4. R

Data Management
Spring & Summer 2018
OSIPP, Osaka U

Shuhei Kitamura

Goal

Get to know about R!

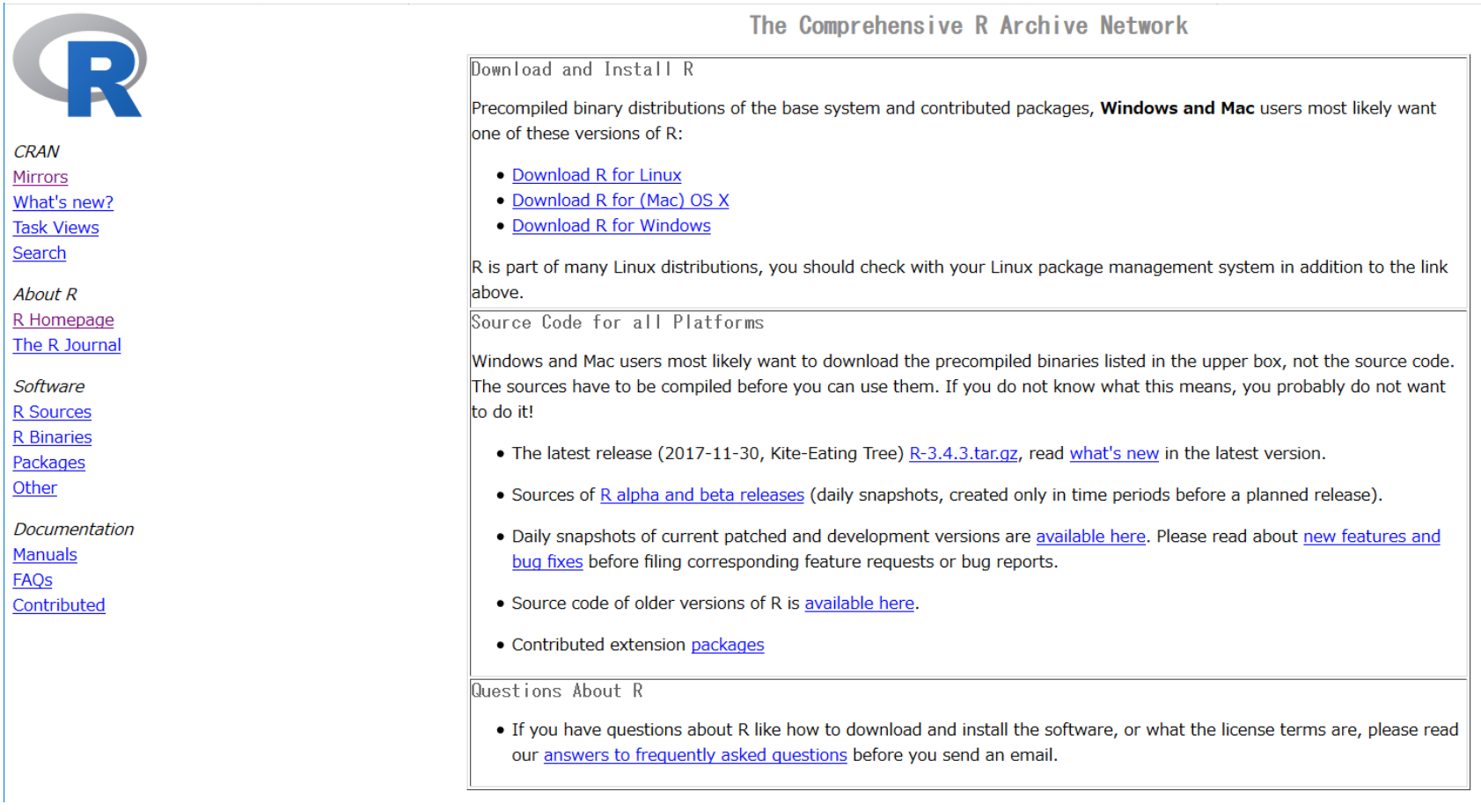
Outline

- R?
- Install R
- R in Jupyter Notebook
- Basics
- Analysis
- Application - Web scraping

What's R?

- R is an implementation of the S programming language. R was created by Ross Ihaka and Robert Gentleman (aka "R & R") at the University of Auckland, New Zealand.
- Currently developed by the R Development Core Team.
- Good for statistical computation. It's free!
- Many packages (more than 11,000 additional packages as of July 2017) available at the Comprehensive R Archive Network (CRAN) and other repositories.

Install R



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!

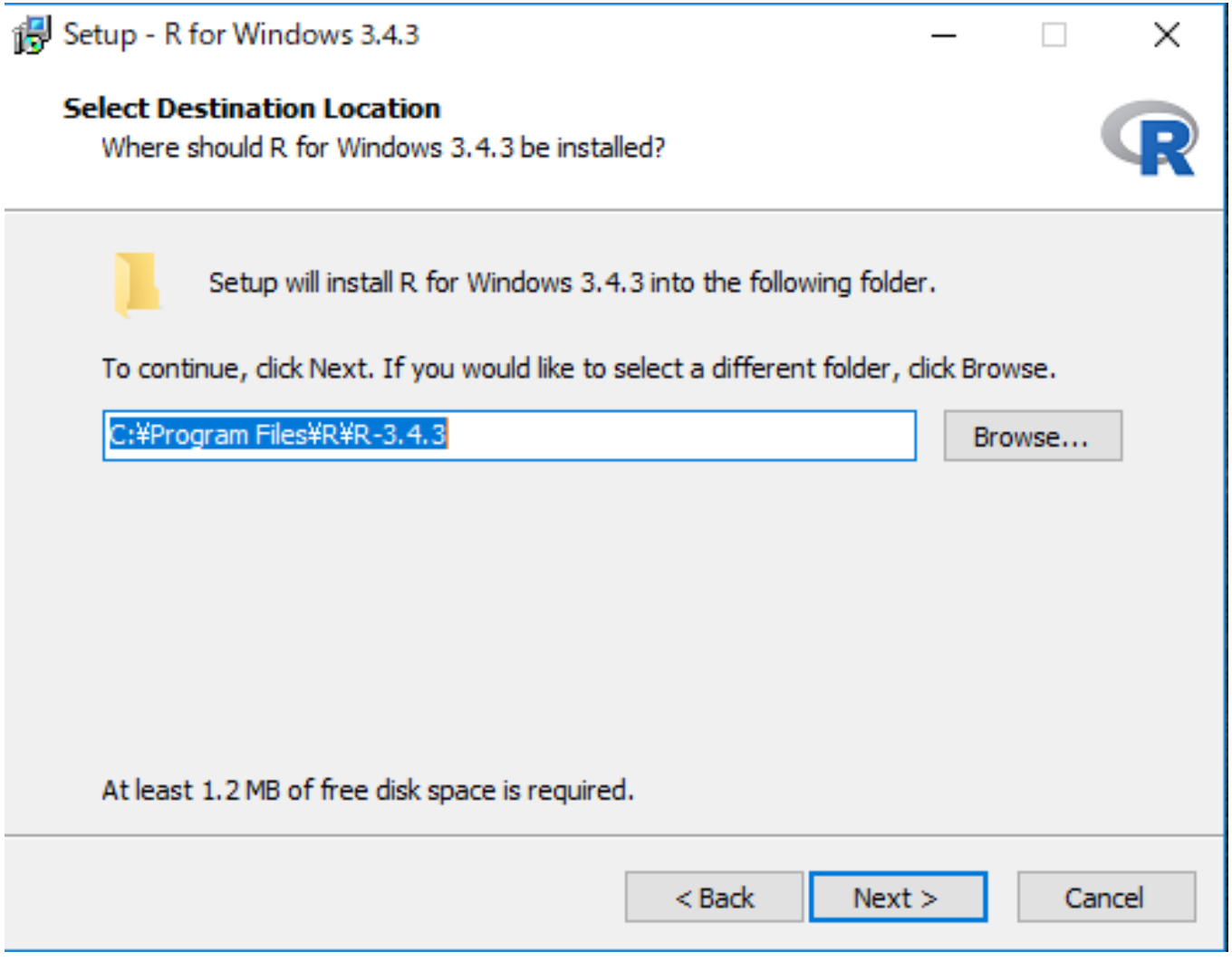
- The latest release (2017-11-30, Kite-Eating Tree) R-3.4.3.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. 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.
- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

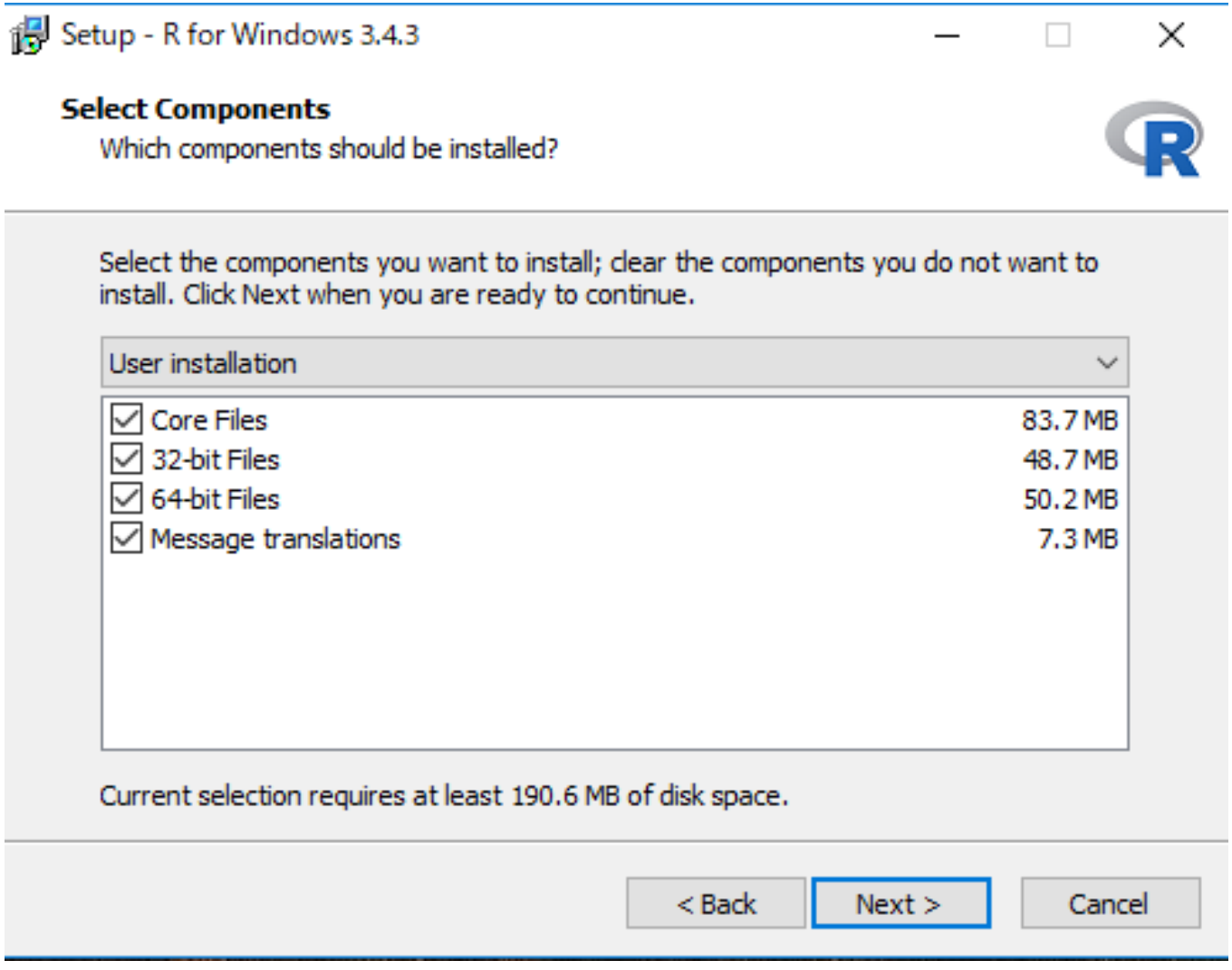
Go to CRAN. → Download & Install.

Install R (cont.)



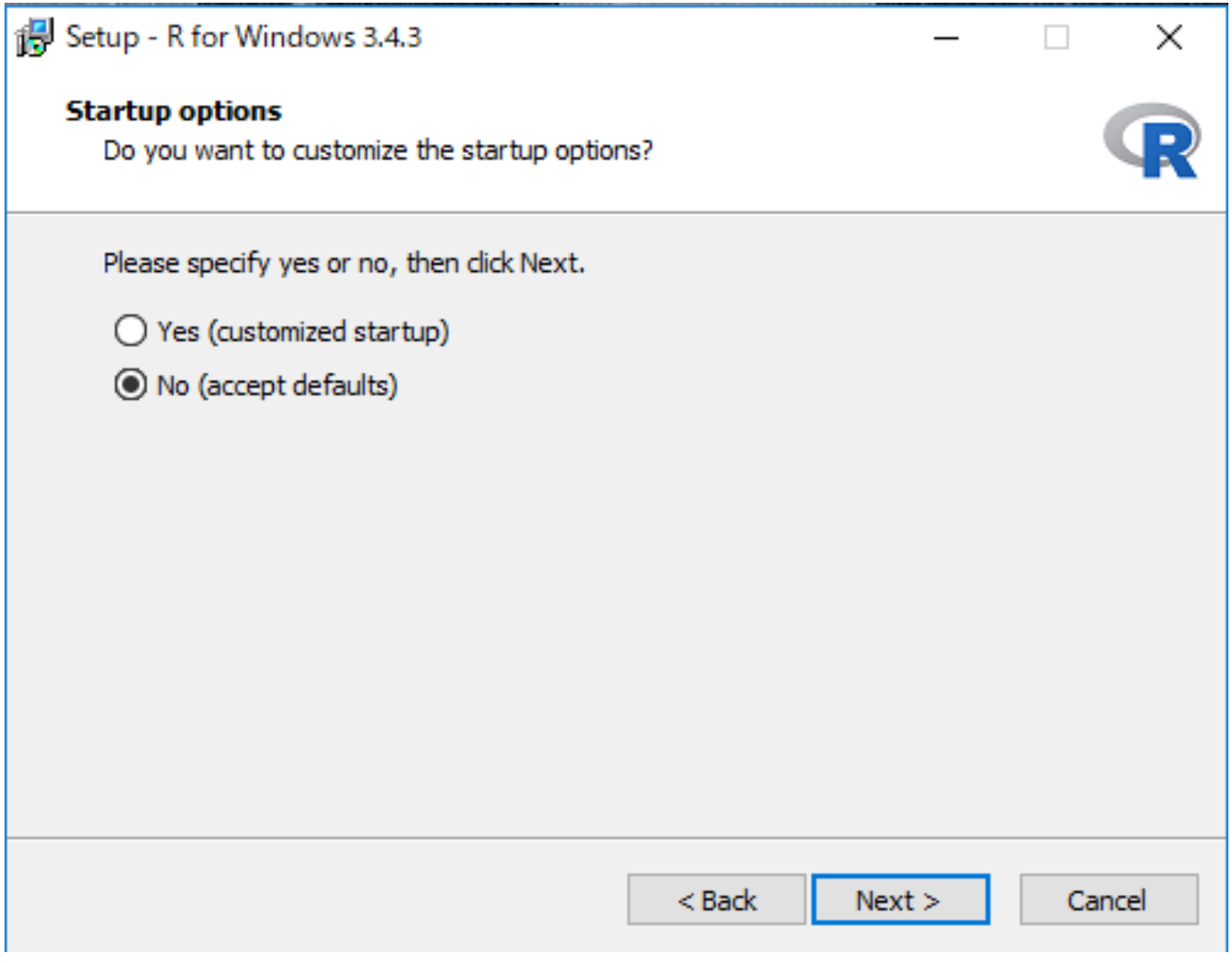
Click Next. (Please remember the location where you install R.)

Install R (cont.)



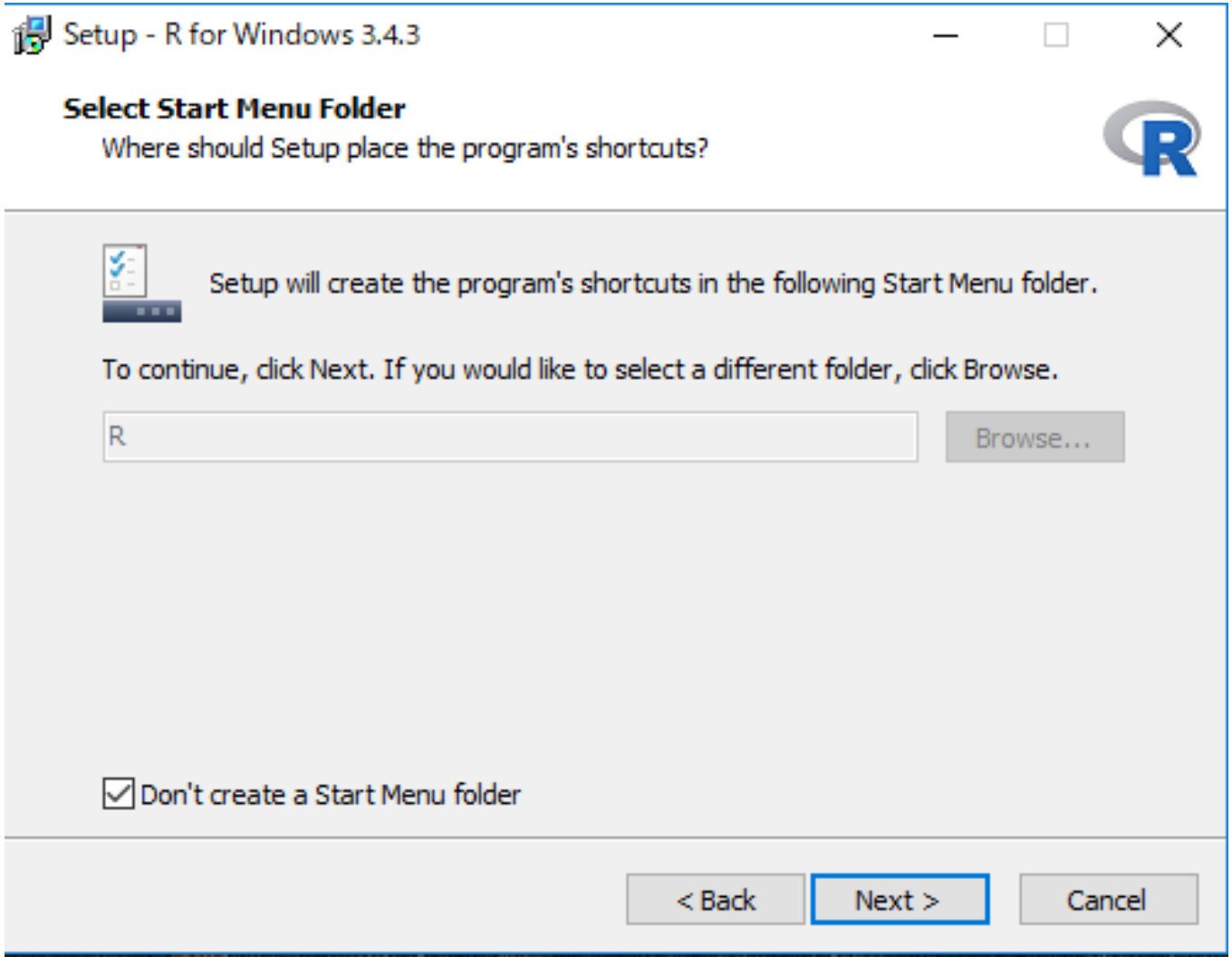
Click Next.

Install R (cont.)



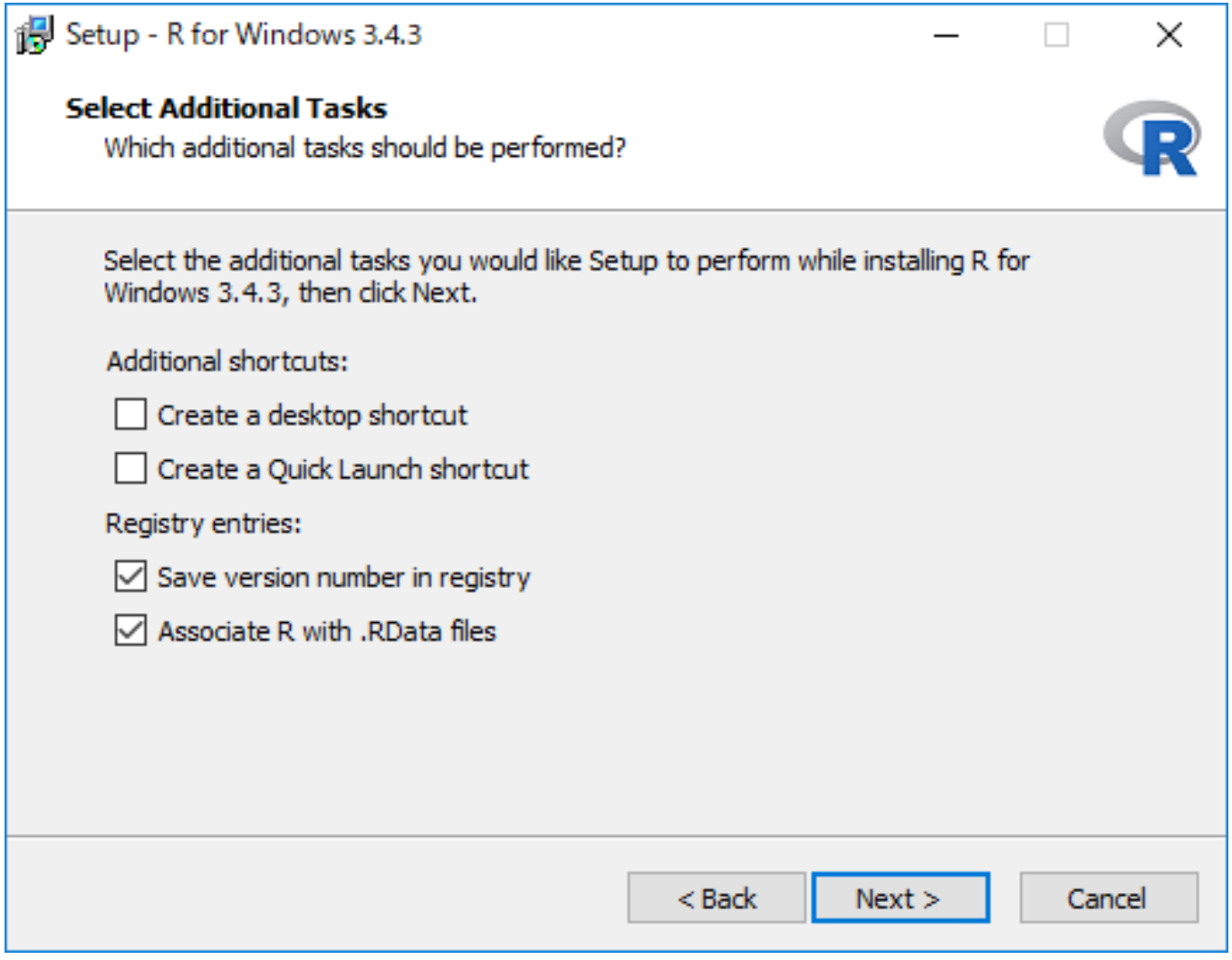
Click Next.

Install R (cont.)



Create a Start Menu folder? $\rightarrow$ No, unless you want it.

Install R (cont.)



Create a Desktop shortcut? $\rightarrow$ No, unless you want it.

Install RStudio

- RStudio is popular software among R users.
- Download an installer from RStudio Desktop and install.

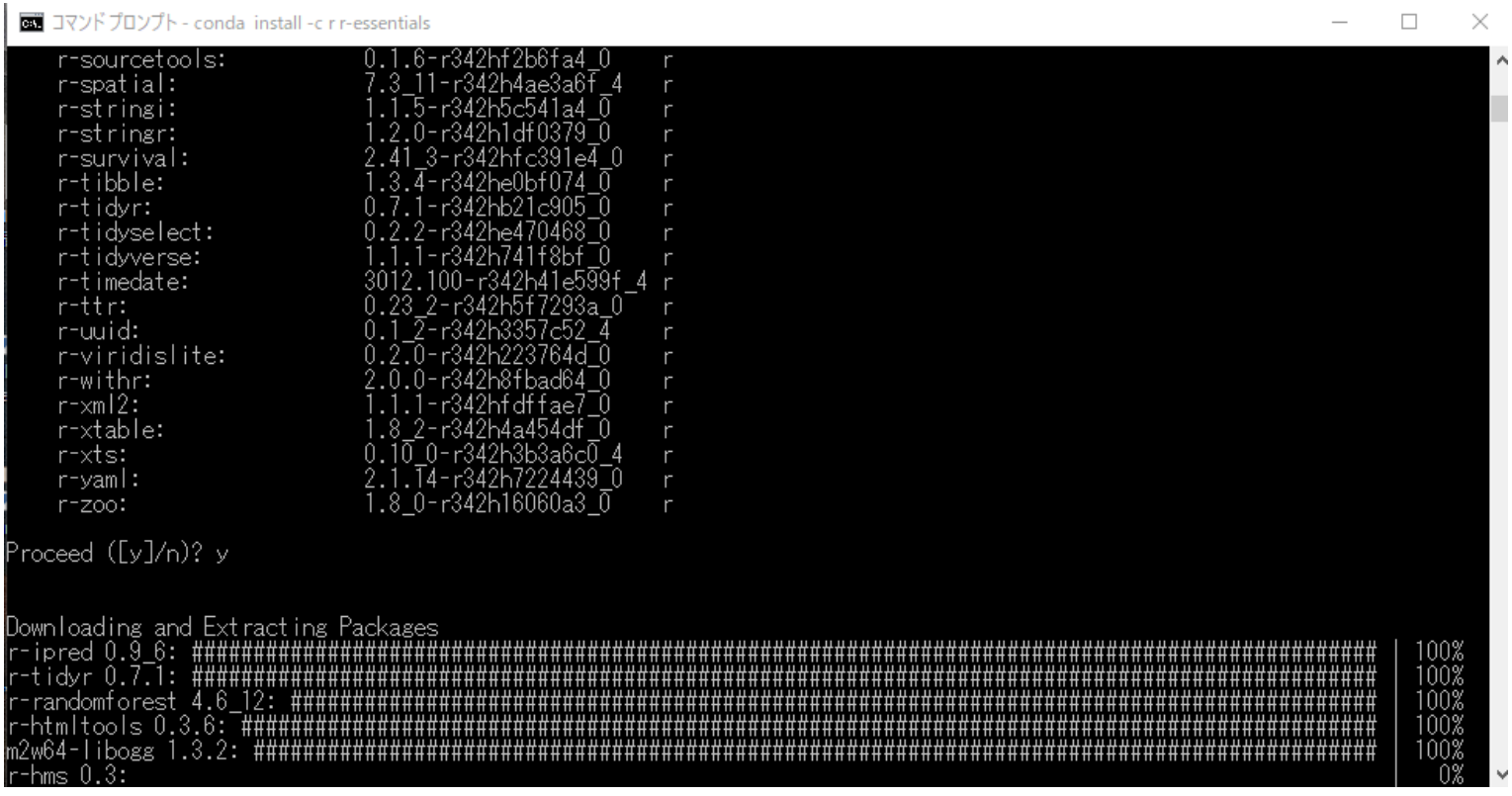
Install the IRkernel

- Let's install the IRkernel so that you can also execute R commands in Jupyter Notebook.
- Go to the folder you installed R (e.g. C:/Program Files/R/R-3.4.4)
- Go to bin/x64 (if you have installed Win 64bit version).
- Double-click Rgui.exe.
- Type/copy the following three lines one by one in command line (You can also copy the commands from install_IRkernel.txt.)

```
> install.packages(c('repr', 'IRdisplay', 'evaluate', 'crayon',
'pbdZMQ', 'devtools', 'uuid', 'digest'))
> devtools::install_github('IRkernel/IRkernel')
> IRkernel::installspec()
```

- Start Jupyter Notebook.
- Click New to see that R also appears in the list.

Install essential R packages



- Start Command Prompt (for win users). Then type:

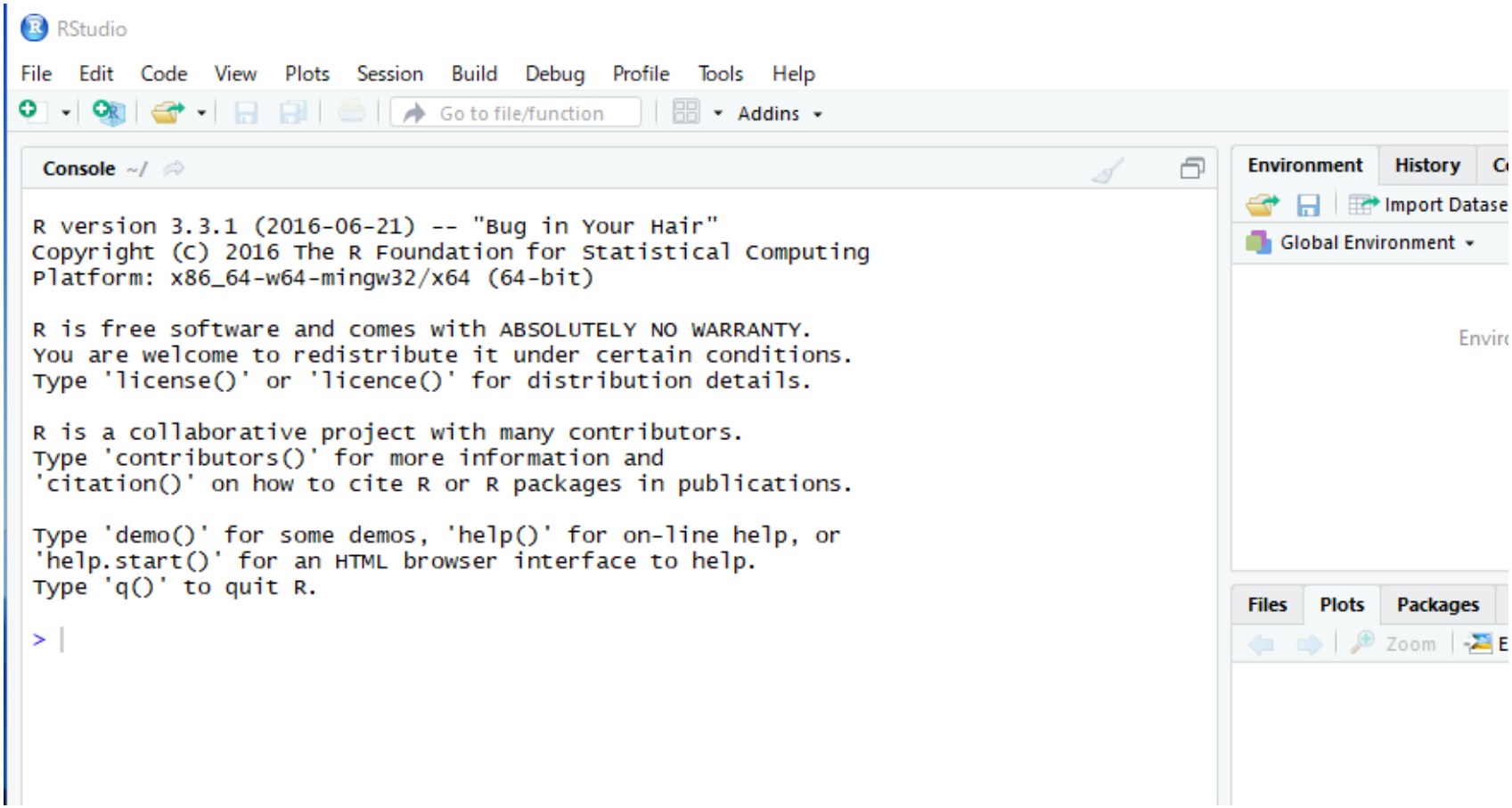
  > conda install -c r r-essentials

  (If "conda" does not work, try "pip" or "pip3".)
- When "Proceed ([y]/n)?" shows up in Command Prompt, type y (means yes), then press Enter. Installation takes a bit of time.

RStudio

- While installation is underway, let's use RStudio for a while.
- Start RStudio.

RStudio



- You type commands in console.
- Data, help files, and outputs (e.g. plot) show up to the right.

Arithmetic Operation

- Main operators are '+', '-', '*', and '/'.
- Write '1 + 2' and execute (press Enter). Next, do the same for 'print(1 + 2)'.
- If you want to write a comment, use #. Write '# print(1 + 2)' in console and execute.
- Play with '+', '-', '*', and '/'.

Arithmetic Operation (cont.)

- '^' or '**' for power (= '**' in Python). Power is right associative. How do you calculate -2^3, 3^{3^3}, and $\frac{3}{3^3}$?
- Note: '^' does not calculate power in Python.
- '%%' for modulus and '%/%' for floor division. What is the outcome of '7 %% 2'? How about '7 %/% 2'?

  - '%' (modulus in Python) and '//' (floor division in Python) do not work in R.
  - Similarly, these R operators do not work in Python.

Run R scripts in RStudio

- Click the plus sign on top. Choose "R Script" (or press Ctrl + Shift + N).
- Type:

```
print(1 + 2)
print("Hello world!")
hp <- mtcars$hp
mpg <- mtcars$mpg
plot(hp, mpg)
```

- Save the file in the local repo (file name = e.g. "myR"). The file extension becomes ".R".
- To run the script, select lines you want to execute, then press "Run" on top or press "Ctrl + Enter".

  - In case you want to run the entire code, you can also use "Ctrl + Shift + Enter" without selecting lines, which is equivalent to "Source with Echo" under "Source" on top.

Run R scripts in RStudio (cont.)

- Difference between "Run" and "Source":

  "The difference between running lines from a selection and invoking Source is that when running a selection all lines are inserted directly into the console whereas for Source the file is saved to a temporary location and then sourced into the console from there (thereby creating less clutter in the console)." (Web source)

- Difference between "Source" and "Source with Echo":

  The latter shows outputs in console while the former doesn't.