

# data and description

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Monday 3<sup>rd</sup> February, 2020 17:48

# outline

misc

basic research design [repetition? making sure basics covered]

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

LQ

## 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

- BLAKELY, E. AND N. LEIGH (2009): Planning local economic development: Theory and practice, Sage Publications, Beverly Hills CA.
- FLORIDA, R. (2008): Who's your city?, Basic Books, New York NY.
- MACKIE, J. (1980): The cement of the universe, Clarendon Press Oxford.

# introductions again!

- ◇ half of class missing last week

## listserv

- ◇ best way to communicate outside of classroom, use it a lot
- ◇ if you didnt get welcome message, pls email me and i'll add you, very important you're on it
- ◇ also do email me if you cant send messages to the listserv!

## make sure sakai's dropbox works

- ◇ lets open it up and make sure it went through on my end

# outline

misc

basic research design [repetition? making sure basics covered]

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

LQ



## **presentation: be thorough, but simple; tell a story**

- ◇ 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 tell a story or if you do not discuss them or if they do not help with argument

## presenting results

- ◇ eg [https://sites.google.com/site/adamokuliczkozaryn/pubs/livability-nov19\\_aok.pdf?attredirects=0&d=1](https://sites.google.com/site/adamokuliczkozaryn/pubs/livability-nov19_aok.pdf?attredirects=0&d=1)
- ◇ 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 in tables
- ◇ 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)
- ◇ annotate/label patterns in graphs

# outline

misc

basic research design [repetition? making sure basics covered]

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

LQ

# a research design is a class itself

- ◇ 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

## spurious correlation

- ◇ 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
- ◇ need theory and mechanism, so called “causal path”
- ◇ <https://www.tylervigen.com/spurious-correlations>

## 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  
eg we cannot randomly assign kids to bad school or to smoking
- ◇ <http://www.socialresearchmethods.net/kb/desexper.php>

## 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, then your story is right...until disproved



## INUS condition (Mackie, 1980)

- ◇ a useful way of thinking about causality

Insufficient but Non-redundant part of Unnecessary but Sufficient Condition

- ◇ many, if not most causes are INUS conditions
- ◇ eg a cigarette as a cause of forest fire
  - it's Insufficient, because by itself it is not enough, eg you also need oxygen, dry leaves, etc
  - it is contributing to fire, hence Non-redundant
- ◇ and along with other stuff (oxygen, dry leaves etc) it constitutes Unnecessary but Sufficient Condition
  - it's not necessary for fire, it can be lightning, etc
  - but it's sufficient – it's enough to start the fire

## two basic designs

- ◇ you can look over time (PRE, POST) (draw a graph)
  - eg 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, eg shift away from manufacturing during the same time, etc etc
- ◇ and you can look across space
  - eg you can compare Philadelphia to Camden

## comparing Camden, NJ and Plano, TX

- ◇ a quick way is to use QuickFacts
- ◇ <https://www.census.gov/quickfacts/>
- ◇ <https://www.census.gov/quickfacts/fact/table/planocitytexas,camdencitynewjersey,TX,NJ/PST045217>
- ◇ what's interesting here?
- ◇ Camden: about 7x more Blacks and 8x fewer Asians
- ◇ homeownership rate: 20% lower in Camden
- ◇ Plano: only 7% of population in poverty; Camden: 36%
  - TX v NJ: almost 2x people in poverty: 17% vs 9%

## full census data

- ◇ census is a good source of data, even at neighborhood level!
- ◇ for city/neighborhood level probably want 5-yr ACS
- ◇ <https://geomap.ffiec.gov/FFIECGeocMap/GeocodeMap1.aspx>  
[find census tract]
- ◇ <https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>
- ◇ say topics-people-poverty
- ◇ for 2 census tracts in Philly: 137 and 138 (Brewerytown)
- ◇ always show map of an area! eg  
<https://www.policymap.com/maps>

## 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

## different levels, different effects

- ◇ variables at different levels may have opposite effects
- ◇ eg 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:
  - 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: 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

# outline

misc

basic research design [repetition? making sure basics covered]

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

LQ



## data and development

- ◇ development planning begins with understanding of the local economy
- ◇ if you cannot measure it, express it in numbers, your knowledge is of 'meager kind' (Lord Kelvin)
- ◇ 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?

## some performance measures

- ◇ population, employment, income (Census Quick Facts)
- ◇ earnings, wages, labor force participation  
(<http://www.bls.gov/bls/blswage.htm>)
- ◇ firm births, deaths, and relocations  
(<http://www.bls.gov/web/empst/cesbdhst.htm>)
- ◇ property values and tax revenues  
(<https://www.zillow.com/research/data/>, census, <https://taxfoundation.org>)
- ◇ analyze over time; and across space:
  - compare to state, metro area, nearby cities
  - variation among demographic subgroups and sub-areas
  - link indicators to key goals & track over time

## 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
- ◇ a great resource is BLS occupation outlook:  
<http://www.bls.gov/ooh/>

## 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 a 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>

- ◇ a terrific website!
- ◇ under regional data you will states and metros
- ◇ and even some smaller areas like counties!
- ◇ <http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=5#reqid=70&step=1&isuri=1>

## 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.
- ◇ 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:
  - can afford Manhattan housing if have a Manhattan job
  - can afford Camden housing if have a Camden job

## property values

- ◇ <http://www.zillow.com/local-info/> interactive
- ◇ <https://www.zillow.com/research/data/> download
- ◇ <http://www.city-data.com/>
- ◇ a useful calculator  
<https://money.cnn.com/tools/homepricedata/>

# tax revenues

- ◇ a measure of local economy health
- ◇ state and local taxes <https://www.census.gov/programs-surveys/ntax.html>

## ecology: land, agriculture, etc

- ◇ everything takes place in some ecology
- ◇ and ecology matters
- ◇ <https://www.ers.usda.gov/data-products/>

## basic analysis: understanding

- ◇ look by industry/sector over time and across space
- ◇ understand local economy's strengths and weaknesses and think about what may be driving them
- ◇ a useful concept is that of outcome line <http://books.google.com/books?id=GBxh0T8btFYC&lpg=PA16&pg=PA15#v=onepage&q&f=false>  
see general one, and then example

## 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 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.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 [repetition? making sure basics covered]

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

LQ



## **basics 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 effects: how growth/decline in one part affects other parts
  - eg organize LPGA tournament, superbowl etc, and bars, airbnb, trains etc will boom
- ◇ 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, eg creative class (Florida, 2008) – an occupation-centered economic base

## LQ (Loc Quotient) (specialization index)

$$\diamond LQ = \frac{\frac{e_i}{E_i}}{\frac{e}{E}}$$

◇  $e_i$  local employment in industry  $i$ ;  $e$  tot loc empl

◇  $E_i$  national employment in industry  $i$ ;  $E$  natl tot empl

◇ 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`][enlarge table, also perc change is useful]

◇ http:

`//data.bls.gov/location_quotient/ControllerServlet-try`

to calculate LQ yourself `https://data.bls.gov/cew/apps/`

`data_views/data_views.htm#tab=Tables` eg 2018 Annual Avg

Private:

• 101 Goods-producing, Autauga County AL

LQ • v 10 Total, all industries, All Counties

## BLS LQ

- ◇ another example: eds&meds Camden county v NJ
- ◇ note: also useful to over time: eg employment this year v 10 years back

## 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
- ◇ 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  
eg universities, venture capital

## paper

- ◇ again, a useful trick is to combine different types of data to come up with a contribution
- ◇ talk to your classmates!
  - eg food deserts and crime
  - eg weather and migration, etc, etc
- ◇ your paper does not have to be quantitative
  - but a good idea to approach your topic from different angles, eg quant and qual

## next week

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