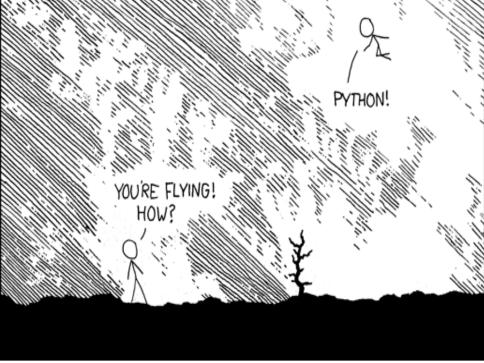# Python intro

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`
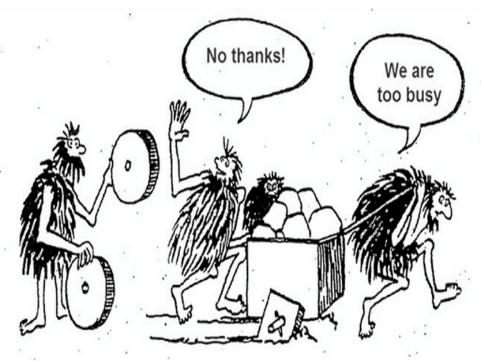
this version: Wednesday 7th April, 2021   13:39

outline

py is fun! and can accomplish so much more!

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

references/tutorials

- com sci, quick, readable! `http: //code.google.com/edu/languages/google-python-class/`
- looks superb!
  `https://automatetheboringstuff.com/#toc`
- general, complete, lengthy
  `http://www.diveintopython3.net/`
- for soc sci, comprehensive
  `http://nealcaren.github.io/python-tutorials/`
- looks like for soc sci learning py, haven't try it though:
- `http://www.data-analysis-in-python.org/`

cool stuff we'll skip

- can email and text from Py :)
- https://automatetheboringstuff.com/chapter16/
- can manipulate images
- https://automatetheboringstuff.com/chapter17/
- in general https://automatetheboringstuff.com has many good ideas

approach

- don't have to do everything in Python
- 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 is very versatile!

why another software?

- well, this is 2021
- we have internet, social media, smartphones
- vast majority of data is there
- cannot have data management class just doing spreadsheets
- and just numbers
- need to be able to process text
- and need to be able to get internet data
- statistical software like Stata is behind
- hence, Python

why not?

- although most user friendly general language,
○ still more difficult than qgis, Stata
- because of the generality you need to code more
○ it is not so nicely customized
○ more flexibility requires more code
- with Py, can do almost everything just in one lang
○ but need to type more, and some functionality not (easily)
  available
○ as in specialized lang like sql, stata, hlm, lisrel, etc

Python for a social scientist

○ 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, scrappy, 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
- or easier can 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 aditional packages and may
  be clunky in general

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

approach

- we'll skip a lot of stuff
- (for more see references/tutorials on first slide)
- I won't bother you with dictionaries, tuples, etc
- I will introduce Python as for
- social data science or computational social science (CSS)
- today we'll cover basics
- ok, let's fire up colab
- see link on syllabus