Python intro

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

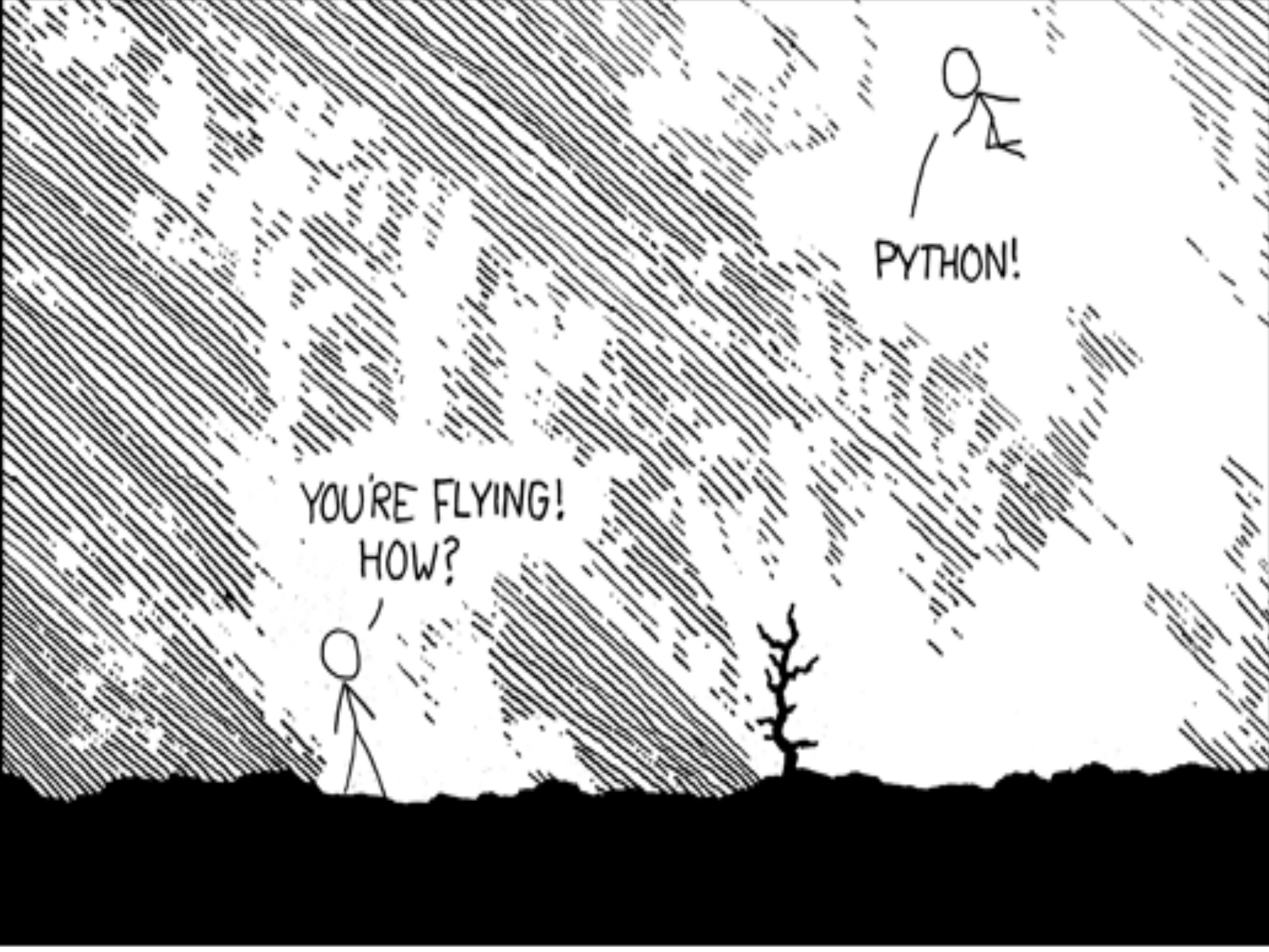
this version: Thursday 21st November, 2019 12:58

outline

why?

python basics

IDE or GUI or Interface

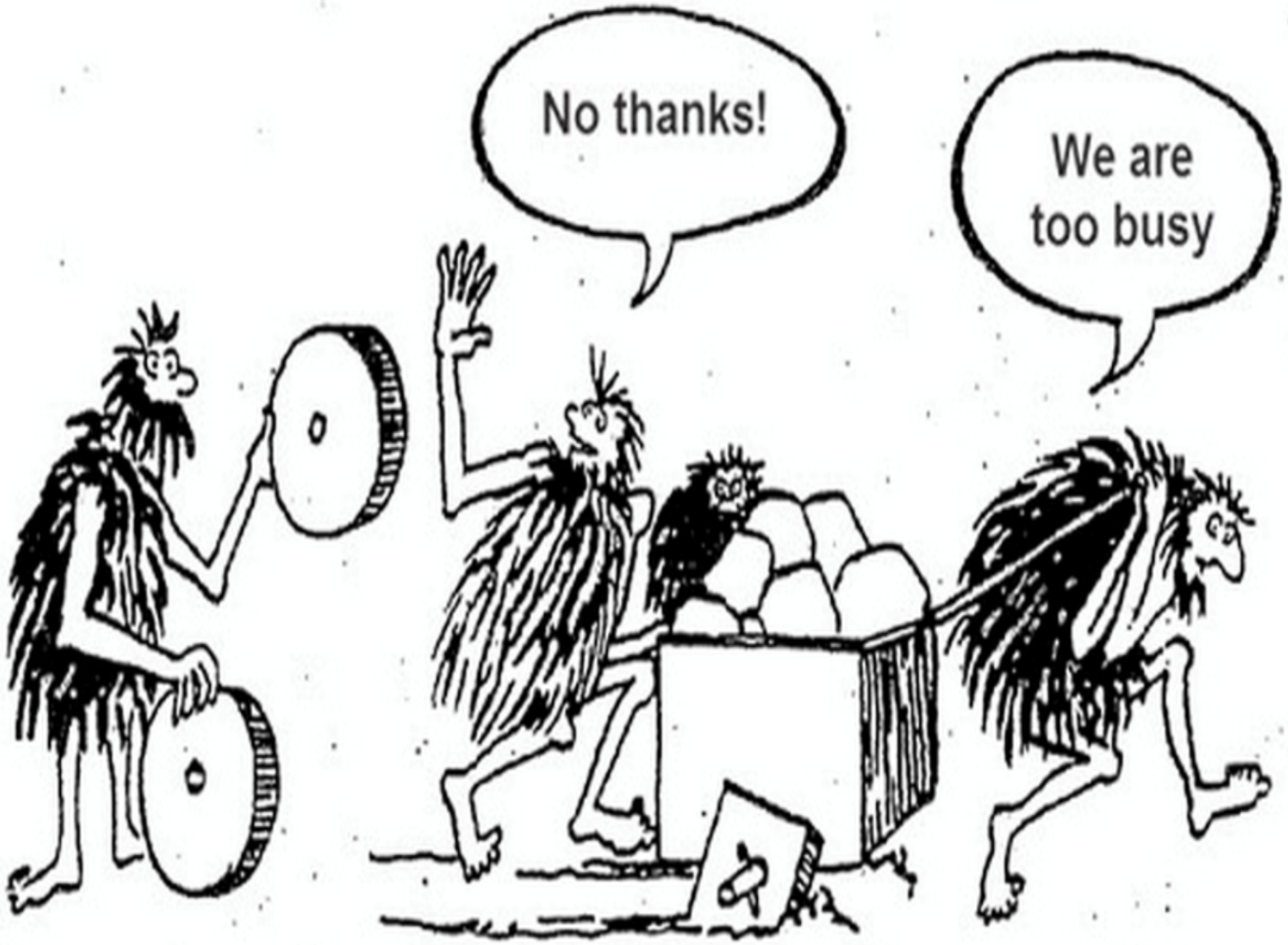

PYTHON!
YOU'RE FLYING!
HOW?

python is so much fun!

- you almost feel like flying
- `http://xkcd.com/353/`
- remember “the Flow” from intro class?

and you 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"


No thanks!
We are too busy

another visualization

- http://lpycot.appspot.com/img?img_id=aghzfmxweWNvdHIKCxIDQXJOGNEPDA

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!

outline

why another software?

- well, this is 2017
- 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 cannot do it
- hence, Python

why Python?

- most user friendly, fun to use, 'glue', and general
- unlike qgis, sas, Stata, R,
  - Python is a regular (hi-lev) programming (scripting?) lang
- does data mgmt and stats (but Stata and R are better)
- but Py can do all sorts of other things
  - that gis/stats software cannot do at all
- a good idea to learn a general language
  –it can do everything that computers can do!

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 things are not 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 !

outline

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`

Python versions

- there are two major releases/versions of Python:
  - 2.x and 3.x
- we'll use py $>$3.3

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 Python using Spyder
  - it's on desktop, icon is a spider web

outline

Spyder

- kind of like Stata or RStudio
- on the left we have code
- and it runs on the right in 'console'
- and some other less useful windows that you can drop
- at the top a bunch of menu items
- but most of the time just highlight code you want to run
  - and hit:
  - green 'play' triangle icon with blue rectangle to the left
  - or better just press Ctrl-Enter to run selection
  - note: for win paths may need double backward slash \\

Ipython v regular Python console

- default is Ipython
- but can open console from under 'Consoles' menu
- unlike Stata, can have multiple interpreters running
- the cool thing about Ipython is that it holds all output
- incl graphs
- and also gives the [In] [Out] line numbering
- but sometimes want to use console
  - say to pop out graph in new window that allows manipulation
  - especially useful for zooming in; and in 3d graphs can rotate axes!