MODULE 5: Spatial Statistics in Epidemiology and Public Health
Lecture 2: Spatial Questions and Answers

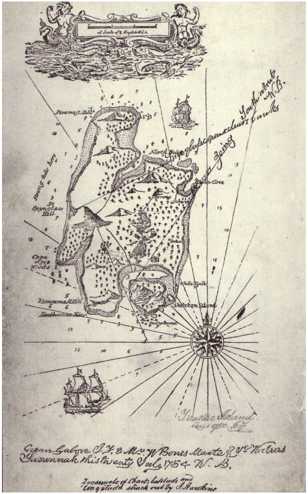
Jon Wakefield and **Lance Waller**

What can we do with a map?

- Most adventures begin with:

  "In my possession, I have a map..."

Treasure Island



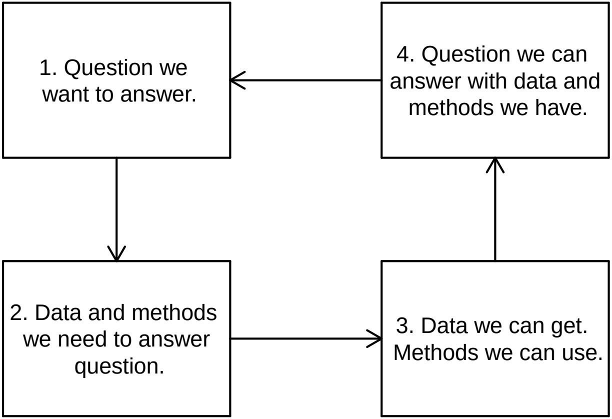
King Kong

Wishful thinking too (1772)

How can maps help us with spatial statistics?

- Spatial questions require:
  - Spatial data
  - Spatial methods
  - Spatial answers
- Maps frame questions, data, methods, answers in a spatial setting

The whirling vortex



Maps hold clues...

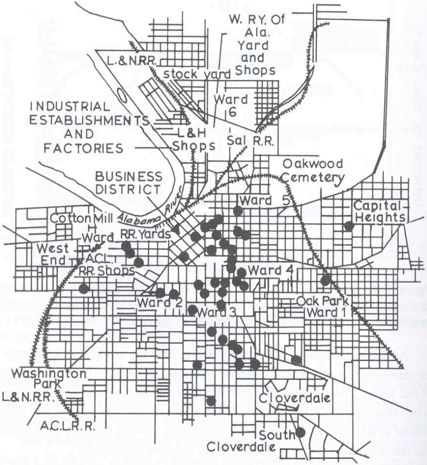


Figure 7–5. Distribution of cases of endemic typhus fever by residence, Montgomery, Alabama, 1922–1925. *Source:* Maxcy (1926).

...but may not reveal them immediately.

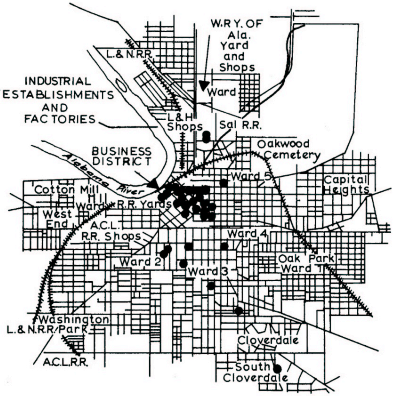
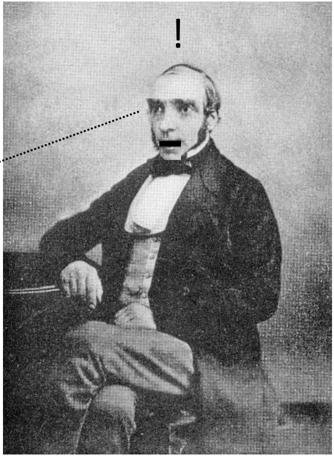


Figure 7–6. Distribution of cases of endemic typhus fever by place of employment or, if unemployed, by place of residence, Montgomery, Alabama, 1922–1925. *Source:* Maxcy (1926).

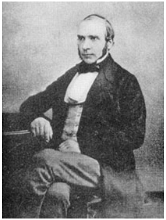
- Lillenfeld and Stolley (1994, *Foundations of Epidemiology, 3rd Ed.*. Oxford pp. 136-140).

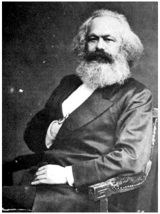
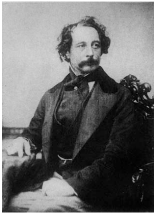
Maps and Health: John Snow, MD

Snow, J. (1949) *Snow on Cholera.*
Oxford University Press: London.



Aside: Victorian Portraiture, circa 1854



Short version of the John Snow, MD story

- "In 1854, Londoners were dropping like flies from cholera until Dr. Snow figured out that the bacteria were carried by water. The water pump he turned off, thereby saving countless lives, was near the site of this pub."
- John Snow Pub entry in *Access London* tour guide, Harper-Collins, 2005.

Truth a little more complicated and fascinating

- Brody et al. (2000) Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. *Lancet*
- Koch (2005) *Cartographies of Disease: Maps, Mapping, and Medicine*. ESRI Press
- Johnson (2006) *The Ghost Map*. Riverhead Books

Big data circa 1854 (from Koch, 2005)

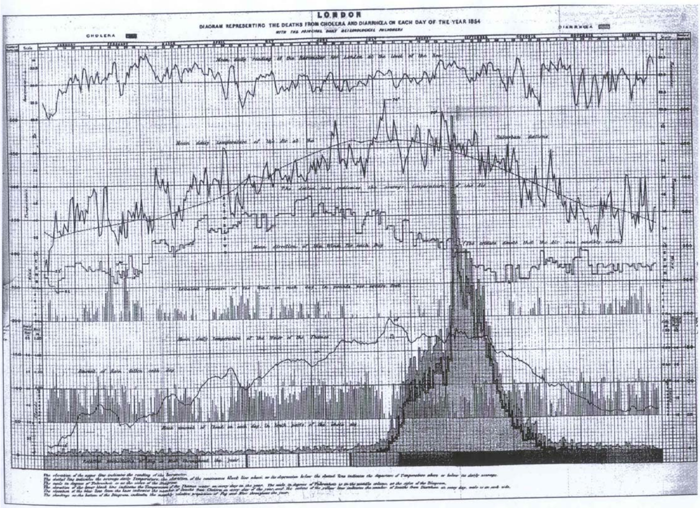
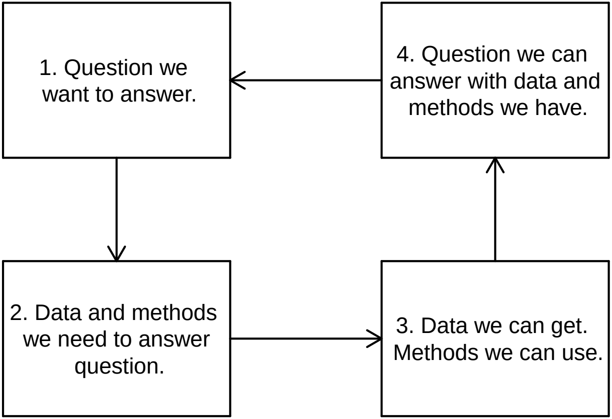


Figure 5.6 A graph of climatic variables joined to incidence of cholera (blue) and chronic diarrhea (yellow) in London, 1854. The map was based on readings from twenty-four urban recording stations in London and prepared by the General Board of Health for a report to both houses of Parliament.

Do maps tell the whole story?

- Contemporary interpretation of Snow's map: "On examining map given by Dr Snow, it would clearly appear that the centre of the outburst was a spot in Broad-street, close to which is the accused pump; and that cases were scattered all round this nearly in a circle, becoming less numerous as the exterior of the circle is approached. *This certainly looks more like the effect of an atmospheric cause than any other*; if it were owing to the water, why should not the cholera have prevailed equally everywhere where the water was drunk?" (Parkes, 1855).

The whirling vortex



What about today? COVID-19

- What questions do we *want* to answer?
- What data do we *need*?
- What data do we *have*?
- What questions *can* we answer?
- What do we need to know when allocating scare resources?
- What do we need to know when monitoring local risk of disease?

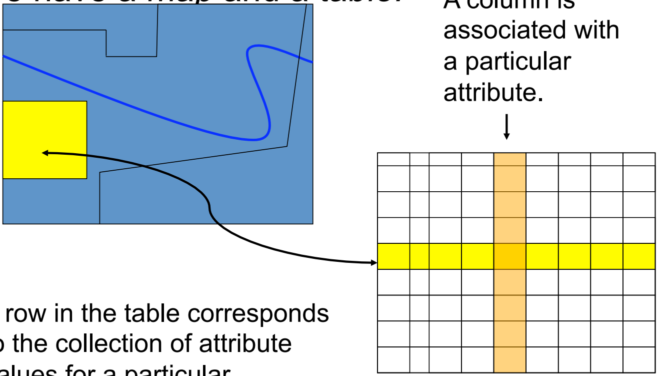
What questions can we answer with a map?

- Merriam-Webster online: Map = "a *representation* usually on a flat surface of the whole or part of an area."
- Note "representation" means "not an exact duplicate"!
- *Thematic* maps include *locations* and *attributes* associated with the locations.
- Think of a *map* of locations linked to a *table* of attribute values.

Maps and tables

We have a *map* and a *table.*

A column is associated with a particular attribute.



a row in the table corresponds to the collection of attribute values for a particular geographic feature in the map.

Geographic Information Systems (GIS)

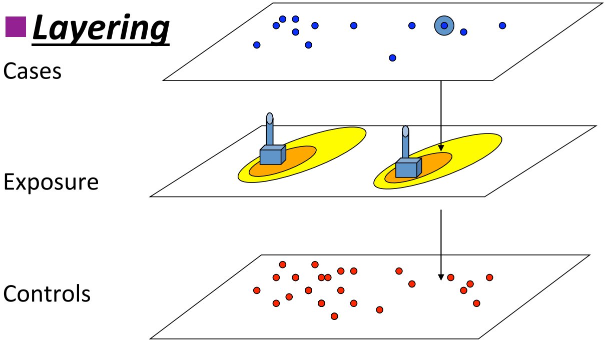
- A *geographic information system* is "a technology designed to capture, store, manipulate, analyze, and visualize georeferenced data" (Goodchild, Parks, and Steyaert 1993).
- GIS is a database system containing locations for every value and allowing operations (search, sorting, etc.) based on locations as well as attributes.
- Allows maps of attribute values.

What does a GIS do?

- Think of data sets as “layers”.
- For example:
  - One layer of case locations (points).
  - One layer of road locations (lines).
  - One layer of population levels (areas).
  - One layer of vegetation type (satellite image (raster)).

Basic GIS operation 1: Layering

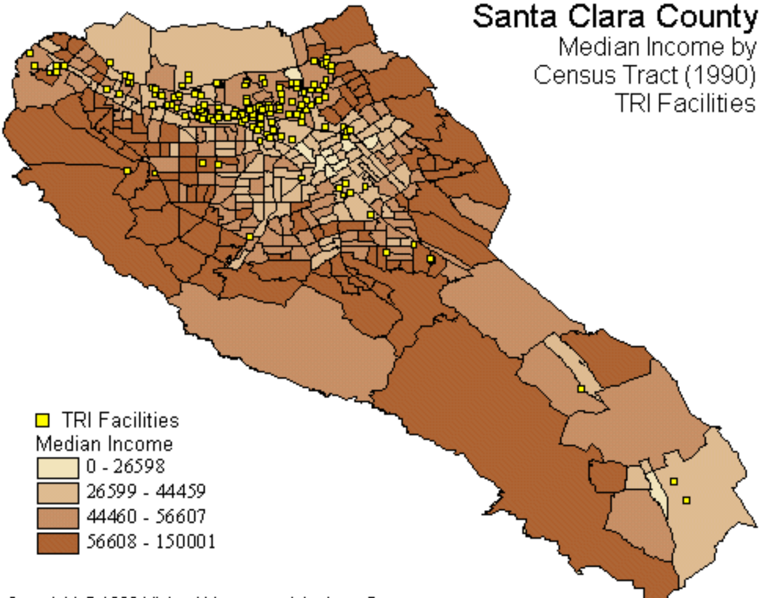
Layering



What questions can we answer with layering?

- Do certain features in layer A occur in the same (or similar locations) as features in layer B.
- Examples
  - Spatial case-control study.
  - Bars and DUI arrests.
  - Library locations and school performance.
  - Environmental justice.

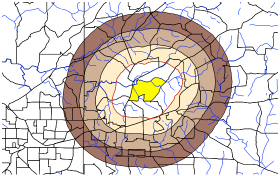
Layering example: Environmental justice



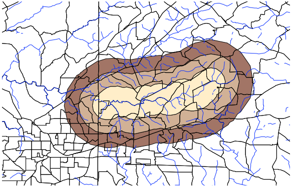
Basic GIS operation 2: Buffering

- ***Buffering***
  - Find areas within a user-specified distance of:
    - points
    - lines
    - areas

Buffers around an area



Buffers around a line feature



What questions can you answer with layering and buffering?

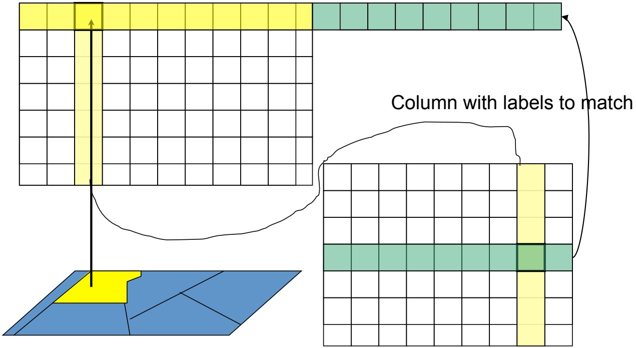
- Layer 1: Pollution sources
- Layer 2: Residents experiencing health effects (cases)
- Layer 3: Residents without health effects (controls)
- Question 1: What fraction of cases are within a given distance of a pollution source?
- Question 2: What fraction of controls are within a given distance of a pollution source?
- Question 3: Are these the same?
- This is the quintessential GIS environmental health study.

Basic GIS operation 3: Joining

- The spatial "join":
- Have:
  - Attribute table linked to map
  - 2nd table of data over same features
  - Common identifier in both
- Want:
  - Add (join) attributes in 2nd table to first table
  - How: Link tables based on common attributes
  - Need: One-to-one correspondence

Basic GIS operation 3: Spatial Join

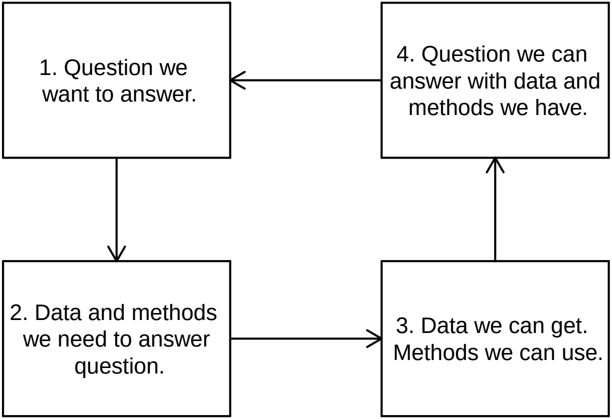
Visually...



Basic GIS operations

- Layering
- Buffering
- Joining
- All GISs do these. They do more, but all of these basic operations are included.

The whirling vortex



GIS analysis

- What can you do with these three operations?
- The key to GIS analysis is to break your problem down into steps consisting of these operations.
- What question(s) do you have?
- What data would you need?
- What data can you get?
- Can you layer, buffer, join data to enable summaries relating to your question (or parts of it)?
- What answers can you provide?

Scenario 1: COVID surveillance

- Your group is tasked with providing a more accurate map of current risks of COVID infection and mortality at the county level.
- What questions?
- What data do you want?
- What data can you get (in what time)?
- What questions can you answer (in what time)?

Scenario 2: Testing Site Selection

- You are working with your local Department of Public Health, several healthcare systems, private companies, and the CDC to assess testing coverage in your city. You want to (a) assess equity in access across sociodemographic subpopulations, and (b) you want to decide where to place a limited number of mobile testing sites.
- What questions?
- What data do you want?
- What data can you get (in what time)?
- What questions can you answer (in what time)?

GIS and statistics

- Some statistical tools and toolboxes available in GIS, but few and specific.
- GIS and statistical languages based on objects and operations, but different objects and operations.
  - ArcGIS: geodatabases, shapefiles
  - SAS: SAS data sets
  - R: objects, datafames, etc.
- Python scripts, R bridge to ArcGIS, pipelines...still evolving.

R, mapping, and GIS

Basic tools:

- Reading GIS data (e.g., shapefiles) into R.
  - `maptools` package allows you to read in shapefiles directly to R.
- Mapping using R graphics
  - Examples using `ggplot`, `leaflet`, `mapview`, `tmap` packages in Moraga (2020).
- Exporting data to GIS.
  - Inelegant but effective solution: `csv` file via `write.table`, but must include identifier to match to map file.
  - More direct sharing of data objects between QGIS (qgis.org) and R

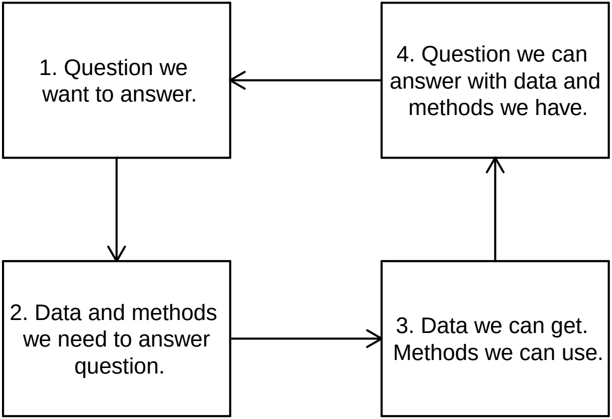
Disciplines and spatial statistics

- Each disciplines has own rules of thumb with spatial analysis.
- Key questions, methods, and *data* vary between disciplines.
- Geography: Spatial autocorrelation (Moran's I, LISAs, spatial regressions).
- Ecology: Associations and diffusion (Mantel tests).
- Criminology: Hotspots.
- Epidemiology: Clusters, Poisson/logistic regression.
- We'll see examples from each of these over the next 2 days.

Role of statistics

- Different methods are fine for different questions, or different data restrictions.
- *Spatial thinking* uses location to identify processes driving pattern.
  - Why did this happen *here* vs. *there*?
  - What data can I link by location?
- *Statistical thinking* places question in probabilistic setting, and builds inference on data-based summaries.
  - Estimation of associations and prediction of new observations (with uncertainty).
  - Detection and power (finding effect when it is there) while avoiding false positives (not finding effect when it is not there).
- *Spatial statistical thinking* (Waller, 2014, *Spatial Statistics*)
  - Estimates/predictions vary with location (and so does uncertainty).
  - False positive rates and power vary by location (I have a better chance of finding an effect *here* rather than *there*.

The whirling vortex



Summary

- Maps are cool.
- Maps place data spatially.
- Spatial data enable answers for spatial answers.
- Spatial data also allow spatial statistics.
- So what spatial statistics can we do?

References

- Koch T (2005) *Cartographies of Disease: Maps, Mapping and Medicine*, Redlands: ESRI Press.
- Moraga P (2020) *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Boca Raton: Chapman & Hall/CRC.
- Waller LA (2014) Putting spatial statistics (back) on the map. *Spatial Statistics*. **9**, 4-19. DOI: 10.1016/j.spasta.2014.03.007.
- Waller LA (2017) Mapping in Public Health. In *Mapping Across Academia*, Brunn, S.D. and Dodge, M., eds. Dordrecht: Springer.