gis

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Tuesday 17th April, 2012 18:52

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

integrity/homesty

- data integrity/honesty, or more broadly:
- intelectual integrity honesty
- be explicit about problems in your data
  - e.g. nonmerges, missing data, miscodings
- be explicit about problems in your models:
  - e.g. don't throw away variables from models just because they contradict your story
  - discuss it: how, why; ask readers to email you ideas
- instead of forcing data to tell your story, listen carefuly; let data tell you her story !
- if you work for somebody: e.g. a bank or NGO: they will ask you to find something; use a disclaimer saying that

scatter plots

- use graphs ! e.g. scatterplots
- but stata defaults are not great...; try:
- display different groups, mark data points
- empty hollow circle; jitter
- lowess
- `http://www.stata.com/support/faqs/graphics/gph/graphdocs/scatter3.html`
- `http://www.indiana.edu/~jslsoc/ftp/stat503/Lab%20guide/iucdaF09_lab_pt2-5_graphs.pdf`
- `http://www.stata-journal.com/sjpdf.html?articlenum=gr0021`

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

maps

- before we begin let's look at some interesting maps
- you'll see that mapping can be useful
  - see patterns that cannot see otherwise
  - absorb easily lots of information
  - compare easily
- examples are supposed to inspire you to produce your own maps

acs

- American Community Survey (acs) is a great source of regular and gis data

  `http://www.census.gov/acs/www/data_documentation/2009_acs_maps/`

  let's see the following by county

- % foreign born
- mean travel to work
- completed hs; have bs/ba
- below poverty level
- housing value

the big sort

- "The big sort
  why clustering of like-minded America is tearing us apart"
- America polarizes by county
  (counties are becoming either R or D)
- `http://www.thebigsort.com/maps.php`

who is your city

- http://utd.edu/~ajo021000/myweb/other.html
- http://www.gallup.com/poll/145913/City-Wellbeing-Tracking.aspx

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

gis

- gis is a rather broad topic, good for many classes or a separate degree
- we will just talk about making maps and producing simple spatial statistics
- this will be very applied, i will skip gis theory
- there will be links for those of you who are really into gis

useful !

- gis is really useful for everybody
- no matter what you study it takes place somewhere, and the place matters
- you always have U/A, e.g. persons, schools, hospitals, states, countries
- all of them are located somewhere
- and usually you can map them
  - no need for exact U/A location, (e.g. address of person, school)
  - just use a larger geographical location, e.g. county, state, country

relatedness

- the first law of geography:
  "Everything is related to everything else, but near things are more related than distant things
  (Tobler)
- in practice it means that things cluster together (have positive spatial autocorrelation)
  - poor people live close to poor people
  - high-crime neighborhoods are close to high-crime neighborhoods
  - hi-tech firms are located close to other hi-tech firms (Silicon Valley, Boston)

mapping

- mapping is fun – it gives a different representation of numbers
- mapping is revealing
  - there are things that you would never notice untill you map them
  - e.g. deVeaux devices in Arizona
  - `http://web.williams.edu/Mathematics/rdeveaux/`

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

setup

- to produce maps/generate spatial statistics we need:
- gis data, i.e. mappable data (more in a second): .shp, .kml, etc
- regular data

howto map it

- ok you have some data, and it would very likely have some geo id:
  - county name/id
  - ISD name, etc
  - then get a shapefile that you can merge with your data (see links at the end or google e.g. "us county shapefile")
- you may also have addresses
  - school address
  - hospital address
- you need to geocode them – the easiest to use `http://batchgeo.com`

  once geocoded it can be displayed in a map

howto map it

- think of geography in your data; usually you have it
- google “geo in you data, shapefile” or visit links from this class
- and merge ...
- beware of representativeness of your data of geo...
- i spent months coding provinces from WVS; then emailed Inglehart and found out that they are not representative....

spatial merge

- even if things do not fit geographically...
  - say you have zip codes in one data, and counties in another data
- you can map both and merge based on spatial location
- so called "spatial join"
  - pick sum, avg or total...
  - qgis - vector - data mgmt tools - join by spatial location
- as any data question, ask gis questions on listserv...

some theory: projection

- we will mostly skip it but just to give you an idea...
- the earth is roughly a sphere (3d)
- map is a rectangular (2d)
- to represent 3d earth in 2d map you need to project sphere onto rectangle
- there are all sorts of projections: conic, cylinder etc...
- software will usually figure it out
- but when your data looks weird (e.g. Dallas in England) you have a projection problem

some theory: data, layers

- gis data is a regular data + location info (lat/long)
- there is always a data table (regular data + location info) that underlies a map
- most of the time you want to superimpose different layers of gis data
  e.g. roads, cities, state boundaries, schools
- often you want to produce thematic (choropleth) maps
  thematic maps use different symbols/colors to show variation in data

some theory: gis files

- gis data may be in many formats
- gis data have location info that allows mapping
- gis data can be points, lines, polygons
- usually, you want to overlay several layers...
- the most popular format is called “shapefile” .shp (comes with .dbf and others...)
- another popular format: google .kml (basically xml)

combining shapefile with data

- you can merge in arcgis, r, qgis, geoda
- it is easy in qgis: vector-data management tools-join attributes

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

software

- there is lots of gis software...
- stata – poor (about 5 people developing code)
- R – pretty good, but cannot point-and-click (actually good with maps); need to write code
  `https://stat.ethz.ch/mailman/listinfo/r-sig-geo`
- arc map – probably most popular, many tools, expensive
- geoda – great for spatial statistics
- qgis – great general software
- geocommons – online mapping tool
- etc...

maps in stata

- you are already familiar with it
- it produces maps, but they are rather poor and the process is very clunky
- `http://www.stata.com/support/faqs/graphics/tmap.html`
- `http://huebler.blogspot.com/2005/11/creating-maps-with-stata.html`
- i will not talk more about maps in stata (unless you insist)

R for mapping

- we already talked about R pros/cons in general
- much better than stata for maps, but still clunky
- problem with GIS and R is that much of GIS is point and click...
  - you want to zoom in/out maps
  - you want to click maps to change layers
  - you want to click maps to see features, etc
- solution: R and google maps (advanced)(CLICK links)
  - `http://www.omegahat.org/GoogleEarth/CityTemperatures/`
  - `http://cran.r-project.org/web/packages/RgoogleMaps/vignettes/RgoogleMaps-intro.pdf`
  - `http://cran.r-project.org/web/packages/RgoogleMaps/`
- rscript

arc map

- arc map is probably the most popular (we have it at lab)
- it is like ms word for writers
  - there are lots of features – you can do most of usual gis with it
  - it is quite visual / point-and-click
  - it often crashes, become slow, uses lots of memory, and it is expensive
- still it is ok software and industry standard
- great for specific tasks: e.g. sewer system, political redistricting
- we'll use geoda and qgis instead

geoda

- easy to use, point and click, free, spatial statistics
- LAB: on desktop; HOME: sign up and download from here
  `http://geodacenter.asu.edu/software/downloads`
- LAB: just open sample data.. HOME: columbus data from
  `http://geodacenter.asu.edu/sdata`
- you can open it in geoda
- then you can open columbus.dbf with say oo calc (excel may work too)
  and you can add a new var say "NEW_VAR,N,11,6"
  i guess, N is number and 11 is # of digits and 6 is # of decimal places

geoda – thematic maps

- table - merge table data
- map - percentile/quantile/std/cartogram/new map window
- explore - histogram/scatter plot
- note: you can click on polygons/bars/points AND select a bunch of them with rectangular-select

geoda – spatial statistics

- tools - weights - create
- id var is a unique id like state
- queen contiguity [draw a picture] (can also have distance or k-neighbors)
- open contiguity matrix in npp
  poly with id 1 has 2 neigh: 2, 3
  poly with id 2 has 3 neigh: 4, 3, 1
  and so on...
- having spatial matrix we can calculate some spatial statistics

geoda – spatial statistics

- space - univariate moran (like corr of i and Wi)
- univariate lisa - significance/values

qgis

- easy to use, point and click, free, good mapping with layers, lots of good plugins
  `http://www.qgis.org/`
- check out plugins !
- check it at home...
- you can use many useful tools: vector-data management tools

online software

- for a list see `http://aok.us.to/class/data_mgmt#gis`
- we will use geocommons
- there are lots of maps uploaded by other users
- you can pick base map (google, mapquest, etc)

geocommons: happiness

- let's reproduce this: `http://utdallas.edu/~ajo021000/tmp2/` (you can embed geocommons maps on your website)
- sign up `http://geocommons.com/`
- get my shapefiles aok2... at `http://aok.us.to/papers/gesis/`
- on geocommons
  - upload data (you may need to click several times)
  - map it - visual theme - inc ppp - colors - quantile - chose style
  - and we will add another layer...
  - add data - and do the same for other vars

geocommons: campaign contributions

- let's reproduce this: `http://geocommons.com/maps/62603`
- let's look at data `http://aok.us.to/tmp`
  - open contr.csv; we cannot use it as it is: run gis.do
  - open districts.kml
- upload both to geocommons
  - geocode csv data; change columns and select US
- and you are ready to map
- you may do it over time, do animations, etc
- search for other user's data on geocommons and add to your map...
- and there is more data `http://www.dalcoelections.org/`

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado

utdallas

- program `http://www.utdallas.edu/epps/gis/contact.html`
- intro `http://www.utdallas.edu/~briggs/`
- applied `http://www.utdallas.edu/~briggs/poec6381.html`
- data/resources

  `http://www.utdallas.edu/~briggs/other_gis.html`

harvard/mit

- main http://gis.harvard.edu
- newsletter
  http://gis.harvard.edu/icb/icb.do?keyword=k235&pageid=icb.page189865
  email to majordomo@mail.hmdc.harvard.edu
  this subscribenewsletter_at_lists_cga_harvard_edu
- data http://gis.harvard.edu/icb/icb.do?keyword=k235&tabgroupid=icb.tabgroup53814
- data http://libraries.mit.edu/gis

harvard free resources

- conferences (slides, video, etc)

  `http://gis.harvard.edu/icb/icb.do?keyword=k235&tabgroupid=icb.tabgroup110805`

- take some classes

  `http://gis.harvard.edu/icb/icb.do?keyword=k235&pageid=icb.page189848`

- blog, interesting things

  `http://gis.harvard.edu/icb/icb.do?keyword=k235&pageid=icb.page189849`

outline

misc

some examples

intuition

setup

software

next steps

if we have time, let's revisit ado