PRACTICAL APPLICATIONS OF GEOSPATIAL ANALYSIS IN RHEUMATOLOGY RESEARCH

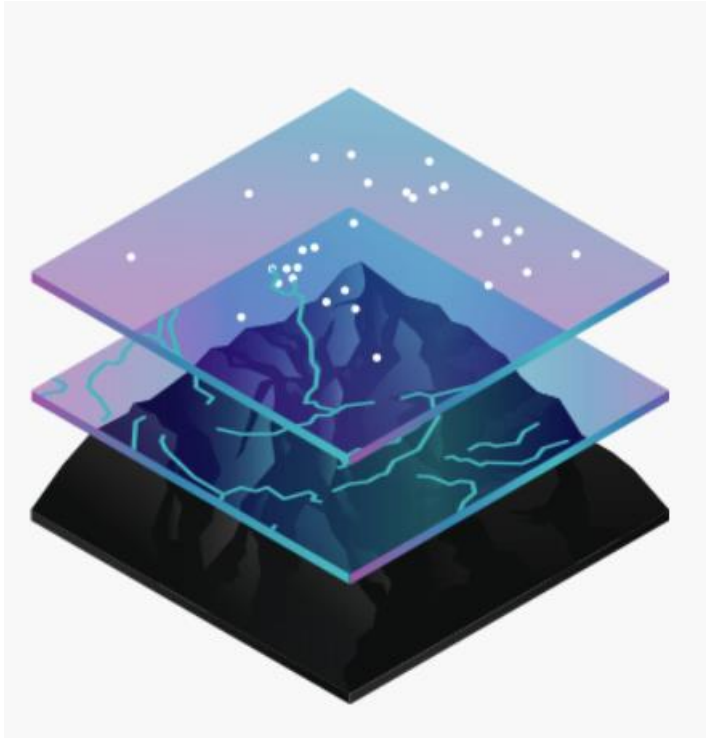
LEAH SANTACROCE

STATISTICAL PROGRAMMER AND ANALYST

BRIGHAM AND WOMEN'S HOSPITAL

INTRO TO GEOSPATIAL ANALYSIS

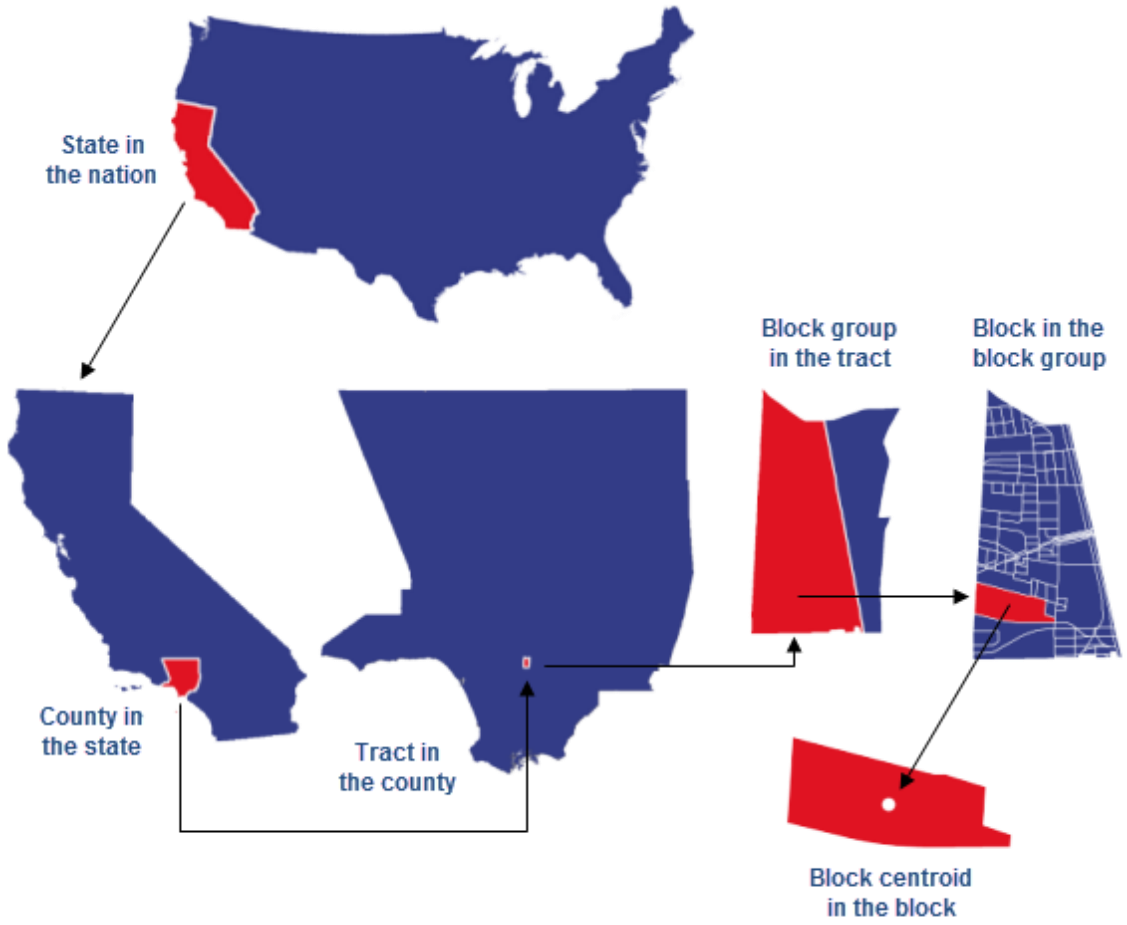
- Geospatial data: Location and geography information
- GIS
  - Geographic Information Systems
  - Software systems to manage and analyze geospatial data
    - ArcGIS, GeoDa, QGIS
- Many types of Geospatial Analysis
  - Geocoding, Network analysis, Spatial Autocorrelation



UNITS OF ANALYSIS IN GEOGRAPHY

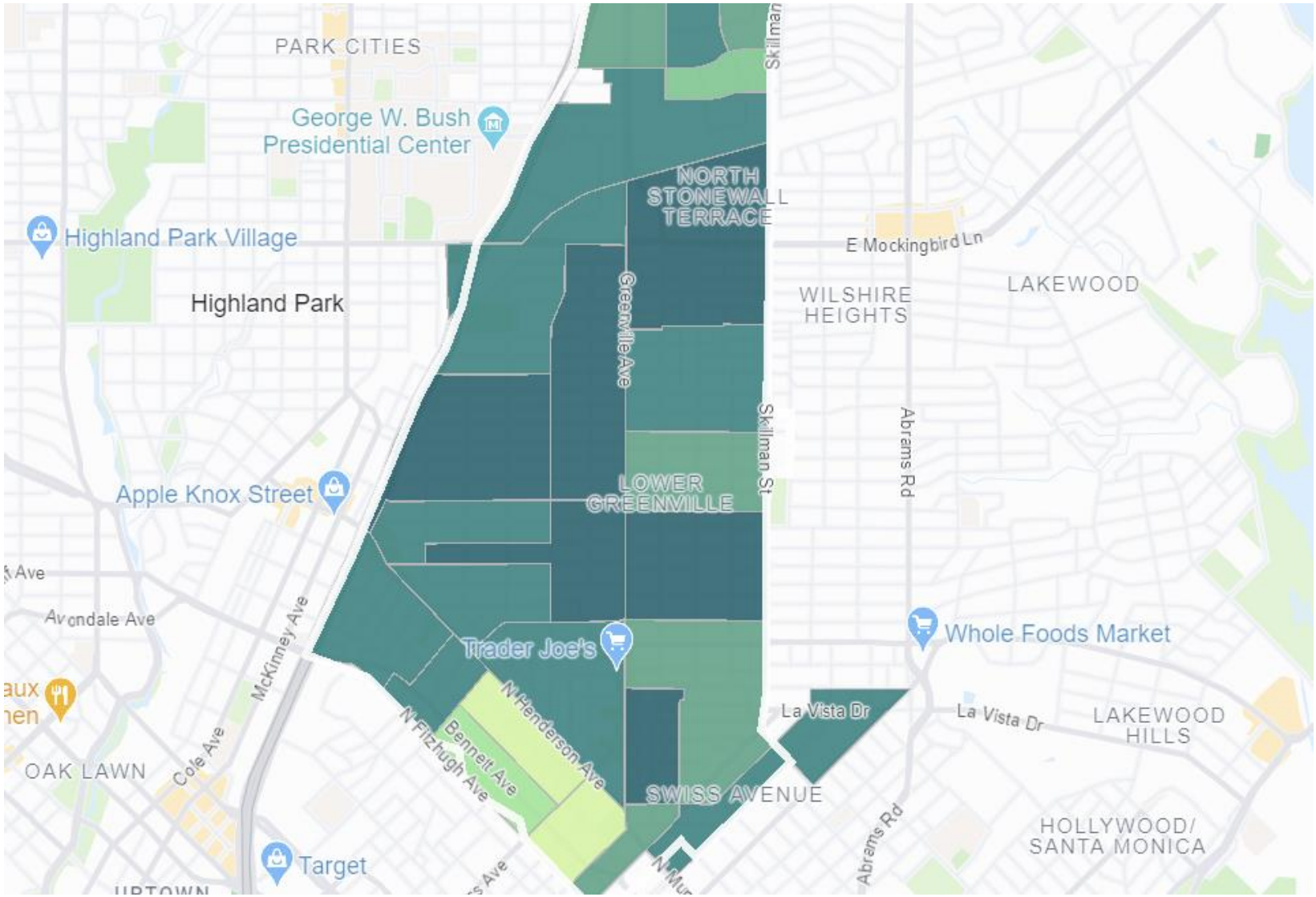
- Census tracts in a city, state, province
- Neighborhoods in a city
- States/Provinces in a country

- These will differ greatly depending on what country you are looking at!



A NOTE ON ZIP CODES

- Does not represent an area, rather a collection of roads
- Includes many census tracts, but not perfectly
  - Worse in urban areas



https://carto.com/blog/zip-codes-spatial-analysis

GEOCODING

- Process of transforming a description of a location to geographic coordinates on the Earth's surface
- Example Inputs
  - Boston, MA; Copenhagen, Denmark (not precise)
  - 123 Commonwealth Avenue, Boston MA (precise)
- Outputs
  - X,Y (Latitude and Longitude coordinates)
  - Point on map

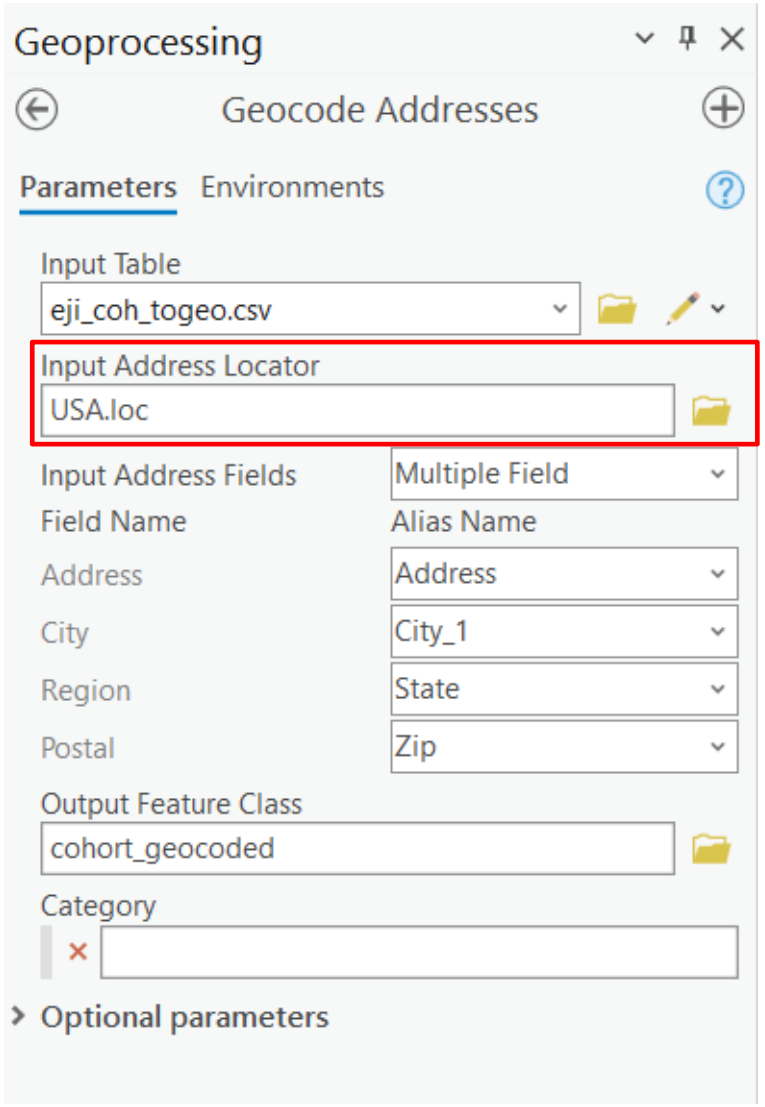
GEOSPATIAL ANALYSIS WITH SENSITIVE DATA

- Traditional software sends addresses online to third parties to geocode
  - NOT HIPAA compliant
- Options:
  - Get access to a local geocoder
    - Many institutions have these
  - DeGAUSS geocoder
    - https://degauss.org/
- Software itself (ArcGIS) runs on local computer
  - https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm

GEOCODING STEP 1: CLEANING ADDRESSES

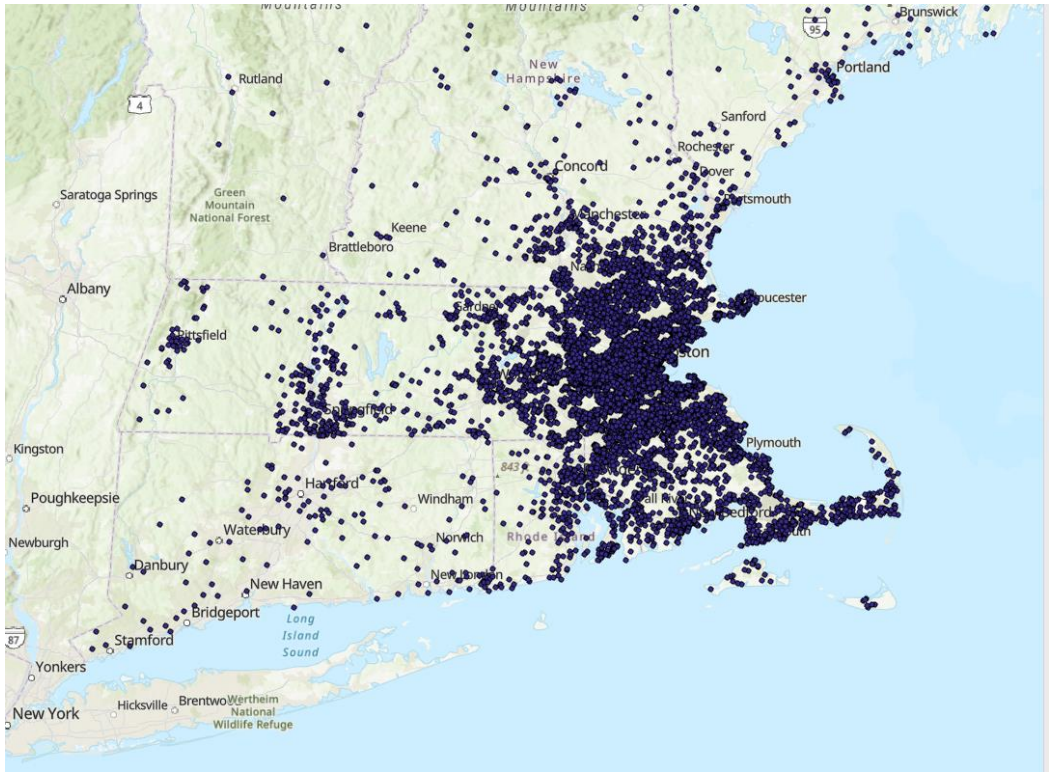
- Most time-consuming part of geocoding!
- Sort your addresses to get an idea of your data
  - If the address doesn't start with a number, it likely won't geocode properly

Original Address	Cleaned Address	Action
Apt 51 1325 Commonwealth Ave	1325 Commonwealth Ave	Removed apartment number
The Landmark Building, 401 Park Drive	401 Park Drive	Removed building "name"
PO Box 152		Delete row, do not geocode
/123 Beacon Street	123 Beacon Street	Removed mistyped character



GEOCODING STEP 2: UPLOAD TABLE AND GEOCODE

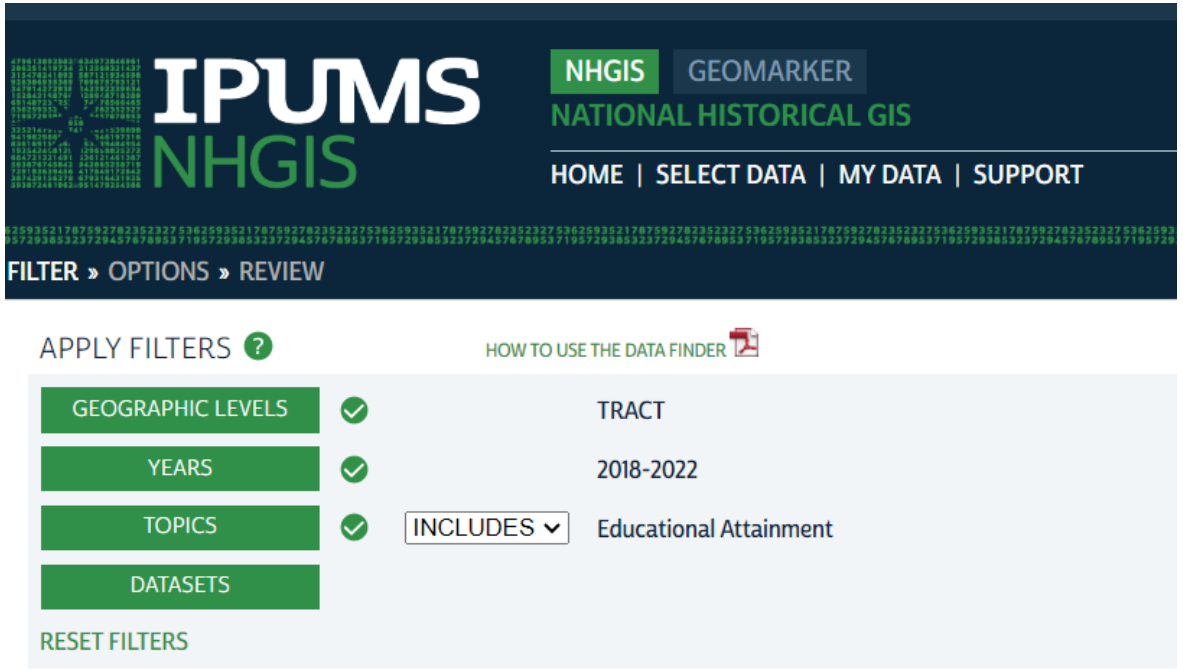
GEOCODING STEP 3: INSPECT GEOCODED POINTS



- Geocoded points have an "attribute table"
- Inspect "Loc_name"
  - What type of address was the point matched to?
    - Street address, Postal address, etc
- Inspect "Status"
  - M = Matched, T = Tied, U = Unmatched
- Re-clean addresses as needed

HOW TO FIND AREA LEVEL DATA

- Data Sources
  - US Census Data
    - Updated every 10 years
  - American Community Survey (ACS) Data
    - Updated every year, 5-year summaries
- IPUMS
  - GIS data from around the world
    - US and International datasets
  - Find individual Census and ACS data
- Existing Indices
  - Social Vulnerability Index
  - Environmental Justice Index
  - These are already in map format

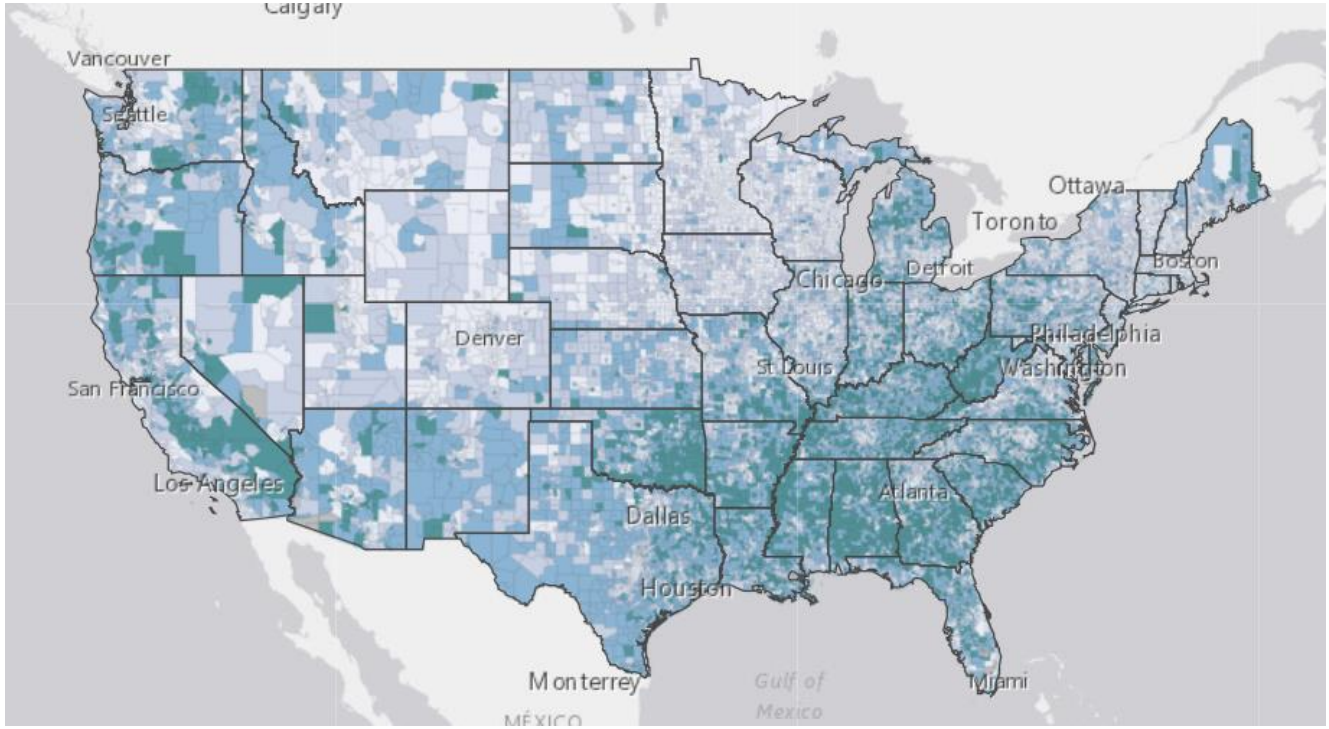


https://www.ipums.org/

https://www.atsdr.cdc.gov/placeandhealth/index.html

ENVIRONMENTAL FACTORS

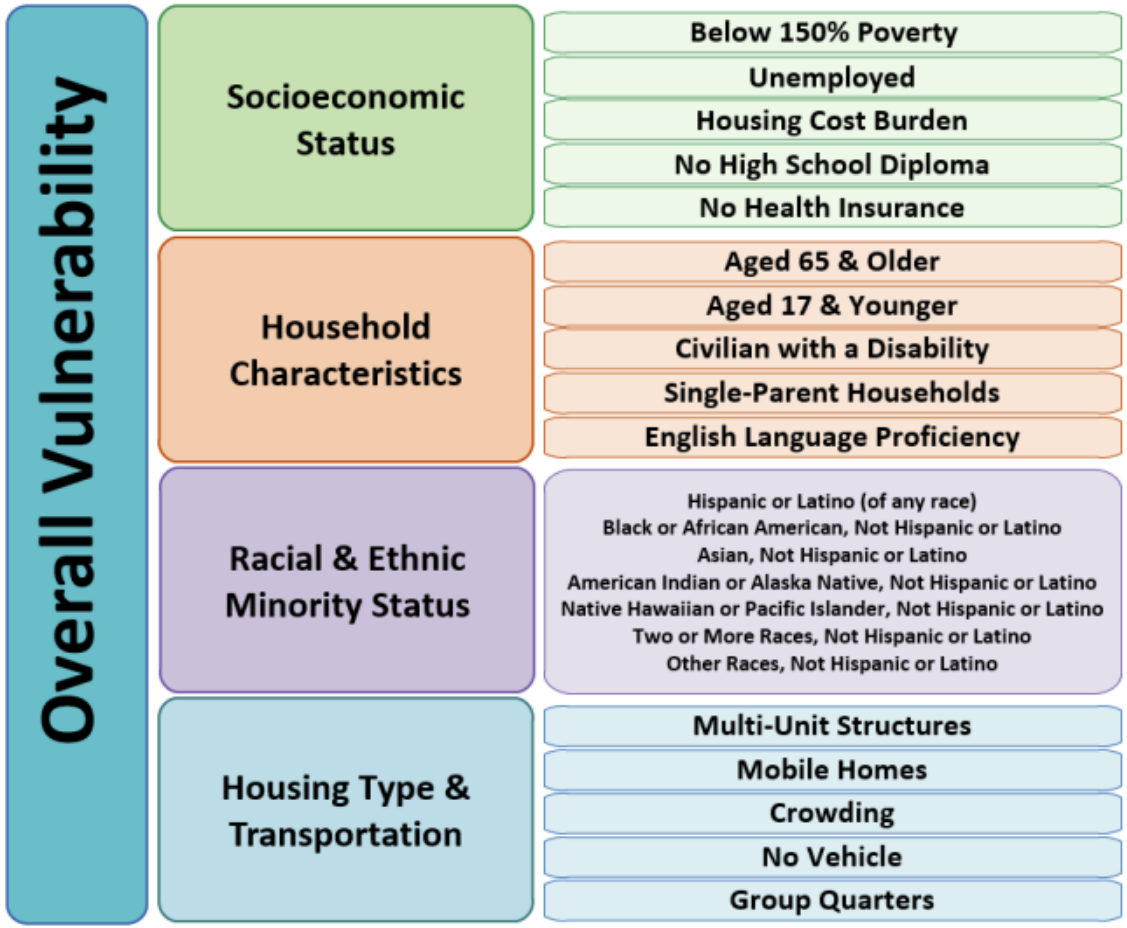
- Environmental Justice Index
  - Ozone, PM2.5, Toxic sites, mines, walkability, water pollution
- Heat Vulnerability Indices
  - Surface temperature, impervious surfaces, open spaces
- Other variables
  - Flood risk
  - Daily temperature



https://www.atsdr.cdc.gov/placeandhealth/eji/index.html

SOCIAL VULNERABILITY INDEX

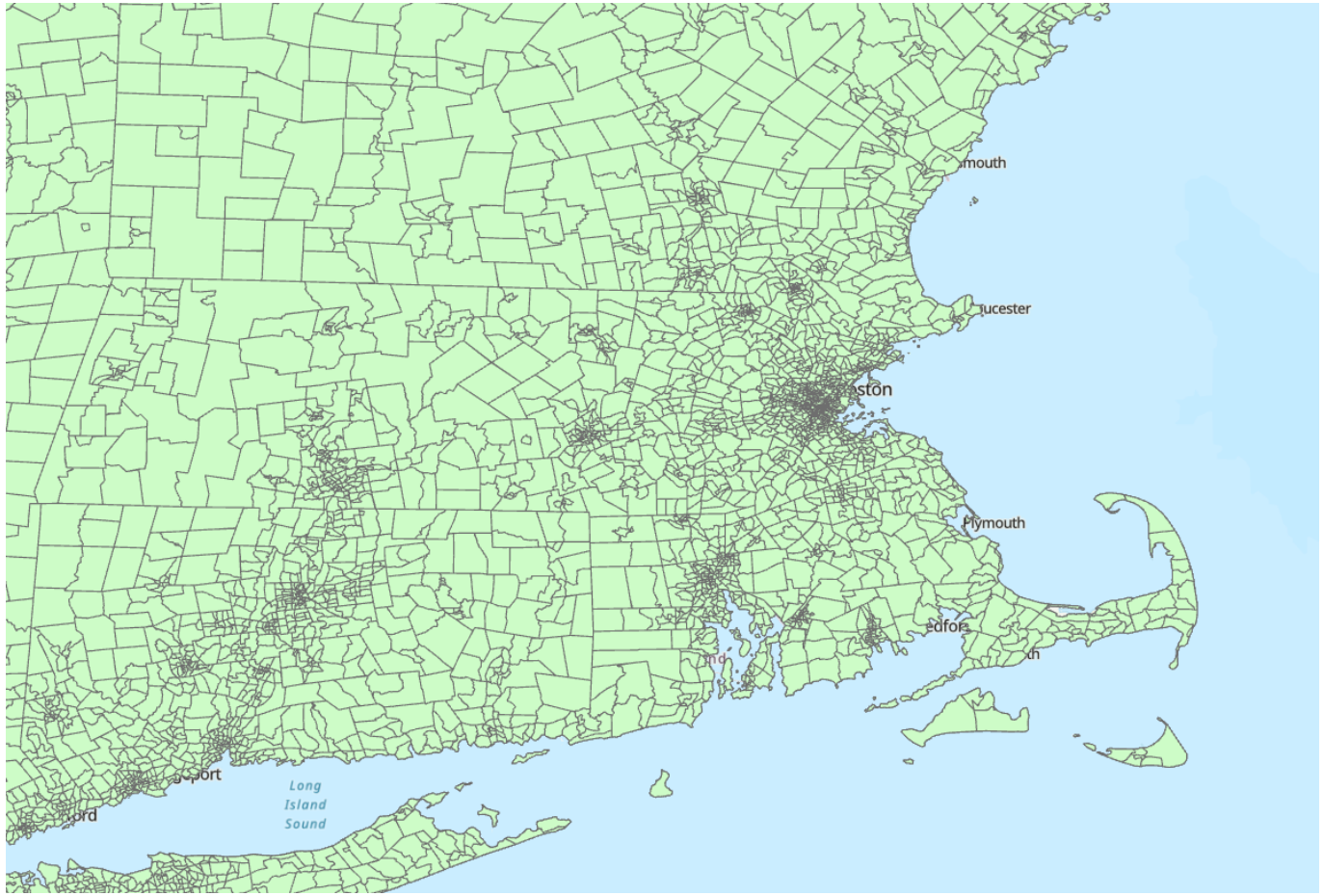
- Social Vulnerability Index
  - 0 to 100th percentile
  - Interpretation: SVI of 40 means the census tract is more vulnerable than 40% of the census tracts in the country/state
  - Often categorized into quartiles
- Overall Vulnerability Ranking or Single Theme



https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

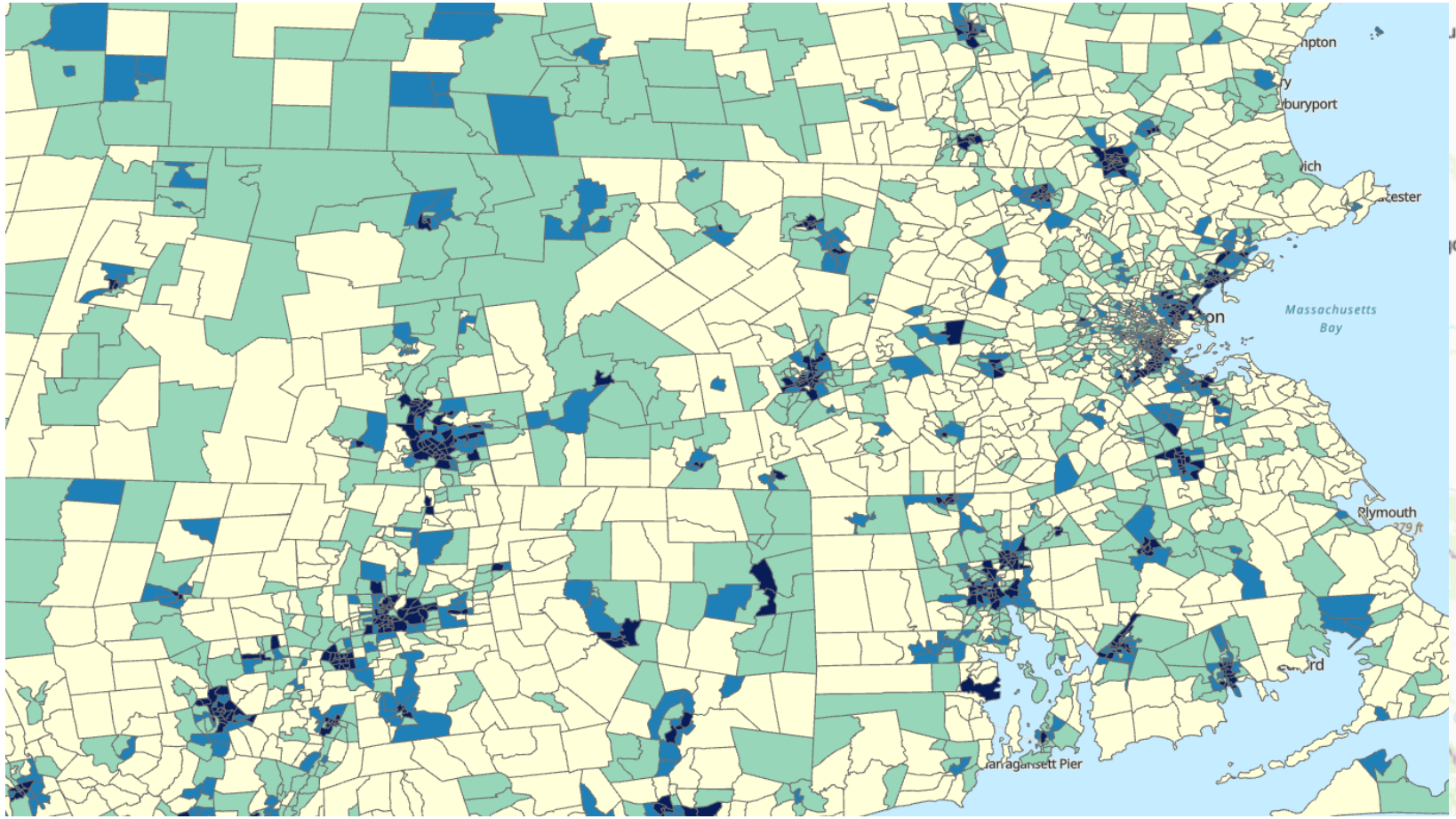
AREA LEVEL DATA TYPES

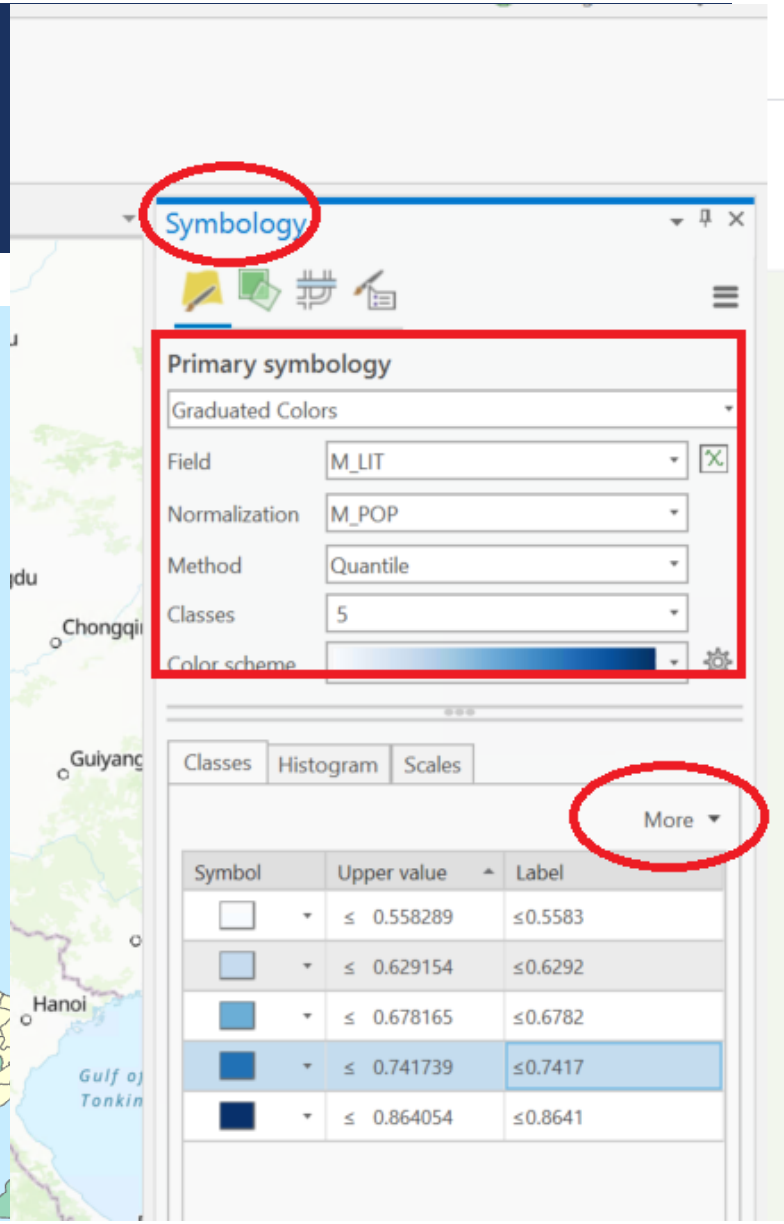
- Shapefile (.shp)
  - ArcGIS file format
- Geodatabase (.gdb)
  - ArcGIS collection of geographic data (.gdb)
- GeoJSON (.geojson)
  - Open geographic data
- Data Tables (.csv, .xlsx)
  - Not maps, can be merged with location data



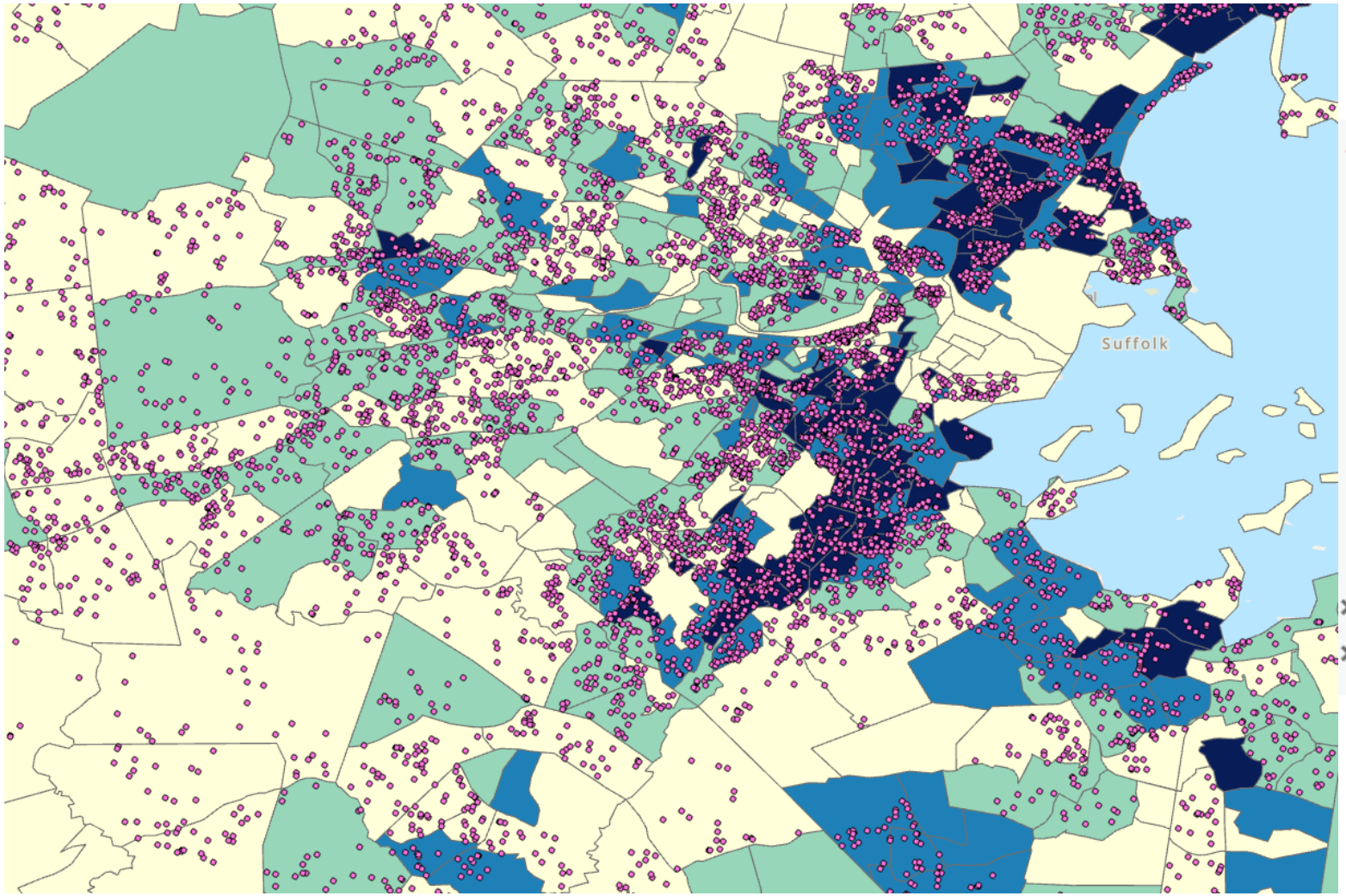
https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

ADD AREA-LEVEL DATA





SPATIAL JOIN



Add Spatial Join

Target Features

Cohort

Join Features

SVI Tract

Keep All Target Features

Match Option

Within

Permanently Join Fields

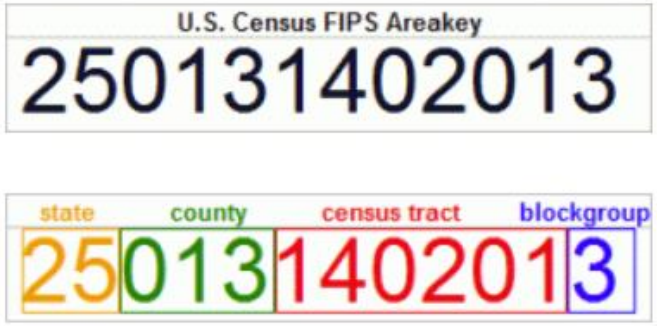
Search Radius

Decimal Degrees

Fields

Matching Attributes

UNDERSTANDING THE OUTPUT GEOGRAPHY



- Each patient has a FIPS and all geographic data from the map
  - Now you can merge any geographic data without going through the mapping process again
- Can export as an excel file and use your favorite data analysis methods to run models, create graphs, etc

https://www.hsph.harvard.edu/thegeocodingproject/

EXAMPLE STATISTICAL ANALYSIS

- Multilevel (hierarchical) regression models
  - Can use individual AND area level metrics in one model
  - Adjust for correlation between people who live in the same census tracts

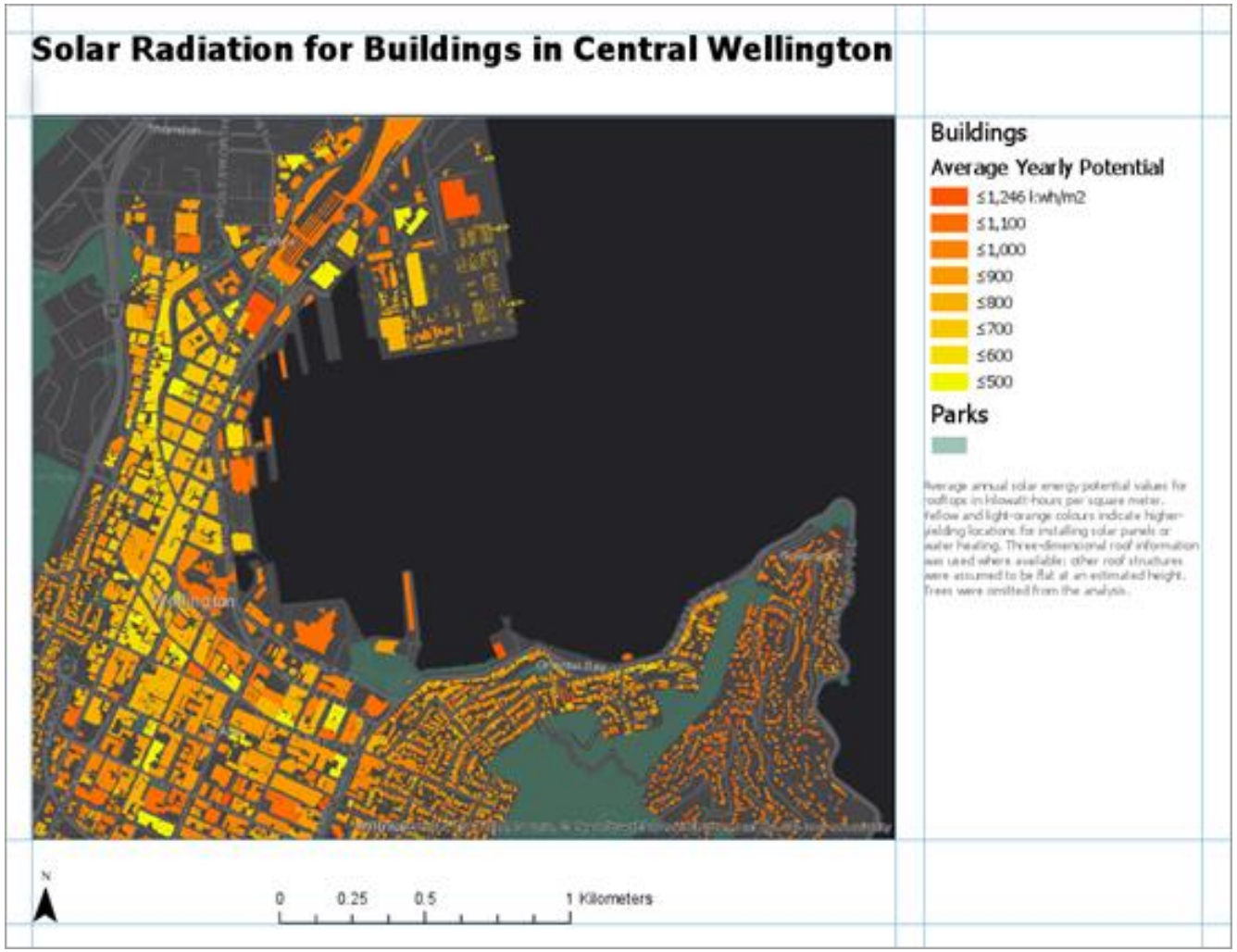
Multilevel model estimating the odds of hospitalization by individual-level factors and census tract–level social vulnerability

Variable	Odds Ratio (95% CI)
Race (ref = White)	
Asian	0.96 (0.61 – 1.51)
Black	**1.50 (1.14 – 1.97)**
Other/unknown	1.43 (0.45 – 4.75)
Gender (ref = Male)	
Female	**0.64 (0.54-0.76)**
SVI (ref = least vulnerable)	
2nd quantile	**1.29 (1.01 – 1.64)**
3rd quantile	**1.43 (1.12 – 1.83)**
4th quantile (most vulnerable)	**1.84 (1.43 – 2.36)**

Also adjusted for age, ethnicity, insurance, comorbidities

Santacroce, L., Dellaripa, P. F., Costenbader, K. H., Collins, J., & Feldman, C. H. (2023). Association of Area-Level Heat and Social Vulnerability With Recurrent Hospitalizations Among Individuals With Rheumatic Conditions. *Arthritis Care & Research (2010)*, *75*(1), 22–33. https://doi.org/10.1002/acr.25015

CREATING A MAP LAYOUT FOR PUBLICATION

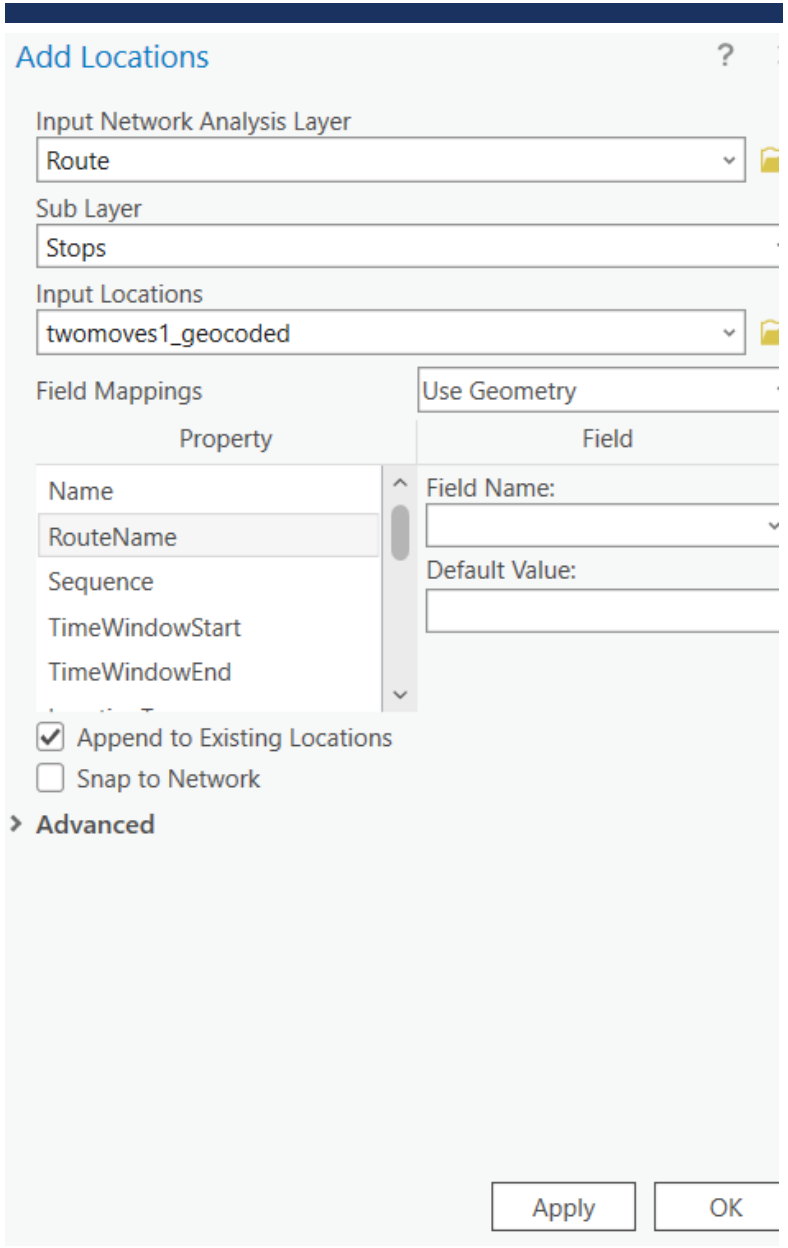


- Map vs Layout View in ArcGIS
- Essential map elements
  - Map frame
  - Legend
  - North arrow
  - Scale

https://pro.arcgis.com/en/pro-app/latest/get-started/add-maps-to-a-layout.htm

TIMING OF GEOSPATIAL DATA

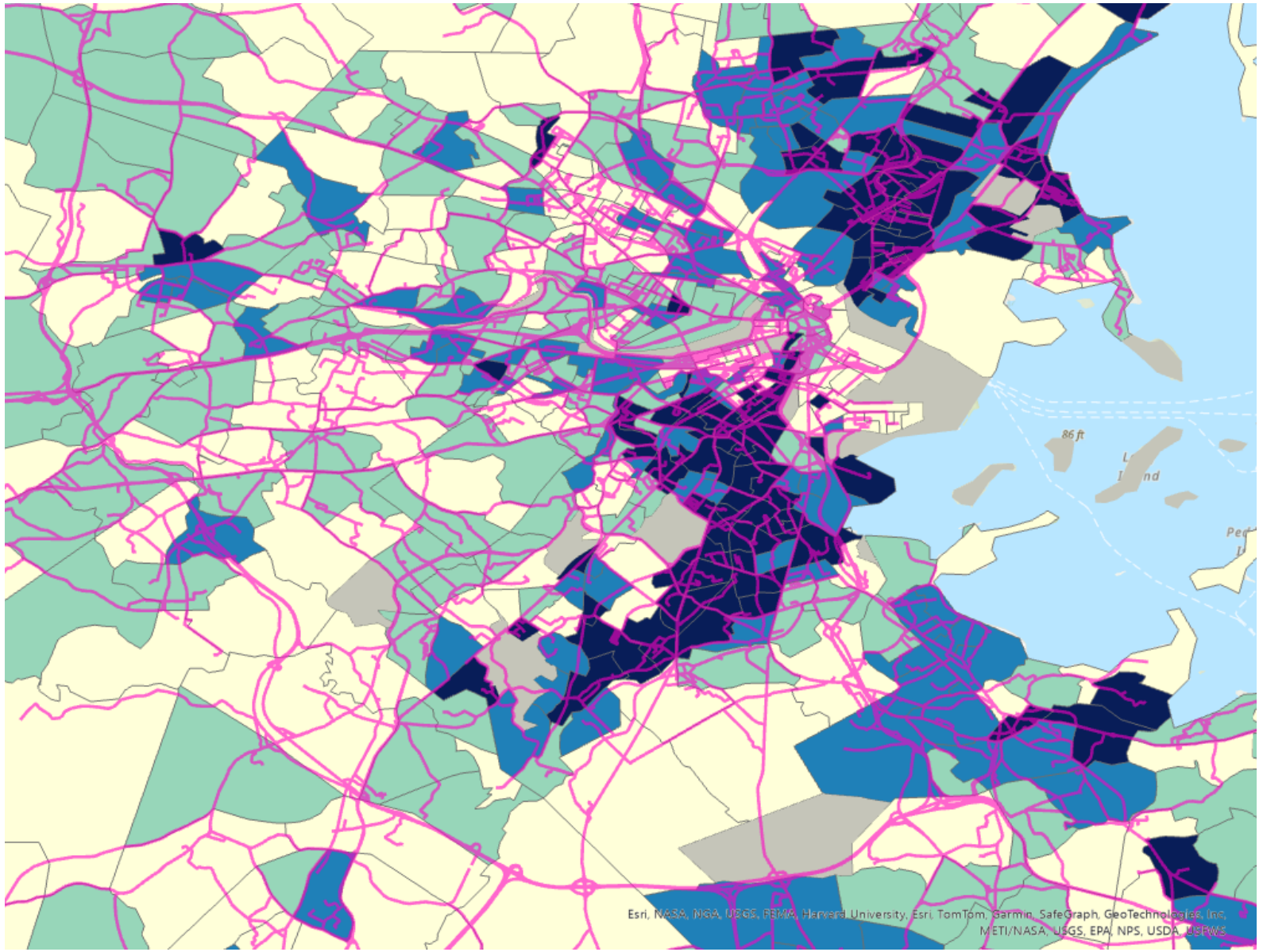
- Timing of address and outcome/other variables of interest are important
  - People can move to areas that have very different characteristics at any time
  - Outcome of interest (i.e. hospital visit) should match address time
- Census tracts boundaries are updated every 10 years
  - Data within census tracts get updated more frequently
    - ACS 5-year summaries



BEYOND GEOCODING: NETWORK ANALYSIS

- Network Analyst
  - ArcGIS extension
- Driving distance or time
  - Between one point and many, one and one
- Distance between home and rheumatologist, home and work, etc

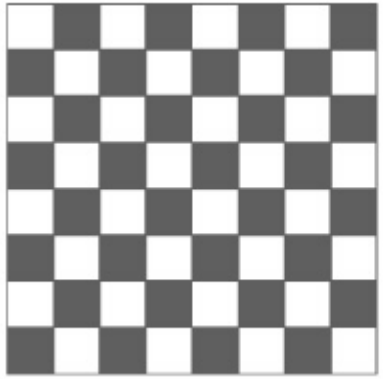
- Must use local network with sensitive data!



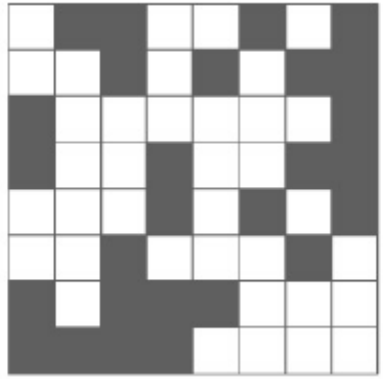
NETWORK ANALYSIS MAP

BEYOND GEOCODING: SPATIAL AUTOCORRELATION

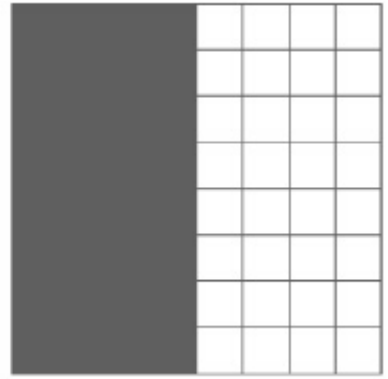
- 1st Law of Geography: Everything is related to everything, but near things are more related
- Spatial Autocorrelation is a quantification of the 1st law
  - Moran's I Statistic
    - Univariate and Bivariate options



Negative
-1



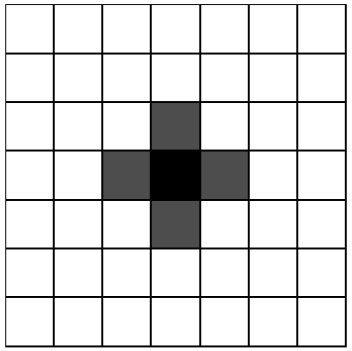
None
0



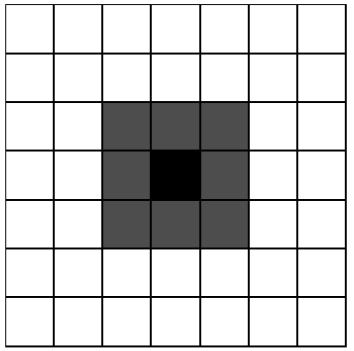
Positive
1

SPATIAL AUTOCORRELATION CONTINUED

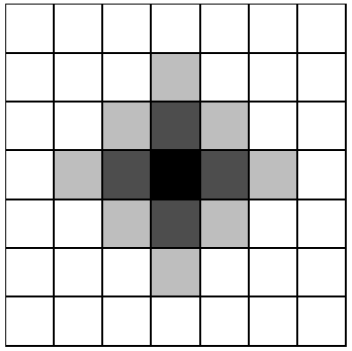
Rook. First order



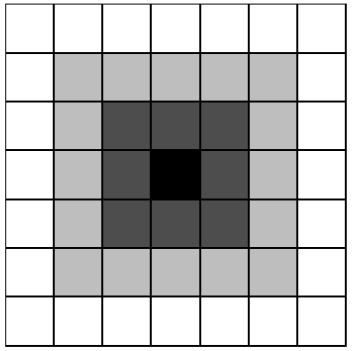
Queen. First order

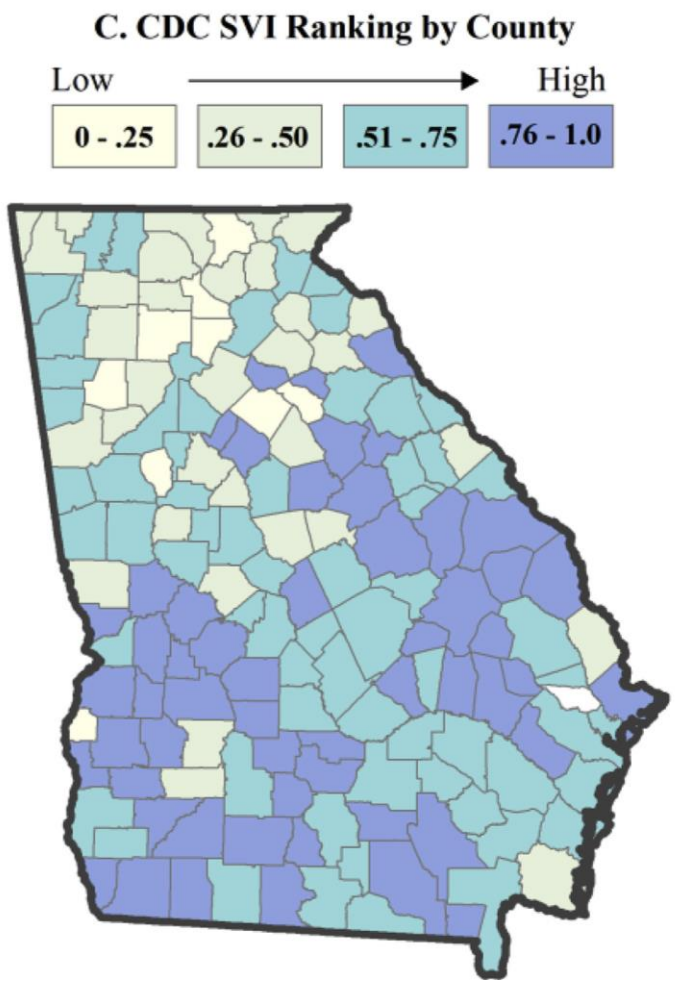


Rook. Second order

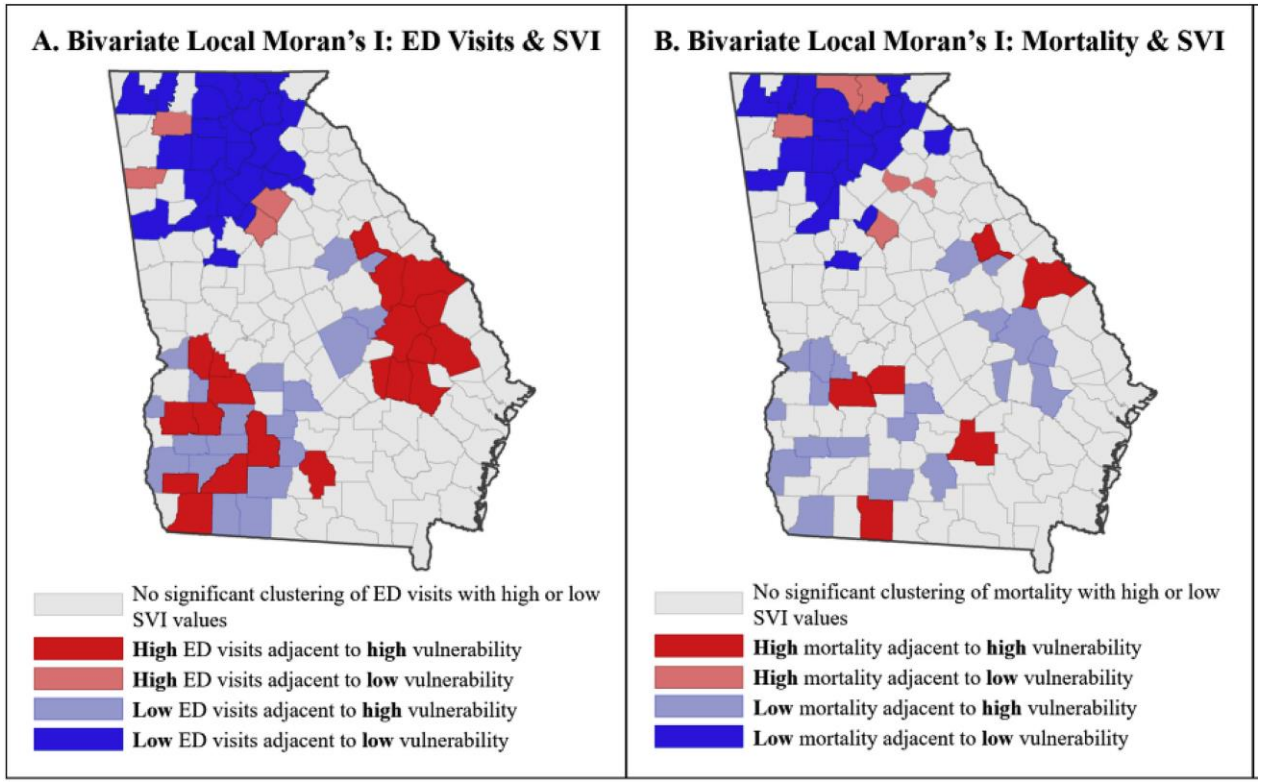


Queen. Second order





Moran's I: **0.43**
P-value: <0.001



Moran's I: **0.21**
P-value: <0.001

Moran's I: **0.04**
P-value: 0.11

Erica Adams Lehnert, Grete Wilt, Barry Flanagan, Elaine Hallisey,
Spatial exploration of the CDC's Social Vulnerability Index and heat-related health outcomes in Georgia, International Journal of Disaster Risk Reduction, Volume 46, 2020, 101517, ISSN 2212-4209, https://doi.org/10.1016/j.ijdrr.2020.101517.

ACKNOWLEDGEMENTS

- Brigham and Women's Division of Rheumatology, Inflammation and Immunity
  - Candace H. Feldman, MD, MPH, ScD
  - Daniel H. Solomon, MD, MPH
- Harvard Center for Geographic Analysis
  - Scott Bell, PhD
  - Connie Chen, PhD
  - Jeff Blossom, MA