



UNIVERSITY OF
PORTSMOUTH

R for Data Analysis

Introduction to R

(TB2 - Week 6)

Atefeh Khazaei

atefeh.khazaei@port.ac.uk



What we will learn this week?

- ❑ Introduction to R
 - ❑ What is R?
 - ❑ Installing R
 - ❑ Basic of R

What is R?

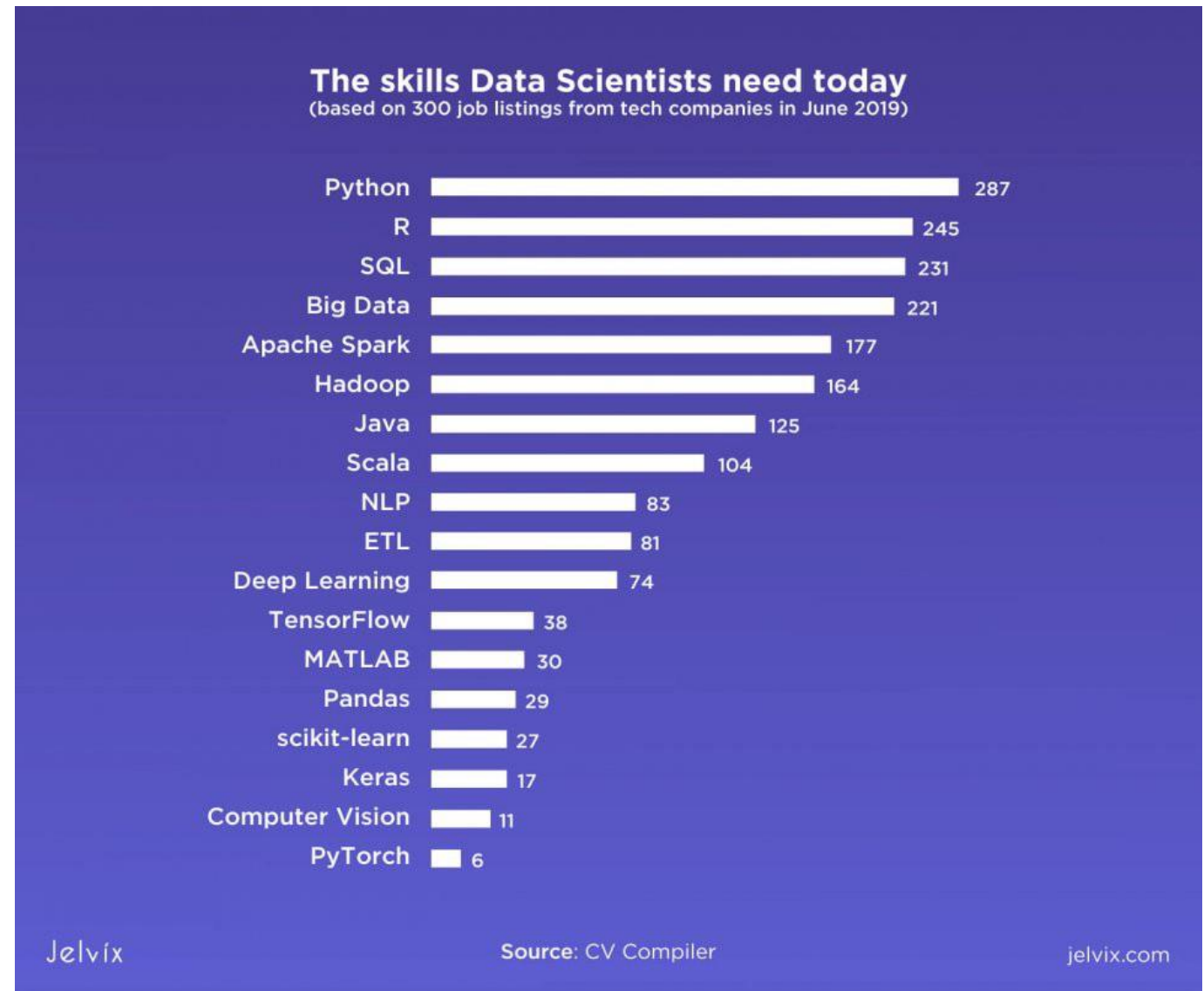


- ❑ R is a language and environment for statistical computing and graphics.
 - ❑ R is open source and free, also the packages are free
- ❑ R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible.
- ❑ One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.

What is R? (cont.)

- ❑ R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:
 - ❑ an effective data handling and storage facility,
 - ❑ a suite of operators for calculations on arrays, in particular matrices,
 - ❑ a large, coherent, integrated collection of intermediate tools for data analysis,
 - ❑ graphical facilities for data analysis and display either on-screen or on hardcopy,
 - ❑ a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

What is R? (cont.)



Ref: <https://medium.com/javarevisited/best-11-data-science-programming-languages-in-2020>

Installing R

Anaconda Navigator

File Help

ANACONDA.NAVIGATOR

Run Anaconda Navigator

Applications on base (root) Channels

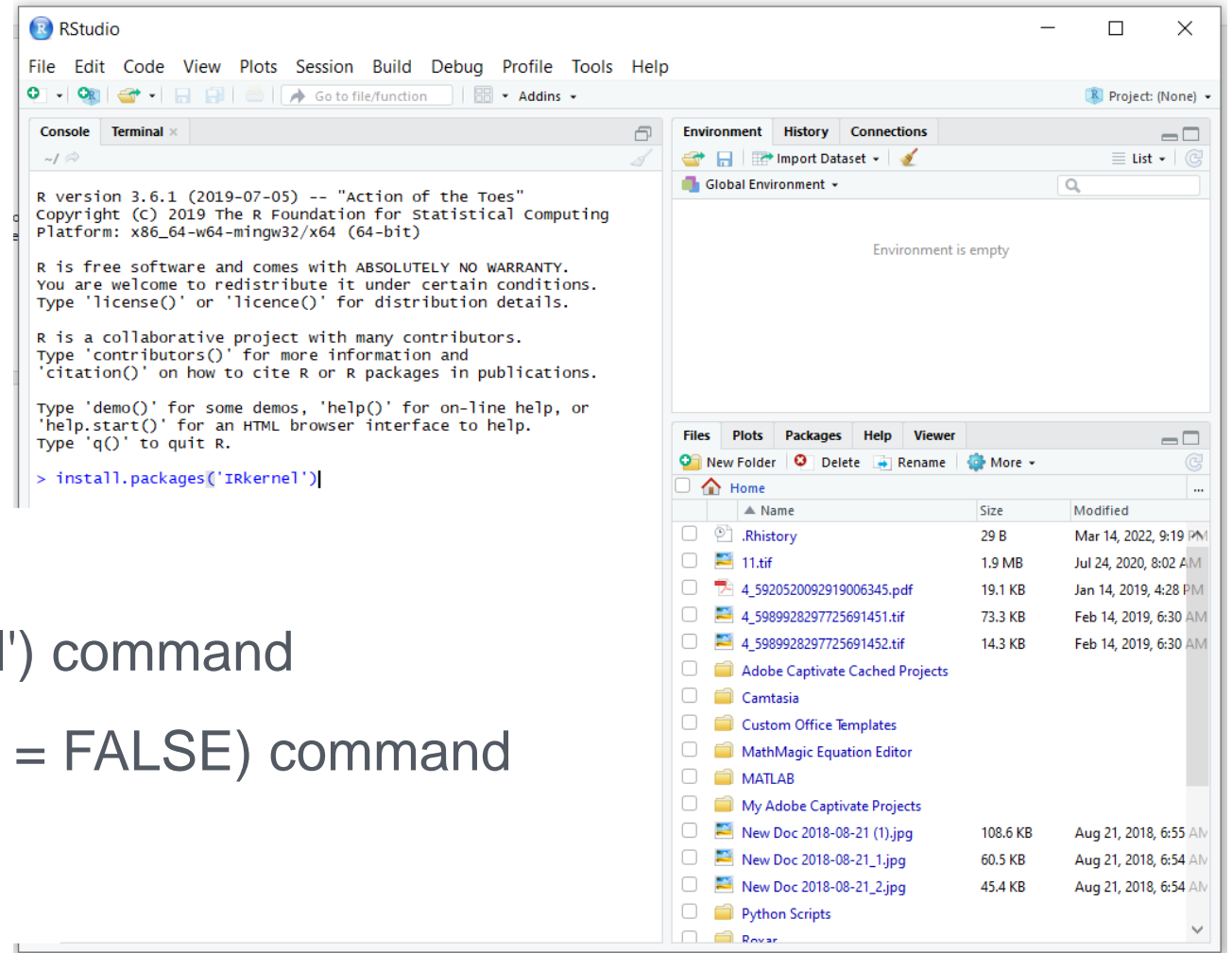
Application	Version	Description	Action
CMD.exe Prompt	0.1.1	Run a cmd.exe terminal with your current environment from Navigator activated	Launch
Datalore		Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.	Launch
IBM Watson Studio Cloud		IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.	Launch
JupyterLab	3.0.14	An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.	Launch
Notebook	6.3.0	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	Launch
Powershell Prompt	0.0.1	Run a Powershell terminal with your current environment from Navigator activated	Launch
Qt Console	5.0.3	PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	Launch
Spyder	4.2.5	Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features	Launch
Glueviz	1.0.0	Multidimensional data visualization across files. Explore relationships within and among related datasets.	Install
Orange 3	3.26.0	Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.	Install
PyCharm Professional		A Full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.	Install
RStudio	1.1.456	A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	Install

Install

Alternatively you can download Rstudio from: <https://www.rstudio.com/products/rstudio/download/>

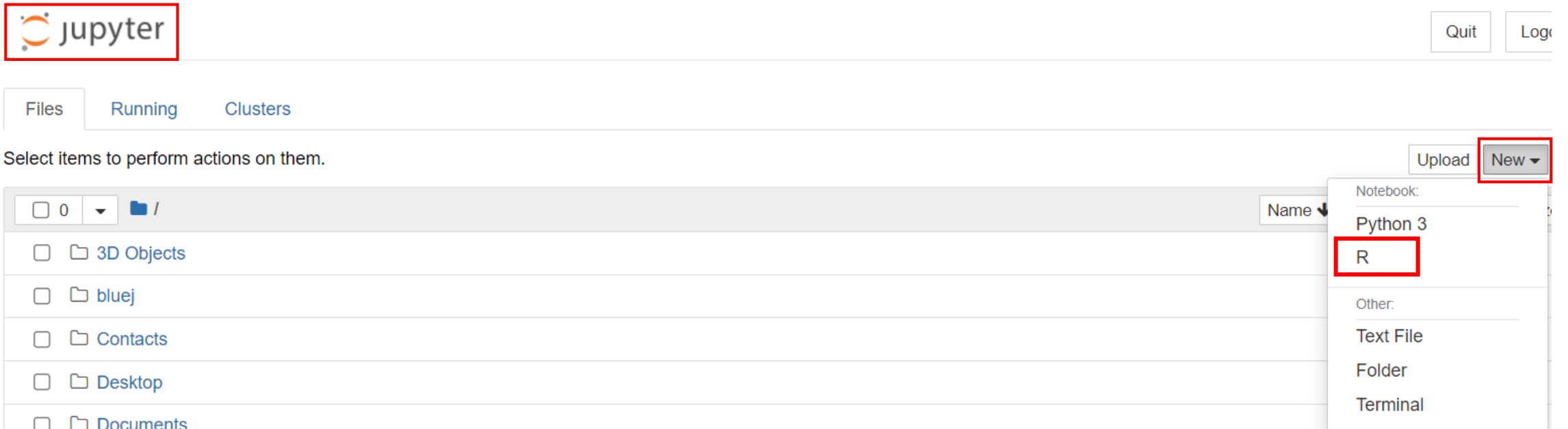
Installing R (cont.)

- ❑ Run RStudio
- ❑ Execute `install.packages('IRkernel')` command
- ❑ Execute `IRkernel::installspec(user = FALSE)` command



Installing R (cont.)

Running Jupyter Notebook with R



The screenshot shows the Jupyter Notebook web interface. At the top left, the Jupyter logo is highlighted with a red box. To the right are 'Quit' and 'Log out' buttons. Below the logo is a tab bar with 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser. A message says 'Select items to perform actions on them.' Below this is a list of files and folders: '0', '/', '3D Objects', 'bluej', 'Contacts', 'Desktop', and 'Documents'. On the right side of the file browser, there are 'Upload' and 'New' buttons. The 'New' button is highlighted with a red box, and its dropdown menu is open, showing options: 'Notebook:', 'Python 3', 'R' (highlighted with a red box), and 'Other:', 'Text File', 'Folder', 'Terminal'.

Basic of R Jupyter

jupyter Untitled7 Last Checkpoint: 18 minutes ago (unsaved changes)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

R

Save New Split Cell Run Stop Restart Run All Code

```
In [1]: # It is a comment
```

```
In [2]: # Assignment to a variable (First method)
```

```
a = 5 + 2
```

```
a
```

```
7
```

```
In [3]: # Assignment to a variable (Second method)
```

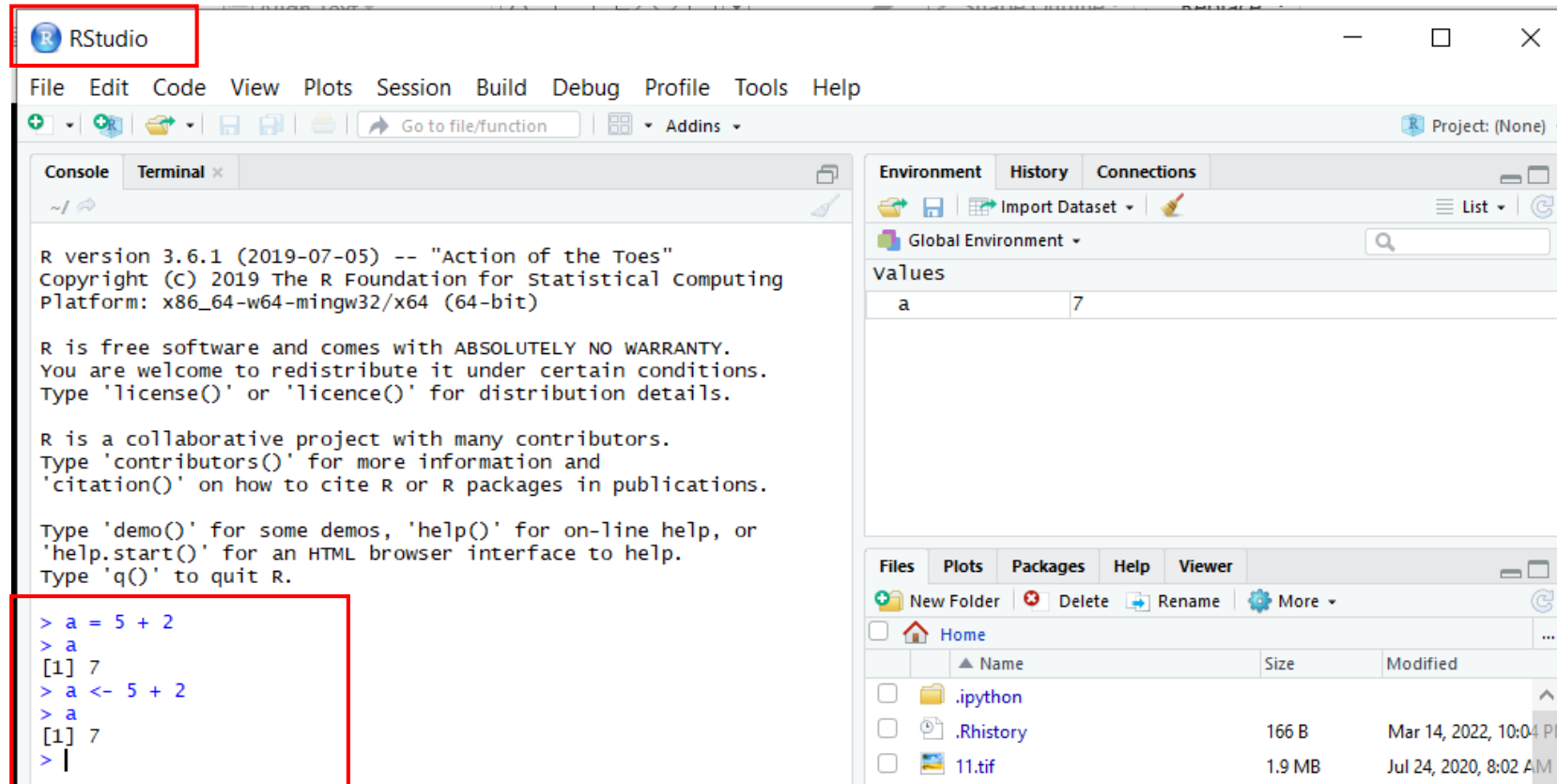
```
a <- 5 + 2
```

```
a
```

```
7
```

Basic of R (cont.)

RStudio



Basic of R (cont.)

Installing Packages / Jupyter

jupyter Untitled7 Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | R O

Run Code

`install.packages("package name")`

```
In [4]: # install.packages("package name")
# "dplyr" package that helps to format data
install.packages("dplyr")

also installing the dependencies 'rlang', 'tibble', 'tidyselect'
```

Example:

"dplyr" package that helps to format data

<https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html>

`install.packages("dplyr")`

```
There are binary versions available but the source versions are later:
  binary source needs_compilation
rlang    0.4.11 1.0.2              TRUE
tibble   3.1.1 3.1.6              TRUE
tidyselect 1.1.1 1.1.2              TRUE
dplyr    1.0.6 1.0.8              TRUE
```

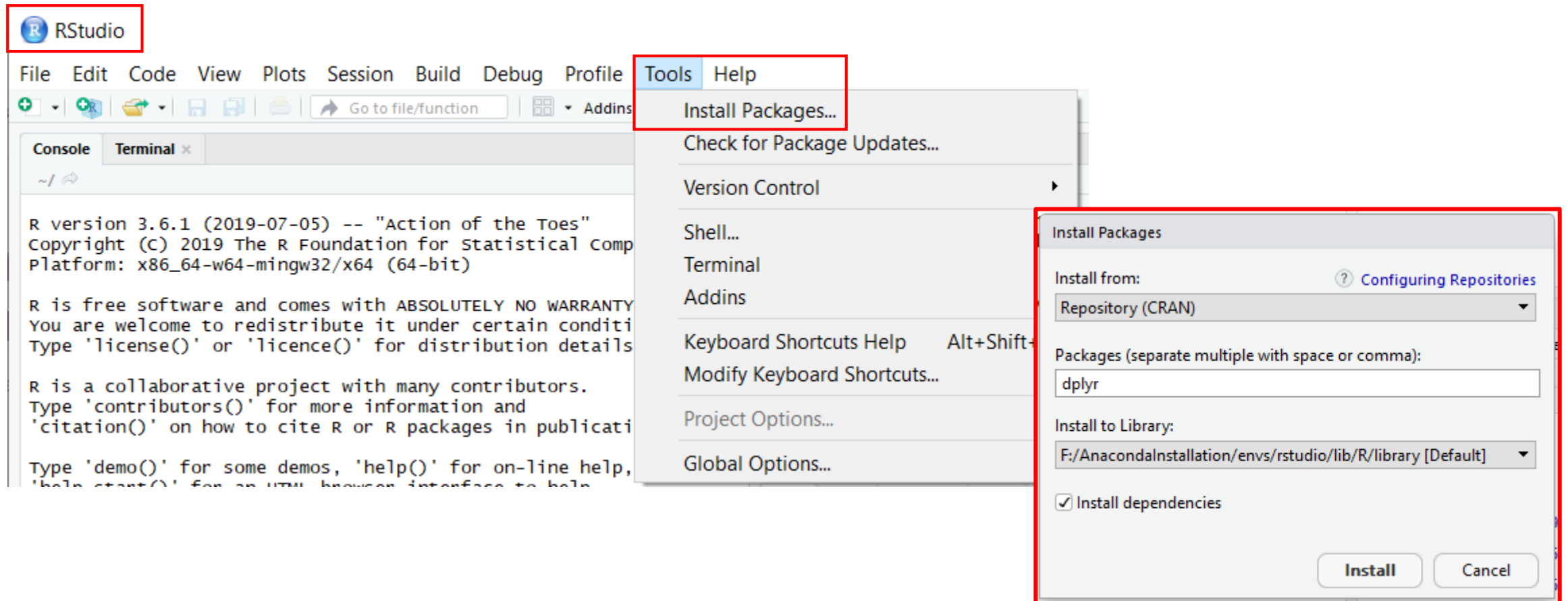
```
Binaries will be installed
package 'rlang' successfully unpacked and MD5 sums checked
```

```
Warning message:
"cannot remove prior installation of package 'rlang'"
Warning message in file.copy(savedcopy, lib, recursive = TRUE):
"problem copying F:\AnacondaInstallation\envs\rstudio\lib\R\library\00LOCK\rlang\libs\x64\rlang.dll to F:\AnacondaInstallation\envs\rstudio\lib\R\library\rlang\libs\x64\rlang.dll: Permission denied"
Warning message:
"restored 'rlang'"
```

```
package 'tibble' successfully unpacked and MD5 sums checked
package 'tidyselect' successfully unpacked and MD5 sums checked
package 'dplyr' successfully unpacked and MD5 sums checked
```

Basic of R (cont.)

Installing Packages / RStudio



Basic of R (cont.)

Arithmetic with R

- ❑ Following arithmetic operators are the basic forms of Math calculation with R:
 - ❑ Addition: +
 - ❑ Subtraction: -
 - ❑ Multiplication: *
 - ❑ Division: /
 - ❑ Exponentiation: ^
 - ❑ Modulo: %%
 - ❑ `sqrt(a)` # *'a is a variable or a number'*
- ❑ **Try some examples with these operators.**

Basic of R (cont.)

Practical Sheet

- ❑ There are some examples on practical sheet:
 - ❑ Basic commands
 - ❑ Assigning values (`=` `->` `<-`)
 - ❑ Sequences
 - ❑ Math operators
 - ❑ Data Types (Numeric, Character, Logical, `typeof()` function)

References & More Resources

□ References:

□ Learning R:

<https://www.linkedin.com/learning/learning-r-2/>



□ R Programming in Data Science: Setup and Start

<https://www.linkedin.com/learning/r-programming-in-data-science-setup-and-start/>

□ To use LinkedInLearning, you can log in with your university account:

<https://myport.port.ac.uk/study-skills/linkedin-learning>



Practical Session

- ❑ Install R.
- ❑ Try installing packages (Jupyter and Rstudio).
- ❑ Try different examples with arithmetic operators (Slide 13).
- ❑ Open the practical file, run the commands in Jupyter and Rstudio.