Three Capstone Project Proposals – 01-19-2016 – Brendan Hill

1. The Long term unemployed - Who are they, and have they changed over time

In the Great Recession, long term unemployment was a serious problem. Long term unemployment was especially bad in the Great Recession. In fact the Bureau of Labor Statistics change the definition of long term unemployment to include longer spells of joblessness.

For the long-term unemployed, being without jobs over long periods of time can have serious consequences. One is the perceived loss of their skills by potential employers (“deskilling”).

For a capstone project, I want to examine and compare the long term unemployed during the Great Recession and after. This includes looking at what are the factors were highly predictive of long-term unemployment during both snap shots of time and see if they have changed. If similar people are unemployed for long periods of time regardless of the performance of the economy, more uniform policies over time (possibly at different intensities) might suffice to help the longer term unemployed. If they are different over time, effective policies might need to change if the health of the economy changes. This can be much more difficult.

The data set I would use is the Census/Bureau of Labor Statistics Current Population Survey (CPS). It is one of the most oft cited government data source because it produces in the unemployment rate. It's a monthly survey of 6,000 household. At the sub-state level, BLS only reports data on 54 Metros. Maybe we could produce data for more areas with pooling over multiple time periods, but I don't know what to do that.

In addition to employment status, the CPS also collects data on socioeconomic and demographic variables like race, ethnicity, age, and income. It also asks about health care and family structure. To see more, here's a codebook; the variables start on page 18

<https://cps.ipums.org/cps/resources/codebooks/cpsmar14.pdf>

I must admit, doing a project with complexly structured microdata is intimidating. I’m a little nervous it would require a lot of time and work, and I'm not sure if it's above my level. However, although its incomplete), I saw one person who's worked with CPS data in Pyton (<https://github.com/TomAugspurger/blog-source/blob/master/content/python_for_cps.md>) so I might be able to work off of that, but some of that code went beyond my head. I saw similar sites to wrangle the data in R and Stata.

1. What amenities and other characteristics are associated with the neighborhood diversity?

Ever since the Immigration and Nationality Act of 1965 became law, America has continuously become a more diverse nation. Despite a recent backlash against immigration, this trend is not likely to dissipate. However, despite nationwide increases in diversity, in many cities across the country, many neighborhoods are segregated into different ethnic groups. In fact, as Nate Silver pointed out in a recent article, overall diversity in cities and segregation are often correlated. (<http://fivethirtyeight.com/features/the-most-diverse-cities-are-often-the-most-segregated/#fn-6>)

Across this country, many other characteristics of America's cities and their neighborhoods are as diverse the race and ethnicities of their population. Whereas Nate Silver's article focuses mostly on cities, I would look more at the neighborhood level. What I want to know is what characteristics do the most diverse neighborhoods share? Are there common characteristics that other neighborhoods could cultivate and increase their diversity? If so, this might help guide policies to diversify neighborhoods.

The data I would start with is Brown University’s American Communities Project

(<http://www.s4.brown.edu/us2010/index.htm>) to identify the most diverse city neighborhoods. This data produces statistics on diversity at the city level, the tract/neighborhood level, and at the city vs neighborhood level (index of dissimilarity, neighborhood diversity index, integration-segregation index).

I would then merge that data with other data sources at the tract level, although some of this data is pooled multiyear data, and I’m not sure how annual and multiyear data could work together. Some data sources I already know how to use, some I'd need to learn. I'd also look for alternative sources I don't know yet.

One I know well is the American Factfinder. It has great demographic and quality of life data at the tract level, especially from American Community Survey (ACS).

<http://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2014/>

<http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>

Census's TIGER also has some statistics at the tract level (<https://www.census.gov/geo/maps-data/data/tiger.html>) as does the Federal Financial Institutions Examination Council (<https://www.ffiec.gov/geocode/>)

1. Why were some cities more able to weather the Great Recession than others?

Although only a small number of cities--particularly the ones with the largest housing bubbles--were shedding jobs as the Great Recession began, by the end of 2009, nearly all U.S. metros were suffering. All cities were affected, but there was a huge disparity in the severity (how many jobs were lost) and the length of time (how long it took for cities to regain the jobs lost during the recession) cities faced before they recovered.

What characteristics did cities that were relatively resilient share? Do they have similar population sizes, industrial make up, geographic location, or income? Were there other factors? If it is predictable which cities might fare worse during economic downturns, more federal aid could be directed towards cities that need it most.

The data I would start with is the Bureau of Labor Statistics (BLS), State and Metro Area Employment, Hours, & Earnings (SAE) Current Employment Statistics (SES) (<http://www.bls.gov/sae/>). I would use BLS’s SAE to calculate jobs and job loss by metro area, and calculate the amount of time job loss is negative compared to the beginning of the recession. For the other major statistics, income and total city production, I'd use the Bureau of Economic Analysis’s (BEA) Regional Economic Accounts ( <http://bea.gov/regional/index.htm>).

I would then merge these sources with other data sources at the metro level that might have other predictive power about the effects of recession. First would be demographic data, and that could come from State & County QuickFacts (<http://quickfacts.census.gov/qfd/states/00000.html>), or more likely the ACS data described in proposal #2. Except for this proposal, the data would be at the metro level, not the tract level.

Then, I'd look at some less used but still interesting data sources. I don’t know them all well (or at all). Some I just found while researching this proposal, but most look promising:

* Metrotrends (<http://www.metrotrends.org/data.cfm>) has data on the arts and culture, local financial institutions and behavior health of the population.
* The State of the Cities Data Systems (<https://www.huduser.gov/portal/datasets/socds.html>) has great data on housing, construction and crime.
* Finally I'd mine the Internet Data Sources for Social Scientists (<https://www.ciser.cornell.edu/ASPs/datasource.asp>) for datasets like
  + City Data ([City-Data.com](http://www.city-data.com/)) for environmental profiles, and
  + Kids Count Data Center (<http://datacenter.kidscount.org/>) for detailed health and other well-being outcome information.