Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

Introduction to R and RStudio

2022 Virtual Training on Data Disaggregation and Small Area Estimation for the SDGs –
22-25 November 2022, *INStaD Benin, INSTAT Mali, Stats South Africa, Statistics Botswana*

General overview

- **R** is a programming language for statistical computing, data analytics and scientific research. It is one of the most widely used languages by statisticians, data analysts and researchers to manage, manipulate, analyze and visualize data.

- **Rstudio** is an integrated development environment for R that allows users to interact more easily with R by integrating different aspects of scripting, from code completion to debugging.

Note: On July 27, 2022, it was announced that Rstudio was changing its name to Posit to move from an R-exclusive tool to a language agnostic system.

In order to use RStudio, R needs to be installed first.

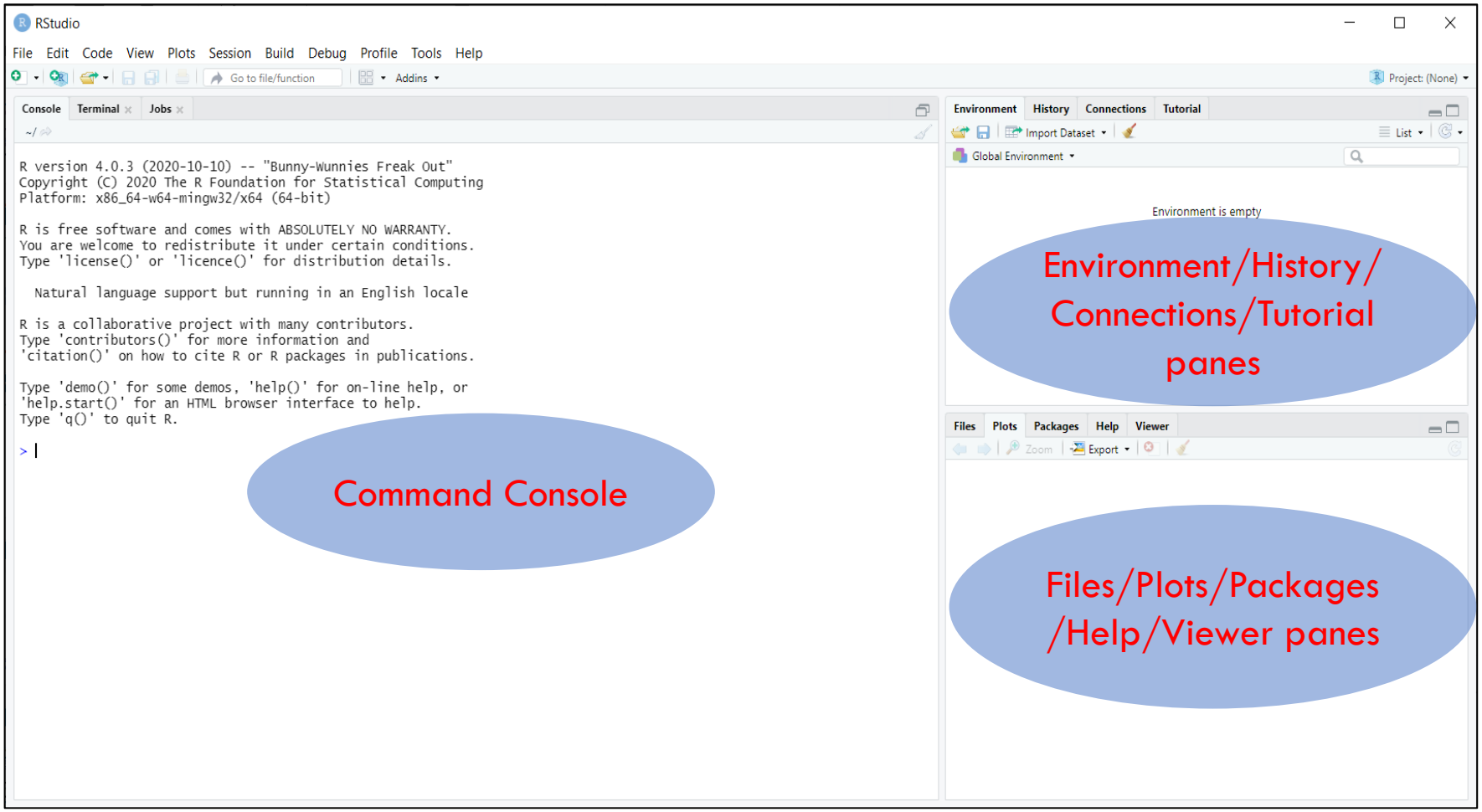
R and RStudio Installation

- **Step 1:** Go to the Comprehensive R Archive Network (CRAN) website and click on "Download R for Windows" (or *"Download R for MacOS"*).
- **Step 2:** Click on the "*base*" subdirectory link (or the package link, *.pkg file*).
- **Step 3:** Click on "*Download R-4.2.1 for Windows*" (the R version available might be different depending on updates produced after the compilation of this ppt). The link allows downloading an installer extension (*.exe file*).
- **Step 4:** Run the .exe file and step through the installation wizard accepting the default settings.

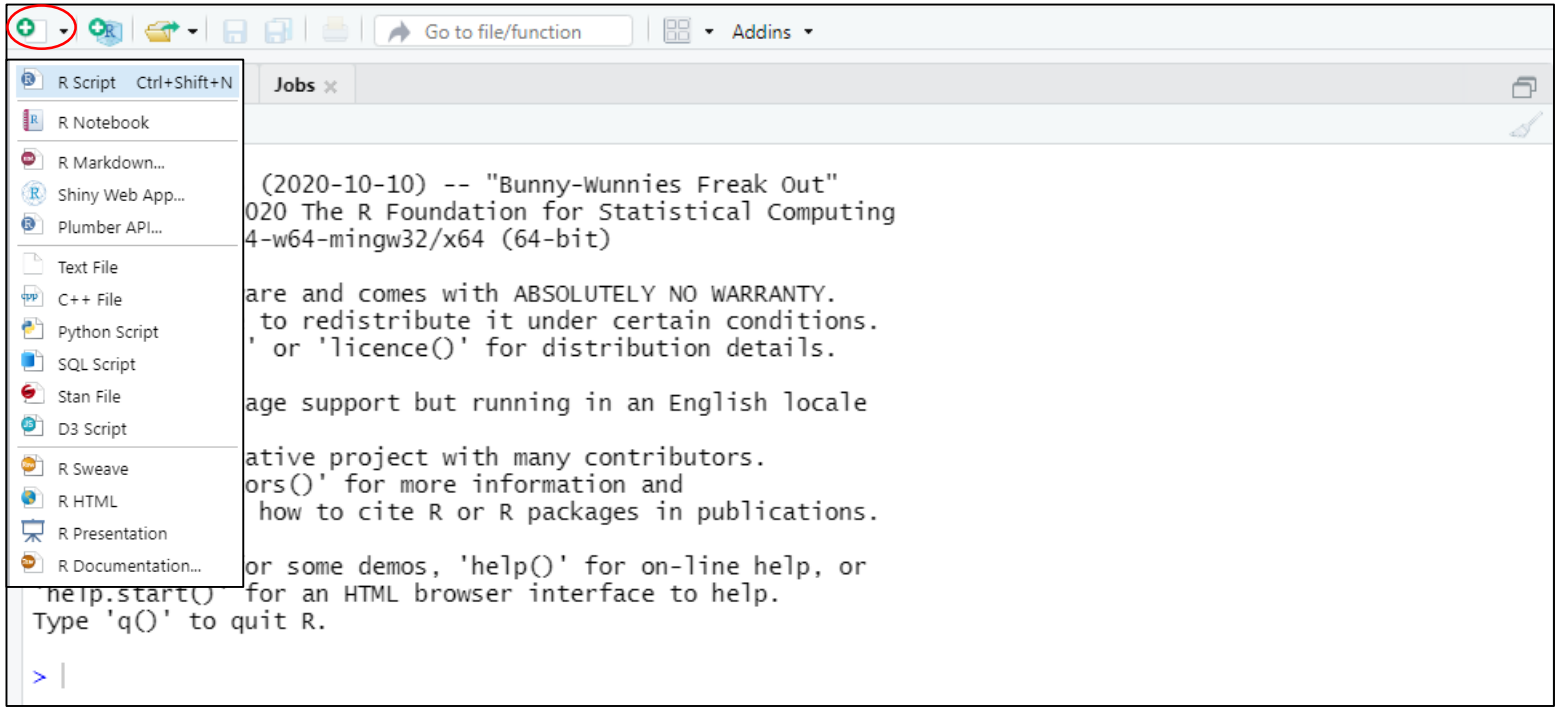
Once R is installed, you can then proceed to the installation of RStudio.

- **Step 5:** Go to RStudio download website and click on the "Download RStudio for Windows" button (or the link for the MacOS version).
- **Step 6:** Run the .exe file and follow the installation instructions.

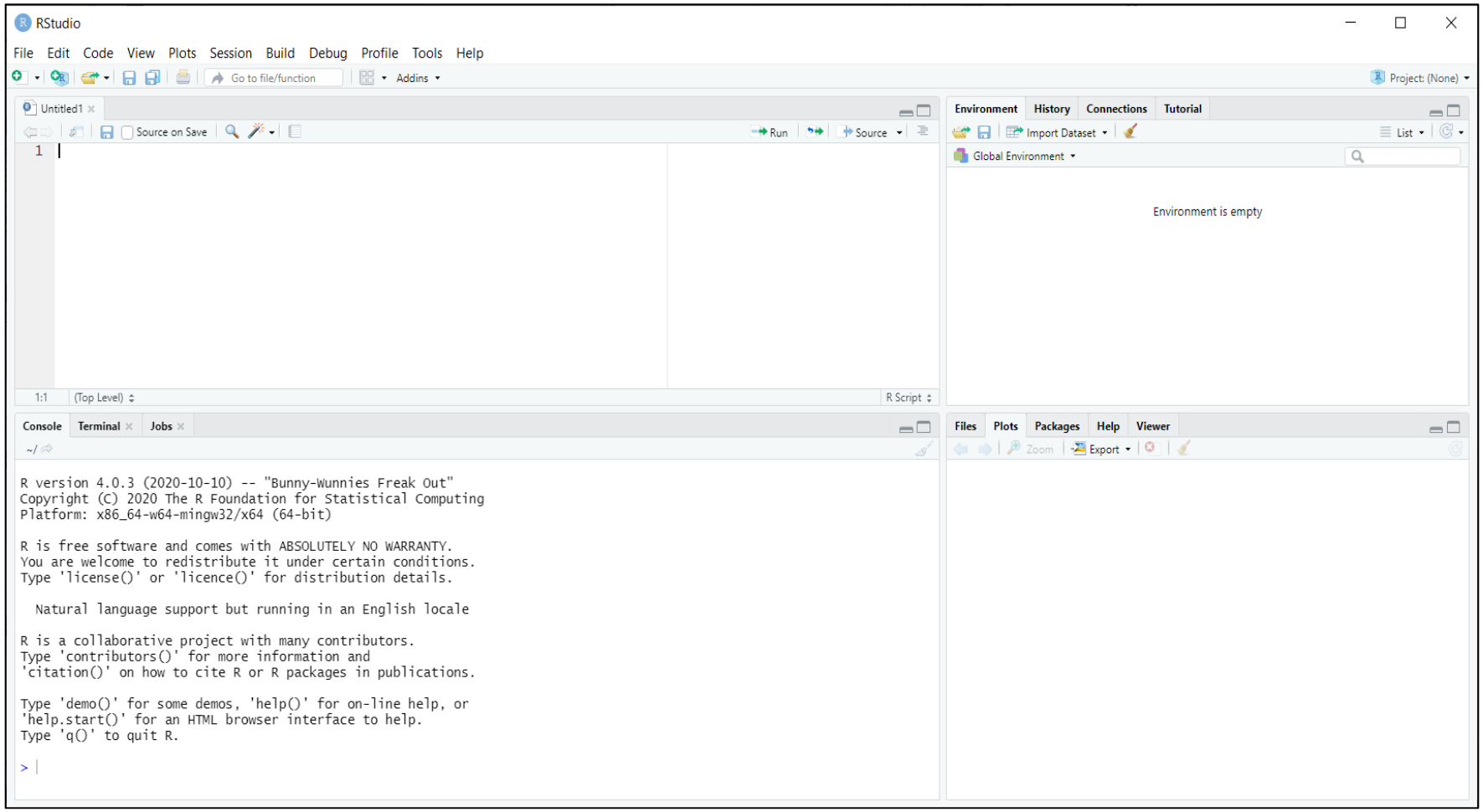
RStudio interface



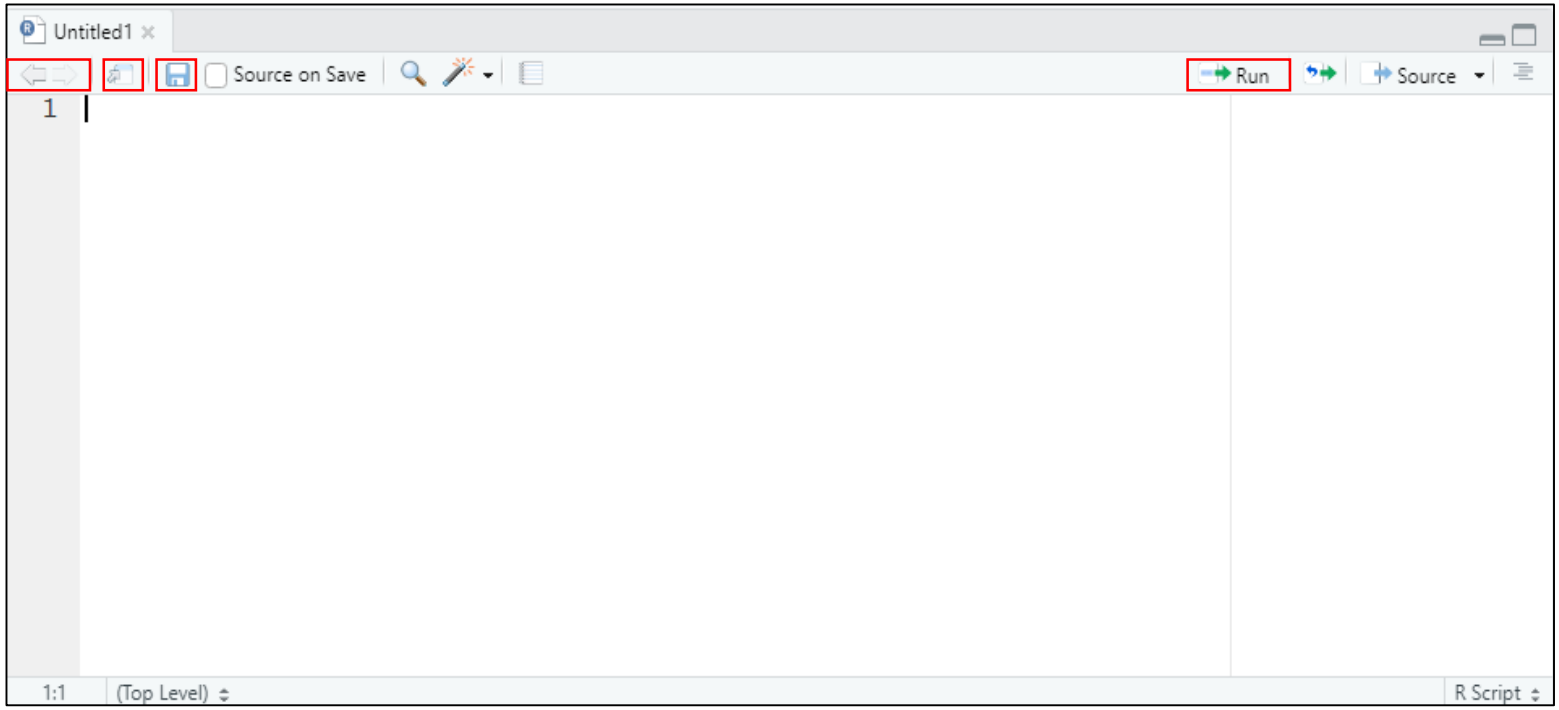
RStudio: command console



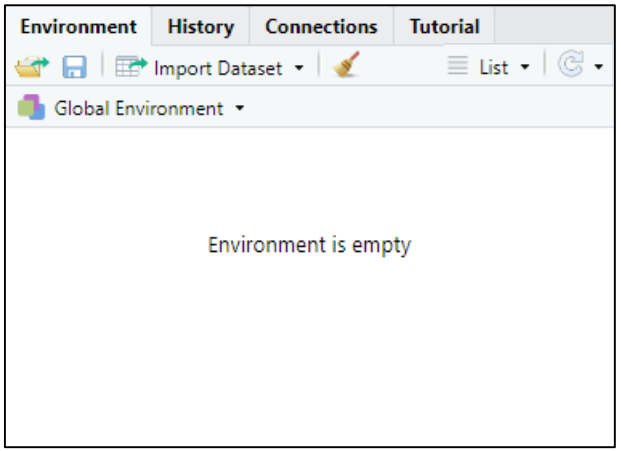
RStudio interface with editor window open

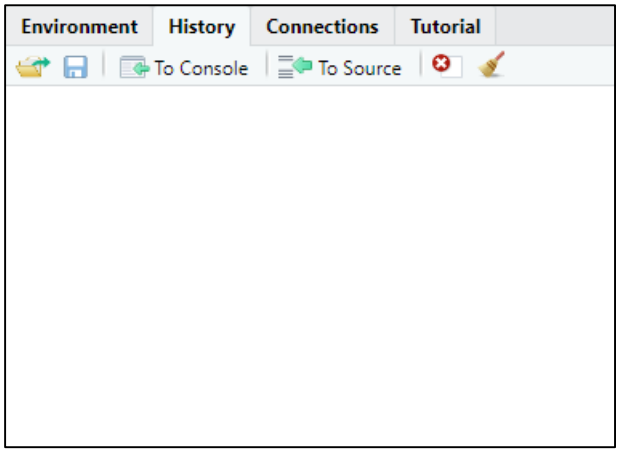


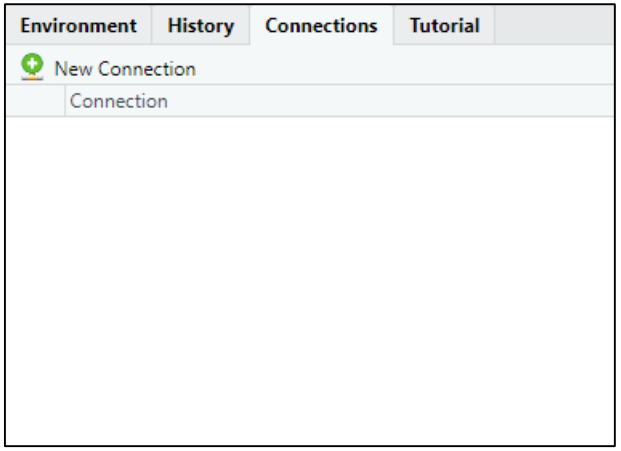
RStudio: editor window

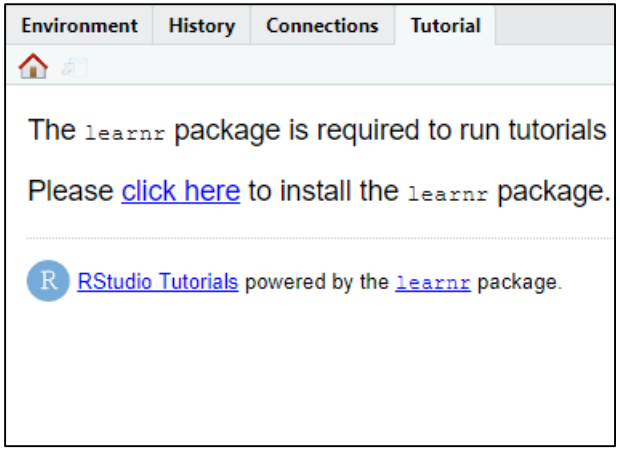


RStudio: Environment/History/Connections/Tutorial panels

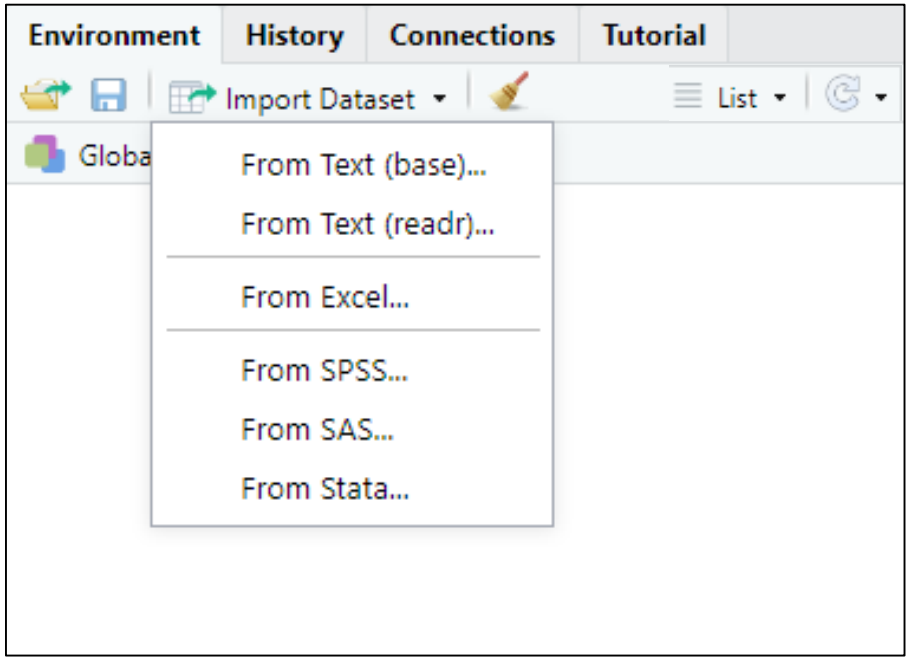




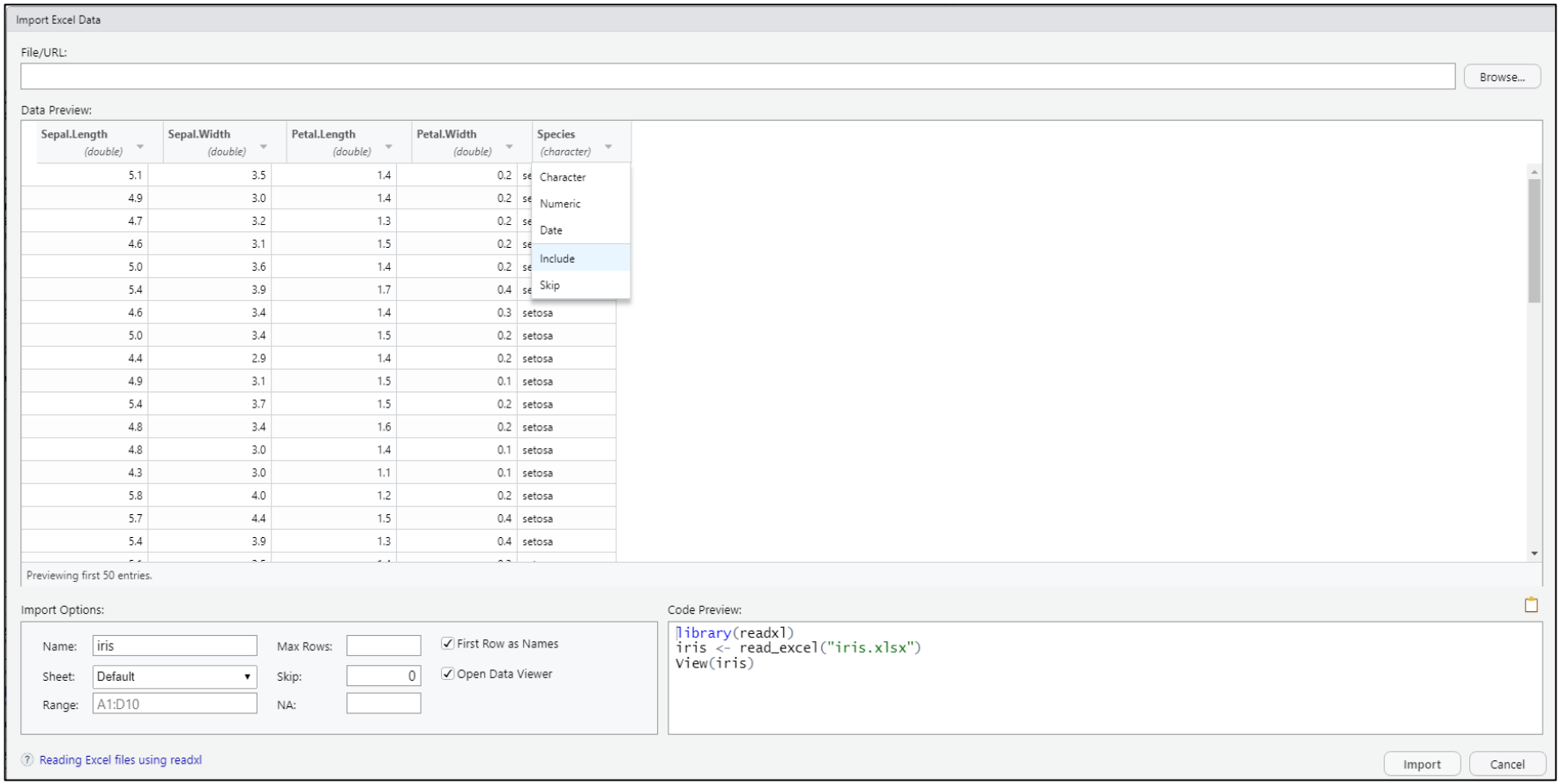




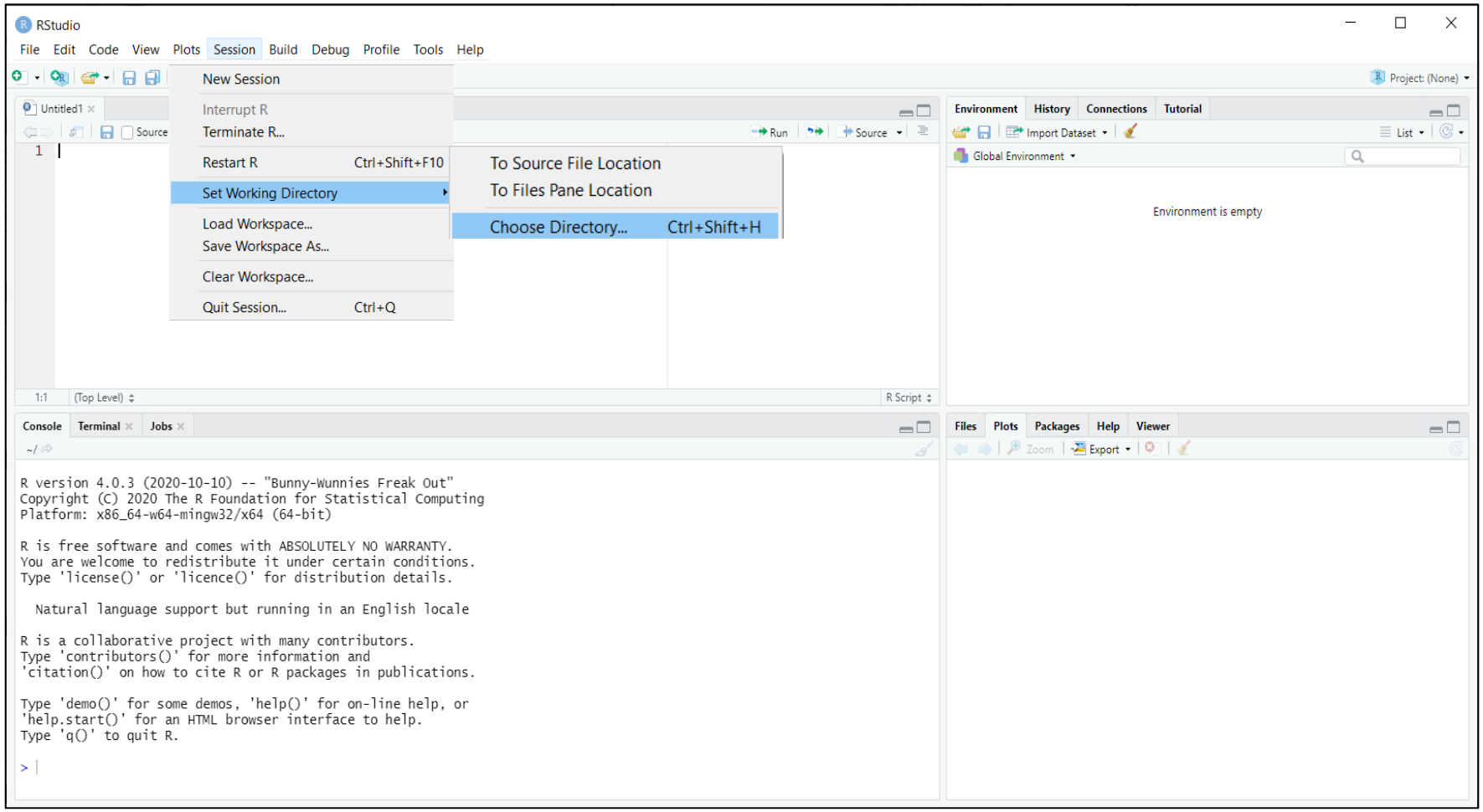
Import dataset tab



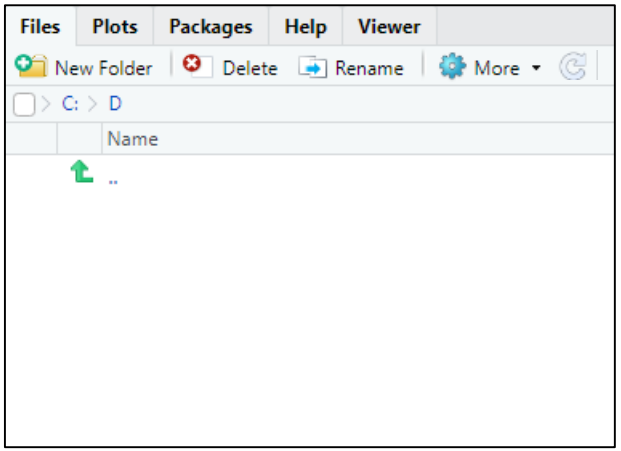
Import dataset tab (2)

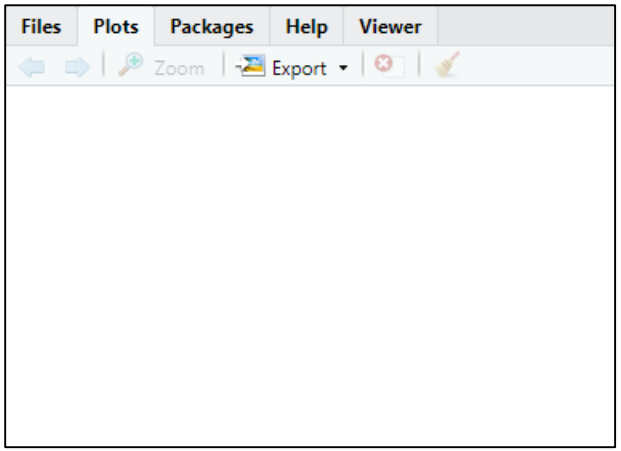


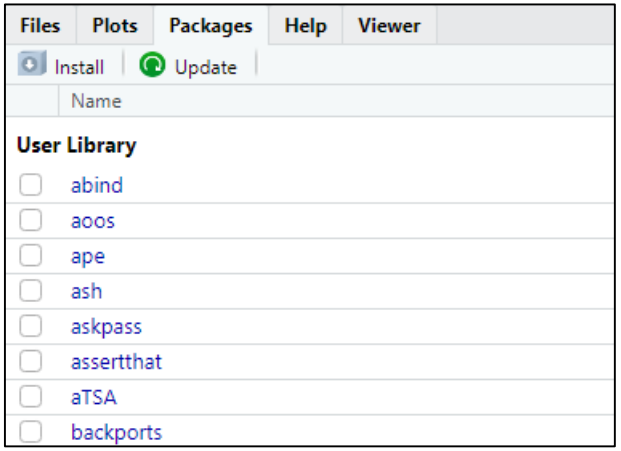
Setting the working directory

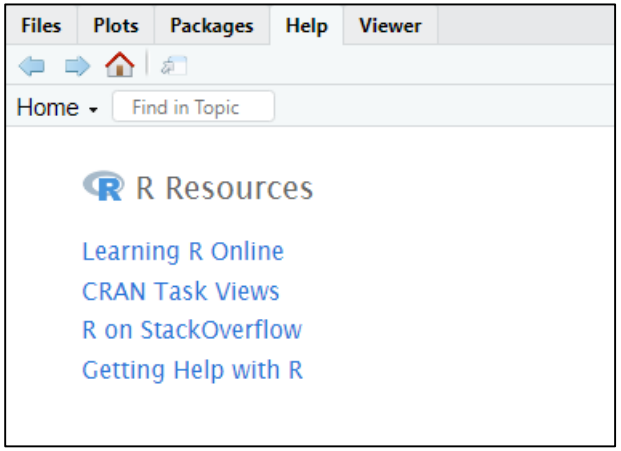


RStudio: Files/Plots/Packages/Help/Viewer panels









Packages

- **Packages:** extensions that contain code, data, and documentation in a standardized format that can be installed and used by users of R to solve specific analytical problems.
- The base version of R already includes many useful packages that allow performing elementary tasks such as simple calculations, data exploration, and loading of text data files.
- Complex tasks may require intensive coding using base packages functions, but are made easier by user-supplied packages.

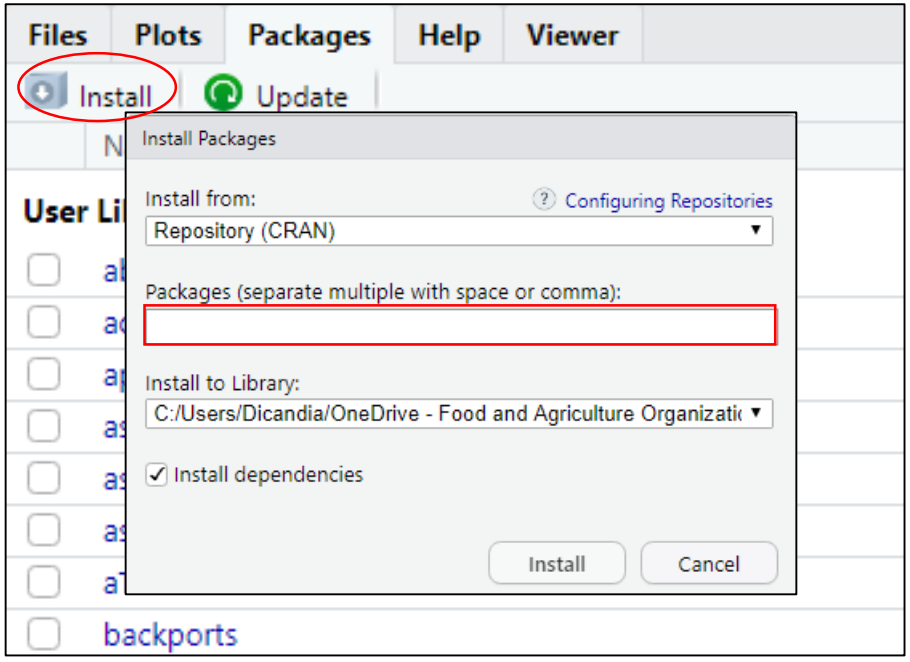
Packages (2)

To **install a package** from CRAN (the official repository for user contributed R packages) and then load it, use the following commands:

install.packages("name of the package")

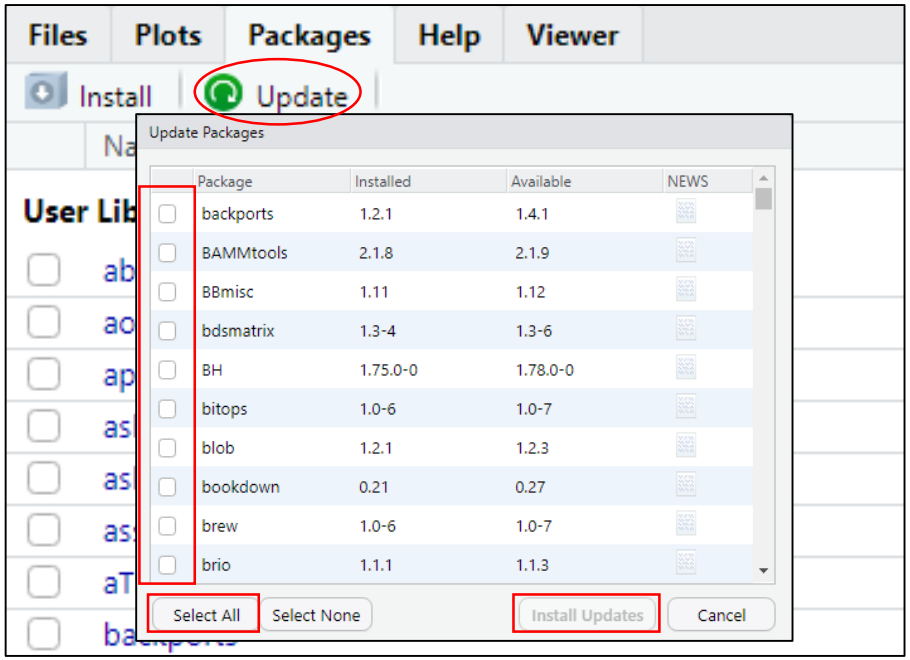
library("name of the package")

Another way to install packages is to use the Install Packages tab:



Packages (3)

- R packages are sometimes **updated** to improve or modify functionality. It is advisable to occasionally update the packages installed on your computer.
- You can update your installed R packages in RStudio by clicking the Update button in the toolbar in the Packages panel.



List of packages used in the training

For importing	Description
openxlsx	*Read, Write and Edit xlsx Files*
readxl	*Only read excel files*
data.table	*Fast file reader, writer*
haven	*Import and Export 'SPSS', 'Stata' and 'SAS' Files*

For data manipulation	Description
dplyr	*A fast, consistent tool for working with data frame*
tidyr	*Tools to help to create tidy data*

For various uses	Description
Boruta	*Wrapper algorithm for all relevant Feature Selection*
relaimpo	*Relative Importance of regressors in linear models*
car	*Companion to Applied Regression*
lme4	*Fit linear and generalized linear mixed-effects models*
RLRsim	*Exact (Restricted) Likelihood Ratio Tests for Mixed and Additive Models*
RColorBrewer	*Provides color schemes for graphics*

List of packages used in the training

For analysis of complex survey samples
survey
ReGenesees

For Small Area Estimation	sae	emdi	mind	SAEval

For data visualization	ggplot2

For spatial data	Description
raster	*Reading, writing, manipulating, analyzing and modeling of spatial data.*
tabularaster	*Tidy Tools for 'Raster' Data*
sp	*Classes and methods for spatial data*
rgdal	*Bindings for the 'Geospatial' Data Abstraction Library*
sf	*Support for simple features, a standardized way to encode spatial vector data*
spdep	*Provide tools to study spatial dependence*
geodata	*Functions for downloading of geographic data for use in spatial analysis and mapping.*

PRACTICAL EXAMPLES IN RSTUDIO

Thank you!