

Python intro

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Friday 1st April, 2022 14:55

outline

A simple stick figure is shown in mid-air, with its arms and legs outstretched as if it is flying or falling. The figure is positioned in the upper right quadrant of the frame.

PYTHON!

YOU'RE FLYING!
HOW?



py is fun! and can accomplish so much more!

- why com sci while we're already busy with soc sci?
- so that you can do more soc sci and stop wasting time!
- don't say "no thanks! we are too busy"



No thanks!

We are
too busy

references/tutorials

- we'll skip a lot of stuff, for more see:
- com sci, quick, readable! <http://code.google.com/edu/languages/google-python-class/>
- looks fun <https://automatetheboringstuff.com/#toc>
- eg email/text <https://automatetheboringstuff.com/chapter16/>
- manipulate images <https://automatetheboringstuff.com/chapter17/>
- general, complete, lengthy
<http://www.diveintopython3.net/>
- for soc sci, comprehensive
<http://nealcaren.github.io/python-tutorials/>

strategy/approach

- I won't bother you with dictionaries, tuples, etc
- I will introduce Python as for
 - social data science or computational social science (CSS)
- don't have to do everything in Python
 - stata's (still) probably better for general dat man
- fine to skip most of the stuff
- but do pick up things that will be useful to you
- i am sure there will be some—Py can do just about anything!

but why another software for dat man?

- it's 2022!
- we have internet, social media, smartphones
- vast majority of data there
- cannot have data man class just doing spreadsheets
- and just numbers
- need to process text, make geo maps, etc
- stat software like Stata is behind

why not?

- although most user friendly general language,
 - still more difficult than Stata
- with Py, can do almost everything just in one lang
 - but need to type more
 - and much traditional soc sci data/stats functionality not (easily) available
- likewise, most of Py is for com sci, but quite a lot for soc sci!

Python for soc sci

- automating os, like BASH scripts (os, system)
- data management (pandas)
- text processing (re, nltk)
- statistics (pandas, scipy, statsmodels, scikit-learn)
- API (urllib, rauth)
- gis (pysal)
- web scrapping (beautifulsoup, scrapy, html5lib)
- beautiful graphs, I'd say THE BEST (pandas, matplotlib)
- and much more !

feel

- very clean
- transparent, easy to figure out what is going on
- almost like human language
- I find it much more user friendly than R

where is it?

- can install on your PC: free and cross-platform
 - <https://www.anaconda.com/download>
- we will just run online on goog colab notebook
- can even run from within stata
 - but need stata ≥ 16 , eg <https://apps.camden.rutgers.edu/novnc>
 - but may be difficult to install additional packages and may be clunky in general

py from within stata

- either say in command win “python” and it turns into Py
- or in dofile, eg:

```
python
print("hi I'm python")
end
```

- there are basic packages but installation of additional ones may be tricky on public machine
- but if doing on your machine, just install conda,
- and install packages with pip (like on colab)
- and then can import these packages within stata
- tho I will only show in stata basic stuff like re

indentation

- indentation is part of syntax!
- with loops, “if else” statements, programs, etc
 - you have to indent
- so no braces { and } like in Stata foreach loop
- just indentation
- actually makes sense—you should indent anyway
 - so the braces are unnecessary
- this is Python: cleanliness first!
- we'll see examples