Setting up Python and reticulate locally and on HPCC

Tim B

This note describes how to set up Python and reticulate. These steps should get Python and reticulate working both locally and on Wharton's HPCC grid. If on HPCC, be sure to include the following in your .bashrc (and exit and log back in after updating your .bashrc).

```
module load python/python-3.10.4
module load R/R-4.1.2
module load gcc/gcc-11.3.0
```

If working locally, ensure that your R and Python versions are up-to-date. For convenience, my local .bashrc aliases "python" to "python3" as follows:

```
alias python=python3
```

where "python3" is the standard Python 3 executable (located at

`/Library/Frameworks/Python.framework/Versions/3.10/bin/python3`;

check via "which python3").

First, create a project-specific Python virtual environment. I will create a virtual environment called "~/py/lowmoi-venv."

```
cd
mkdir py
python -m venv py/lowmoi-venv
```

Next, activate the virtual environment.

```
source py/lowmoi-venv/bin/activate
```

Upgrade the base installer packages to latest.

```
python -m pip install -U pip
python -m pip install -U setuptools wheel
```

Next, download and install project-relevant Python packages. I will install `numpy`, `statsmodels`, `scikit-learn`, and `scipy`.

IMPORTANT NOTE: The up-to-date version of `numpy` (i.e., 1.22.4) does not appear to work on HPCC with the current versions of Python, R, and reticulate (3.10.4, 4.1.2, and 1.25, respectively). Instead, install `numpy` version 1.21.5, which is the latest version of `numpy` that works on HPCC (as far as I can tell).

```
python -m pip install numpy==1.21.5
python -m pip install scipy
python -m pip install statsmodels
python -m pip install scikit-learn
```

Open R and install `reticulate`.

```
R
> install.packages("reticulate")
```

Point `reticulate` to your Python virtual environment. One way to do this is to add the line `Sys.setenv(RETICULATE_PYTHON = "/Users/timbarry/py/lowmoi-venv/bin/python")` to your .Rprofile (replacing the file path here with the file path to the Python executable in your virtual environment).

Finally, check that reticulate works, both locally and on HPCC. (According to Hugh, the best practice is to run reticulate while the Python virtual environment is active, although I am not sure this is strictly speaking necessary.)

```
R # launch R
> library(reticulate)
> py_config() # verify correctness of configuration
> np <- import("numpy") # import numpy
> np$zeros(10L) # call a numpy function
```

If the above lines of code execute without issue, then you should be good to go.