Table of Contents

[Overall Information 1](#_Toc129705079)

[**Required Python Packages (520)** 1](#_Toc129705080)

[**Required Python Packages (540)** 2](#_Toc129705081)

[**Required R Packages (540)** 2](#_Toc129705082)

[Using RStudio as IDE 2](#_Toc129705083)

[**Install R** 2](#_Toc129705084)

[**Install Rstudio** 2](#_Toc129705085)

[**Install Python (suggestion miniconda or connect to your python)** 2](#_Toc129705086)

[Using the High-Performance Computing Lab from HU 3](#_Toc129705087)

[**Fill out an IT Ticket** 3](#_Toc129705088)

[**How to Install Packages** 4](#_Toc129705089)

[Using Google Colab (520 Recommendation Only) 5](#_Toc129705090)

[**Go to your Google Account** 5](#_Toc129705091)

[**Install Packages** 8](#_Toc129705092)

[**Get More Help** 8](#_Toc129705093)

## **Overall Information**

* Please note: you only need to use one of the installation options.
* You can use any set up you would like, but you should ensure you have a working set up.
* Other options include using Anaconda to run Jupyter notebooks, Spyder, or PyCharm (Professional version, which isn’t free, see below) from Jetbrains. This guide just covers a few options. For these options, you can simply install the software.
* The guides below indicate how to set up an IDE or account with Python running. Each option will have different rules for installation of packages, and you should learn how to install them.
* You will **need to be able to make reports** – usually this is done through markdown or jupyter notebooks. You should be able to run code and write text within the file and compile the file into a report (html preferred).
* To work with tesseract, you will need to install the software (depends on operating system).
  + You may not need this software, but if you want to do any OCR, it is free to use.
  + <https://tesseract-ocr.github.io/tessdoc/Installation.html>

## **If you have spaces in your username**

* You will have issues using miniconda and Rstudio.
* You can solve some of them by:
  + <https://stackoverflow.com/questions/74151654/error-cannot-install-miniconda-with-rstudio>
  + Set your Python for every time you use Rstudio:
    - py\_config()
    - Copy down the path for the miniconda install.
    - usethis::edit\_r\_environ()
    - RETICULATE\_PYTHON=”PATH”
    - Paste the path into PATH.
  + You will then need to install the spacy language model by clicking on Terminal and typing (not using spacyr)
    - python -m spacy download en\_core\_web\_sm

### **Required Python Packages (520)**

* Python 3+ required.
* Packages to install:
  + bs4, datetime, requests, pytesseract, tesseract, nltk, spacy, textacy, gensim, sumy, scikit-learn, fasttext, lime, pandas, keras, tensorflow, network, eli5, pyspellchecker, pysrt, PyPDF2
  + I recommend doing these one at a time to see what errors you get.
* Install spacy language module: en\_core\_web\_sm.
  + This installation will depend on your choice of IDE/Computer. See sections below.
* Things I think come with base python or base miniconda:
  + urllib, json, PIL, string, sqlite3, re, html, itertools
  + Make sure you can import these.

### **Required Python Packages (540)**

### **Required R Packages (540)**

## **Using RStudio as IDE**

### **Install R**

<https://cran.r-project.org/>

### **Install Rstudio**

<https://posit.co/download/rstudio-desktop/>

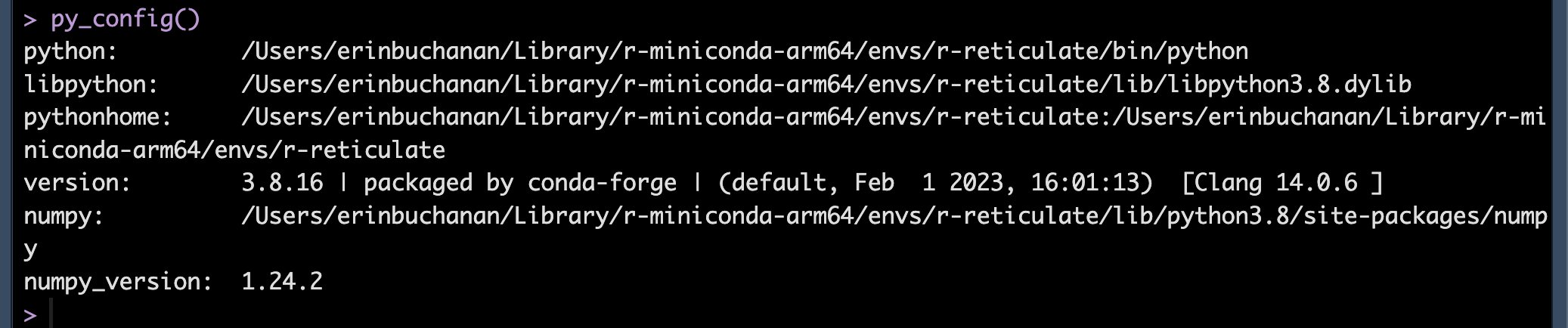
* MAC: You may have to install command line tools.
* Windows: You may need to install Rtools.
* Windows: You will need to install Visual C++ <https://visualstudio.microsoft.com/visual-cpp-build-tools/>

Graphical user interface, application

Description automatically generated

### **Install Python (suggestion miniconda or connect to your python)**

* Install R-package **reticulate**
* Load the reticulate library.
* Run **install\_miniconda()**
* Check your version of python by running **py\_config()**
* Make sure it says there’s a numpy installed!



**Install Python Packages**

* In the ***R* console**, use py\_install(“PACKAGE”, pip = T)
* Replace package with the name of the package. For example py\_install(“scikit-learn”, pip = T) will install scikit-learn.
* If you have a problem with a package using pip, try setting pip to false.
* Installing spacy language model in the right place:
  + install.packages("spacyr")
  + library(spacyr)
  + spacy\_download\_langmodel(

model = "en\_core\_web\_sm",

envname = "r-reticulate",

conda = "auto")

**Known Issues**

* Mac: tensorflow – install without pip.

## **Using the High-Performance Computing Lab from HU**

### **Fill out an IT Ticket**

* https://ithelp.harrisburgu.edu/
* Click login.
* Create a new support ticket.
* Be sure to click the HPC option on the support ticket.
* The subject should be Datalore.
* Ask for access to the HPC-GPU for ANLY 520 with Dr. Buchanan.
* <https://datalore.harrisburgu.cloud/>

Graphical user interface, application, email, website

Description automatically generated

### **How to Install Packages**

* You will want to install packages based on the notebook you are using. Each notebook is new environment (I believe).
* Create a new notebook.

Graphical user interface, text, application, chat or text message

Description automatically generated

* Choose Python.
* Different sections will require different power. I will share notebooks with the class, so you can know which versions are necessary. Start by creating a small CPU to test the installation of some packages.
* Click the environment tab on the top left side.

Graphical user interface, text, application, email

Description automatically generated

* Search for an install a few packages (spacy … be sure to use 3.5+ not dev 4.0, bs4).
* Restart the kernel!
* Install spacy language model.
  + In a chunk type:

**import** spacy

*#%%*

**import** subprocess

*#%%*

print(subprocess.getoutput(**"python -m spacy download en\_core\_web\_sm"**))

* You should see a success message at the end. You need to put this code at the top of every notebook that includes spacy. Only run it ONCE per session, but when you restart the notebook, it will not be there again.

Text

Description automatically generated with medium confidence

* Note that this system means that **each notebook is a different environment** – you will have to install all the packages you need for that notebook to run each time. The packages per notebook do stay installed.

## **Using Google Colab (520 Recommendation Only)**

### **Go to your Google Account**

* Click on New > More > Connect More Apps

Graphical user interface, application

Description automatically generated

* Search for Colaboratory in the apps (note it’s one L)

Graphical user interface, application, website

Description automatically generated

* Install the Colaboratory app.

Graphical user interface, text, application, email

Description automatically generated

* You can then create Google Colab files, which are Jupyter notebooks.

Graphical user interface, application

Description automatically generated

### **Install Packages**

* Type in a chunk: !pip install PACKAGENAME
* !pip install advertools
* Be sure to use the ! point.

### **Get More Help**

* <https://colab.research.google.com/notebooks/welcome.ipynb>