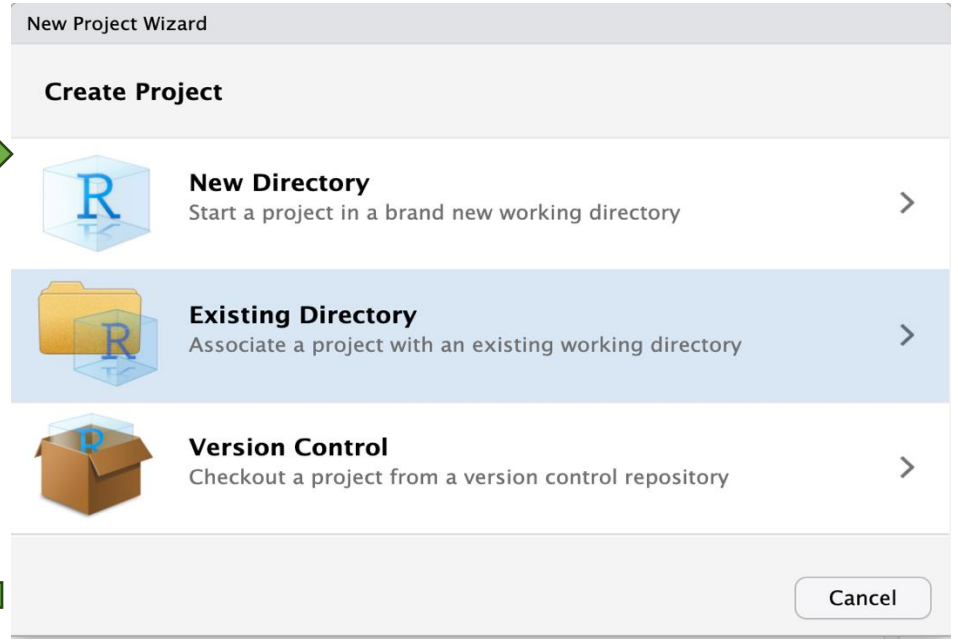
R studio: User interface, Customizable interface, Code autocompletion, write code in other languages

Creating an R project

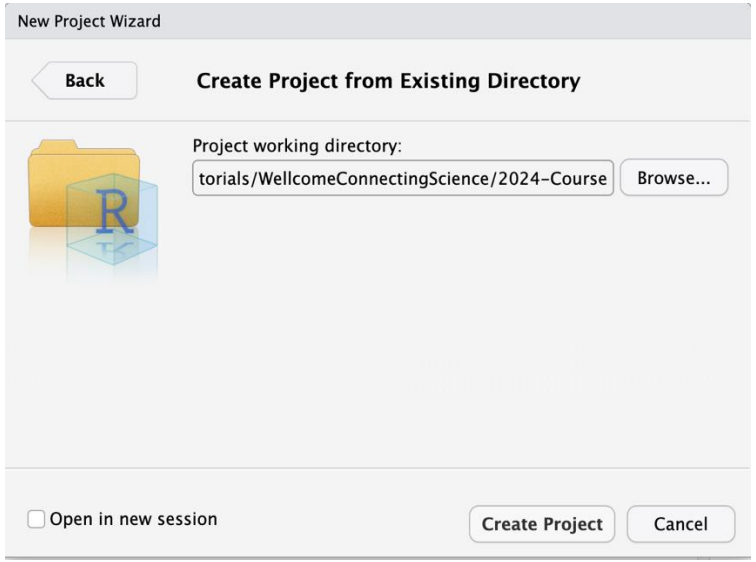
Step 1



Step 2

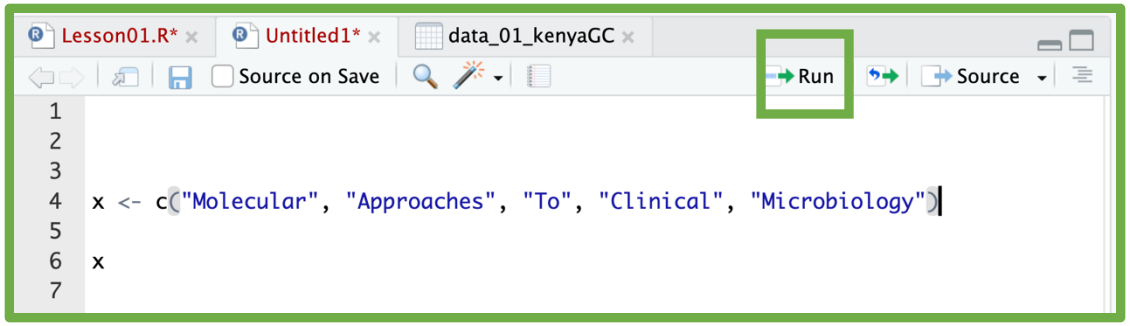


Step 3

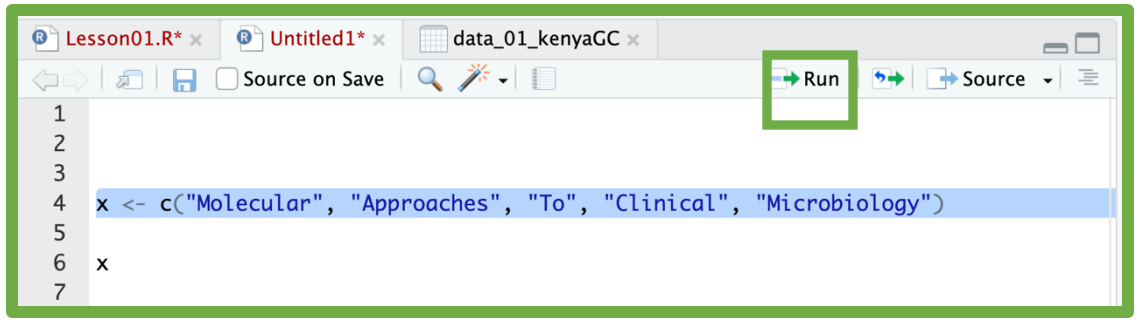


Running your Code

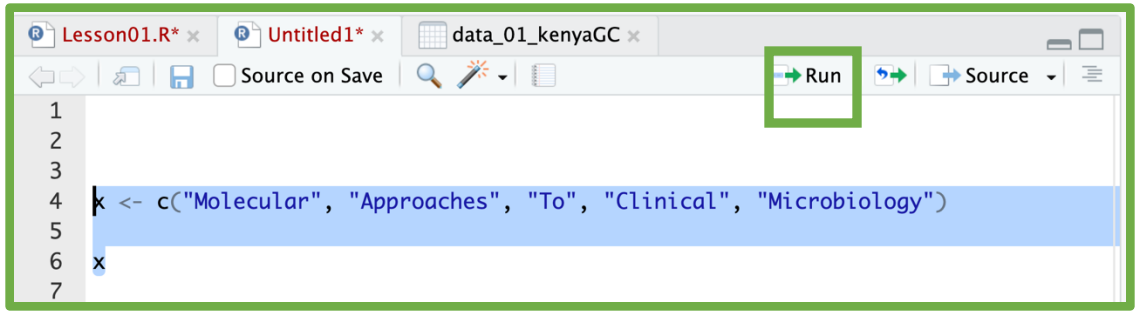
1



2



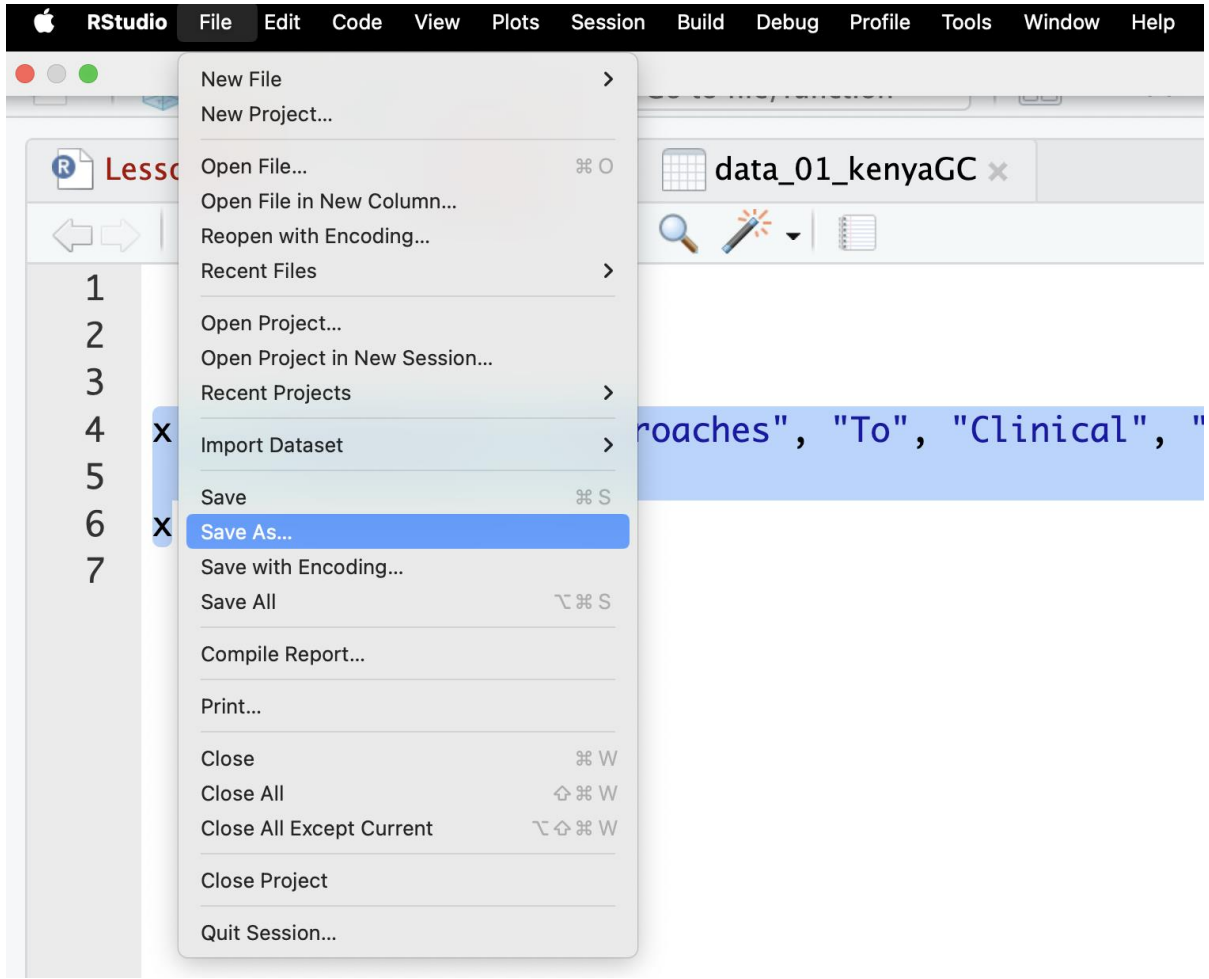
3

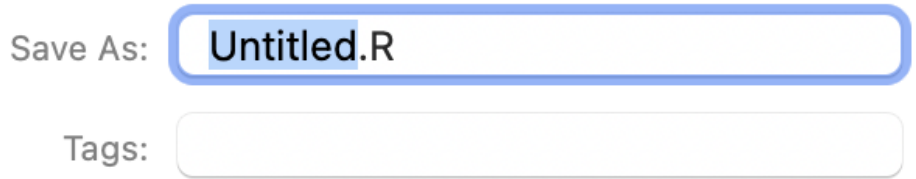


Or use:

Ctrl + Enter
instead of pressing the ”Run” button

Saving your code





Basic but Important Commands

- # - not a command, but used for commenting you code

```
# Molecular Approaches 2024
# Date: 3rd September 2024
# Name: Leo
```

Basic but Important Commands

1. **#** - not a command, but used for commenting you code
2. Installing and loading packages: **install.packages("name_of_package")** then **library(name_of_package)**

```
# Installing the required packages
install.packages("") # all required packages have been pre-installed


# Loading required libraries
library(data.table)
library(dplyr)
```

Basic but Important Commands

1. **#** - not a command, but used for commenting you code
2. Installing and loading packages:
   install.packages("name_of_package") then
   library(name_of_package)

Use the function **search()** to know which packages have been loaded

3. setwd() & getwd()

Assigning variables (working on the console)

```
> x <- 10
> x
[1] 10
```

Method 1: Most preferred

```
> x=10
> x
[1] 10
```

Method 2

```
> assign("x", 10)
> x
[1] 10
```

Method 3

Assigning multiple values to a variable (working on the console)

```
> x <- 10, 20, 30, 40
```

```
> x <- 10, 20, 30, 40
Error: unexpected ',' in "x <- 10,"
```

```
> x <- c(10,20,30,40)
> x
[1] 10 20 30 40
```

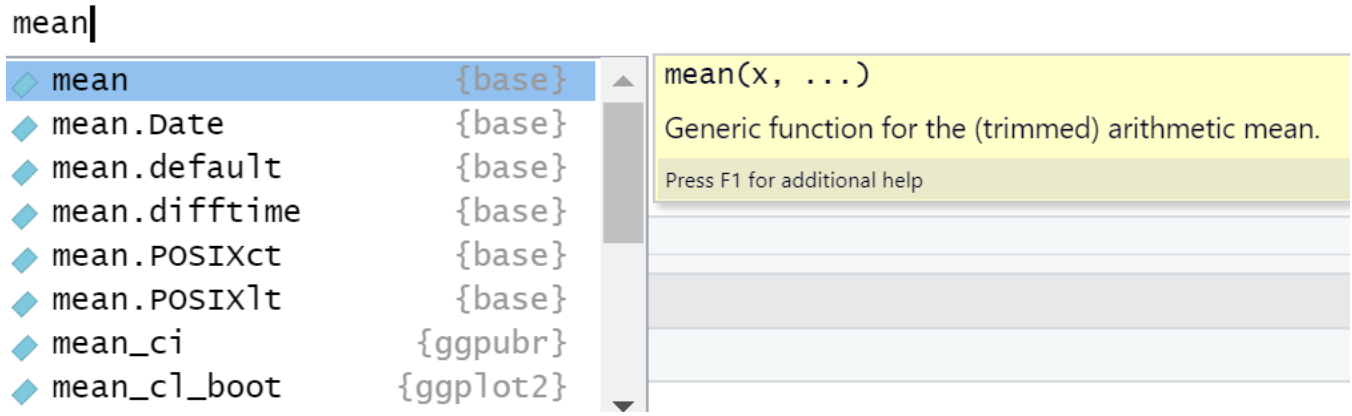
Creating a data frame (working on the text editor)

month	male	female
Jan	20	40
Feb	30	30
March	40	20

```
male <- c(20, 30, 40)

female <- c(40, 30, 20)

month <- c("Jan","Feb", "March")

df <- data.frame(male, female, month)
```

function

Simple Statistics (working on the text editor)




```
mean(df$male)
```

Task: Try and get the **median** and **sum**. Use the function ‘summary’ to summarise the entire dataset

Bioinformatics Eyesight Test

✗
```
install.package(data.table)
```

✗
```
install.packages(data.table)
```

✓
```
install.packages("data.table")
```

✗
```
libary(data.table)
```

✓
```
library("data.table")
```

✓
```
library(data.table)
```

✓
```
require(data.table)
```

```
x <- 20
y <- 30
z <- 40
```

✗
```
female -> c("x", "y", "z")
```

✗
```
female -> c(x, y, z)
```

✓
```
female <- c("x", "y", "z")
```

✓
```
female <- c(x, y, z)
```

✗
```
month <- c("Jan" "Feb" "March")
```

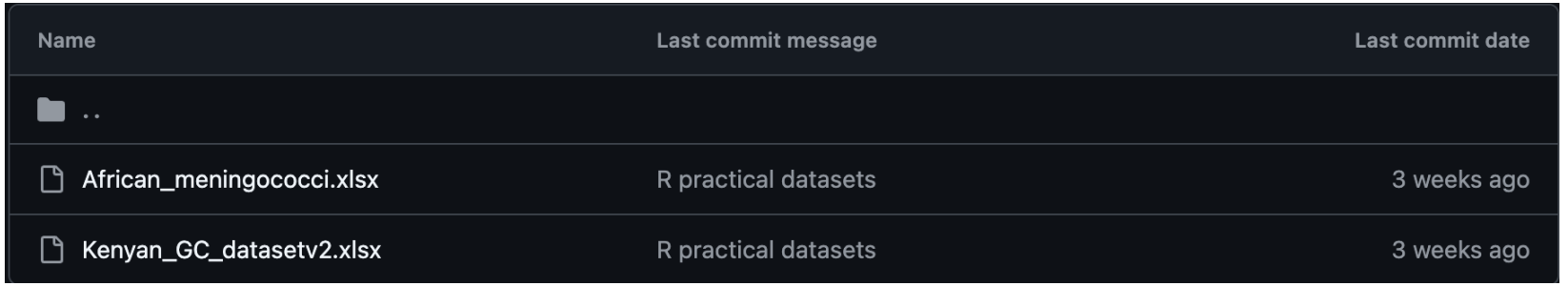
✗
```
month <- c(Jan Feb March)
```

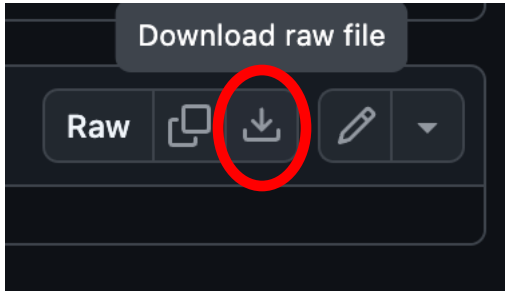
✓
```
month <- c("Jan", "Feb", "March")
```

✓
```
female <- c(40, 30, 20)
```

Download the Datasets (while inside you VM)

- Github page- https://tinyurl.com/2dwx5jmy
- course_data/Rstudio/*.xlsx





Load datasets

```
read.xlsx(path = "PATH_TO_FILE")
```

```
read.xlsx(xlsxFile = "datasets/African_meningococci.xlsx")
```

```
data_01_Africa <- read.xlsx(xlsxFile = "datasets/African_meningococci.xlsx")
```