Using Geospatial Methods with Other Health and Environmental Data to Identify Populations

Ellen K. Cromley, PhD
Consultant, Health Geographer

ellen.cromley@gmail.com

Purpose and Outline

To illustrate the use of geospatial methods to identify populations from data sources other than electronic health records or the U.S. Census

- The spatial view of data and spatial sampling
- Locating populations--
  - From their places of residence
  - From administrative records
  - From their social networks
  - From their activity locations and activity spaces
  - Taking into account environmental exposure
- Where can we go from here?

Views of Data: Tabular

Tabular View of Two Databases (The Same)

NAME	GRADE	NAME	GRADE
Andy	B	Andy	B
Bob	B	Bob	B
Carmela	C	Carmela	C
Dave	A	Dave	A
Ed	A	Ed	A
Felicia	C	Felicia	C
Gordon	A	Gordon	A
Hank	B	Hank	B
Inez	C	Inez	C

Views of Data: Statistical

Statistical View of Two Databases (The Same)

Grade	Frequency
A	3
B	3
C	3

Grade	Frequency
A	3
B	3
C	3

Views of Data: Spatial

Spatial View of Two Databases (Not the Same)

Teacher				Teacher		
B	C	A		A	A	A
C	A	B		B	B	B
A	B	C		C	C	C

- The locations of the observations are part of the data record
- The location data are necessary to support spatial data analysis

Implications for Sampling

- Samples in geographic space
  Such samples select observations from a population that is itself geographically distributed

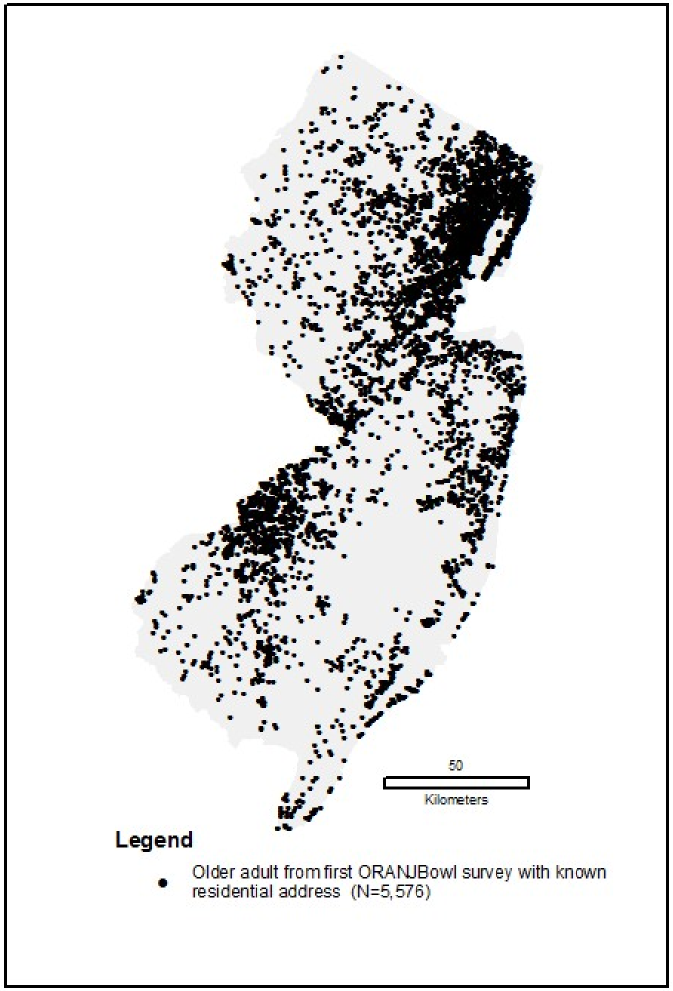
  Every sample of people or other entities located on the earth's surface is implicitly a spatial sample

  A random sample of all people is not a random sample of all places unless people are uniformly distributed

- Geospatial technologies make the spatial basis of evidence explicit so that the sample captures the spatial characteristics of the population

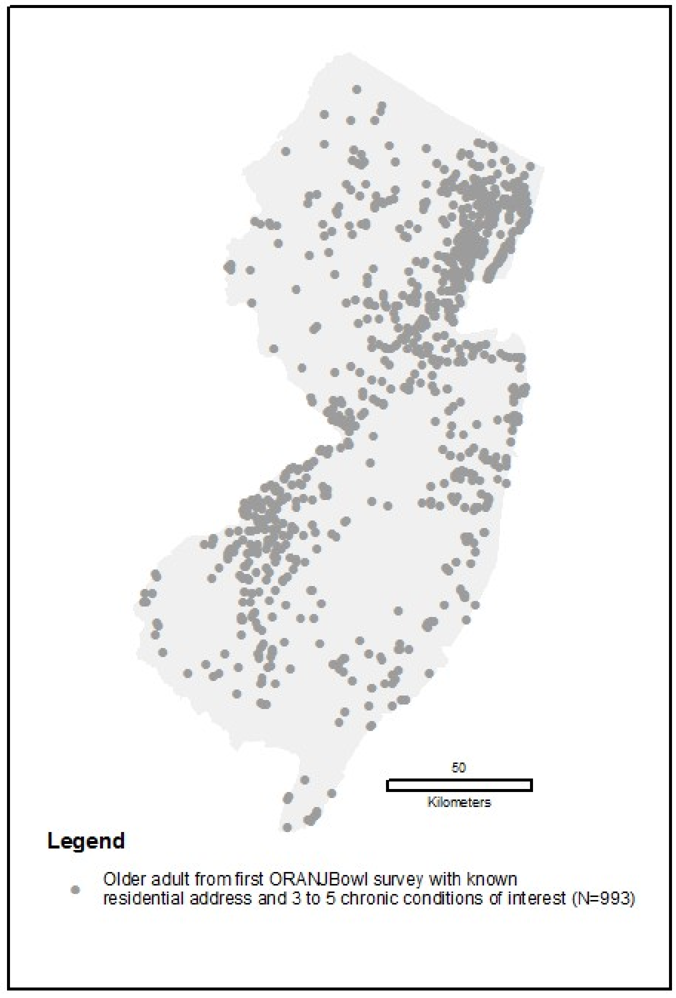
Studies Based on Residential Locations

- Locating older adults with multiple chronic conditions
- ORANJBowl telephone survey rachelpruchno.net/OB.html
- Residential locations of participants from first survey geocoded to census block centroids
- Combinations of three of the following: arthritis, diabetes, heart disease, hypertension, pulmonary disease

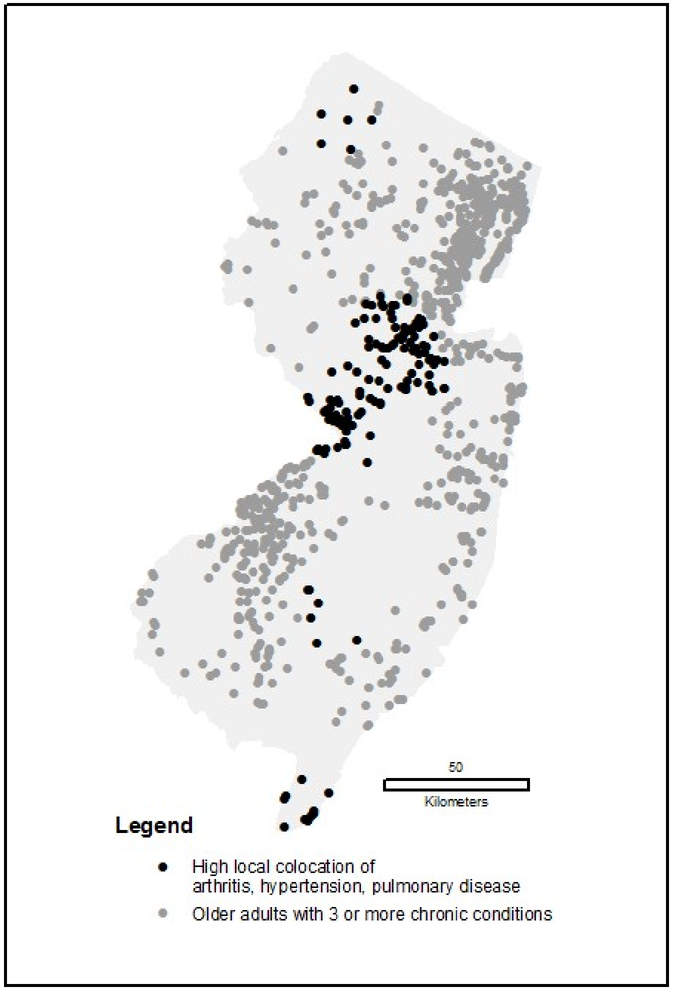


Locating People with Multiple Conditions

- Identified older adults with 3-5 chronic conditions
- Calculated local colocation quotients as a local measure of spatial association of conditions among older adults
- Mapped and tested the significance of any observed areas of spatial association



Spatial Patterns of Multiple Chronic Conditions



- Statewide proportion of older adults with 3-5 chronic conditions including arthritis-hypertension-pulmonary disease was 38 %
- Proportion of highlighted 155 older adults with 3-5 chronic conditions including arthritis-hypertension-pulmonary disease was 50 %

Studies of People in Group Quarters

- People who do not live in housing units reside in group quarters (Census Bureau definition)

- Types
  Institutional—prisons, nursing homes, inpatient mental health facilities
  Non-institutional--college dormitories, military barracks, group homes, shelters

- Size and distribution vary widely in space

Long Term Care Facilities in Massachusetts

- Data from MassGIS
  www.mass.gov/orgs/massgis-bureau-of-geographic-information

- "Global" view of capacity
  N = 740 facilities
  Total beds = 65,272

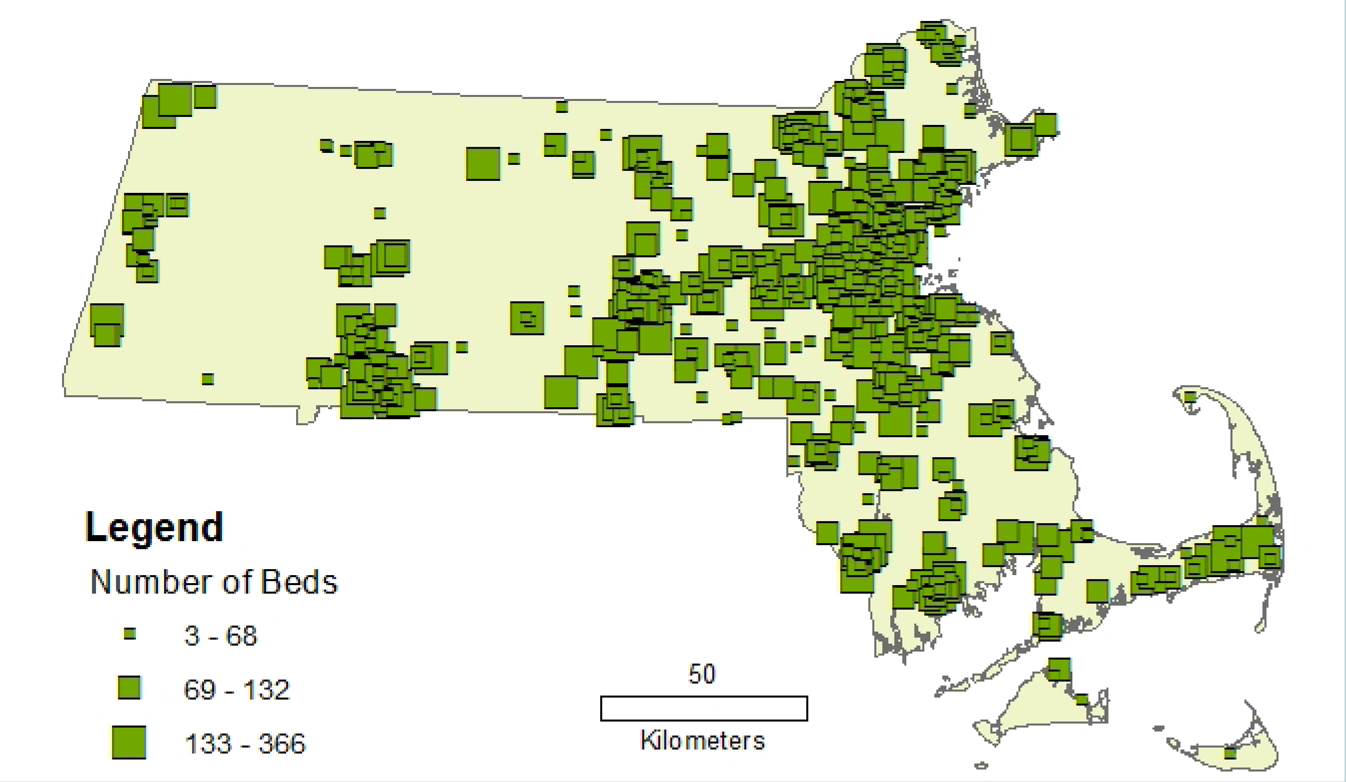
  Mean = 88.2
  Median = 83
  Min = 3, Max = 366
  Standard deviation = 52.4

Spatial Distribution of Facilities by Size

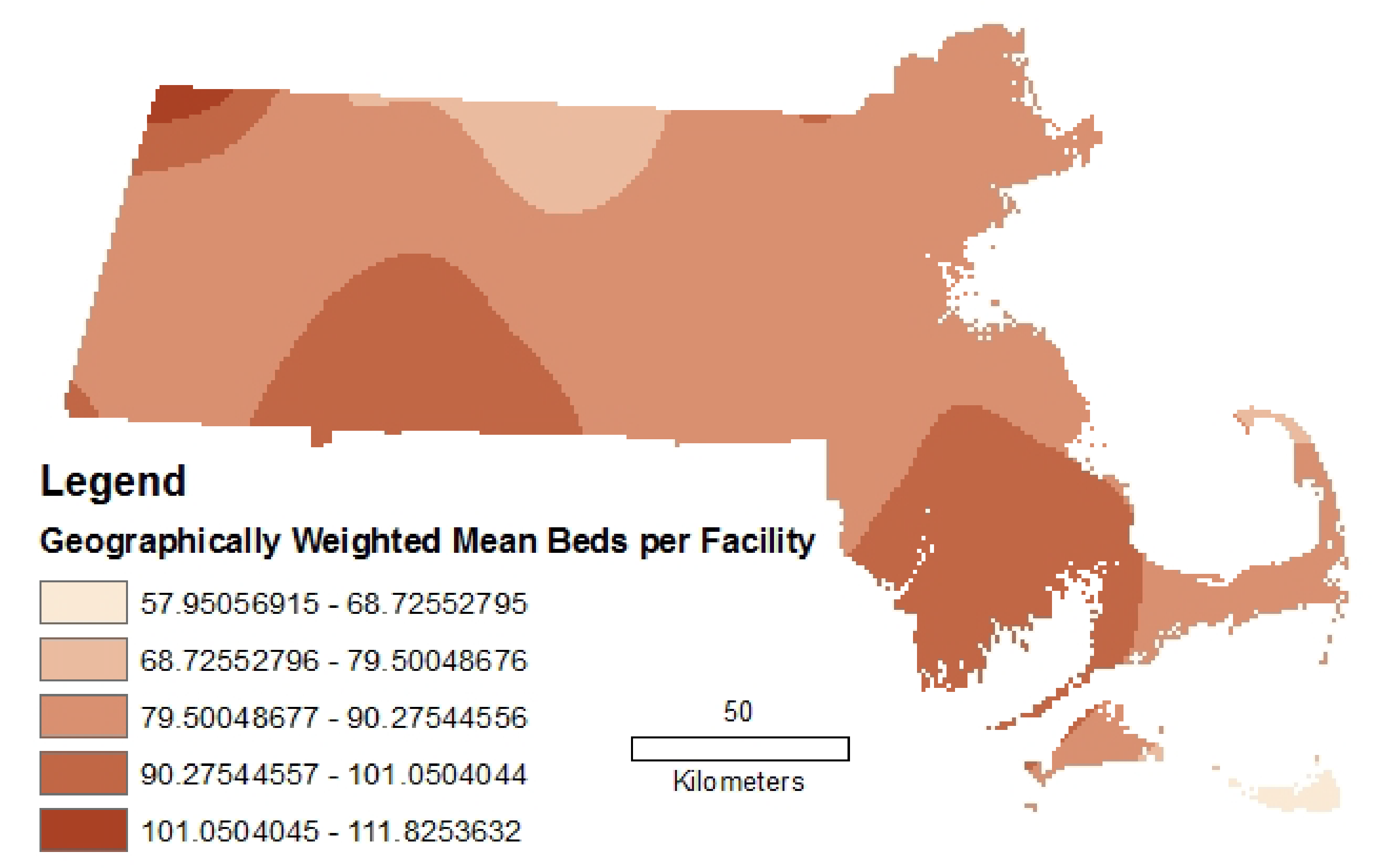
- “Local” view of capacity

Is facility size important in sampling?

Is capacity the same in every region?



Geographically Weighted Mean Capacities



Studies Based on Administrative Records

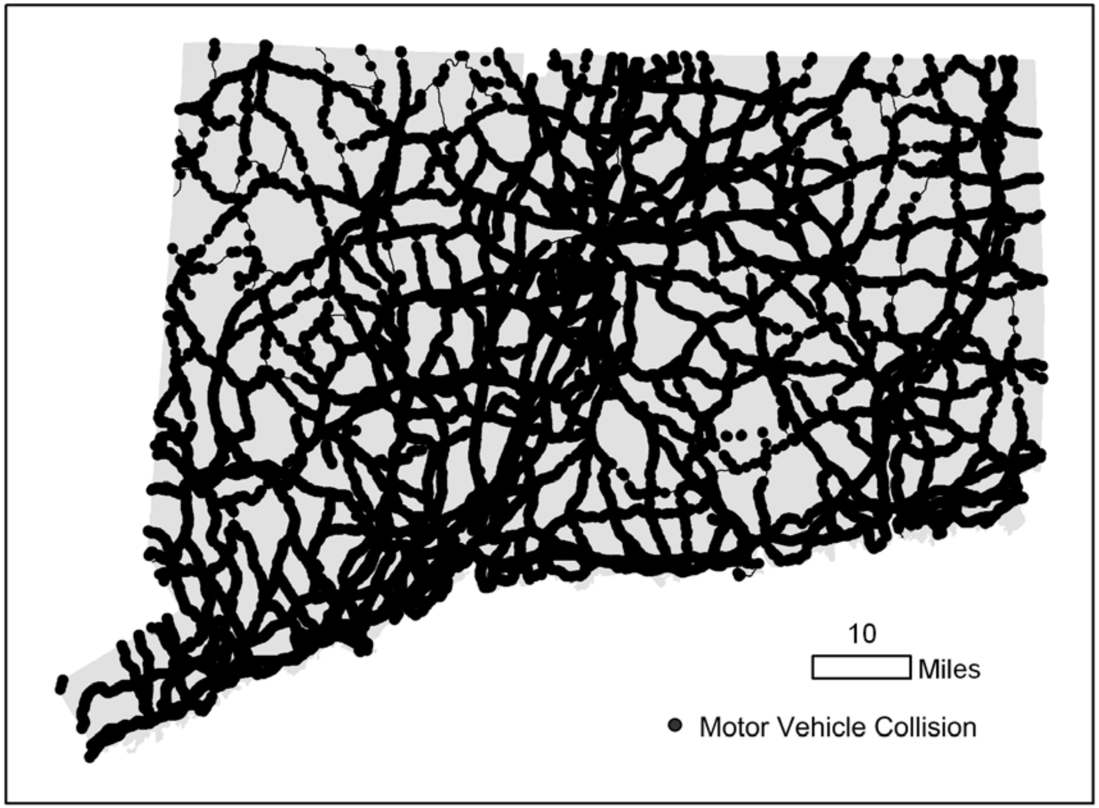
- Definition (Statistics Canada | Statistique Canada)
  Data collected for the purpose of carrying out various non-statistical programs

- Examples
  Vital records (births and deaths)
  Registries (immunization records, tumor registries)
  Health and social services records
  Public safety reports (collisions, shootings, fires)

- Administrative record linkage

- Data privacy and confidentiality

Connecticut CODES Project



- Police accident reports from state DOT
- Map shows collisions on federal and state roads 1995-1996 (N=124,053)
- Where to target an intervention to reduce fixed-object collisions?

Identifying Collisions in Local Areas

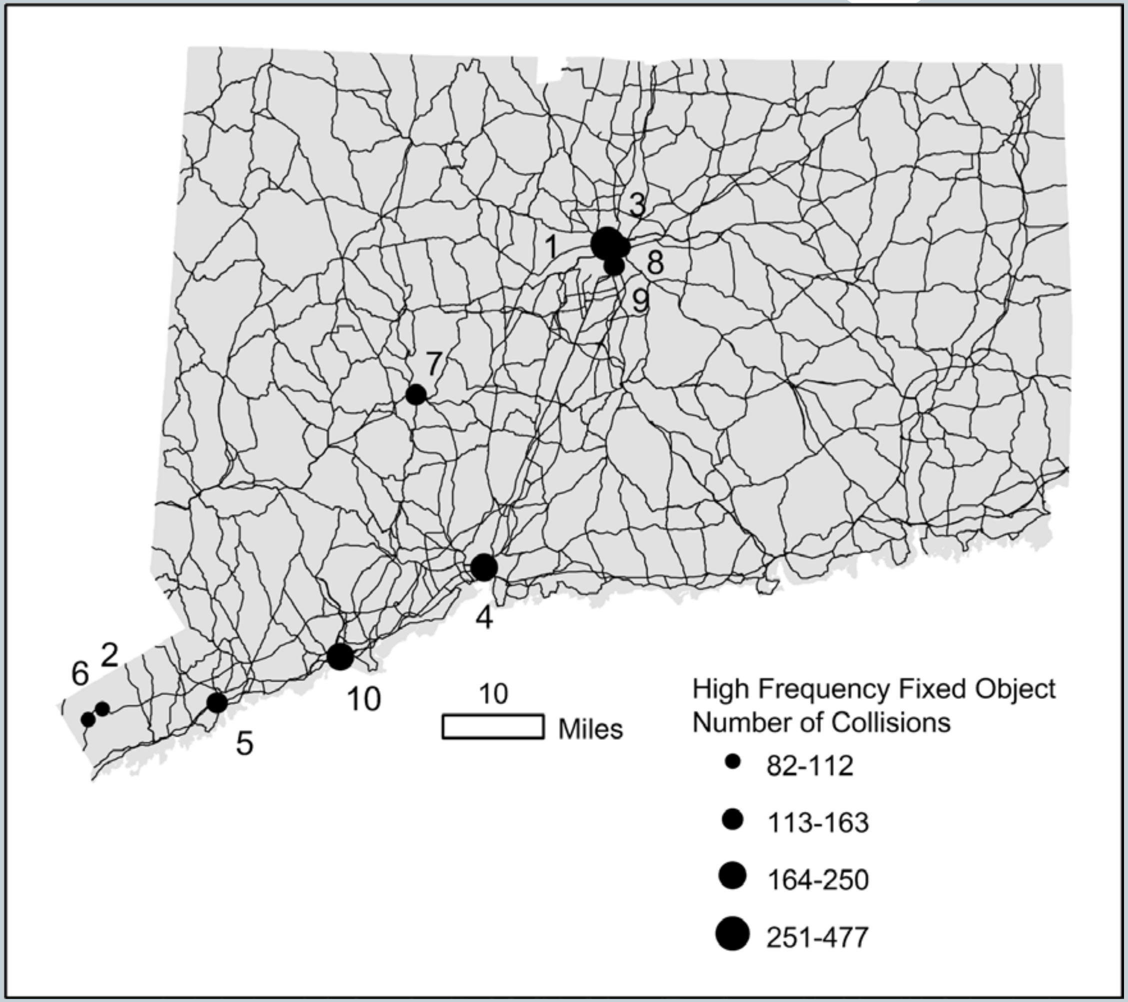
- **Geospatial methods**

  Used a box-shaped kernel based on stopping distance to group collisions along road segments

  Identified places with high numbers of collisions by type of collision

  Eliminated overlapping “places”

Fixed Object Collision Places



- 10 places with highest frequency of fixed object collisions
- Geographically distinct

Local Proportions

Place ID	Fixed Object	Rain or Snow	Dry Road	Daylight	Age 25-44	Male	Too Fast for Conditions
1	**0.29 z=5.23**	**0.46 z=11.19**	**0.46 z=-8.86**	0.71 z=.65	**0.52 z=2.82**	0.60 z=-.09	**0.27 z=9.75**
2	**0.91 z=16.48**	**0.66 z=8.87**	**0.10 z=-10.55**	0.76 z=-.86	0.51 z=1.16	0.58 z=1.43	**0.62 z=13.61**
3	**0.55 z=9.75**	**.39 z=3.82**	**0.48 z=-3.99**	0.66 z=-.86	0.51 z=1.25	0.68 z=-.61	**0.62 z=9.87**
4	**0.29 z=3.48**	0.25 z=.38	0.65 z=-.18	0.67 z=-1.02	0.49 z=1.10	0.58 z=-.73	**0.26 z=5.81**
5	**0.40 z=6.65**	**0.40 z=4.61**	**0.48 z=-4.73**	**0.52 z=-4.78**	**0.54 z=2.24**	0.61 z=.14	**0.29 z=6.36**
6	**0.70 z=11.61**	**0.69 z=9.52**	**0.20 z=-8.56**	0.78 z=1.71	0.54 z=1.58	0.54 z=-1.18	**0.57 z=12.16**
7	**0.37 z=5.48**	**0.38 z=3.95**	**0.57 z=-2.14**	0.73 z=.84	0.48 z=.69	0.62 z=.26	**0.31 z=6.93**
8	**0.35 z=4.82**	**0.39 z=4.42**	**0.49 z=-4.28**	0.71 z=.37	0.52 z=1.52	**0.52 z=-2.32**	**0.29 z=6.27**
9	**0.44 z=6.78**	**0.39 z=3.83**	**0.54 z=-2.62**	0.64 z=-1.45	0.51 z=1.17	00.65 z=.90	**0.27 z=5.03**
10	0.22 z=1.09	0.29 z=1.75	0.62 z=-1.01	0.67 z=-.89	0.51 z=1.58	0.66 z=1.81	**0.19 z=2.84**
State	0.19	0.24	0.65	0.70	0.46	0.61	0.12

Local Odds Ratios

Place ID	Rain or Snow	Dry Road	Daylight	Age 25-44	Male	Too Fast for Conditions
1	**2.79 z=4.91**	**0.29 z=-5.58**	**0.45 z=-3.76**	0.73 z=-1.56	0.87 z=-.69	**11.09 z=10.08**
2	2.83 z=1.30	0.22 z=-1.63	1.27 z=.27	0.00 z=.00	1.70 z=.66	4.71 z=1.78
3	**4.80 z=3.75**	**0.17 z=-4.46**	**0.36 z=-2.52**	1.34 z=.80	0.56 z=-1.54	**17.50 z=5.65**
4	**2.89 z=3.21**	**0.19 z=-5.16**	**0.30 z=-3.82**	1.25 z=.73	0.63 z=-1.53	**8.62 z=6.14**
5	**3.06 z=3.33**	**.28 z=-3.74**	**.48 z=-2.27**	1.32 z=.87	1.03 z=.09	**3.11 z=3.17**
6	1.05 z=.09	0.54 z=-1.10	0.23 z=-1.86	1.14 z=.27	1.14 z=.27	**2.64 z=1.98**
7	**5.53 z=4.65**	**0.16 z=-4.87**	1.03 z=.08	1.24 z=.65	**.36 z=-2.90**	**16.55 z=6.41**
8	**3.04 z=3.21**	**0.32 z=-3.27**	**0.35 z=-.292**	1.20 z=.55	1.20 z=.55	**10.50 z=5.81**
9	**2.33 z=2.25**	**0.29 z=-3.28**	**0.46 z=-2.02**	0.83 z=-.52	1.02 z=.06	**3.32 z=2.85**
10	0.94 z=-.19	0.78 z=-.81	0.32 z=-3.59	0.80 z=-.72	0.92 z=-.25	**3.86 z=3.81**
State Odds Ratio	1.76	0.50	0.40	1.13	1.10	6.38

Error in Administrative Records

- Challenges
  Errors due to inclusion criteria
  Errors due to coding of thematic attributes
  Errors in coding spatial attributes

- Responses to challenges
  Talk to the people who collect and code the data
  Engage with agencies to improve data quality
  "Adopt robust methods of analysis"–

  Waldo Tobler, Analytic Cartographer,
  Member of the National Academy of Sciences

Studies Based on Social Networks

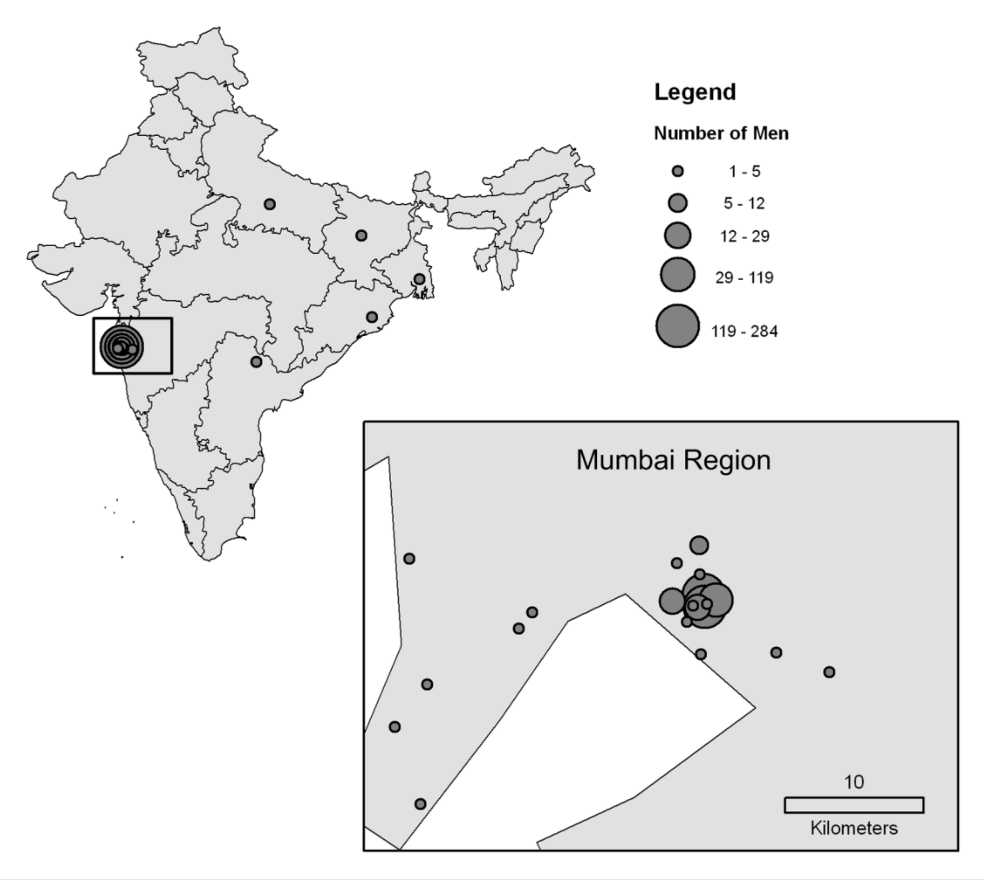
- Social networks function in geographic space and they can be mapped
- In respondent-drive sampling, examine the distribution of locations for recruiting “seed” participants
- Include mechanisms for monitoring the spatial locations of network members recruited into the research study

Studies Based on Activity Sites

- Places outside the home where people engage in activities can be used for locating populations
  - Health service delivery sites
  - Workplaces and schools
  - Bars
  - ....and many more
- Geospatial methods can be used to map these venues to show their spatial distribution
- A critical consideration in venue-based research is understanding who is seen at these venues

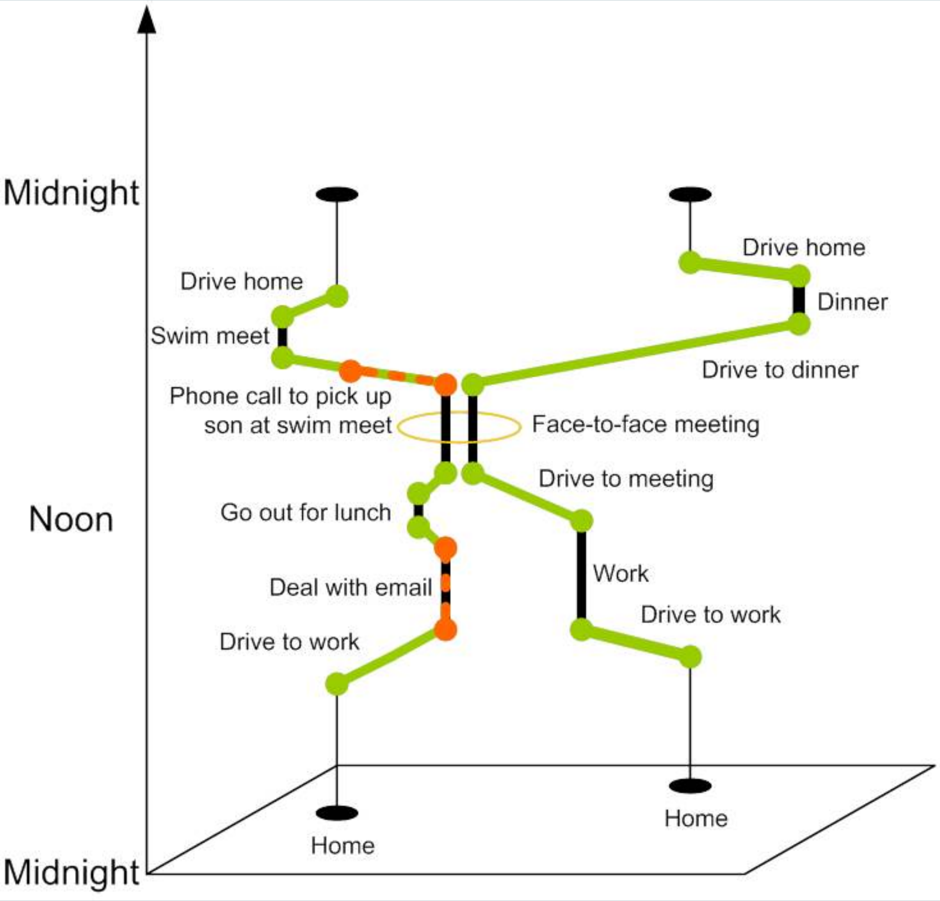
The ASHRA Project

- Study of alcohol and sexual risk in three low-income communities in Mumbai, India
- 640 drank with friends at places in one of the three study communities
- 111 drank in other widely-dispersed locations in the Mumbai region and elsewhere



Locating People Who Share Activity Spaces

- Home base, activity sites, routes of travel
- Various ways we can locate populations for health research are represented in this graphic
- With geo-enabled devices, the information can be collected continuously in real time



Grouping People Based on Activity Sites

- Make pair-wise comparisons of study participants
- Calculate an index of association based on co-location in space or in time and space

$$I = \frac{2 \times (\text{Number of activity sites in common})}{\text{Number of sites Person 1} + \text{Number of sites Person 2}}$$

- Analyze resulting similarity measures by using scaling techniques to group people with similar activity patterns in space or in time and space

Exposure Studies

- Some epidemiological study designs require information on exposure (cohort and case-control studies)
- Dominant view of what geospatial methods contribute to health research
  - Map air quality made by taking samples of space
- Geospatial technology can be used to assess exposure without reference to such "maps"
  - Personal exposure monitoring
  - Monitoring environmental conditions at residences, other activity sites, or travel routes and assigning measurements to study subjects
- Contextual analyses can be "geographic" but not "spatial"

Where Can We Go from Here?

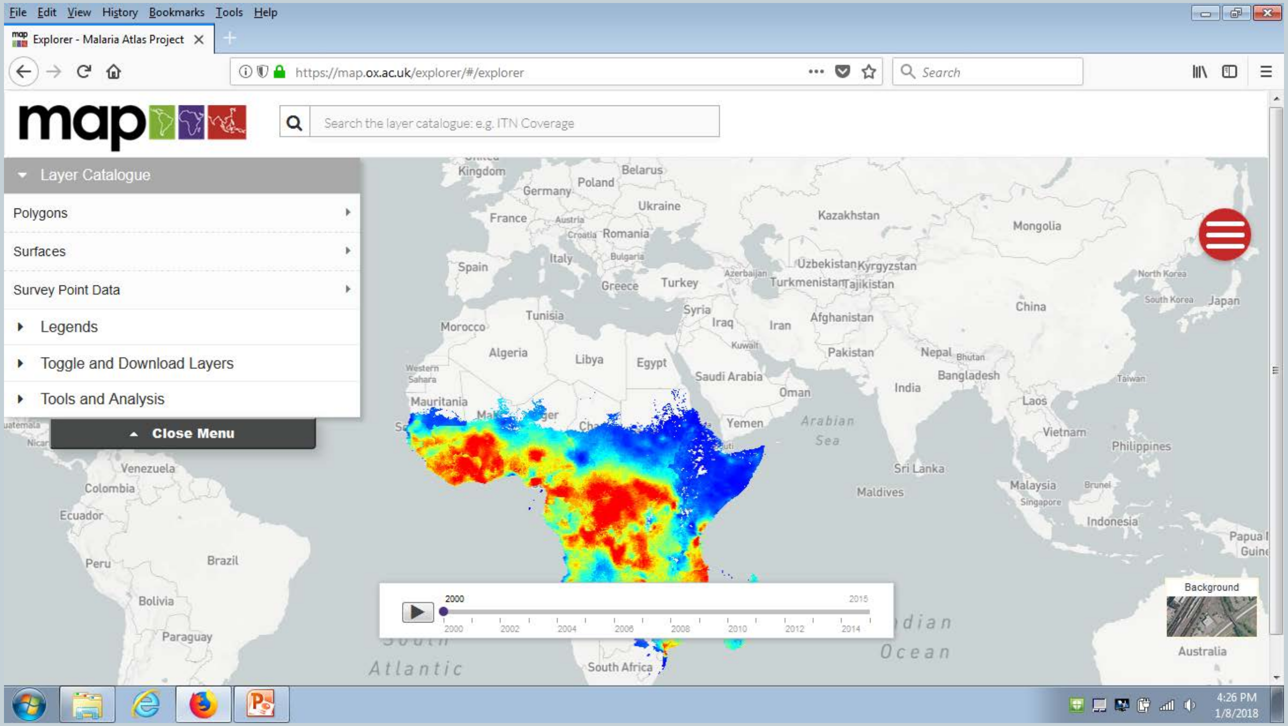
- The key impact of geospatial technologies has been to enable us to study large areas at high levels of detail (maintaining individual locations and attributes)
- Building spatial data commons
- Adopting a spatial analytic framework for health science explicitly addressing spatially varying processes

Building Data Commons

- The Malaria Atlas Project aims to disseminate free, accurate and up-to-date information on malaria and associated topics, organized on a geographical basis (map.ox.ac.uk/)

Open access policies

Subscribes to GATHER



Putting it All Together: Matlab

- Uses geospatial technology to determine
  - How cholera vaccine efficacy varies spatially in the study area
  - What ecological socio-environmental variables are related to cholera vaccine efficacy (which variables are effect modifiers?)
  - How protective efficacy varies with access to treatment facilities (is access a spatial confounder?)
  - Whether cholera incidence in the placebo group is related to vaccine coverage rates (is herd immunity important?)
- Currently integrating social networks into a spatial analytic framework for vaccine trial evaluation
- Cholera vaccine efficacy as a spatially-varying process

Questions?