# Making sense of spatial demographic data

Chandler Armstrong

CERL

Spring 2019

# Agenda

- Human geography
- Human geography and military doctrine
- Human geography at CERL
- Human geography in RAPTOR
- hypothesis testing & simulation
- mapping social vulnerability
  - data sources
  - building the SVI
  - mapping the SVI to space

# Human Geography

**Definition**

Human Geography Human geography is the branch of geography that deals with the study of people and their communities, cultures, economies, and interactions with the environment by studying their relations with and across space and place.

**Definition**

Space Space is the boundless three-dimensional extent in which objects and events have relative position and scale.

# Space (the patterned frontier)

## Corollary

*Patterns are not explanations. Patterns confirm or hint at causes. Space lacks intrinsic properties; is rarely (never?) an explanatory variable, but patterns in space provide clues to fundamental explanations, and classes of spatial patterns are useful for categorizing other phenomena–e.g. scale free, preferential attachment, concentric zones, multiple nuclei; these spatial patterns contain a lot of information about the machinations of the phenomena they describe.*

# Human geography and military doctrine

- Space is universal
- Human geography a natural head for organizing broader social science

# Human geography at CERL

- Population simulation & **interpolation**
- **Statistical hypothesis testing**
- **Data integration**
- **GIS**

# Human geography in RAPTOR

- Disaster impacts and response
- Effects of environmental events on conflict
- Long-term urban change
- social vulnerability
- infrastructure vulnerabilities

# Disaster-driven migration

- data integration
- interpolation
- hypothesis testing

# Data integration

**Definition**

indexing of functions or features–e.g. spatial indexing of data-generating functions.

- preprocessing problem; upfront work
- reduction
- representation
- resampling

# Data interpolation

- data spatially indexed and resampled
- interpolate proportions and averages, not people...
- estimate demographic indices where no survey data exists
- dasymetric: distribute data between zones based on weights over features:
  $z_n = \sum_1^m x_{nm}\beta_m; 1, \ldots, n$
- kriging: distribute data between zones based on weights over functions:
  $z_n = \sum_1^m z_m\lambda_{nm}; 1, \ldots, n$
- regression kriging: kriging with varying mean:
  $z_n = \sum_1^m (E[z_m] - z_m)\lambda_{nm}; 1, \ldots, n$

# Uncertainty

- bayesian (probabilities without experiments)
  - MCMC of posterior
- frequentist

  **bootstrap** Monte Carlo of empirical data

  **interpolation** uncertainty due to distance from observations

# Hypothesis testing

- hypothesis of features; “x has a significant effect on y”
- hypothesis of space; “cases of x are not randomly distributed across space”
- regression unbiased but inefficient under spatial dependence
- mind the level of analysis (individual, family, community, etc.)

# Disaster-driven migration and social vulnerability

**Definition**

social vulnerability One dimension of vulnerability to stress and shocks, including abuse, social exclusion, and natural hazards.

- Social vulnerability is a community-level concept
- ~10 variables compressed into a single factor

**Definition**

migration Human migration is the movement by people from one place to another with the intentions of settling, permanently or temporarily in a new location.

- disaster is an acute stimulus
- most migration a response to economy
- historical patterns and multiple motivations shape disaster-driven migration

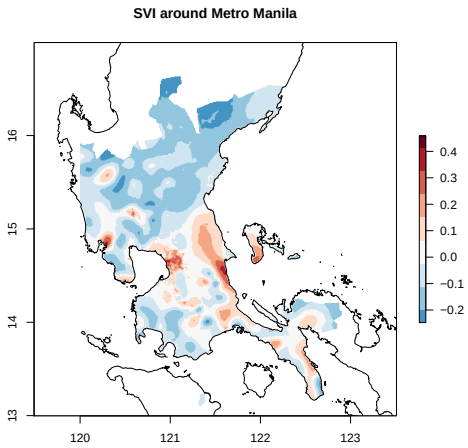
# Propensity to migrate

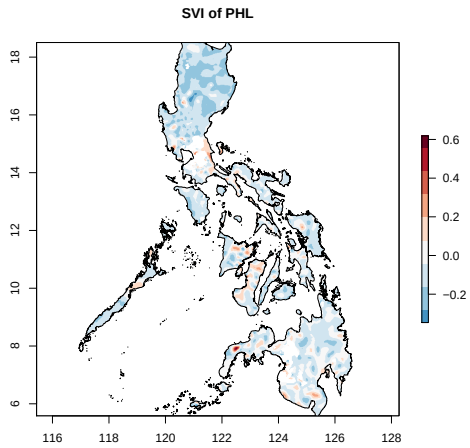
- past migration
- resources to migrate
- distance to potential immigration site

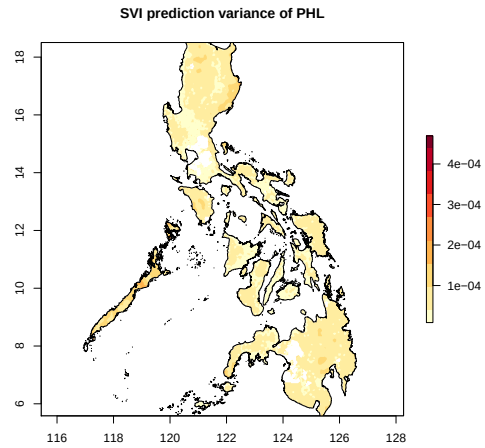
| variable(1=‘yes’) | mean(standard deviation) | n missing |
|---|---|---|
| migration, 5-year | 0.031(0.03) | 1353467 |
| migration, 10-year | 0.04(0.038) | 2280714 |
| native | 0.98(0.02) | 251850 |

Table 1: ({Minnesota Population Center}, 2018)

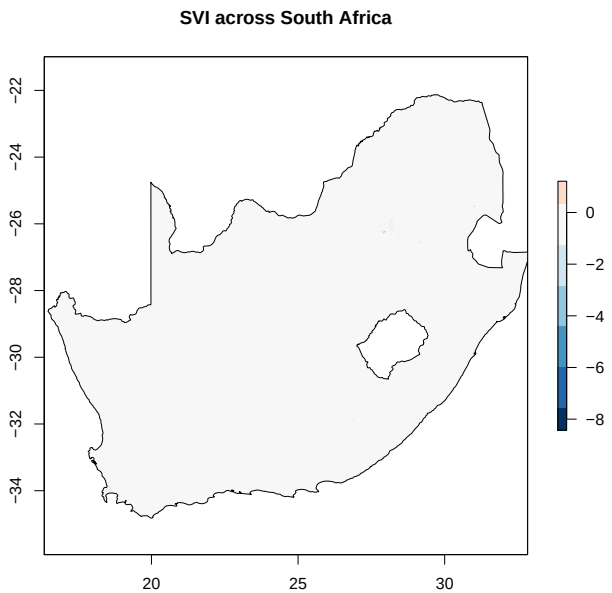
# SVI PHL

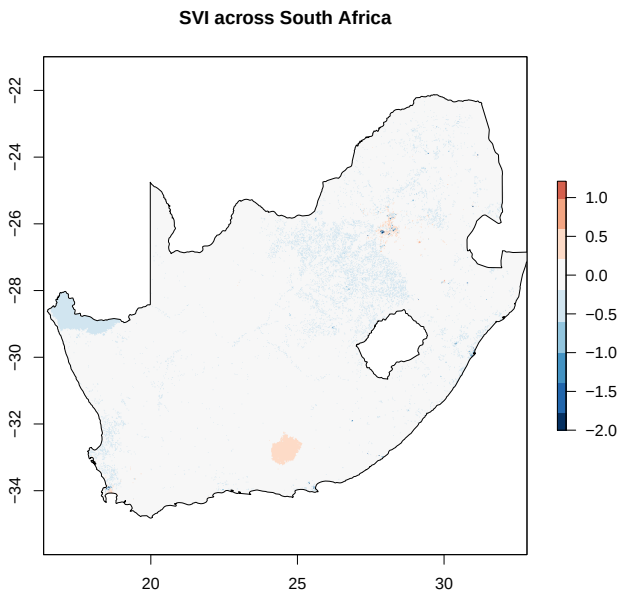

SVI around Metro Manila
16
15
14
13
120
121
122
123
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2


SVI of PHL
18
16
14
12
10
8
6
116
118
120
122
124
126
128
0.6
0.4
0.2
0.0
−0.2

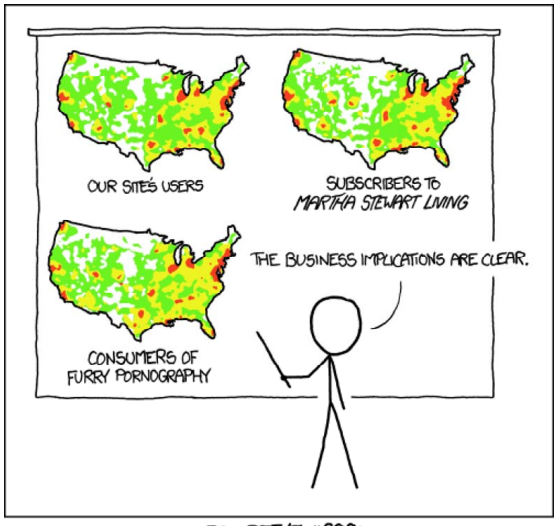

SVI prediction variance of PHL
18
16
14
12
10
8
6
116
118
120
122
124
126
128
4e−04
3e−04
2e−04
1e−04

# SVI ZAF


SVI across South Africa
−22
−24
−26
−28
−30
−32
−34
20
25
30
0
−2
−4
−6
−8


SVI across South Africa
−22
−24
−26
−28
−30
−32
−34
20
25
30
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0

# Challenges & Problems




PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

- Population dominates
- Modifiable areal unit problem
- Embarrasingly parallel, but high space complexity
- Largely deductive, with little to no inductive validation

# Benefits

- Re-expresses all data into a common format and resolution: easy analysis
- Simple to add new data to analysis: resampling & aggregation
- Nonparametric: fast, fewer assumptions, analyze bias, skew, & uncertainty
- Mapped demographics permit geospatial operations: distances, intersections, buffers, etc.
- Many applications: disaster preparedness & mitigation, response, rehabilitation & reconstruction

Gotway, C. A., & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(2), 157-178.

Minnesota Population Center. (2018). *Integrated public use microdata series, international: Version 7.1 [dataset].* doi: https://doi.org/10.18128/D020.V7.1

# POC

Chandler Armstrong, PI chandler.m.armstrong@erdc.dren.mil