

Reflections: Philosophy of Government

Jamie Saxon

Introduction to Programming for Public Policy

November 28, 2016

Using data; connecting datasets

1. Topics we've covered.
 2. Connecting datasets.
- ▶ Philosophy of government
expressed through computing policies.

What We've Covered

	Applications	What now?
Git	Version Control	Use (and contribute!) to packages!
Command Line	Navigation; exploratory analysis, data reduction, and cleaning.	First line of defense with large datasets.
Python	Algorithmic problem solving, data processing. Many libraries: <u>math</u> , <u>bs4</u> , <u>requests</u> , etc.	Languages need practice: try <u>project euler</u> , <u>python challenge</u> , or <u>coding bat</u> (easier).
Pandas	Joining datasets; slicing columns and filtering rows (masks). Plotting and fitting.	Stats (<u>scipy.stats</u> , <u>statsmodels</u> , <u>lmfit</u>), math (<u>numpy</u> and <u>sympy</u>), plots (<u>seaborn</u>).
Databases/SQL	Storing, connecting, and processing data.	Ask! They're behind most serious applications. Be aware of indices.
Scraping/APIs	Use the data on the web!	Learn about <u>schemas</u> !
Web Development	Static and dynamic websites.	Javascript! Databases! Servers!
GIS	Merging datasets in spacetime; beautiful, interactive maps.	Spatial statistics; many classes at Harris/UC.
Machine Learning	Targetting interventions.	Some courses through the CI.

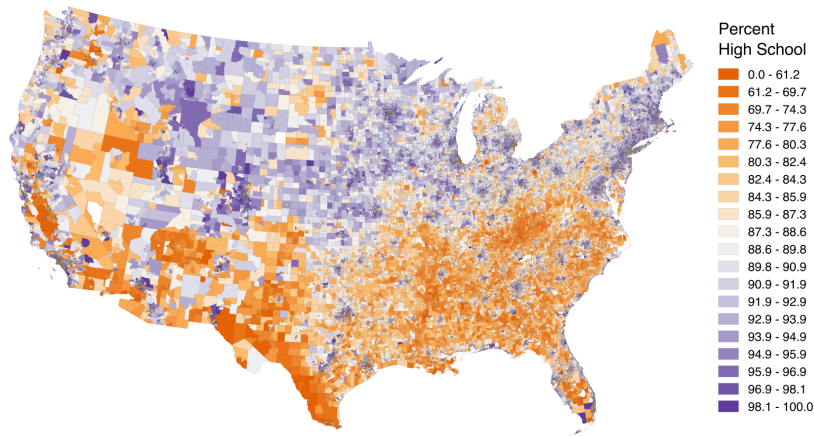
An Amusing Diversion: Python Challenge

- An engaging set of linked puzzles – great for the holidays! 🌿



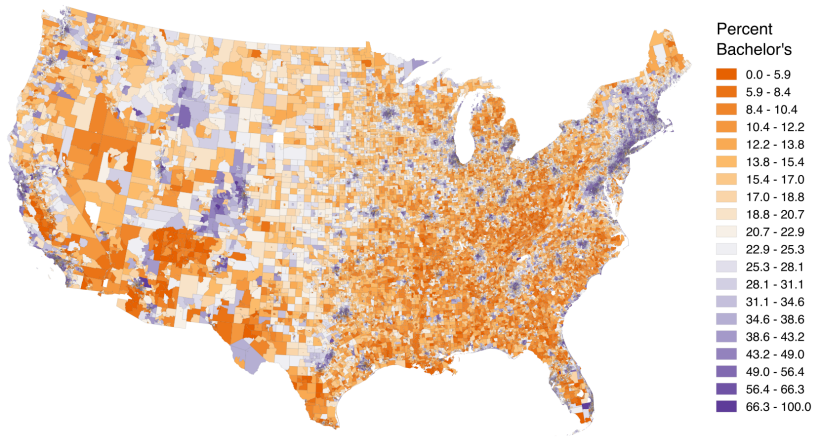
Educational Attainment by Census Tract

- Huge difference in educational attainment maps at high school and college. Do ambitious kids leave the mountain states, ne'er to return?



Educational Attainment by Census Tract

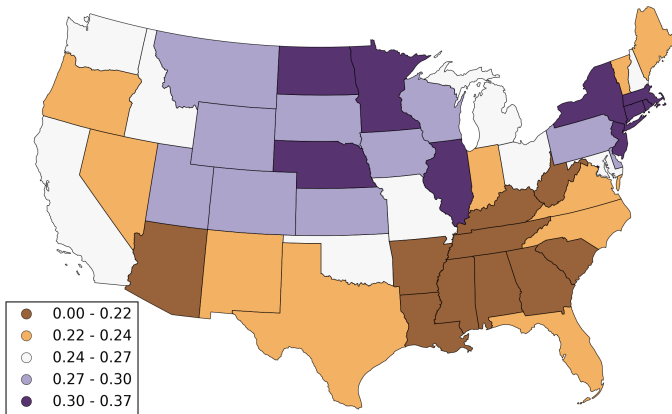
- Huge difference in educational attainment maps at high school and college. Do ambitious kids leave the mountain states, ne'er to return?



Educational Attainment by Birthplace

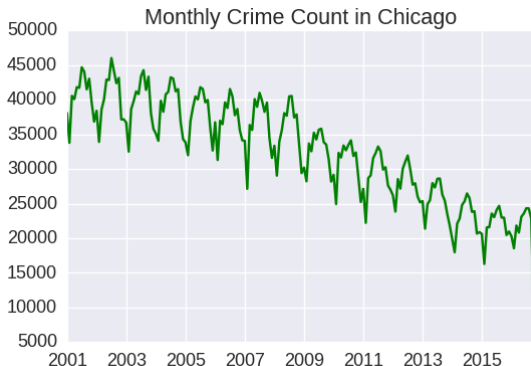
- ▶ Need **IPUMS** (Integrated Public Use MicroSample) to get birthplace.
 - ▶ These are 1% and 5% samplings of the Census and ACS.

Percent College Graduates, Born 1950 - 1970 (by Quintile)



Weather and Crime

- ▶ Long ago, we toyed temperature and crime in Chicago.
 - ▶ But was it driven by temperature, or seasonal?
 - ▶ Let's use APIs, ~SQL, and pandas to check San Francisco!
 - ▶ Is there annual variation in the monthly crime count?
1. No seasonal variation in weather (or crime). Incredibly fast to do this.
 2. But the datasets have superficial differences – code not portable.



Weather (Solution)

```
import pandas as pd

sf_url = "https://data.sfgov.org/resource/cuks-n6tp.json?"
query  = "$select=date_trunc_ym(date)%20as%20month,"
query += "%20count(*)&$group=month"

df = pd.read_json(sf_url + query, convert_dates = ["month"])
df.plot(x = "month", y = "count", c = "r")
```

(Wow, was this fast!)

- ▶ I hope this class has given you greater comfort in retrieving, processing, and sharing data. You know statistics from other classes.
- ▶ If you value data, ask for it, and do your own analysis.
 - ▶ Don't settle for PDFs of other people's ideas!

Aspirations of Open Data

OMB Memo M-13-13: Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public. In order to ensure that the Federal Government is taking full advantage of its information resources, executive departments and agencies must manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information.

PA Data Portal: One of our most valuable and underutilized resources in state government is data. Our goal is to make data available in order to measure government performance, engage citizens, create economic opportunities, and develop innovative policy solutions.

Realities of Open Data: A Disconnect

- ▶ Many flashy sites, powered by Socrata and CKAN (open source).
 - ▶ But most of the datasets contained in these.... suck.
- ▶ Pennsylvania data portal: education datasets.
- ▶ White House: a few datasets of budget and salaries, but mainly questionable ones on National Parks, etc.
- ▶ And data.gov is relatively sparse in integrated data.

- ▶ You will go and change the world: make good data!

Building Technology for Government

Cities, states, and counties like to think they're unique. **They're not.**

- ▶ Make open source the default: responsibility is to the taxpayer, not the contractor.
 - ▶ There are plenty of companies (e.g., Oracle!) that understand the place of open source software.
 - ▶ We typically think of a 'public dollar' as less valuable than a single private one (distortionary, etc.). Fight against that!
- ▶ Intellectual property is unlike other civic infrastructure. Save money by standardizing inputs and building shared tools.
 - ▶ Boston 311 developed with Connected Bits; now licensed to Chicago, SF, Austin, etc. (Though private licensing is a mixed blessing.)
 - ▶ Several parking app platforms shared among cities (pango, passport).
 - ▶ You don't have to contract it out – but you shouldn't build it anew every time! Use existing services if you can! (We don't write libraries.)
- ▶ Consider design needs – what is shared across cities or agencies?
 - ▶ Then write a schema and collaborate!?

An Aside – What is Open Source?

Free as in speech, free as in beer...

- ▶ Open source software may be freely downloaded and used.
- ▶ The source code is also published (often on github, sourceforge, or their predecessors), and anyone can modify it.
- ▶ Alterations and bug reports are pushed back to the source, and a 'maintainer' (or a team of) reviews and approves or rejects each one.
 - ▶ A large community of people may propose and discuss bug fixes.
 - ▶ Makes for stable code, fast fixes, and good security.

Stunningly successful model. Linux runs the majority of supercomputers; Apache dominates in web servers. Many of you have used Firefox.

(The \$1 fix at F18: all bugs are shallow.)

Schemas and Standardization

Great fanfare around posting data online;
inadequate care to their quality and intercompatibility.

- ▶ Obvious candidates for standardization (good quality data):
 - ▶ Crime: [Portland](#), [NYC](#), [Chicago](#), [LA](#), [Boston](#), [Philly](#), [Phoenix](#), [Denver](#), [San Francisco](#), [Dallas](#).
 - ▶ Voting returns: [OH](#), [IL](#), [UT](#), [PA](#), [VA](#).
 - ▶ Budgets, school performance, road work, housing inspections, etc.
- ▶ At [schema.org](#), Google, Microsoft, Yahoo, and Yandex work with an open source community to standardize structured data.
 - ▶ This is what allows for displaying e.g. movie showings or restaurants directly in search results.

Why Does Government IT Stink?

- ▶ Customer Relationship Management from the progressive age: people at desks.
- ▶ Terrible procurement process with fixed vendors. Particular problem when you want things fast.
- ▶ Inadequate entrepreneurial spirit in government.
 - ▶ Consider prizes, novel funding mechanisms.
- ▶ Don't settle! Our relationship and view of government comes from our experience of it.

The Power of Technology

- ▶ Better technology allows government to collect, analyse, and apply data more quickly.
- ▶ Citizens can more-easily contribute information; government workers can *use* more information to make better, more independent decisions.

Data as a Philosophy

- ▶ Expect push-back: lots of people build and defend fiefdoms, and their power is their data.
- ▶ There are legitimate concerns for a lot of data (privacy), but for the time being, we're pretty far from that on the open data front.
- ▶ And everyone is better off if you build and share tools and data.
 - ▶ Freeing data allows new collaborations and shifts development to the private sector where possible (think GPS, NOAA).
- ▶ Transparency and an informed citizenry is an important part of democratic government. It's worth fighting for.
- ▶ Build data collection into your program design, in such a way as to evaluate if they work!

Conclusions

- ▶ Have fun solving puzzles and building things.
- ▶ Look for connections between data sources.
- ▶ Demand and use data; keep learning.
 - ▶ If interest: monthly troubleshooting hour?
- ▶ Build government to be proud of.