

API

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Wednesday 14th April, 2021 17:13

outline

again, data management always comes first

- ◇ for traditional dat man, (still?) Stata is the best
- ◇ but for untypical or internet data, Py is probably the best
 - great for getting data (APIs) and webscrapping
 - and great for text processing

data data everywhere: internet!

- ◇ you've heard about:
 - information age, machine age, web 2.0, social media, etc
- ◇ data are everywhere:
 - blogs
 - wikis
 - government websites, etc
- ◇ soc sci traditionally just use the data that somebody puts in a spreadsheet for download
- ◇ but why not use ANY data that is online?
 - you will see that Python excels at this task and it's pretty easy

API: what is it

- ◇ API=Application Programming Interface
- ◇ basically an interface to get internet data
- ◇ most major websites have API
- ◇ you can connect to API and get data
- ◇ major websites run on top of databases, eg google, twitter
- ◇ API is a way to access that database

API v web scraping

- ◇ web scrping is getting data from internet by hand
 - flexible—can do any website
 - but has to write the code to extract information
 - time consuming, better use API
 - we'll just do API
- ◇ if you want data from website without API
 - let me know and we'll try to scrap it