Foundation of Machine Learning

CSE4032

Lecture 00: Installation guide for R and RStudio

Dr. Kundan Kumar
Associate Professor
Department of ECE

Faculty of Engineering (ITER)
S'O'A Deemed to be University, Bhubaneswar, India-751030

Outline

Introduction to R

- R is a free software environment for statistical computing and graphics.
- R is the most popular language in the world of Data Science.
- It is heavily used in analyzing data that is both structured and unstructured.
- This has made R, the standard language for performing statistical operations.
- R allows various features that set it apart from other Data Science languages.

Why learn R?

- Free and open-source tool
- Large community of users
- Latest cutting edge technology
- Independent platform
- Gateway to lucrative career
- Robust visualization library
- Go to language for Statistics and Data Science
- Used in almost every industry

Setup R environment

- It compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS.
- Download the R installer from https://cran.r-project.org/
- Run the installer and keep default settings.
- Must insure that you have admin rights. Without this, you will not be able to install additional packages later.
- Usually, I prefer two approaches to setup R environment.
  - Independent environment
  - Anaconda environment

From where to download R

- Download the R installer from https://cran.r-project.org/

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

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
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2021-02-15, Lost Library Book) R-4.0.4.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.

From where to download R

- Download the R installer from https://cran.r-project.org/

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

From where to download R

- Download the R installer from https://cran.r-project.org/

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

R-4.0.4 for Windows (32/64 bit)

Download R 4.0.4 for Windows (85 megabytes, 32/64 bit)

Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server. You will need a version of md5sum for windows: both graphical and command line versions are available.

Frequently asked questions

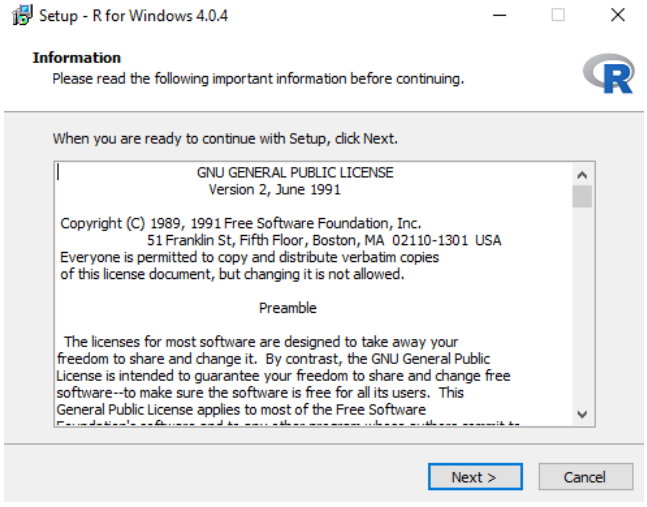
- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

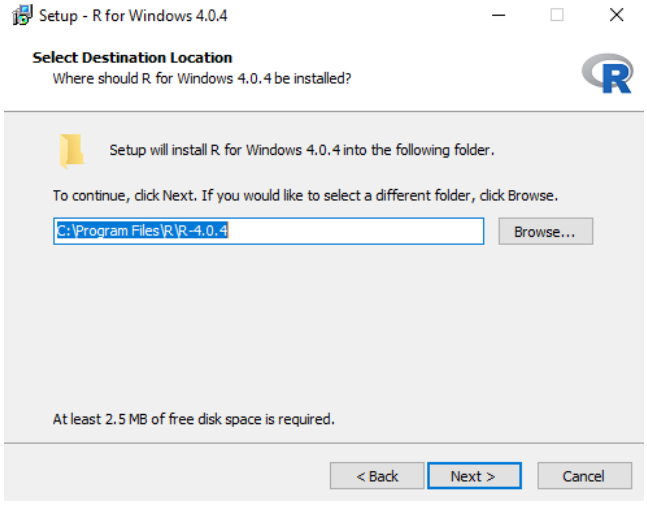
Other builds

- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

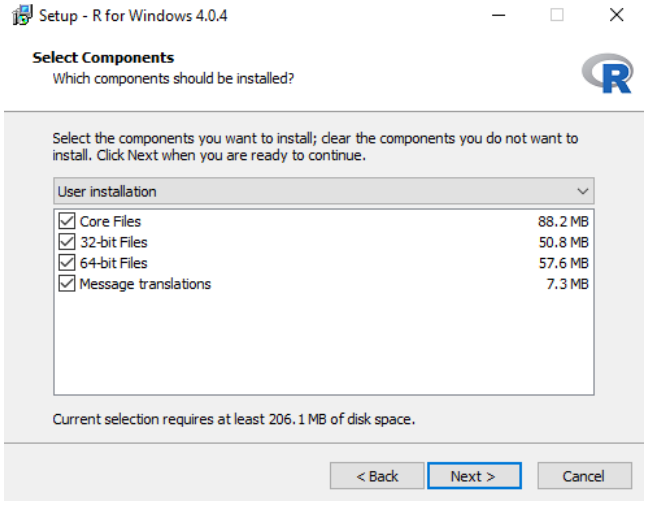
R installation guide



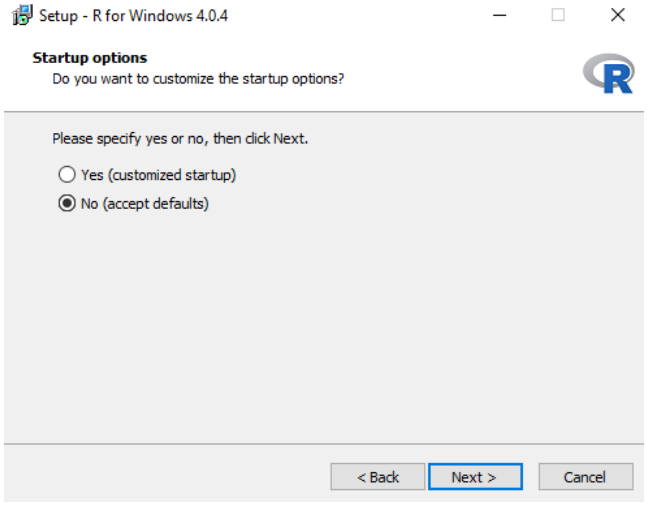
R installation guide



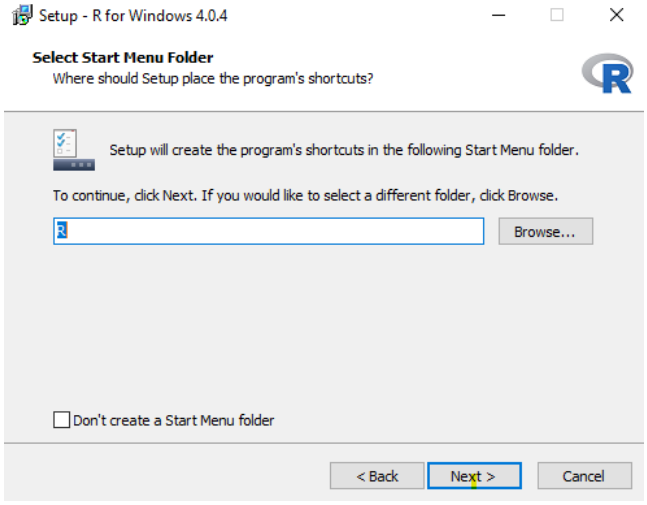
R installation guide



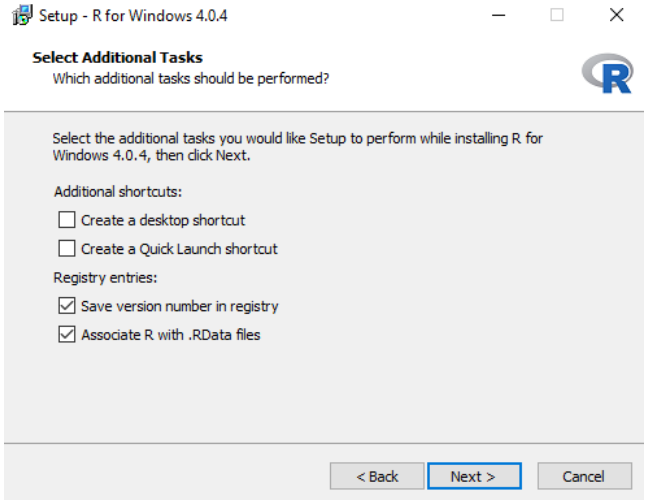
R installation guide



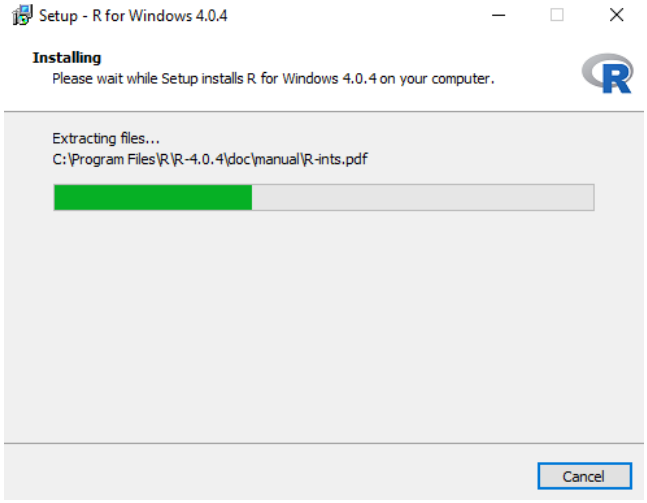
R installation guide



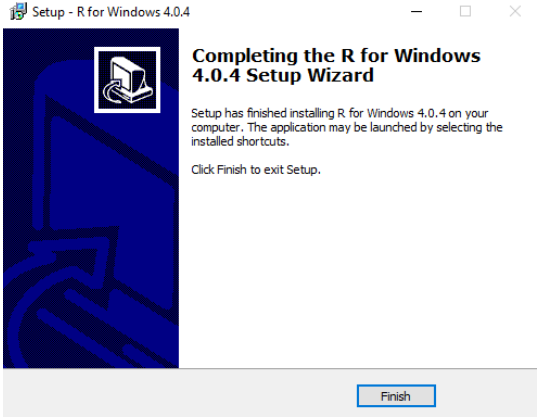
R installation guide



R installation guide



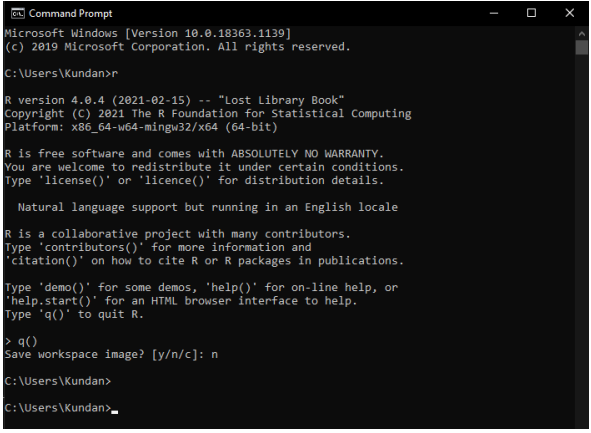
R installation guide



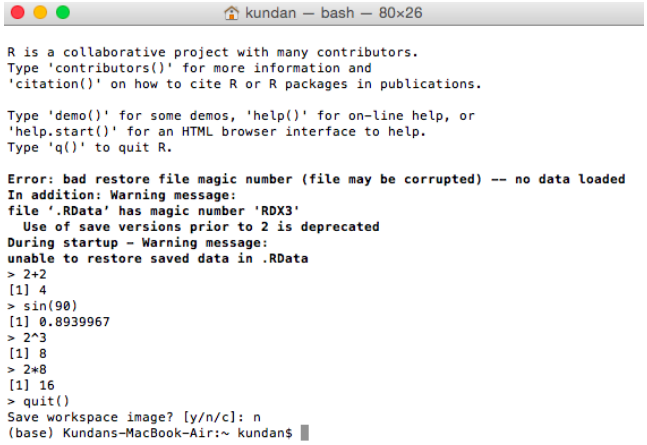
- Don't forget to set the path to use R from command prompt / terminal.

Check R installation

- Type 'r' in Command Prompt to ensure that R is the path.



Use R as a calculator



From where to download RStudio

- Download the RStudio installer from https://rstudio.com/products/rstudio/download/#download

RStudio Desktop 1.4.1106 - Release Notes

1. Install R. RStudio requires R 3.0.1+.

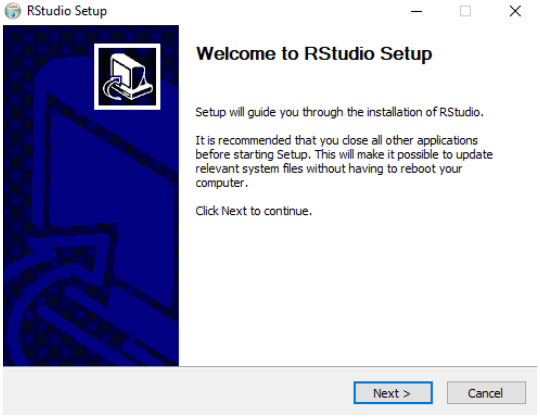
2. Download RStudio Desktop. Recommended for your system:

DOWNLOAD RSTUDIO FOR WINDOWS
1.4.1106 | 155.97MB

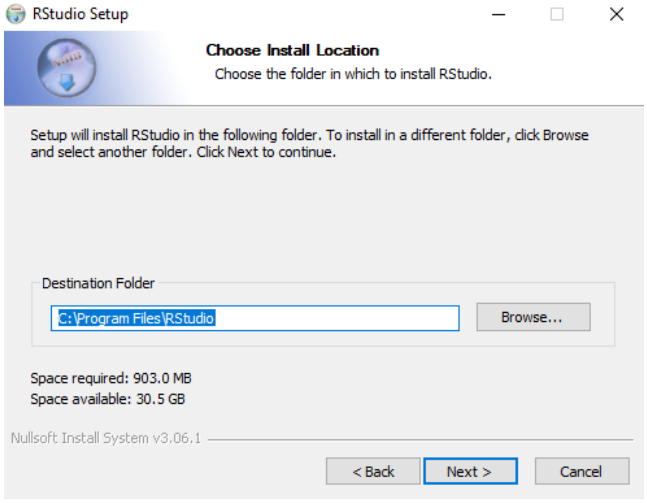
Requires Windows 10/8 (64-bit)

RStudio Installation guide

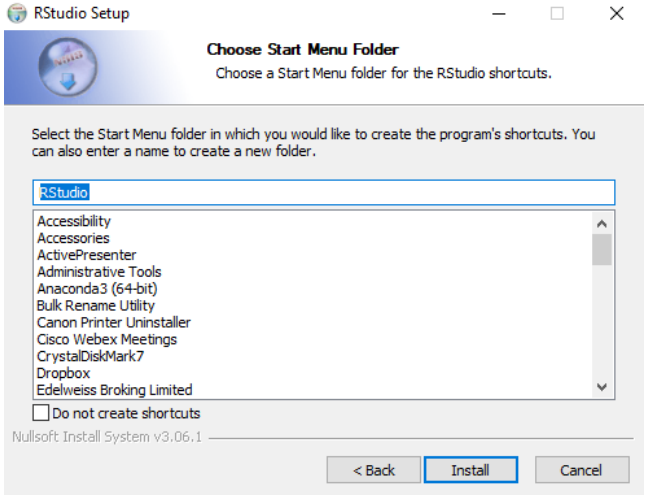
- Run RStudio-1.4.1106.exe as administrator.



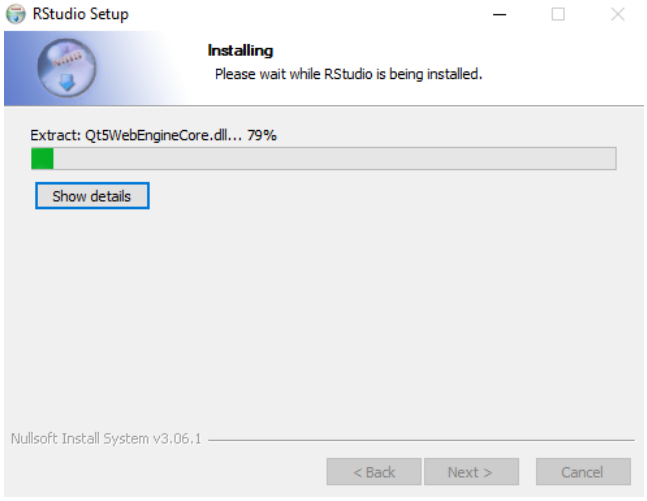
RStudio Installation guide



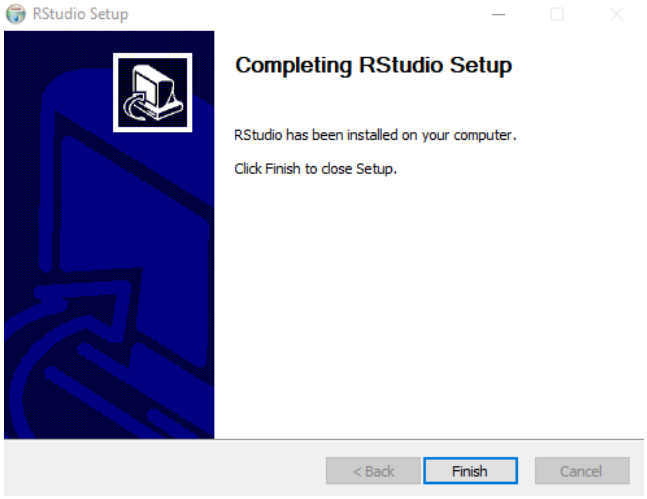
RStudio Installation guide



RStudio Installation guide



RStudio Installation guide



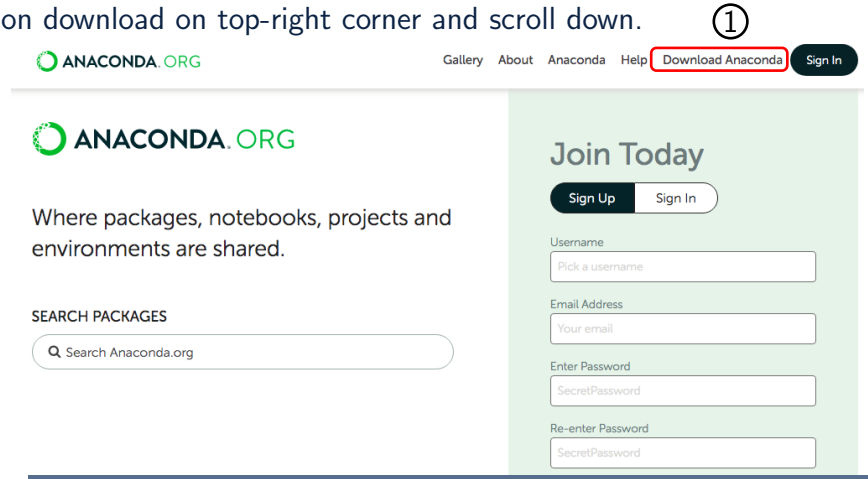
Introduction

- Anaconda is an open-source distribution for python and R.

- It is used for
  - data science, data analytic,
  - machine learning,
  - deep learning, etc.
- More than 300 libraries are available for data science.
- Simplified package management and deployment.
- An easily manageable environment setup which can deploy any project with the click of a single button.

Where to find Anaconda?

1. Go to website: https://www.anaconda.org
2. Click on download on top-right corner and scroll down.

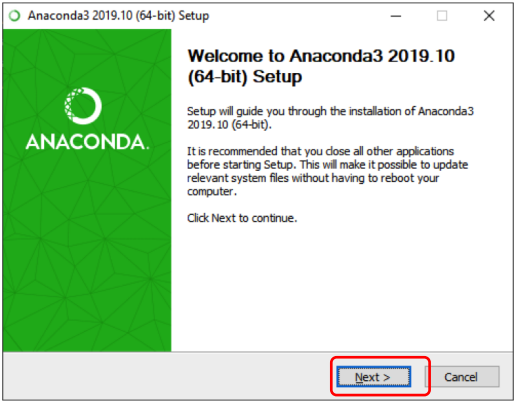


Where to find Anaconda?

3. Scroll down and choose your operating system.
4. Download 64-Bit or 32-Bit Graphical Installer (Python 3.8 version).

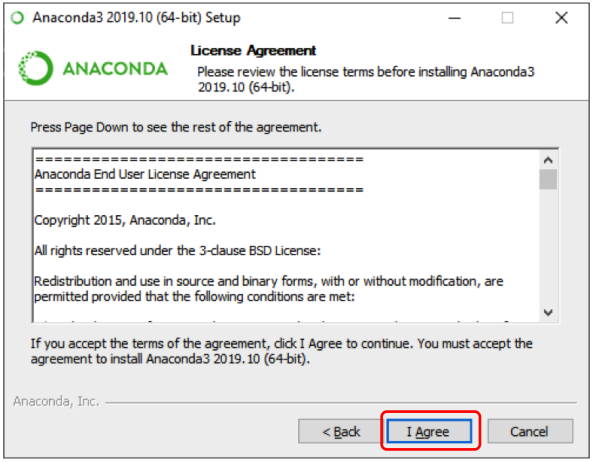
How to install Anaconda?

- In windows, double click the installer to run (you may choose run as Administrator for safe side).
- Click on Next.



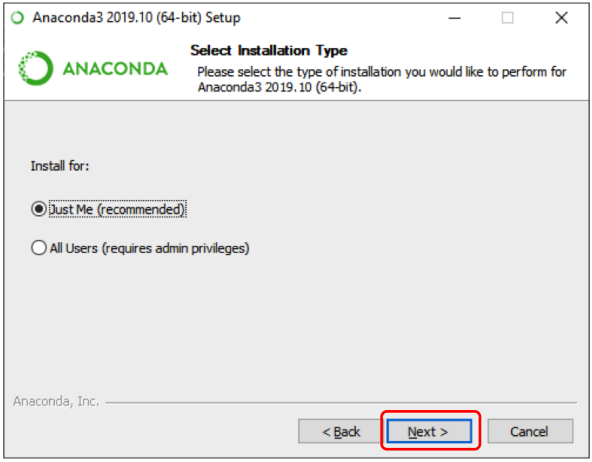
How to install Anaconda?

- Click on I Agree.



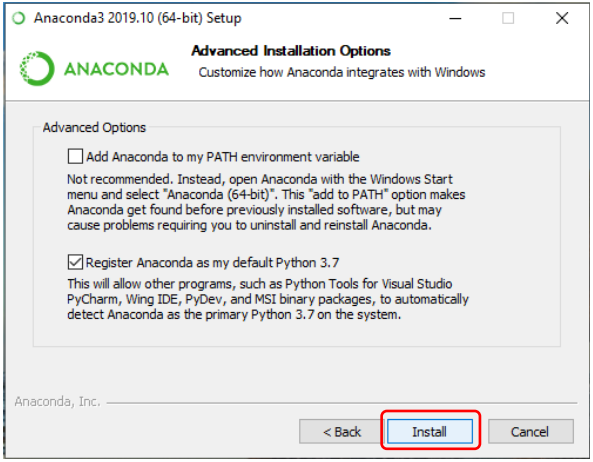
How to install Anaconda?

- Click on Next.



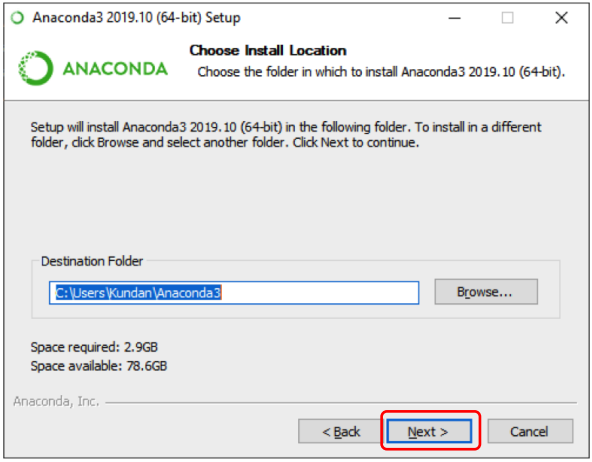
How to install Anaconda?

- Click on Install.



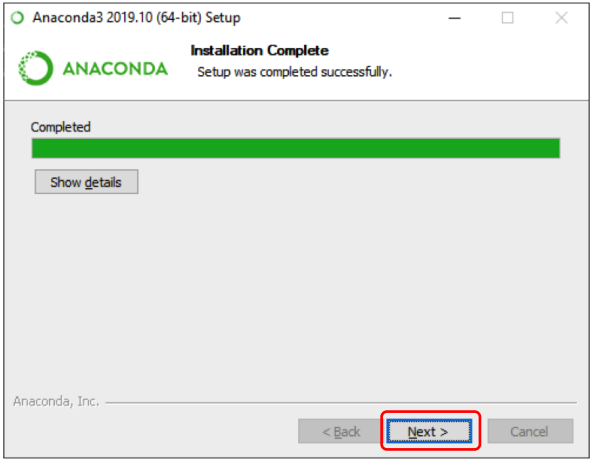
How to install Anaconda?

- Click on Next.



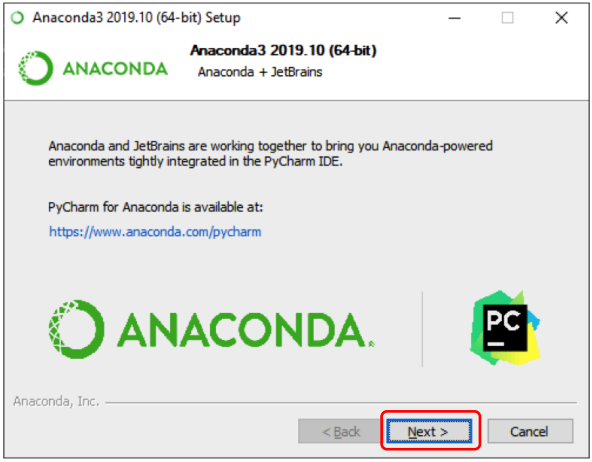
How to install Anaconda?

- Click on Next, when it gets highlighted.



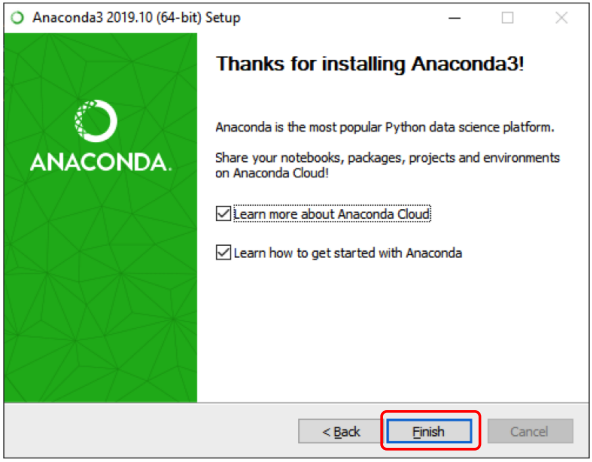
How to install Anaconda?

- Click on Next.



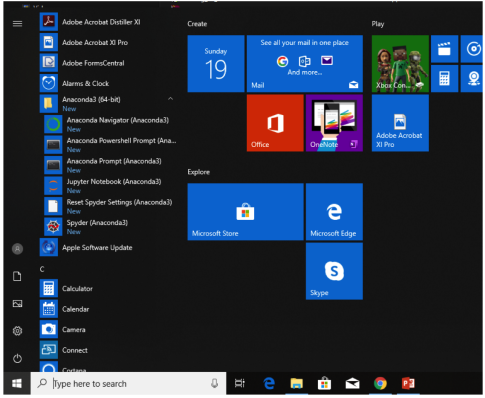
How to install Anaconda?

- Click on Finish.



Open Anaconda Powershell Prompt

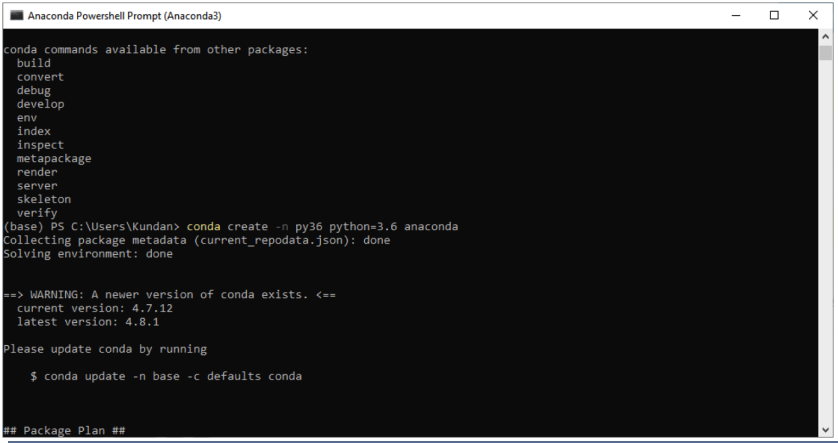
- Go to start (bottom-left corner) and scroll down to find Anaconda3 (64 bit).
- Click on Anaconda Powershell Promt to open it.



Check, is anaconda in path?

- In the powershell prompt run "conda" to ensure that anaconda is in path.

$ conda



How to create a virtual environment for R?

- To create a virtual environment for R, run

  $ conda create -n R4 r-essentials r-base

- press Y to proceed. Wait for complete the installation.
- After the completion of the installation, activate the virtual environment as

  $ conda activate R4

- Ensure that default environment base is changed to R4.
- To deactivate the environment

  $ conda deactivate

NOTE: You can use up and down key in the keyboard to see command history executed.

How to create a virtual environment for Python?

- To create a virtual environment for python, run

  $ conda create -n py36 python=3.6 anaconda

- press Y to proceed. Wait for complete the installation.
- After the completion of the installation, activate the virtual environment as

  $ conda activate py36

- Ensure that default environment base is changed to py36.
- To deactivate the environment

  $ conda deactivate

NOTE: You can use up and down key in the keyboard to see command history executed.

How to install and start RStudio

- Activate the virtual environment as

  $ conda activate R4

- Ensure that default environment base is changed to R4.
- Install RStudio using following command in terminal

  $ conda install -c r rstudio

- Similarly, other library/package/module can be installed.
- You can install packages manually from RStudio.
- You start RStudio from terminal

  $ rstudio

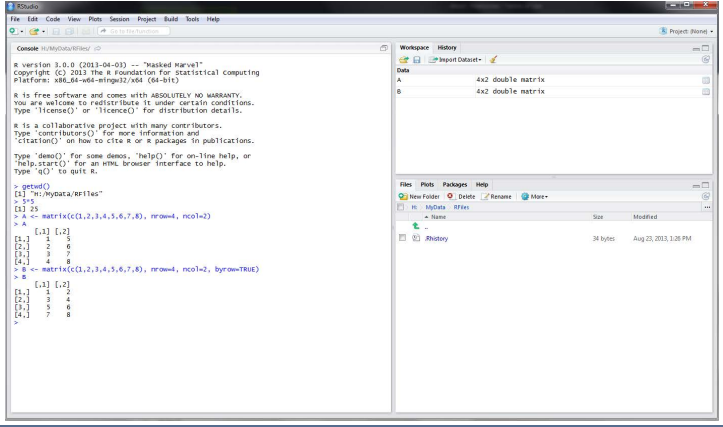
- Alternatively, Jupyter notebook can be used to write your code in R or Python.

How to switch between python and R?

- Deactivate environment if you are already in an environment.
- Then activate environment as per your coding platform.
- Start RStudio or Jupyter Notebook IDE as per you choice.

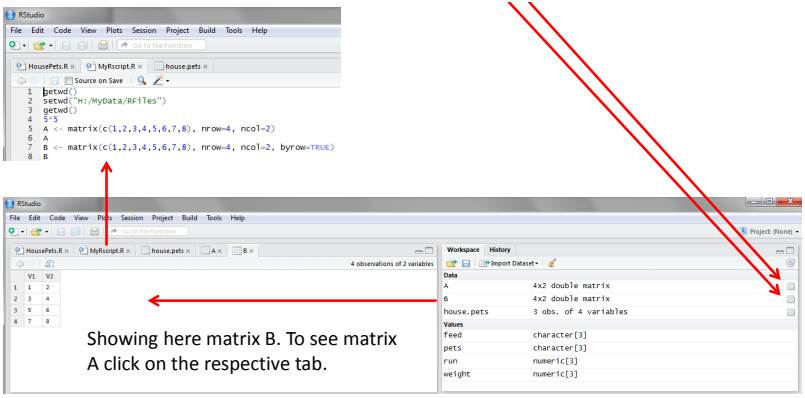
RStudio screen

- RStudio allows the user to run R in a more user-friendly environment. It is open-source (i.e. free).



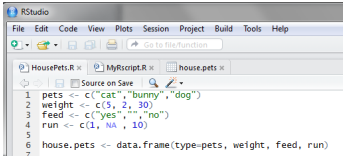
Workspace tab

- The workspace tab stores any object, value, function or anything you create during your R session. In the example below, if you click on the dotted squares you can see the data in specific window.

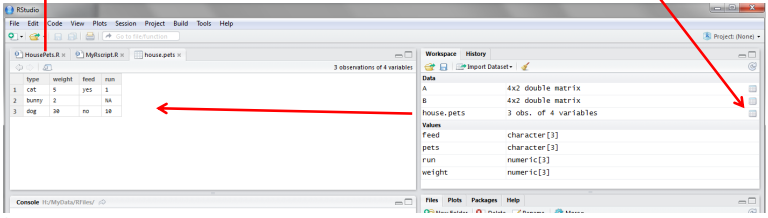


Workspace tab

- Here is another example on how the workspace looks like when more objects are added. Notice that the data frame house.pets is formed from different individual values or vectors.

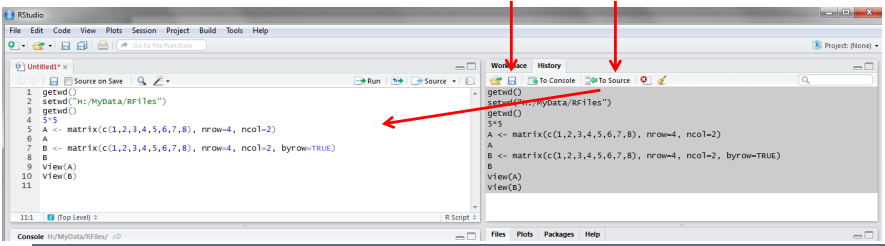


Click on the dotted square to look at the dataset in a spreadsheet form.

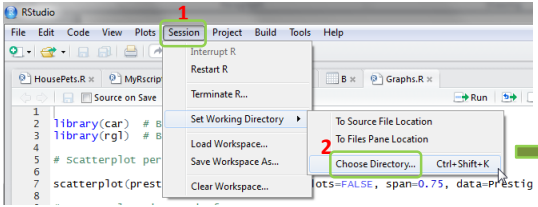


History tab

- The history tab keeps a record of all previous commands. It helps when testing and running processes. Here you can either save the whole list or you can select the commands you want and send them to an R script to keep track of your work.
- In this example, we select all and click on the "To Source" icon, a window on the left will open with the list of commands. Make sure to save the 'untitled1' file as an *.R script.

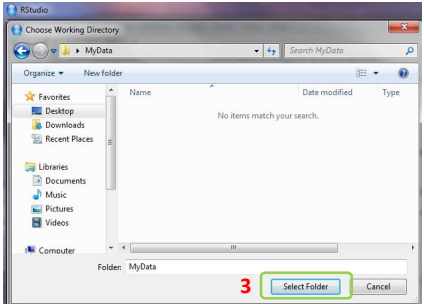


Changing the working directory

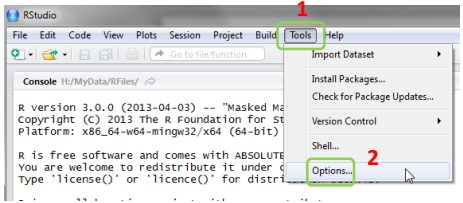


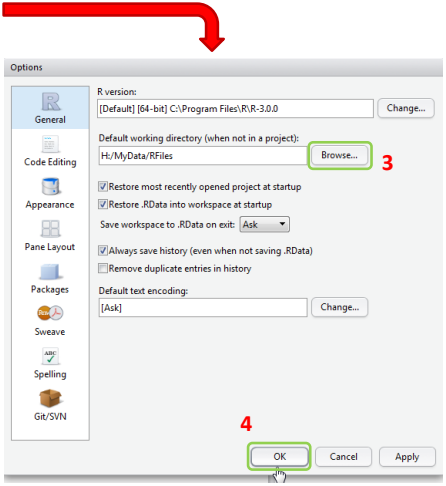
If you have different projects you can change the working directory for that session, see above. Or you can type:

```
# Shows the working directory (wd)

getwd()

# Changes the wd

setwd("C:/myfolder/data")
```



Setting a default working directory

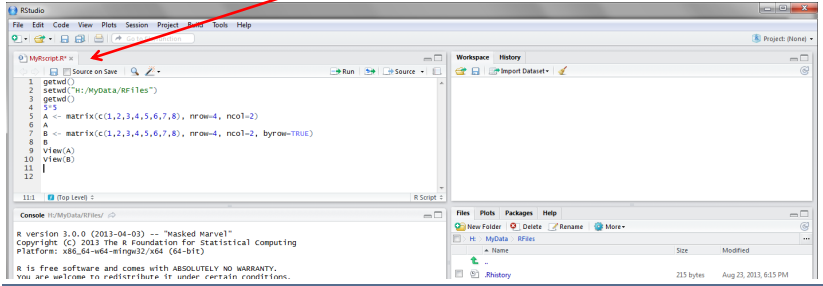




Every time you open RStudio, it goes to a default directory. You can change the default to a folder where you have your datafiles so you do not have to do it every time. In the menu go to Tools->Options

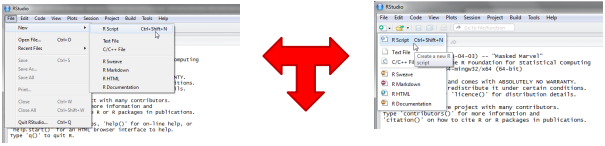
R script

- The usual RStudio screen has four windows:
  1. Console.
  2. Workspace/Environment and history.
  3. Files, plots, packages and help.
  4. The R script(s) and data view. The R script is where you keep a record of your work.



R script

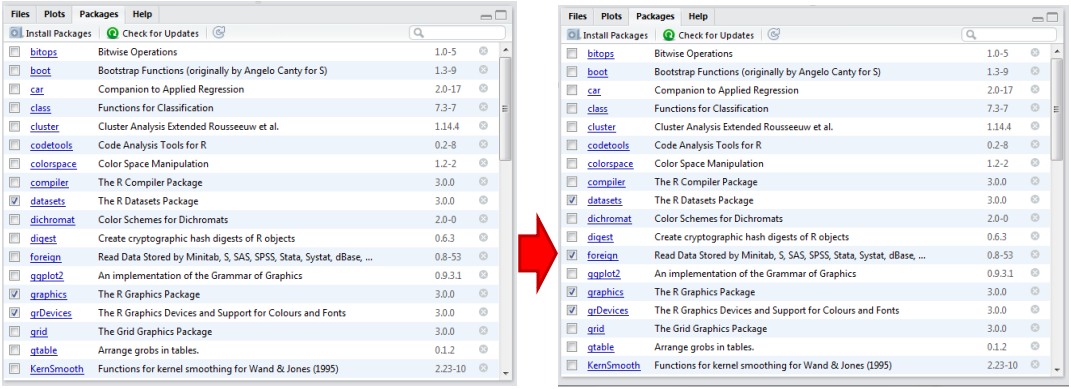
- To create a new R script you can either go to File → New → R Script, or click on the icon with the "+" sign and select "R Script", or simply press Ctrl+Shift+N. Make sure to save the script.



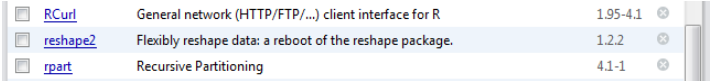
Packages tab

- The package tab shows the list of add-ons included in the installation of RStudio. If checked, the package is loaded into R, if not, any command related to that package won't work, you will need select it. You can also install other add-ons by clicking on the 'Install Packages' icon.
- Another way to activate a package is by typing, for example, `library(foreign)`. This will automatically check the –foreign package (it helps bring data from proprietary formats like Stata, SAS or SPSS).

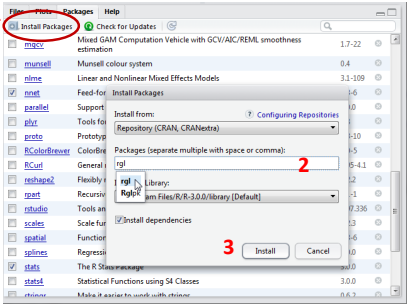
Packages tab



Installing a package

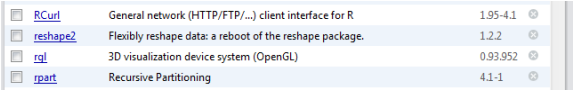


Before

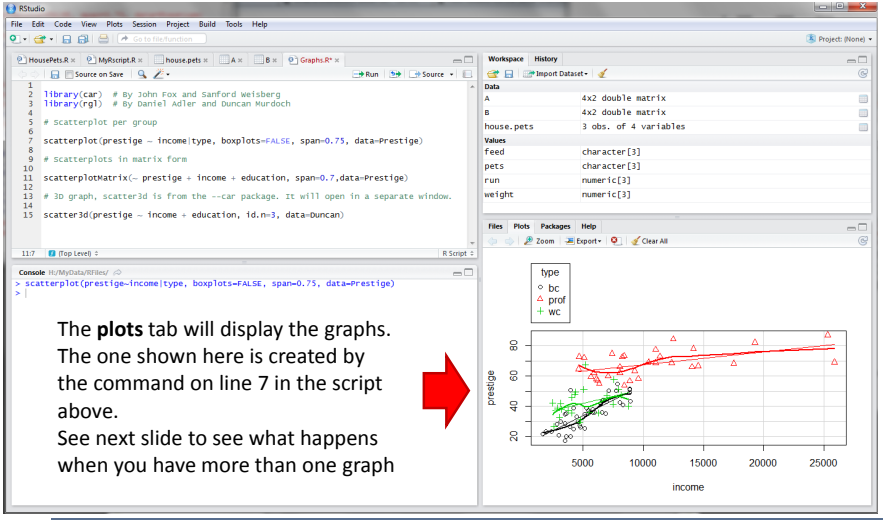


We are going to install the package – `rgl` (useful to plot 3D images). It does not come with the original R install.
Click on "`Install Packages`", write the name in the pop-up window and click on "`Install`".

After



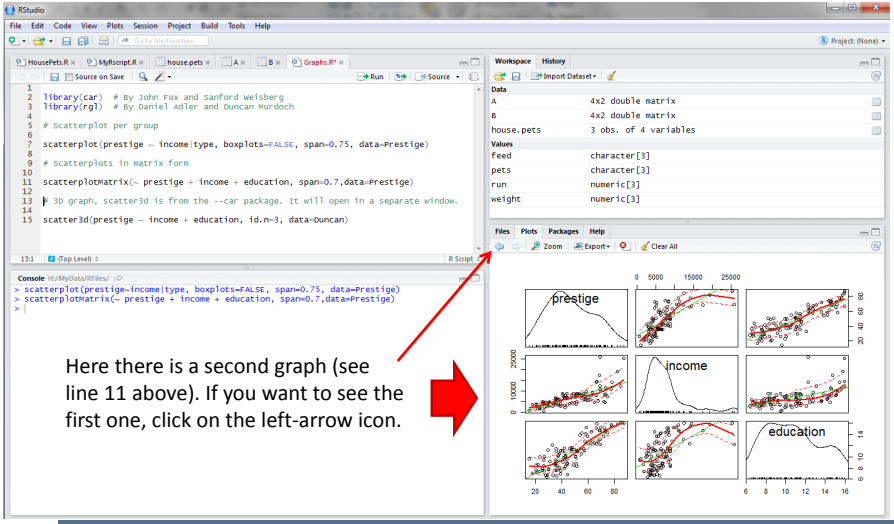
Plots tab



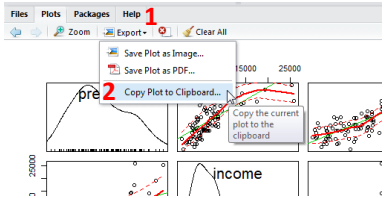
The **plots** tab will display the graphs. The one shown here is created by the command on line 7 in the script above.

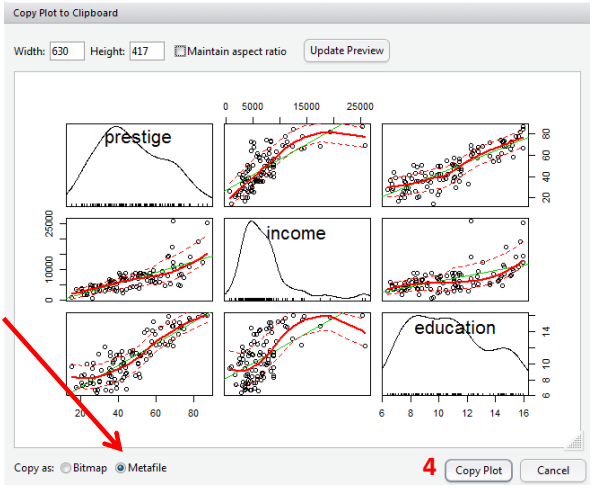
See next slide to see what happens when you have more than one graph

Plots tab


RStudio
File Edit Code View Plots Session Project Build Tools Help
HousePets.R
MyRscript.R
house.pets
Graphs.R*
Source on Save
Run
Source
1 library(car) # By John Fox and Sanford Weisberg
2 library(rgl) # By Daniel Adler and Duncan Murdoch
3
4 # Scatterplot per group
5
6 scatterplot(prestige ~ income|type, boxplots=FALSE, span=0.75, data=Prestige)
7
8 # Scatterplots in matrix form
9
10 scatterplotMatrix(~ prestige + income + education, span=0.7,data=Prestige)
11
12 # 3D graph, scatter3d is from the --car package. It will open in a separate window.
13
14 scatter3d(prestige ~ income + education, id.n=3, data=Duncan)
13:1 (Top Level)
R Script
Console R:/MyData/RFiles/
> scatterplot(prestige~income|type, boxplots=FALSE, span=0.75, data=Prestige)
> scatterplotMatrix(~ prestige + income + education, span=0.7,data=Prestige)
Workspace History
Import Dataset
Data
A 4x2 double matrix
B 4x2 double matrix
house.pets 3 obs. of 4 variables
Values
feed character[3]
pets character[3]
run numeric[3]
weight numeric[3]
Files Plots Packages Help
Zoom
Export
Clear All
prestige
income
education
0 5000 15000 25000
0 10000 25000
20 40 60 80
6 8 10 12 14 16
Here there is a second graph (see line 11 above). If you want to see the first one, click on the left-arrow icon.

Plots tab - Graphs export

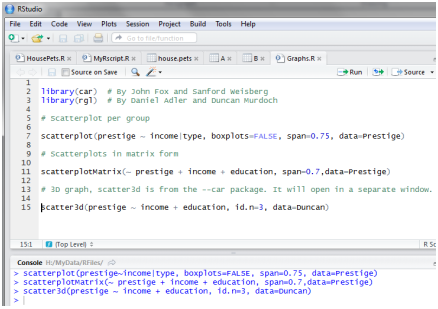




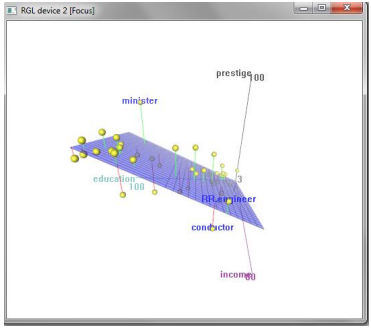
3 Make sure to select '`Metafile`'

5 Paste it into your Word document

Plots tab - 3D graphs



3D graphs will display on a separate screen (see line 15 above). You won't be able to save it, but after moving it around, once you find the angle you want, you can screenshot it and paste it to you Word document.



References

- Introduction to RStudio, https://dss.princeton.edu/training/RStudio101.pdf
- The Comprehensive R Archive Network, https://cran.r-project.org/
- RStudio Desktop, https://www.rstudio.com/products/rstudio/download/#download
- Anaconda, Where packages, notebooks, projects and environments are shared, https://www.anaconda.com/products/individual

Thank you!