UF UNIVERSITY of FLORIDA

Introduction to R

AGR 5266C
Field Plot Techniques
(Sections 0877 and 1933)

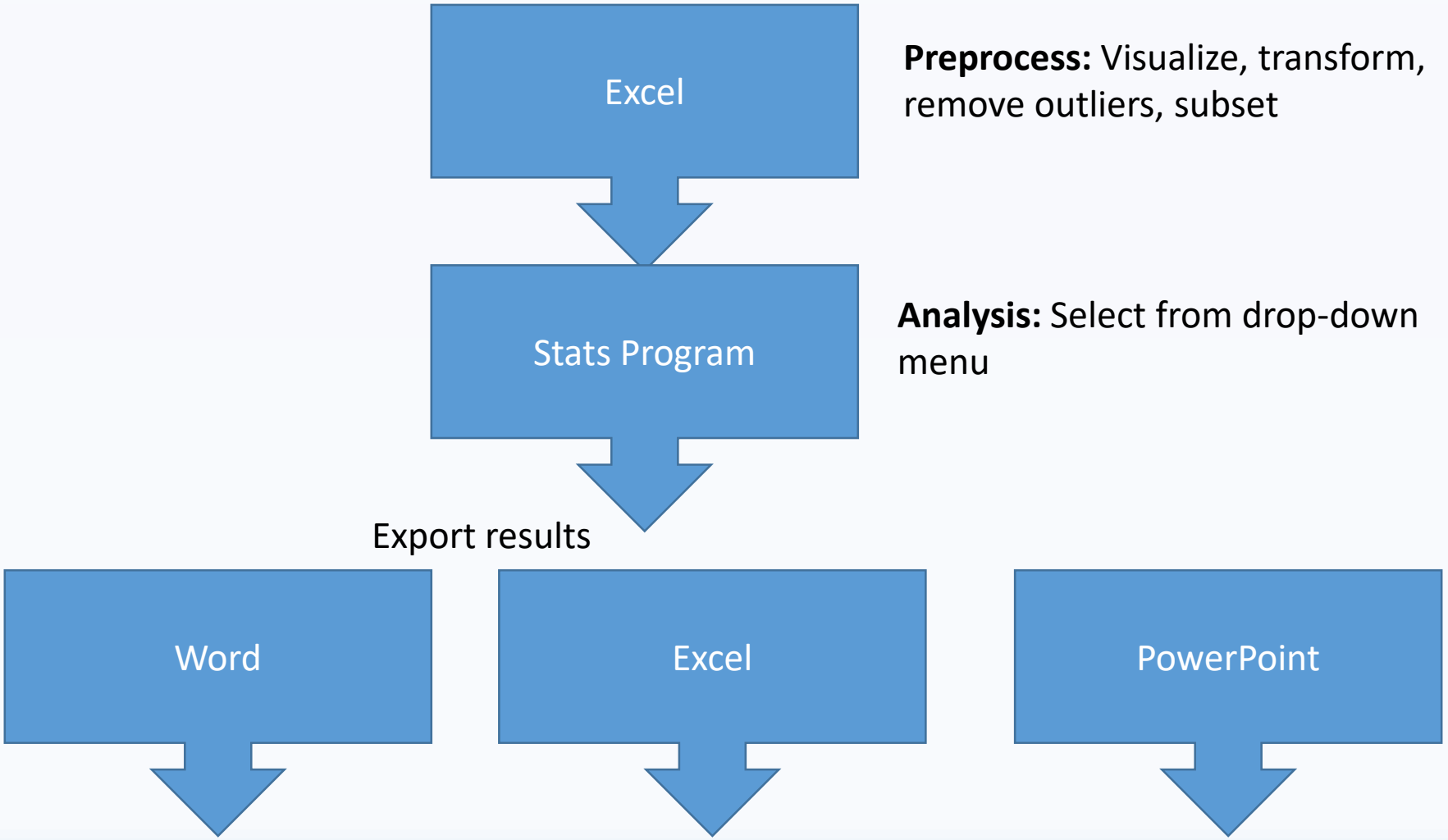
Dev Paudel

RStudio®

Computer programming for statistics

- Current practices
  - Statistics are often done in “canned” programs with dropdown menus eg SPSS, Excel.
  - Involve preprocessing in spreadsheets like Excel
  - Results are exported to make graphics and tables

Multiple steps in analysis



Preprocess: Visualize, transform, remove outliers, subset

Analysis: Select from drop-down menu

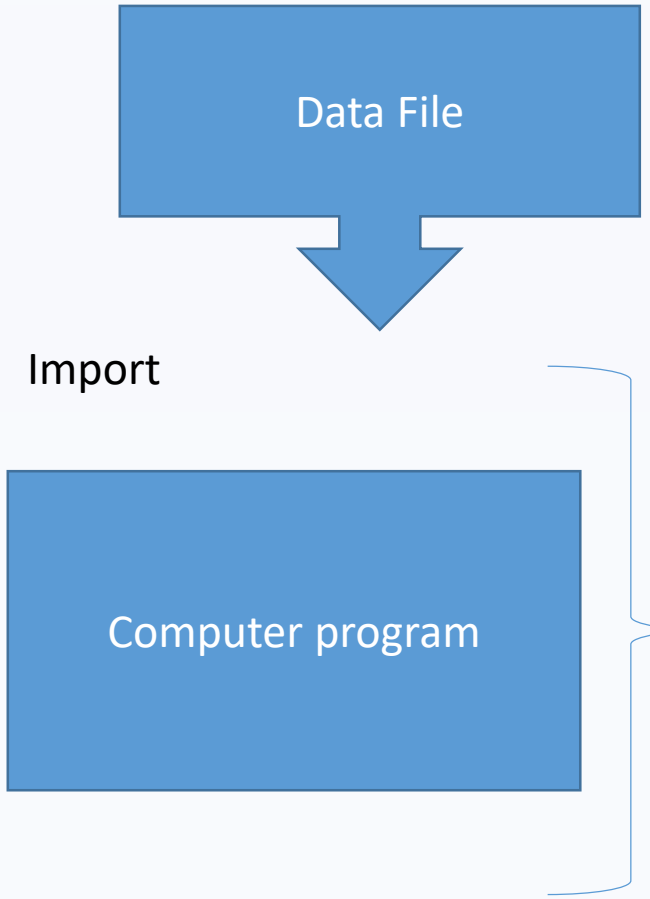
Figures: Make figures or alter figures

Ref: Denis Valle & Benjamin Baiser, Introduction to Applied Statistics for Agricultural and Life Sciences, 2014. ALS5932 Class Notes

How can programming help?

- Allows for flexible and customized analysis
  - In contrast to drop-down menus
- Allows preprocessing in the same software
  - Avoids use of multiple software
- Allows customizable graphs and tables directly

How can programming help?



Preprocess: Visualize, transform, remove outliers, subset

Analysis: Program customized scripts

Figures: Make customized figures

Streamline data analysis

Ref: Denis Valle & Benjamin Baiser, Introduction to Applied Statistics for Agricultural and Life Sciences, 2014. ALS5932 Class Notes

How can programming help?

- Less room for error
- Easily documented analysis
- Allows reproducible science

Experimental design

→ Data collection →

Analysis

Reproducible research

- The goal of **reproducible research** is to tie specific instructions to data analysis and experimental data so that scholarship can be **recreated**, better **understood** and **verified**.

CRAN Task View: Reproducible Research

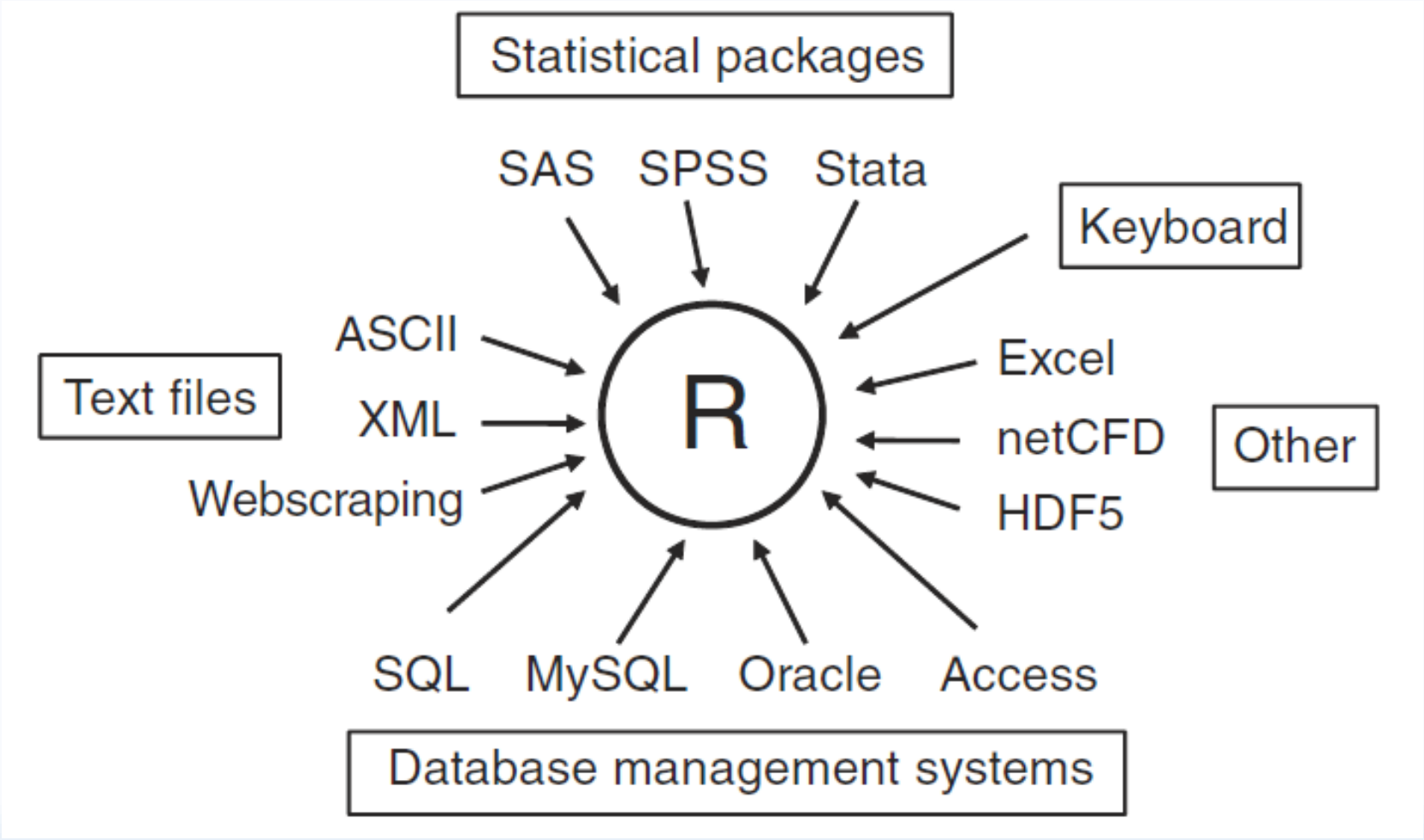
What is R and why use R

- Powerful tool for statistics, graphics, and statistical programming

- R is free
- Contains advanced statistical platform
- Can easily import data from different software

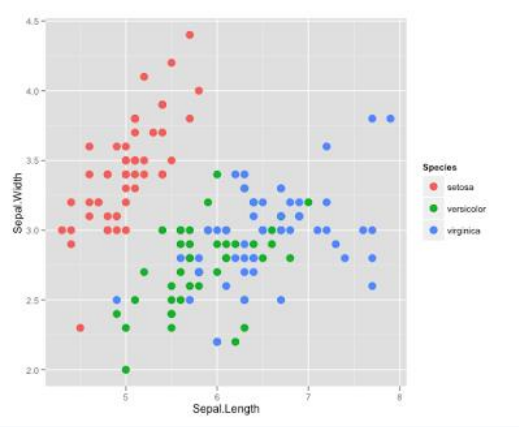
* R in Action

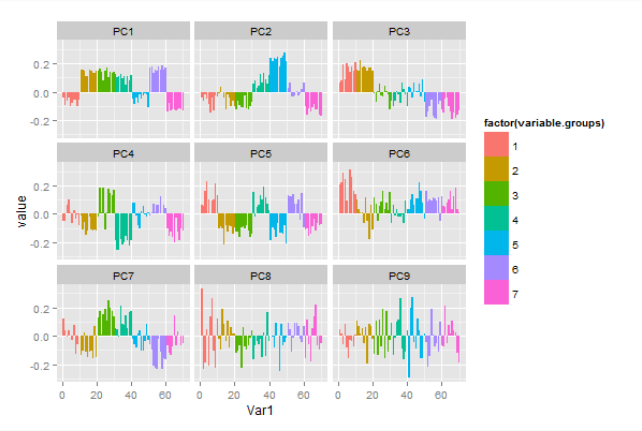
What is R and why use R

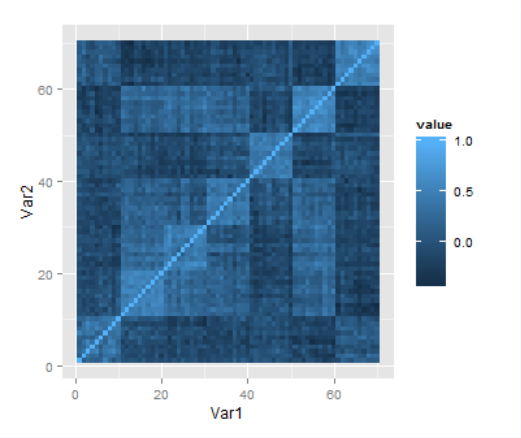


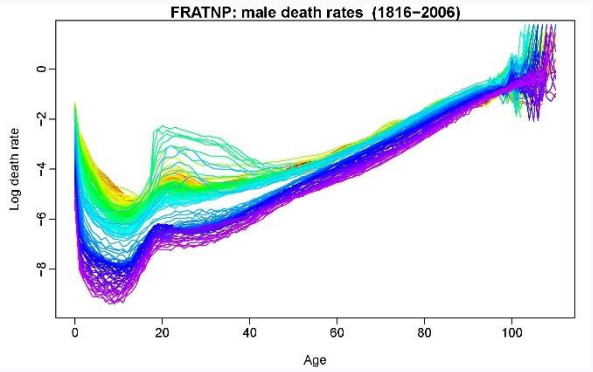
* R in Action

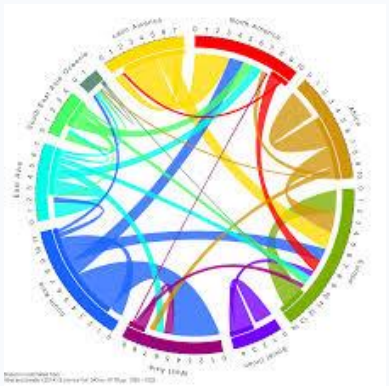
What is R and why use R

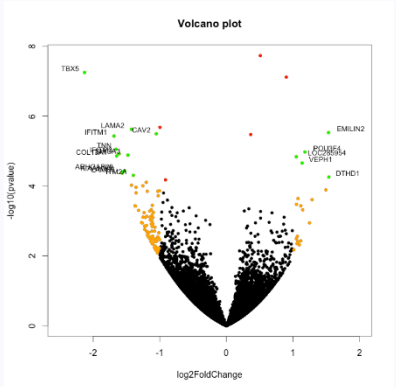


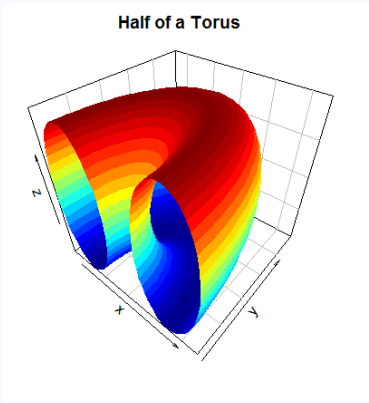


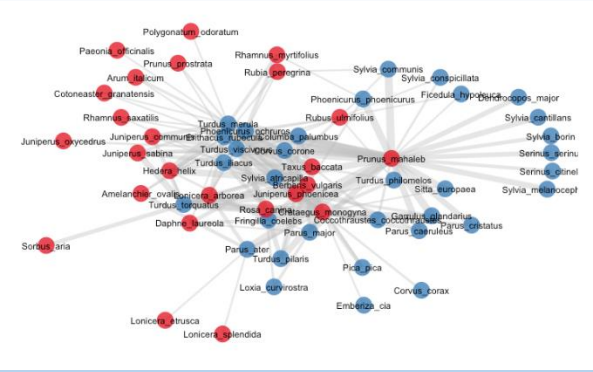


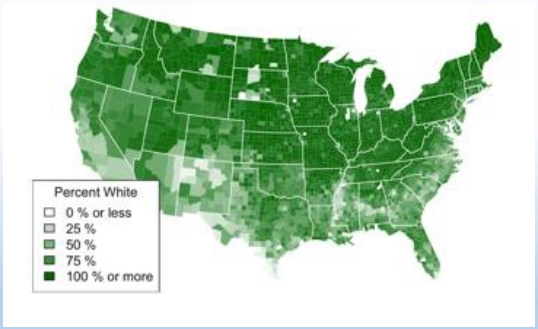


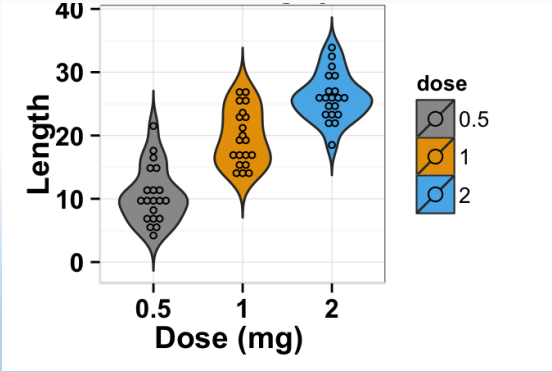




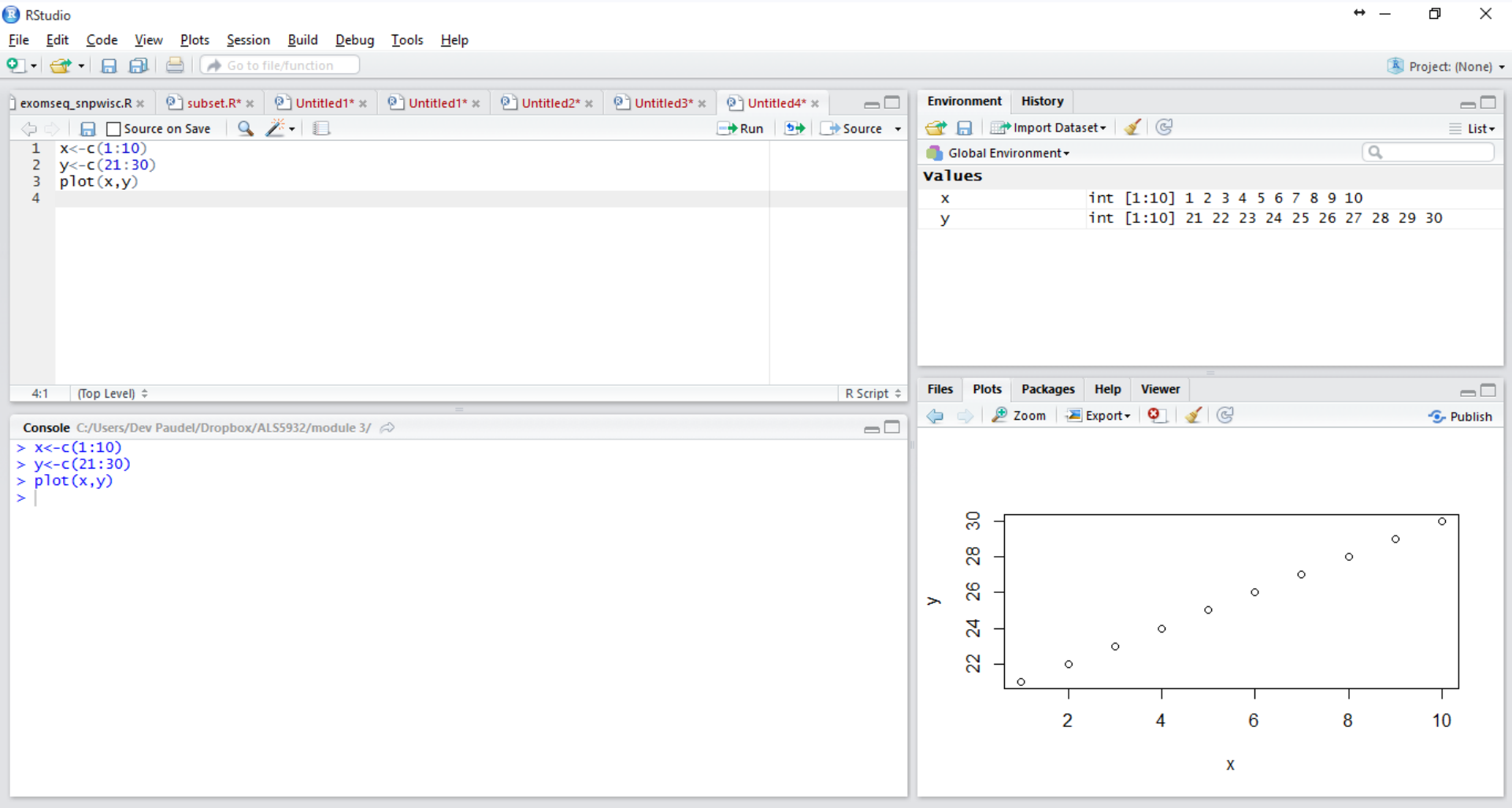




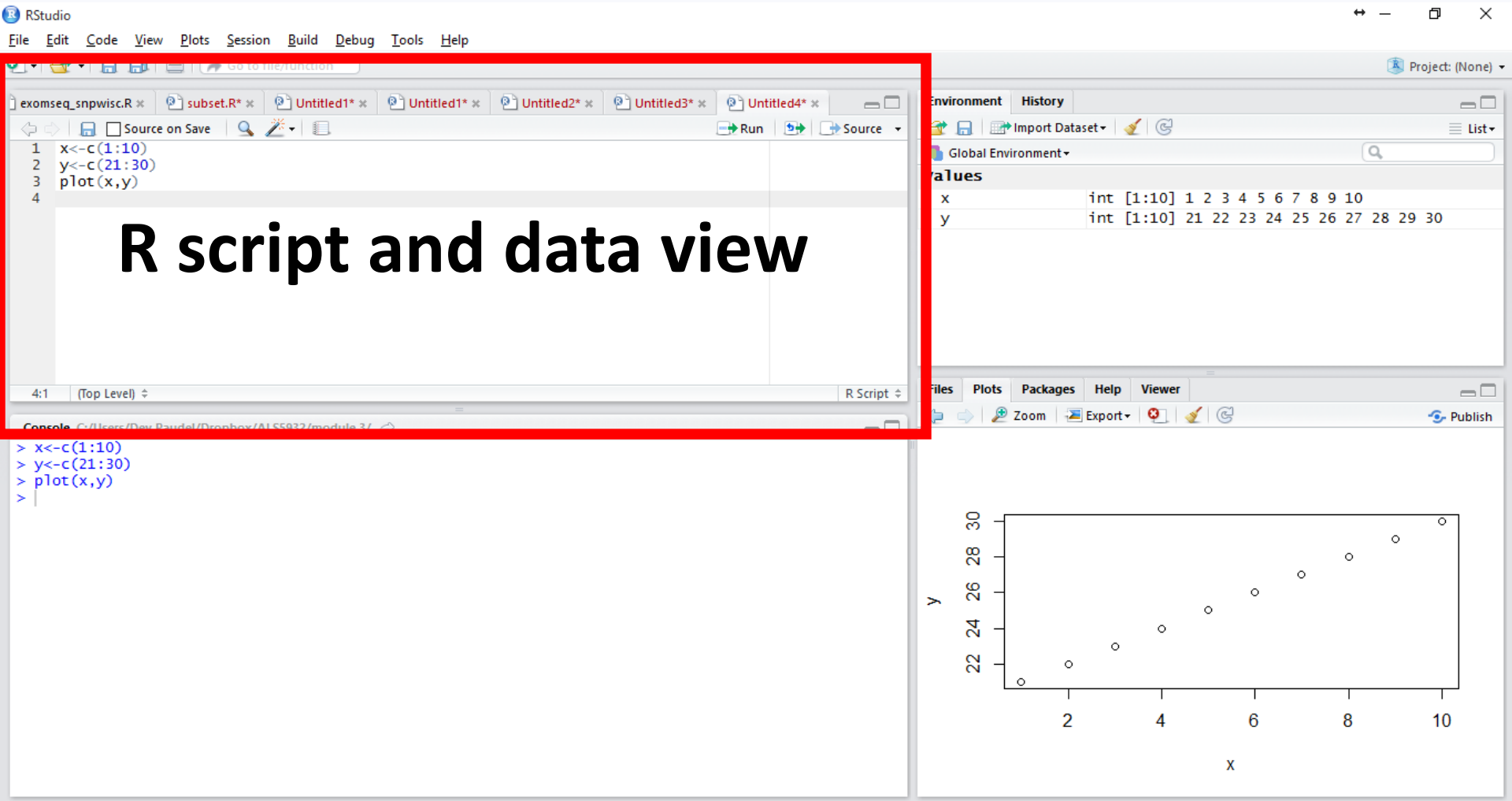




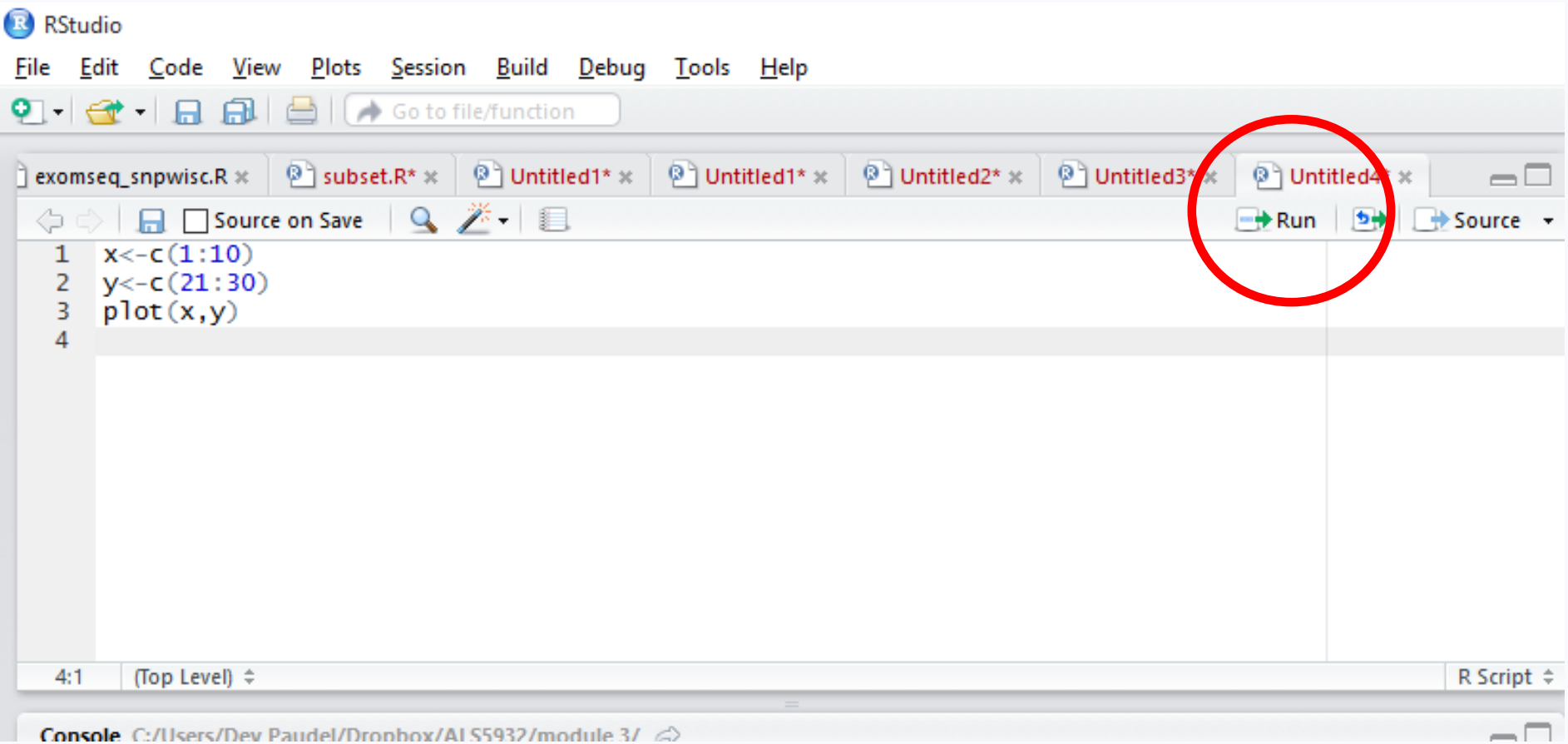
Exploring R studio



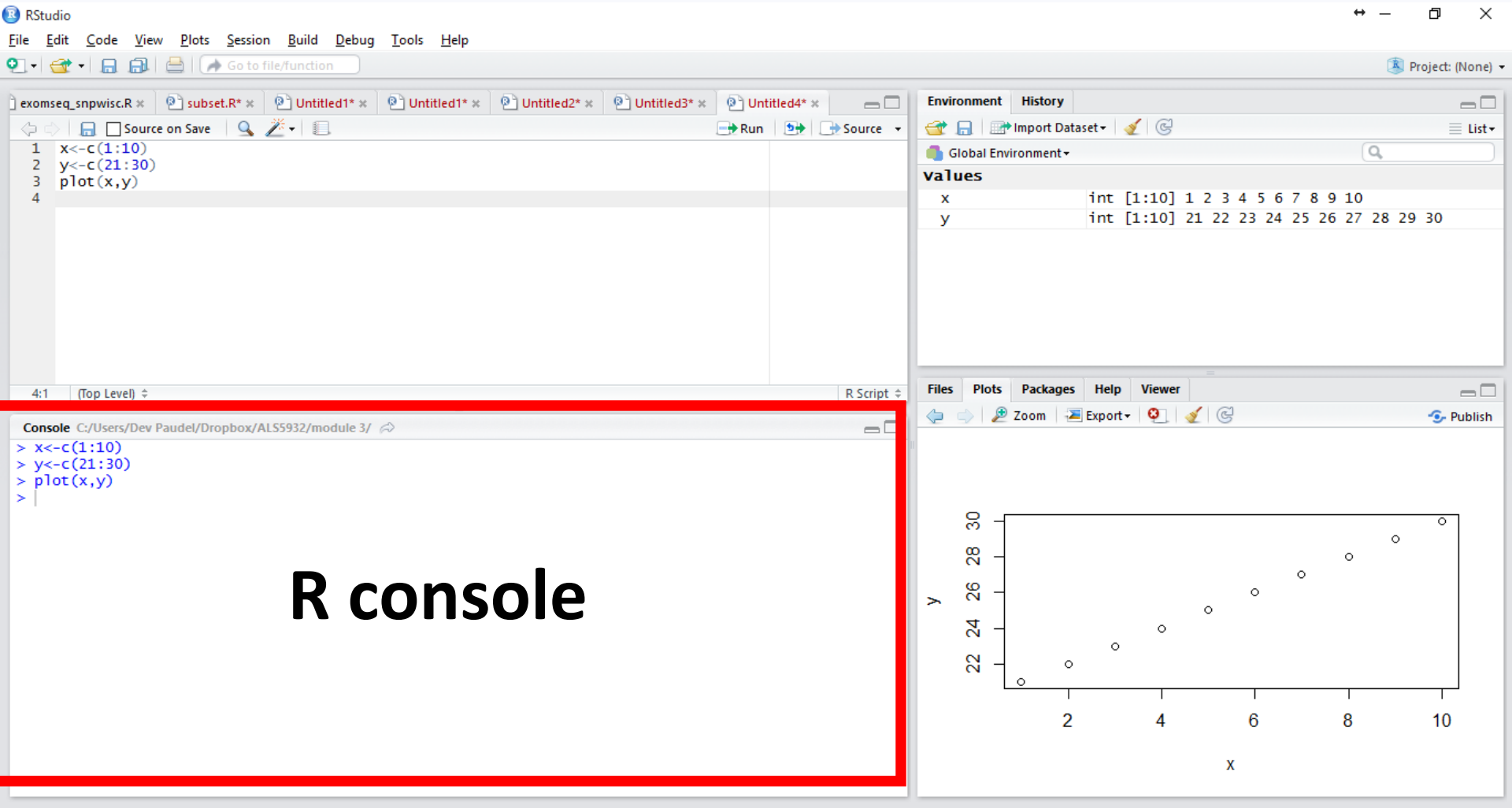
Exploring R studio



Exploring R studio



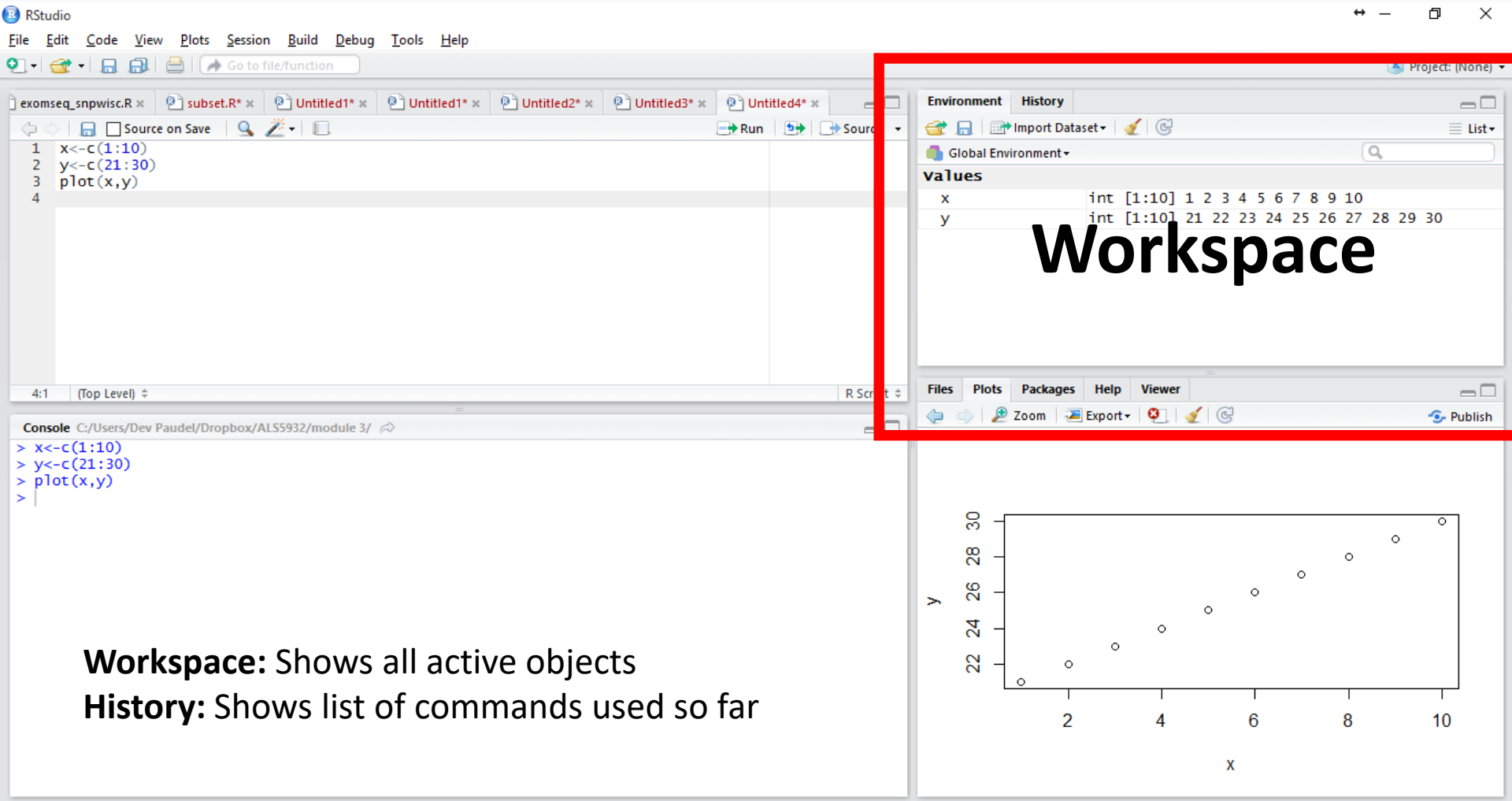
Exploring R studio



Exploring R studio

Console > You can type commands and see output

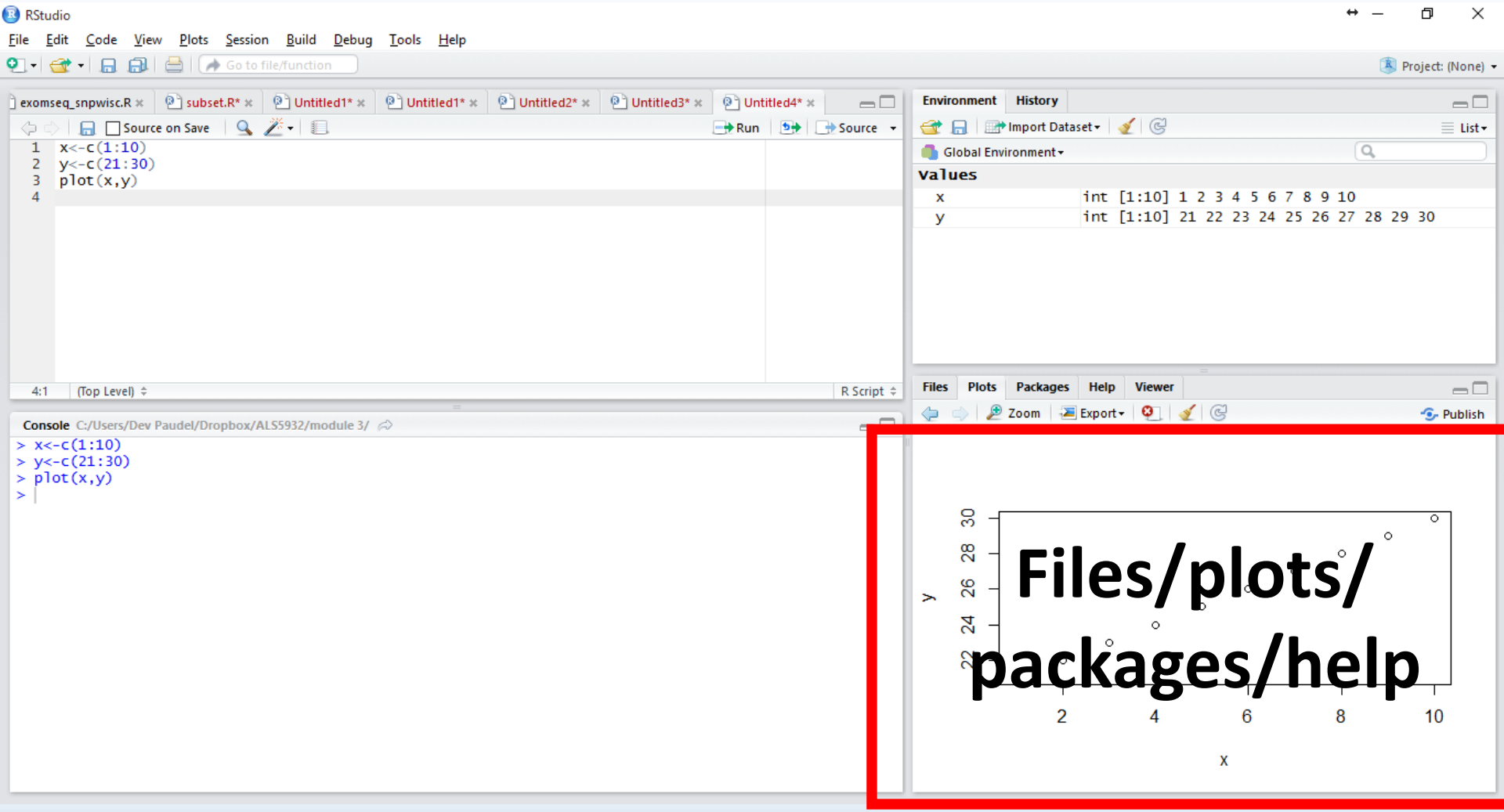
Exploring R studio



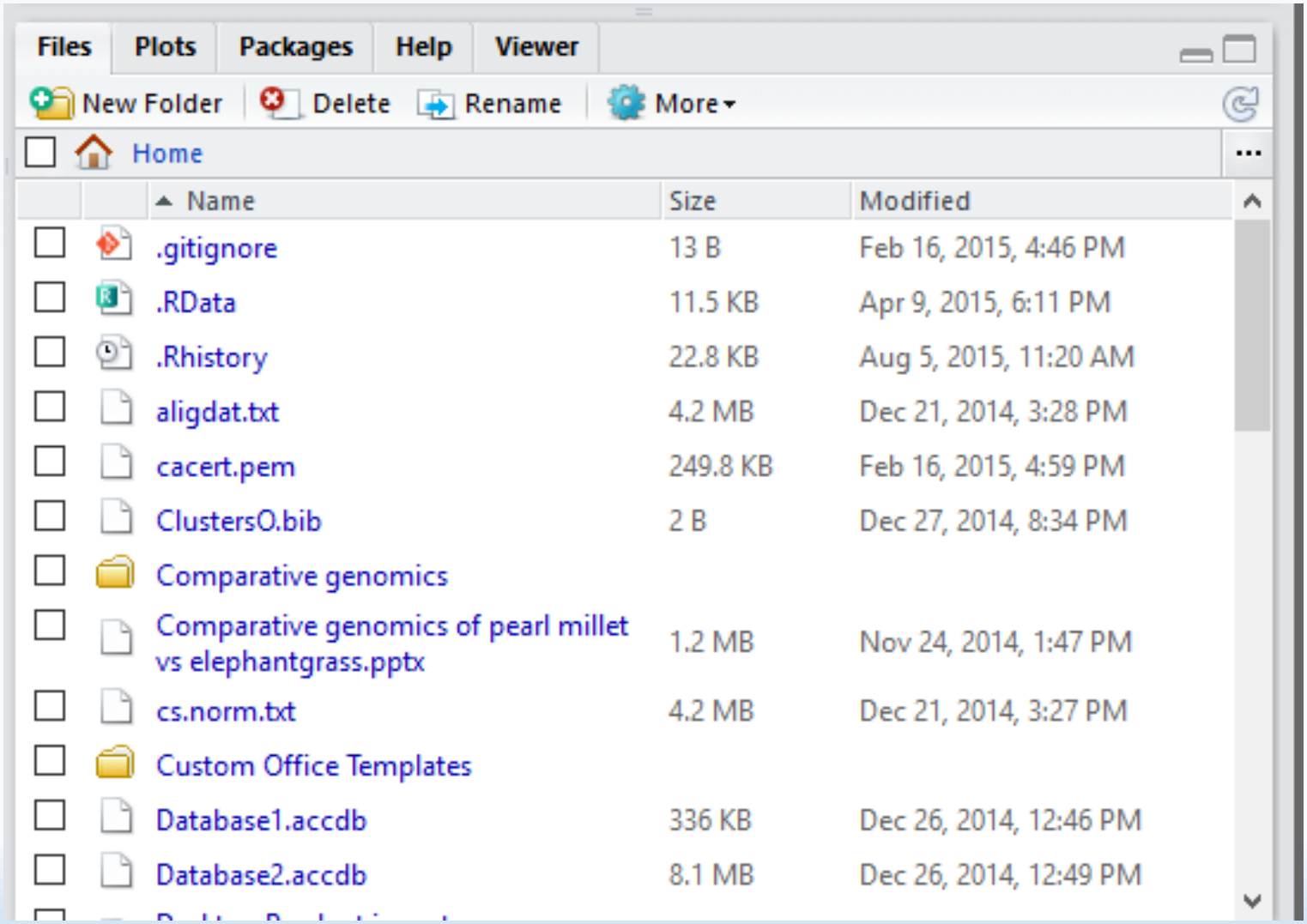
Workspace: Shows all active objects

History: Shows list of commands used so far

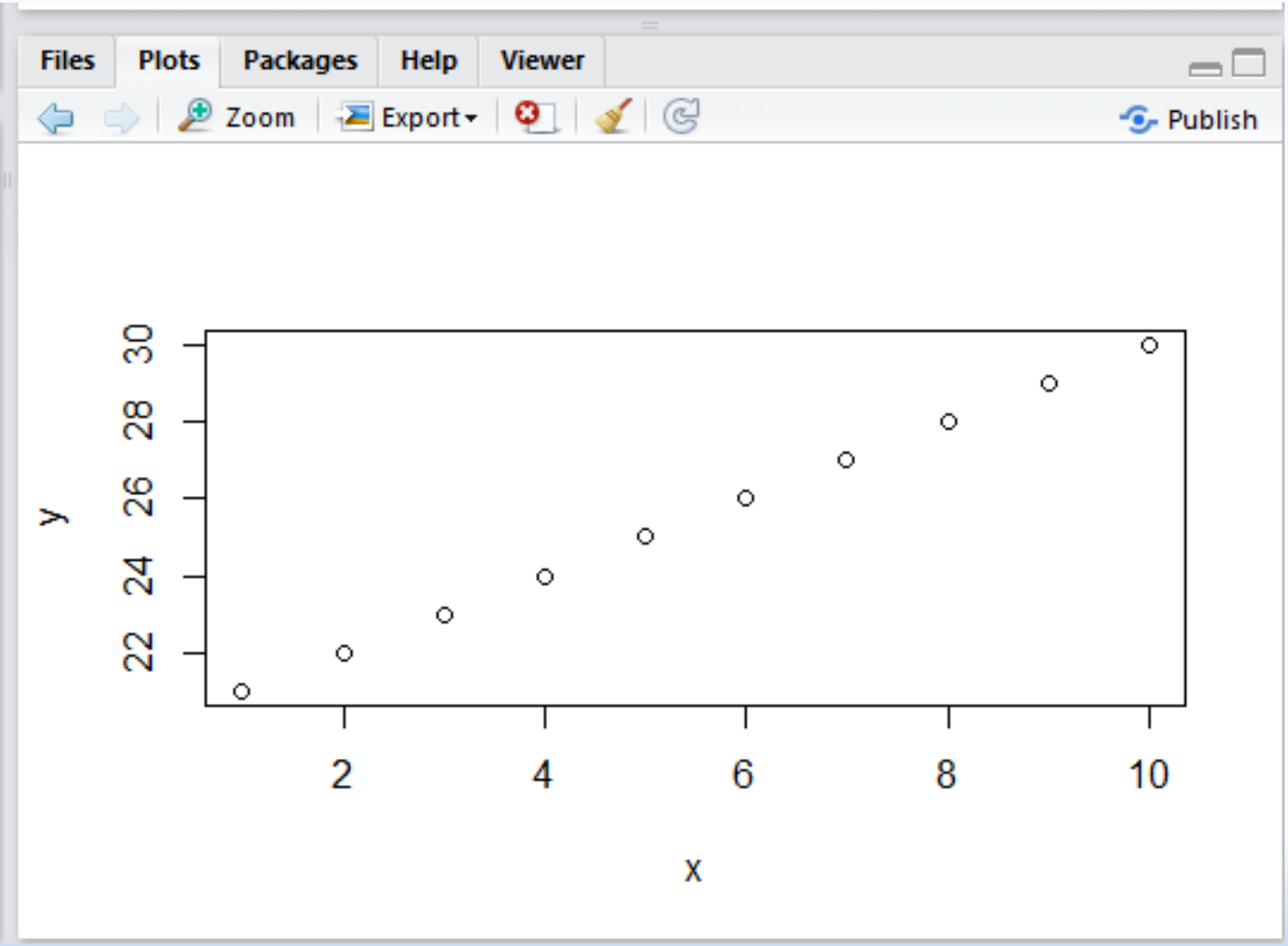
Exploring R studio



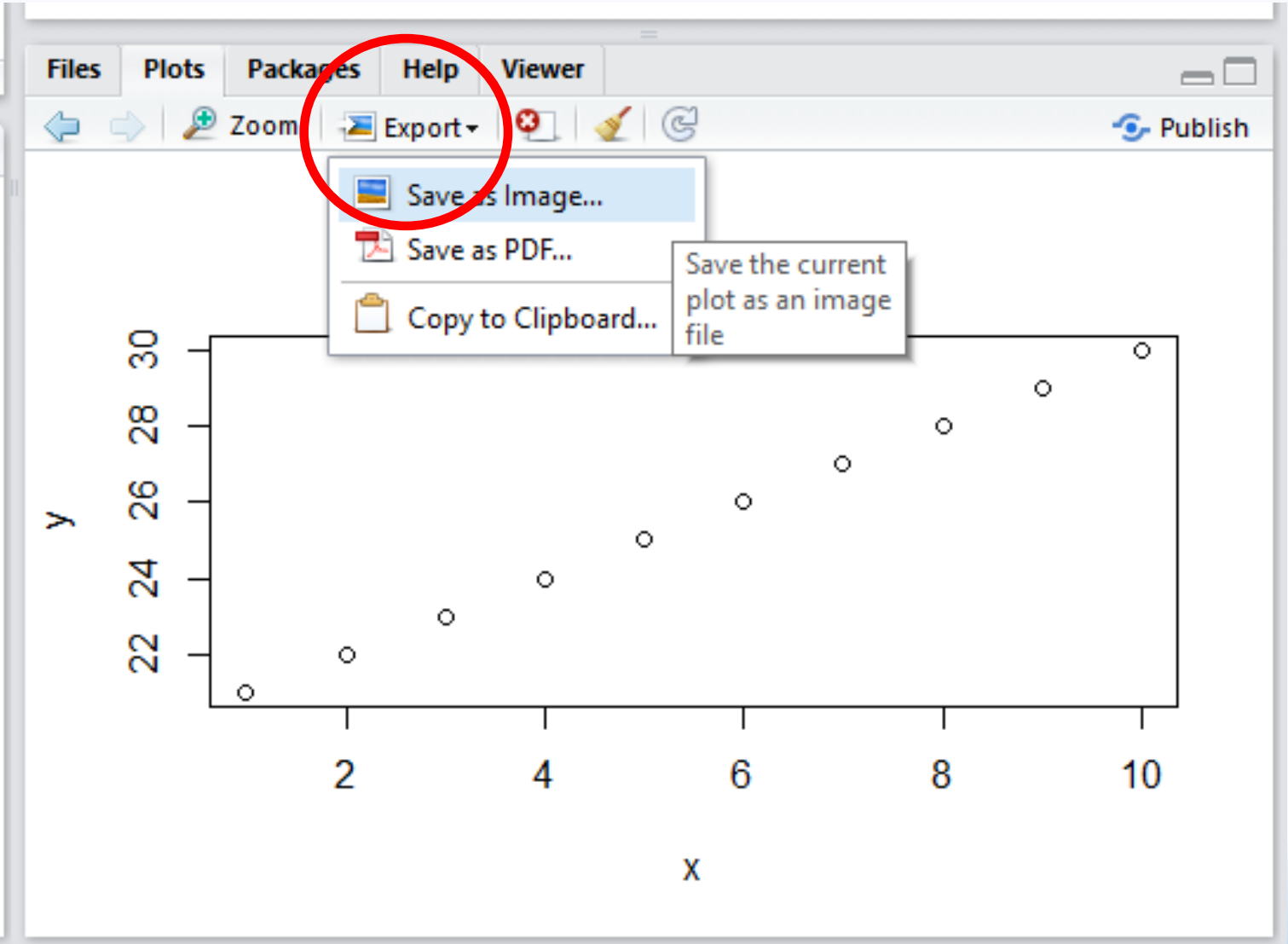
Exploring R studio



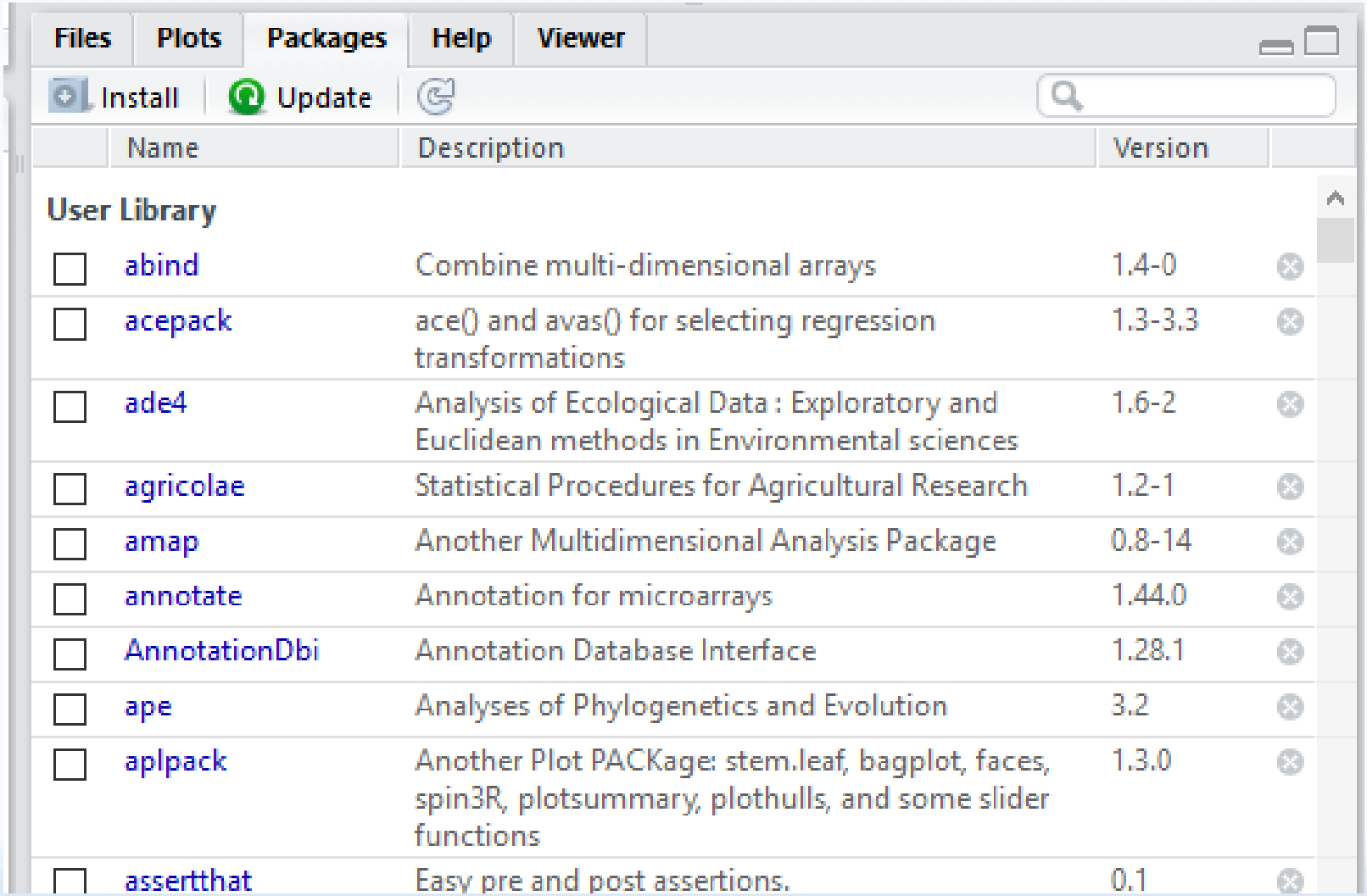
Exploring R studio



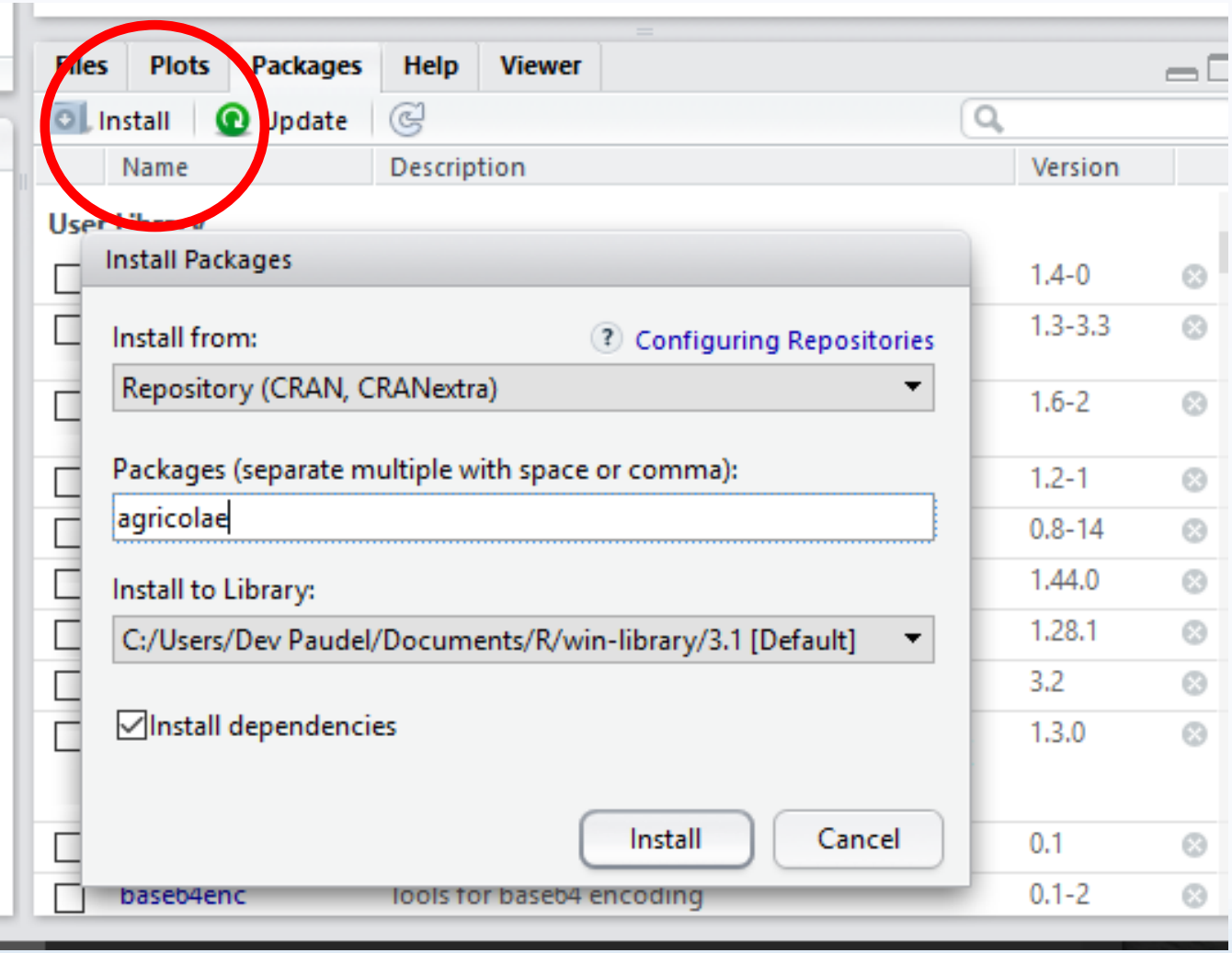
Exploring R studio



Exploring R studio

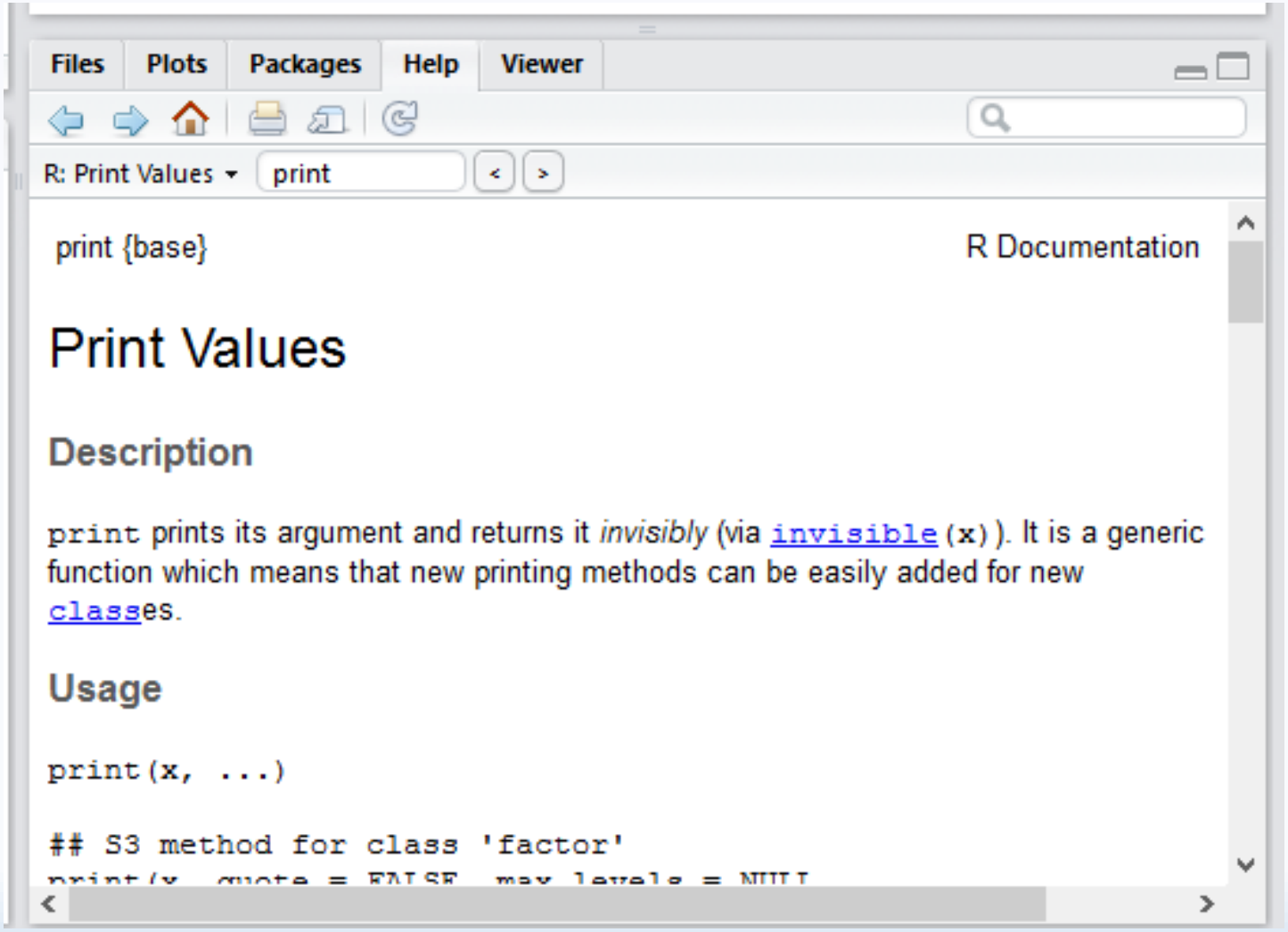


Exploring R studio



install.packages("agricolae")

Exploring R studio



?print

??print

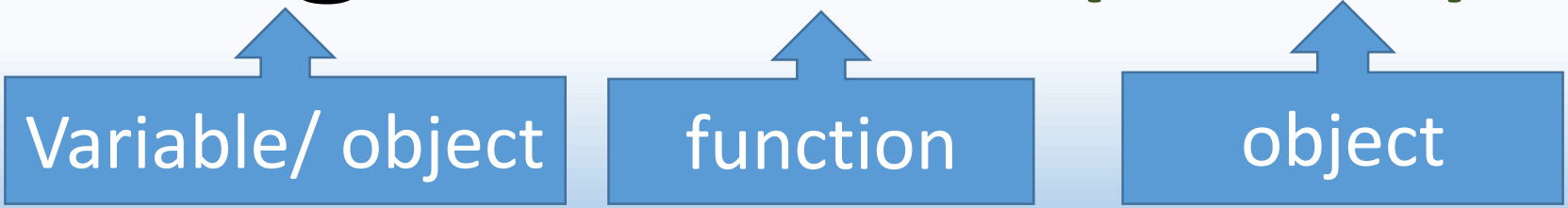
Using R as a calculator

Creating objects

```
## Naming Variables ##
dice = c(1,2,3,4,5,6)    # c is for concatenation
dice  <- c(1,2,3,4,5,6) # it is generally recommend using "<-" to assign
                          the right parts to left object
                        # "=" usually reserve for assign value in the functions
```

- *Everything that exists is an object*
- *Everything that happens is a function call* -John Chambers

```
avg = mean (data)
avg <- mean (data)
```



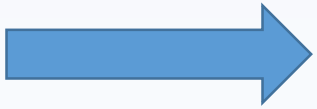
Functions in R

```
dice  <- c(1,2,3,4,5,6)

# Get average of dice
MEAN(dice) # R is case sensitive #
Mean(dice) # R is case sensitive #
mean (dice)

## Function within a function
log (x = sqrt(9))
log(3)
```

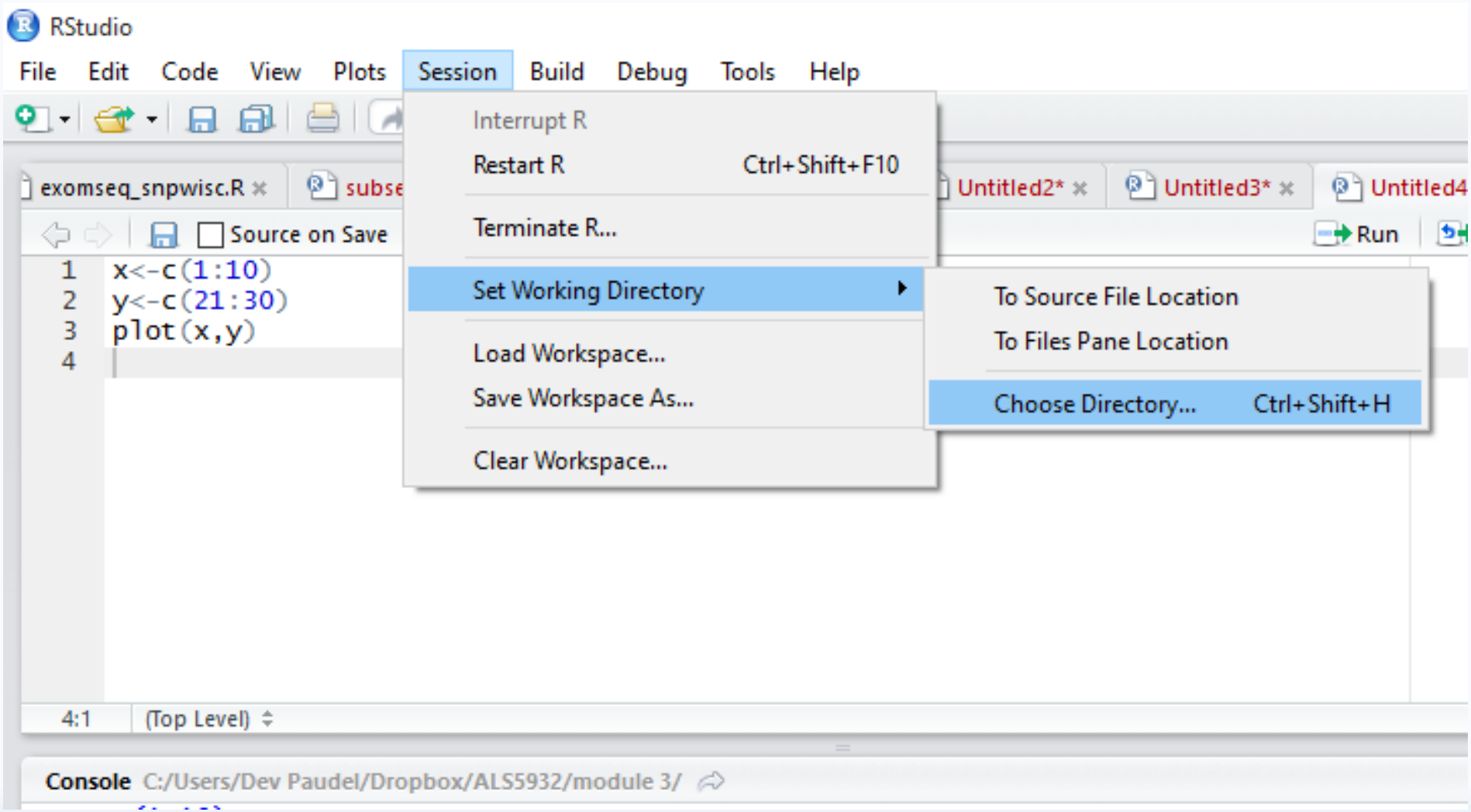
Next: Data Structures



Getting to know your data

- Setting working directory
- Reading data
- Exploring data

Setting up working directory



Ref: Denis Valle & Benjamin Baiser, Introduction to Applied Statistics for Agricultural and Life Sciences, 2014. ALS5932 Class Notes

Importing data

dir()

~/Desktop/RLab

https://raw.githubusercontent.com/dpaudel/cheatsheet/master/Wheat_91.txt

https://raw.githubusercontent.com/dpaudel/cheatsheet/master/wheatc.csv