

Setting up R and Python for web scraping

Dr. Anil Shrestha

Undersecretary (Account)
Financial Administration Section
National Statistics Office

Setting up Python

Install Python 3.12.0 (64-bit)

Select Install Now to install Python with default settings, or choose Customize to enable or disable features.



Install Now

C:\Users\PC\AppData\Local\Programs\Python\Python312

Includes IDLE, pip and documentation
Creates shortcuts and file associations



Customize installation

Choose location and features



☒ Use admin privileges when installing py.exe



☒ Add python.exe to PATH

Cancel

python
for
windows



Install Python 3.12.0 (64-bit)

Select Install Now to install Python with default settings, or choose Customize to enable or disable features.



Install Now

C:\Users\PC\AppData\Local\Programs\Python\Python312

Includes IDLE, pip and documentation
Creates shortcuts and file associations



Customize installation

Choose location and features

☒ Use admin privileges when installing py.exe

☒ Add python.exe to PATH

Cancel

python
for
windows



python
for
windows

Optional Features

☒ Documentation

Installs the Python documentation files.

☒ pip

Installs pip, which can download and install other Python packages.

☒ tcl/tk and IDLE

Installs tkinter and the IDLE development environment.

☒ Python test suite

Installs the standard library test suite.

☒ py launcher ☒ for all users (requires admin privileges)

Installs the global 'py' launcher to make it easier to start Python.

Back

Next

Cancel

Advanced Options


- ☒ Install Python 3.12 for all users
- ☒ Associate files with Python (requires the 'py' launcher)
- ☒ Create shortcuts for installed applications
- ☒ Add Python to environment variables
- ☒ Precompile standard library
- ☐ Download debugging symbols
- ☐ Download debug binaries (requires VS 2017 or later)

Customize install location

C:\Program Files\Python312

Browse

Back

 Install

Cancel

python
for
windows

Setup was successful

New to Python? Start with the [online tutorial](#) and [documentation](#). At your terminal, type "py" to launch Python, or search for Python in your Start menu.

See [what's new](#) in this release, or find more info about [using Python on Windows](#).



Disable path length limit

Changes your machine configuration to allow programs, including Python, to bypass the 260 character "MAX_PATH" limitation.

python
for
windows

Close

Installing necessary packages

- Press “WIN+R” => type “cmd” and press “enter”.
- Enter the following command to install packages.

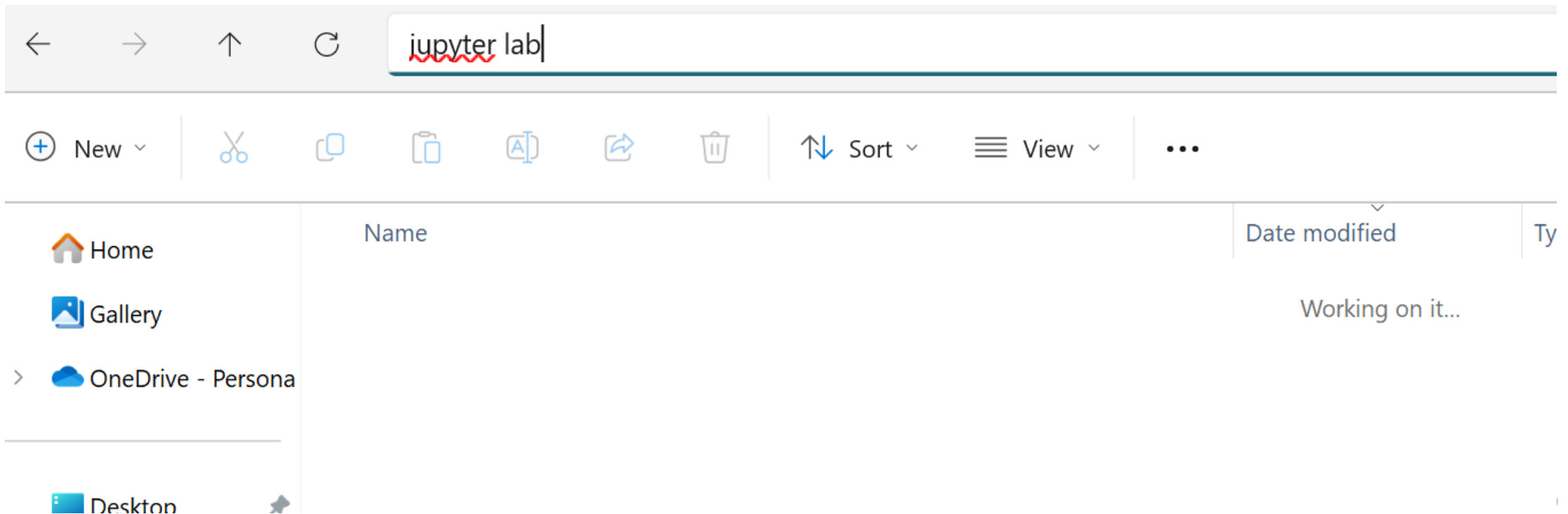
pip install <package_name>

Required packages

***pandas, bs4, requests, jupyterlab, selenium,
webdriver-manager***

Let's check Python installation

- Create a new folder and rename it.
- Open the folder.
- Enter ***“jupyter lab”*** in the address bar.



+

+

⬆

↻


Filter files by name 🔍


📁 /

Name ▲	Last Modified
📄 rclone.exe	4 months ago
• 📄 Untitled.ipyn...	6 hours ago
📄 Untitled1.ip...	6 hours ago


Launcher +


Notebook


Python 3
(ipykernel)



R


▸_ Console



Python 3
(ipykernel)



R


\$_ Other




Terminal



Text File


Markdown File


Python File


R File

Simple ☐ 0  1 

Launcher 1 

The screenshot displays a Jupyter Notebook environment. On the left, a file explorer sidebar shows a directory with files: 'Session 1.p...', 'Session 2.p...', 'Session 3-p...' (selected), 'Session 3-R...', 'Session 3.p...', and 'temp.docx'. The main area shows two code cells. The first cell, labeled '[1]:', contains Python code for setting up a Selenium Chrome driver. The second cell, labeled '[3]:', contains Python code for setting up a Selenium Firefox driver. Both cells use Selenium WebDriver and WebDriverManager to manage the browser services.

```
[1]: from selenium import webdriver
      from selenium.webdriver.chrome.service import Service as ChromeService
      from webdriver_manager.chrome import ChromeDriverManager

      driver = webdriver.Chrome(service=ChromeService(ChromeDriverManager().install()))
      driver.get('https://nsonepal.gov.np/')


[3]: from selenium import webdriver
      from selenium.webdriver.firefox.service import Service as FirefoxService
      from webdriver_manager.firefox import GeckoDriverManager

      driver = webdriver.Firefox(service=FirefoxService(GeckoDriverManager().install()))
      driver.get('https://nsonepal.gov.np/')
```

- We can use either **Chrome** or **Firefox** browser for web scraping.
- We will use **Firefox** and keep the code of the 2nd cell

Setting up R

Select Setup Language



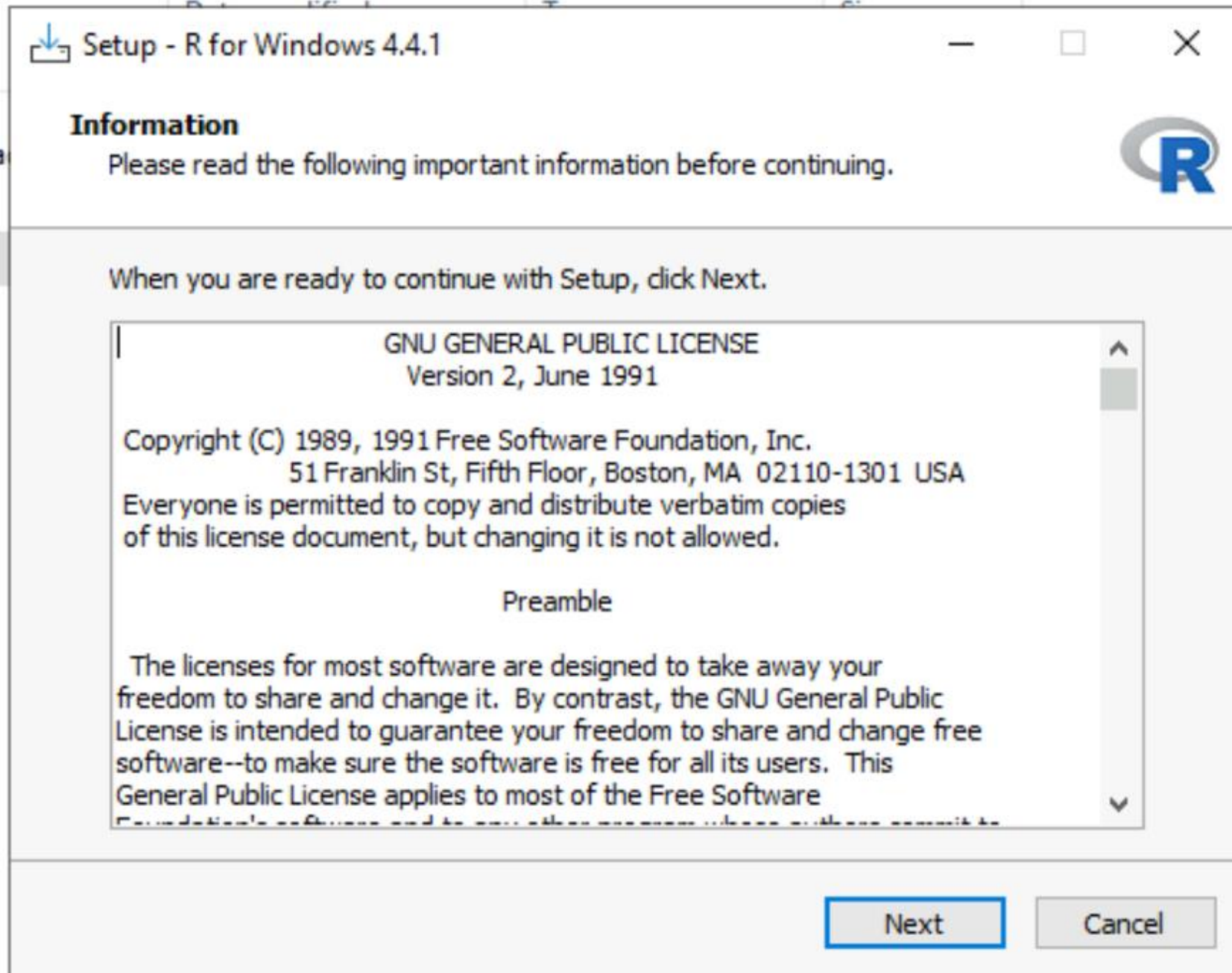
Select the language to use during the installation.

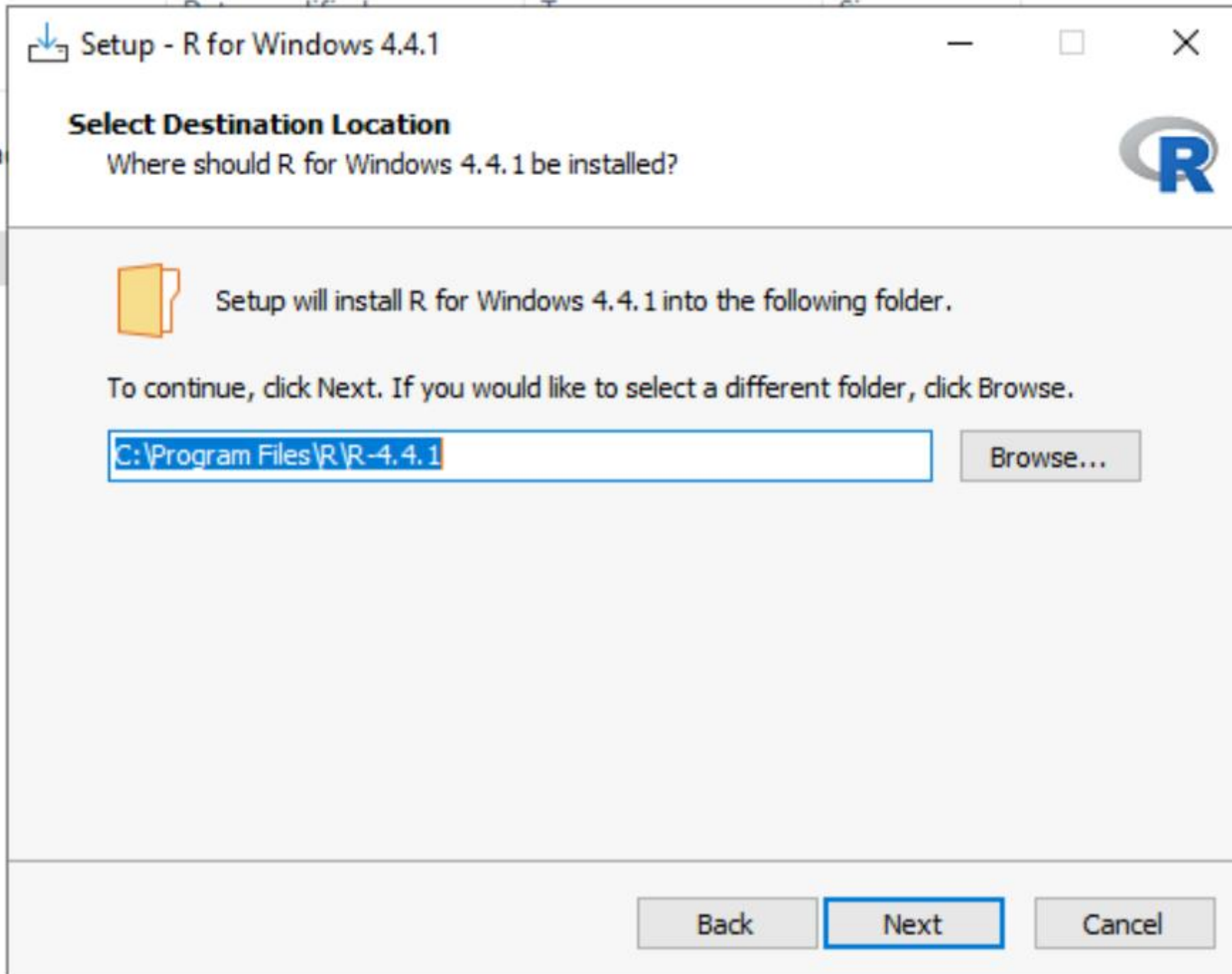
English

▼

OK

Cancel





Setup - R for Windows 4.4.1

Select Components

Which components should be installed?

Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

User installation

<input checked="" type="checkbox"/> Main Files	92.7 MB
<input checked="" type="checkbox"/> 64-bit Files	73.4 MB
<input checked="" type="checkbox"/> Message translations	10.2 MB

Current selection requires at least 179.2 MB of disk space.

Back Next Cancel

Setup - R for Windows 4.4.1

Startup options

Do you want to customize the startup options?

Please specify yes or no, then click Next.

☐ Yes (customize startup)

☒ No (accept defaults)

Back Next Cancel

Setup - R for Windows 4.4.1

Setup will create the program's shortcuts in the following Start Menu folder.

To continue, click Next. If you would like to select a different folder, click Browse.

Browse...

☐

Don't create a Start Menu folder

Back

Next

Cancel


6/19/2024 9:05

18

Setup - R for Windows 4.4.1

Select Additional Tasks

Which additional tasks should be performed?



Select the additional tasks you would like Setup to perform while installing R for Windows 4.4.1, then click Next.

Additional shortcuts:

- ☒ Create a desktop shortcut
- ☐ Create a Quick Launch shortcut

Registry entries:

- ☒ Save version number in registry
- ☒ Associate R with .RData files

Back Next Cancel



Completing the R for Windows 4.4.1 Setup Wizard

Setup has finished installing R for Windows 4.4.1 on your computer. The application may be launched by selecting the installed shortcuts.

Click Finish to exit Setup.

Finish

Setting up R for Jupyter Lab

- Press “**WIN**” key and search **Rgui** app and execute.
- Install “**IRkernel**” package and configure for jupyterlab.

```
> install.packages('IRkernel')  
--- Please select a CRAN mirror for use in this session ---  
also installing the dependencies 'fastmap', 'rlang', 'cli', 'fansi', 'glue', 'l$
```

```
> IRkernel::installspec()  
> |
```

- Now, open **jupyter lab** following the same process as we have done in python setup.

Filter files by name

/

Name	Last Modified
rclone.exe	4 months ago
• Untitled.ipyn...	6 hours ago
Untitled1.ip...	6 hours ago

Launcher

Notebook

Python 3 (ipykernel)

R

Console

Python 3 (ipykernel)

R

Other

Terminal

Text File

Markdown File

Python File

R File

Setting up RSelenium

- Install *java* using *jre-8u411-windows-i586.exe* file.
- Install *firefox* using *Firefox Setup 127.0.exe* file.
- Install the following packages:

```
install.packages(c('RSelenium', 'rvest', 'netstat', 'httr', 'dplyr'))
```

- Run the following code in the jupyter notebook.

```
library(RSelenium)
```


```
library(netstat)
```






```
library(httr)
```


```
rD <- rsDriver(browser = "firefox", port = free_port())
```







```
remDr <- rD$client
```




```
remDr$navigate("https://www.google.com")
```











 File Edit View Run Kernel Tabs Settings Help

 /

Name	Last Modified
 Session 1.p...	22 hours ago
 Session 2.p...	21 hours ago
•  Session 3-p...	3 minutes ago
•  Session 3-R...	now
 Session 3.p...	2 minutes ago
 temp.docx	yesterday

Launcher  Session 3-R.ipynb  


         Code 

```
[ ]: install.packages(c('RSelenium','rvest','netstat','httr','dplyr'))





[ ]: library(RSelenium)
library(netstat)
library(httr)
rD <- rsDriver(browser = "firefox", port = free_port())
remDr <- rD$client
remDr$navigate("https://www.google.com")
```


Setting up RSelenium


- An error will occur due to a bug in recent *RSelenium* version. To solve this bug, go to C:\Users\<username>\AppData\Local\binman\binman_chromedriver\, search for all LICENSE.chromedriver file and delete it.
- Re-run the code in the jupyter notebook again.







File Edit View Run Kernel Tabs Settings Help













Filter files by name 

 /

Name	Modified
 jre-8u411-windo...	18m ago
 R.ipynb	now

R.ipynb  

         Code 

```
[ ]: install.packages(c('RSelenium', 'rvest', 'netstat', 'httr', 'dplyr'))

[ ]: library(RSelenium)
library(netstat)
library(httr)
rD <- rsDriver(browser = "firefox", port = free_port())
remDr <- rD$client
remDr$navigate("https://www.google.com")
```


Thank you