data and description

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Tuesday 23rd January, 2018 10:34

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

NECESSARY readings

- https://data.bls.gov/cew/doc/info/location_quotients.htm
  - [if need more reading, some descriptive examples for Indiana: http://www.incontext.indiana.edu/2006/march/1.asp]
- http://data.bls.gov/location_quotient/ControllerServlet–

  try to calculate LQ yourself

Barro, R. (1999): "Determinants of democracy," Journal of Political Economy, 107, 158–183.

Blakely, E. and N. Leigh (2009): Planning local economic development: Theory and practice, Sage Publications, Inc.

Florida, R. (2008): Who's your city?, Basic Books.

Mackie, J. and J. MacKie (1980): The cement of the universe, Clarendon Press Oxford.

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

presenting results

- quote data source in detail; give url
- define variables; maybe table with definitions in the appendix
- describe sample in detail: time, location, sampling, etc...
- what is your contribution? how come everybody else got it wrong or missed it?
- there has to be some contribution in your paper !! data? method? idea?
- avoid results padding: do not present tables, graphs if they do not mean anything or if you do not discuss them or if they do not help with argument

presenting results

- avoid ugly tables !
- graphs/tables need to have captions that are self-explanatory
- graphs/tables need to be referenced in text
- show 2 or 3 decimal points, no scientific notation, no vertical lines
- do not say "increase by one unit"; what is the unit?
- all vars must be defined clearly (say key vars in text, others in appendix); e.g. `http://www.hks.harvard.edu/fs/rpande/papers/qje_all.pdf`
- may want to label interesting cases in graphs `http://www.wzb.eu/alt/iw/pdf/genecult.pdf`

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

a research design is a class itself

- i will just mention few things that will be important for this class
- a quick, useful and applied reference is `http://www.socialresearchmethods.net/kb/design.php`
- a more in-depth treatment is Lawrence B. Mohr, Impact Analysis for Program Evaluation
- in general: `http://www.socialresearchmethods.net/`
- `https://sites.google.com/site/adamokuliczkozaryn/inf_des/descriptive4.pdf`

spurious correlation

- draw a scatter, fit line of some Y and some X

  X is banana production in Honduras, Y is deaths on US highways
- you think that x causes y, but actually it is z
- say, global warming...
  - we have it–we can measure temperature
  - but the cause: we may think it is CO_2, but actually it is Sun activity
  - or the other way round...
- another way to say it: correlation is not causation

the gold standard

- the experimental design
- only with experimental design you can confidently argue causality
- and it is because randomization takes care of the known and unknown predictors of the outcome (draw a picture of 2 groups of people)
- most of the time we cannot have an experimental design because it is unethical and politically impossible
  e.g. we cannot randomly assign kids to bad school or to smoking
- `http://www.socialresearchmethods.net/kb/desexper.php`

internal and external validity

- internal validity is about causality
- external validity is about generalizability
  - can i say something about Rutgers in general by analyzing you?
  - how about just Rutgers-Camden ?
  - no ! people at Law school, computational biology are likely to be different
  - and even per PA, I would ideally like to have a random sample
  - note, random sample is different from randomization/random assignment

threats to internal validity

- history, maturation, regression to the mean
  - something else happened that caused Y
  - things develop over time in a certain way
- selection bias, self selection
  - does smoking causes cancer ?
  - maybe less healthy people select to smoke ?

you still can have a valid inference

- but you need to do more work...
- essentially you want to exclude alternative explanations
- so you act like a devil's advocate...
- and try to abolish your story / find an alternative explanation
- if you cannot find any, they your story is right ...
- until disproved

two basic designs

- you can look over time (PRE, POST) (draw a graph)
  e.g. you can trace unemployment over time in Camden
  - and, say, you can find that it increased during Reagan administration...
  - but you cannot argue causality right away !
  - there may be lots of alternative explanations, e.g. shift away from manufacturing during the same time, etc etc
- and you can look across space
  - e.g. you can compare Philadelphia to Camden

levels of analysis

- you are probably familiar with term Unit of Analysis (U/A)
- in regional development a peculiar thing is that there are many levels
- there are states, counties, metropolitan areas, cities, etc
- and you often get different and even opposite conclusions depending on what level you are looking at

aggregate data

- in regional development research much of the data is aggregate
- e.g. income, home ownership rate at county level are sums of person-level values divided by population
- with aggregate data you are losing information
  you don't know the variability and the distribution

different levels, different effects

- variables at different levels may have opposite effects
- e.g. if i increase your salary, you'll be happier
- but if i increase salary of everybody in your county you'll be less happy
- would you like to live in a world where you make $100k and the average is $150k
- or would you like to live in a world where you make $75k and everybody and the average is $50k
- people chose the second scenario
- "a rich guy is a one who makes $100 more than his wife's sister's husband"

contextual effects

- a closely related concept is of contextual effects
- whatever you study it takes place somewhere and place matters
- so it is not only characteristics of the U/A that predict your outcome
- but also the context (characteristics of larger units in which U/A is nested)
- student is nested within a classroom, a classroom within school, a school within a district, etc etc
- a firm is nested within a city/metropolitan area/town, which is nested within a state, which is nested within a country

paper

- you should address the above issues in the paper
- again, a useful thing to do is be devil's advocate
  - ask yourself how/why what you are saying is not true
  - think about alternative explanations
  - what are the limitations of your study

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

comparing Camden, NJ and Plano, TX

- First, let's use Census data
- a quick way is to use QuickFacts
- `http://quickfacts.census.gov`
- `http://quickfacts.census.gov/qfd/states/48/4858016.html`
- `http://quickfacts.census.gov/qfd/states/34/3410000.html`
- what's interesting here?
- Camden has about 7 times more Blacks and 8 times fewer Asians
- homeownership rate is 20% lower in Camden
- Plano has only 7% of population in poverty, while Camden has 36%
- note that TX has almost twice as many people in Poverty as NJ: 17% vs 9%

little more scientific way

- http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml
- can play around–many ways to go about it
- let's do guided search
- and pick SELECTED ECONOMIC CHARACTERISTICS DP03
- for 2 census tracts in Philly: 137 and 138 (Brewerytown)
- https://geomap.ffiec.gov/FFIECGeocMap/GeocodeMap1.aspx looks like a great way to find census tracts;
- always show map of an area!
- but use caution with their data!
- i found it differed from census! best use census whenever

paper

- again, a useful trick is to combine different types of data to come up with a contribution
- talk to your classmates!
  - e.g. food deserts and crime
  - e.g. weather and migration, etc, etc
- and your paper does not have to be quantitative
  - still, you can approach your topic from different angles

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

data and development

- development planning begins with understanding of the of the local economy
- if you cannot measure it, your knowledge is of 'meager kind'
- and you also want to keep on measuring to see what is going on
  - is the situation improving?
  - any interesting trends ?
  - how are we doing compared to other similar localities?

BEA

- a terrific website!
- under regional data you will states and metros
- and even some smaller areas
- `http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=5#reqid=70&step=1&isuri=1`
- let's have a look at some data!

some performance measures

- population and employment growth (Census Quick Facts)
- unemployment rate (CQF)
- income levels and poverty rates (CQF)
- earnings and wage levels (`http://www.bls.gov/bls/blswage.htm`)
- labor force participation (blswage)
- firm births, deaths, and relocations

  (`http://www.bls.gov/web/empsit/cesbdhst.htm`)

some performance measures

- new development and investment
- property values and tax revenues
- analyze trends over time
- compare to state, metro area, nearby cities
- variation among demographic subgroups and sub-areas
- link indicators to key goals & track over time

wages

- http://www.bls.gov/bls/blswage.htm
- by census division
- http://www.bls.gov/ncs/ocs/compub.htm#Division
- by state http://www.bls.gov/oes/current/oessrcst.htm
- metro http://www.bls.gov/ncs/ocs/compub.htm
- http://www.bls.gov/oes/current/oessrcma.htm

living wage, poverty

- Families working in low-wage jobs make insufficient income to live locally given the local cost of living.
- Recently, in a number of high-cost communities, community organizers and citizens have successfully argued that the prevailing wage offered by the public sector and key businesses should reflect a wage rate required to meet minimum standards of living.
- MIT has a living wage calculator
- by county `http://livingwage.mit.edu/`

property values

- an indicator of place desirability
- low in Camden – nobody wants to live here
- high in Manhattan – everybody wants to live there...
- also reflect job opportunities:
  - you can afford Manhattan housing if you have a Manhattan job
  - you can afford Camden housing if you have a Camden job

property values

- federal housing finance agency
- http://www.fhfa.gov/Default.aspx?Page=87
- land prices over time http://www.lincolninst.edu/subcenters/land-values/price-and-quantity.asp
- http://www.standardandpoors.com/indices/sp-case-shiller-home-price-indices/en/us/?indexId=spusa-cashpidff--p-us----
- http://www.zillow.com/local-info/
- http://www.city-data.com/
- a useful calculator

  http://cgi.money.cnn.com/tools/homepricedata/

tax revenues

- a measure of local economy health...
- state and local taxes `http://www.census.gov/govs/qtax/`

cost of living calculators

- http://cgi.money.cnn.com/tools/costofliving/costofliving.html
  - compare Philly to NYC
- http://www.payscale.com/cost-of-living-calculator
- http://swz.salary.com/costoflivingwizard/layoutscripts/coll_start.aspx

basic analysis

- compare economy to region, state, and nation
- compare economic sectors
- identify which sectors are growing fast and slow
- identify most important industries within key sectors

basic analysis

- show how industry mix varies with that of region
- determine relative wages of major industries
- look at recent trends
- determine which industries are fastest growing
- identify declining industries
- compare local and regional industry growth trends

understanding

- you need to understand economy's strengths and weaknesses
  and think about wheat may be driving them

the outcome line

- a useful concept is that of outcome line read `http://books.google.com/books?id=GBxhOT8btfYC&lpg=PA16&pg=PA15#v=onepage&q&f=false` and put more text here

think of the larger context

- where are we in the business cycle
- what are the global trends ?
  - they do affect the local economies
  - outsourcing manufacturing jobs to China
- local economy is not simply a fraction of the national economy, though
  for instance if the there is drought in Latin America, Iowa will benefit more than Nevada (it produces more food)
- new police lowered crime? crime declining everywhere!

standardize

- if you go over time, you need to deflate dollar amounts

  `http://www.bls.gov/data/inflation_calculator.htm/`

  `http://www.duke.edu/~rnau/411infla.htm`

- if you go across divide by population: otherwise you cannot compare, say Philadelphia to Camden

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

labor force characteristics

- this is key ! jobs are key !
- especially in those difficult times
- key in attracting new employers
- you want to have people in occupations that have good prospects

labor force characteristics

- a great resource is BLS occupation outlook:

  `http://www.bls.gov/ooh/`

  how does your occupation prospects stack up ?

- you'll find a ton online, e.g.

  `http://www.theatlantic.com/business/archive/2012/02/americas-10-fastest-growing-and-fastest-shrinking-jobs/252712/` scroll down and click on jobs

  `http://blog.linkedin.com/wp-content/uploads/2012/06/fluctuations.png`

labor force characteristics

- low labor force participation for a specific demographic group

  may suggest lack of opportunity, discouraged workers, discrimination, etc
- median commute time is interesting indicator
  - if high it suggests a mismatch between housing and job markets
  - and it produces congestion, pollution and unhappiness (people are most unhappy when commuting)

businesses, job supply

- you also want to look at job suppliers–businesses
- interesting thing is that many businesses cannot find people to fill open jobs
- and there is unemployment and underemployment of course, so there is s mismatch
- `http://www.forbes.com/sites/jacquelynsmith/2012/05/29/the-10-hardest-jobs-to-fill-in-america-2/`
- `http://www.nytimes.com/2012/06/28/business/smallbusiness/even-with-high-unemployment-some-small-businesses-struggle-to-fill-positions.html?pagewanted=all`

basic things to understand (Blakely and Leigh, 2009, p164)

- which local parts of the economy are most valued by locals
  - how locals compare themselves to others
  - (can do a survey, interview, focus group)
- what's the local economic base (LQ)
  - what accounts for most jobs and wealth
  - and what's growing/declining most rapidly
- multiplier: how growth/decline in one part affects other parts
- which firms are a part of interdependent cluster

economic base

- exporting industries are important–they bring the money to the locality
- imports are important to look at, too, there may be an opportunity for substitution
- it's businesses that generate wealth that should be targeted for attraction and nurtured
- we used to focus on industries, but now focus on people, e.g. creative class (Florida, 2008) – an occupation-centered economic base

LQ (Loc Quotient) (specialization index)

- $LQ = \frac{\frac{e_i}{e}}{\frac{E_i}{E}}$
- e_i local employment in industry i
- e total local employment
- E_i national employment in industry i
- E national total employment
- see data_sources.csv for a link

  (`http://data.bls.gov/location_quotient/ControllerServlet`)

LQ examples

- http://mailer.fsu.edu/~tchapin/garnet-tchapin/urp5261/topics/econbase/lq.htm
- http://mailer.fsu.edu/~tchapin/garnet-tchapin/urp5261/topics/econbase/lq-ex.htm
- (also some data sources) http://faculty.washington.edu/krumme/350/exercises/lq.html

BLS LQ

- `http://data.bls.gov/location_quotient/ControllerServlet`
- let's compare Camden, NJ to Collin, NJ
- and Camden, NJ to US in different time periods
- also play with sectors, supersectors, etc at the bottom
- ex: eds&meds Camden county v NJ (2013): (40/160)/(600/3240)=1.36

where jobs will be in the future?

- again, `http://www.bls.gov/ooh/`
- but also take into account local conditions
- how is it changing ? look at trends, over-time difference
- "shift-share"(Blakely and Leigh, 2009, p):
  - how's overall eco doing: "rising or falling tide raises or lowers all boats"
- proportion shift: change in industry relative to overall growth
- differential shift change in industry relative to the same industry nationally

shifts formulas (Blakely and Leigh, 2009, p182)

- proportion shift $= \frac{emp10_i}{emp00_i} - \frac{ref10}{ref00}$
- differential shift $= \frac{loc10_i}{loc00_i} - \frac{emp10_i}{emp00_i}$
- $ref00$ 2000 employment in reference economy
- $ref10$ 2010 employment in reference economy
- $emp00_i$ 2000 employment in reference economy in industry i
- $emp10_i$ 2010 employment in reference economy in industry i
- $loc00_i$ 2000 employment in local economy in industry i
- $loc10_i$ 2010 employment in local economy in industry i

interconnectedness

- most things are produced from things that somebody else produces
- and hence my performance affects that of my suppliers and people whom i supply
- there is some specialized software
- beyond the scope of this class
- a similar idea is that of clusters

clusters are..

- geo concentrated
- have competitive advantage because they are concentrated
- share supplier and buyer (marketing) advantages
- are supported by advantageous infrastructure in a region
  e.g. universities, venture capital

cluster table

- to visualize clusters you can produce a following table

not competitive (declining local shift)	competitive (growing local shift)	
transforming industries	growing base industries	high local concentration ($LQ > 1$)
declining industries	emerging industries	low local concentration ($LQ < 1$)

use maps and census data

- `http://www.socialexplorer.com/` a wonderful tool to easily get census data (there is a lot of data!)
- it can make maps, too !
- `geocommons.com/` user-firendly online tool to easily make maps

outline

misc

basic research design

examples

analytical methods for regional development (Blakely and Leigh, 2009, ch1, 6)

what to look at?

map it

always a good idea to provide a map

- it's easy, just google whatever you study
- let's do an example
- go to google maps and e.g. say "university city, philadelphia"
- can also try goog images and say the same
- map is worth 1,000 words–it sets the context etc etc

open city data

- just few examples but trend is that more and more local state fed govt dept etc open their data
- `http://phlapi.com/` , `https://data.cityofchicago.org/` , `http://opencityapps.org/` , `https://www.metrochicagodata.org/` , `http://www.opendataphilly.org/` , `http://www.phila.gov/data/Pages/data.aspx`

next week

- we will always end the class by having a quick look at the next class