

where do i get data?

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Sunday 27<sup>th</sup> October, 2024 10:53

## outline

where?

census data

good practices

# outline

where?

census data

good practices

## no need for data collection

- it used to be that you had to collect data yourself
- some people still do it
- yet, almost always the data that you need was already collected
- it is not easy to collect/produce data
- it is a separate degree or at least a separate class on how to do it

# data data everywhere

- data are everywhere
- whatever you study, you can get data for it
- but wait, where actually is it?

## find data—the fast way

- email me or GA
- again, data questions are perfectly fine
- yet, no one asks them !
- google
- requires a little skill in knowing what to google

## eg googling local educ data

- <http://nces.ed.gov>
- NJ educ data, should be easier to use, but i do not have experience with it <http://www.state.nj.us/education/data/>
- same for NY <http://data.nysed.gov/>
- and PA <http://www.portal.state.pa.us/portal/server.pt/community/dropouts/7396>
- and there is stuff for Philly;; this should be relatively easy to use <http://webgui.phila.k12.pa.us/offices/o/open-data-initiative>  
[https://webapps.philasd.org/school\\_profile](https://webapps.philasd.org/school_profile)  
<http://technical.ly/philly/2014/07/15/salary-data-school-district>

## find data—the slow (classic) way: a list

- there are many excellent datasets ! let's go over them
- i posted them in data\_sources.csv
- and many other lists like this; eg:

[http://journalistsresource.org/tip-sheets/research/  
websites-u-s-federal-government-administrative-datasets](http://journalistsresource.org/tip-sheets/research/websites-u-s-federal-government-administrative-datasets)



- there are many datasets...
- some are international (for ipsd people)
- some are lower-level: eg person level

# outline

where?

census data

good practices

# socialexplorer

- i like `http://socialexplorer.com`
- (need to access it on campus to get it free)

## census website

- actually now can have a similar interface on census website
- `http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml`
- and click geographies on the right
- then at the top can search in box 'topic or table name'
- say for 'income'

# outline

where?

census data

good practices

## cite it!

- always cite your data
  - at least full name and **specific** url [not eg census.gov]
- i do not care about the format; eg
  - cite at the end among regular references
  - cite in text
  - have an appendix
  - cite in footnote

## be careful about confidential data!

- anything that can be used to identify a person
  - ssn, dob, name, any ID #, etc
- is a red flag!
- be extremely careful!
- just de-identify:
  - either drop sensitive vars
  - or create 2 files with your own new ID
  - one with sensitive info and ID, and keep it secure
  - the other with other info that you can easily carry around

## be careful about confidential data!

- be very careful—almost nothing is “secure” these days
- best stay away from sensitive data, or work with
  - someone who knows what she is doing!
  - very easy to make mistake and get into big trouble
- tell a story about what happened to a prof
  - who is my wife's boss...
  - very experienced and still got in trouble!