

People, Patterns, and Populations

Considering multiple perspectives in the spatial analysis of VGI

Colin Robertson¹, Rob Feick²

¹Wilfrid Laurier University, Waterloo, ON, Canada

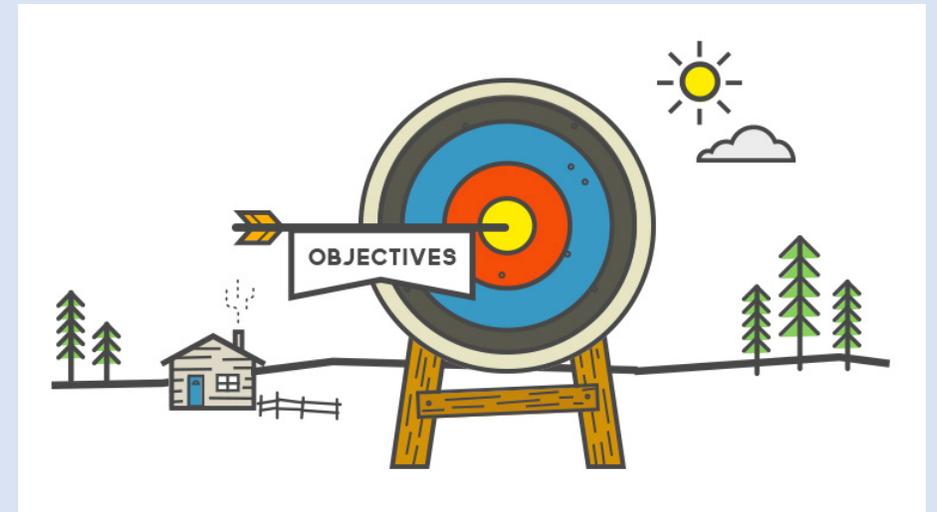
²University of Waterloo, Waterloo, ON, Canada

Towards Volunteered Geographic Analytics?

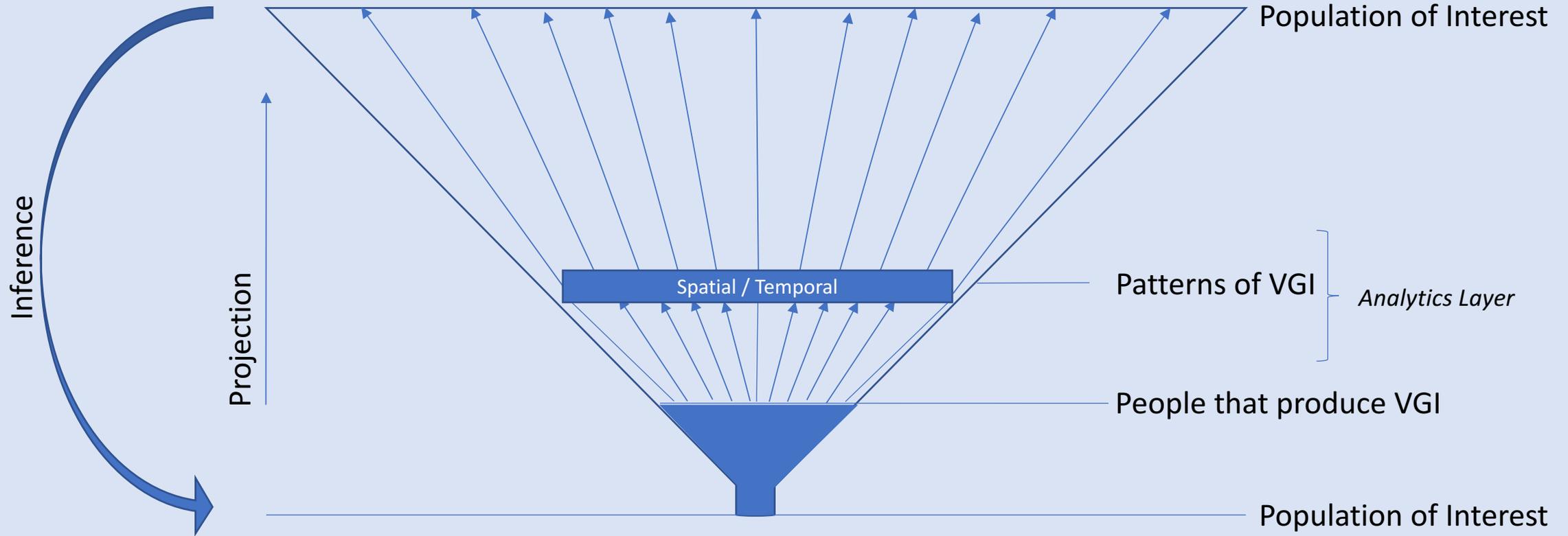
- VGI has changed dramatically over the last 10 years
- Increasingly, what constitutes VGI is...
 - Authored by diverse and heterogeneous users
 - Multiplatform
 - Linked
- What does this evolving “datascape” mean for analytical approaches?
 - Identifying spatial patterns and areal differentiation across regions
 - Characterizing urban places
 - Event detection and monitoring

Talk Objectives Today

- Revisit VGI from the ground up – is VGI still special?
- Identify key challenges underlying the development of a VGI Analytics
- Spur discussion and dialog on VGI from an analytical perspective
 - Specifically through the lens of three key perspectives
 1. People: Who produces VGI
 2. Patterns: Identifying and interpreting meaning in VGI
 3. Populations: Creating new geographic knowledge with VGI



People, Patterns and Populations



What is VGI?

- Goodchild (2007)
 - Geographic data created and shared freely by people with varying knowledge of formal geographic principles and practices
- Distinguished from expert-GI
 - Heterogeneity - media, thematic foci
 - Lack of, or loose, standards
 - Multi-authored data sets ...
- Analytical value
 - Relative to its quality and novelty to expert GI



But what is VGI, really?

- We know its not necessarily volunteered (or information)
- VGI = new forms of geographic data (10 years ago)
 - Defined in part by WHO creates it
 - Peculiar category for information
- Often Social data
 - data created through social exchange / interactions
- Increasingly hybrid



Figure 2.5: VGI framework

Scientific knowledge VGI Type 1	Local knowledge VGI Type 2	Personal knowledge VGI Type 3
<ul style="list-style-type: none"> - volunteered - Objective - Structured - 2 way or n way - Digitizing, GPS, twitter - Only points or points, lines, polygons 	<ul style="list-style-type: none"> - facilitated-VGI - Subjective - Unstructured - 1 way, 2 way or n way - Selection, Drawing, geocoding - Points, lines, polygons 	<ul style="list-style-type: none"> - Kept private - Subjective - Unstructured - n way - Location through networks, geocoding - Only points

Is VGI fit for analysis?

- VGI with objective comparators (e.g. OSM, some citizen science)
 - Assess internal characteristics relative to expert alternatives or specifications
- VGI without authoritative comparators
 - Infer fitness from data authors' qualifications, credibility, motivations, etc.
- “Informational trust” - Bishr & Kuhn (2013)
 - Transitivity of trust of author to their data
 - Keßler & De Groot (2013) – OSM feature level



Changing nature of geographic knowledge

- Gollege (2002)
 - Shift from inventorying and describing geographic facts
 - “analyzing those facts to produce new information and knowledge ...”
- VGI research seems to be on a similar path, with inventorying focus
- Goodchild & Li (2012): geog. knowledge concepts to assess data quality
- Can individuals’ differing expertise to contribute locale-specific data provide an avenue for a geographic knowledge approach?

Geographical expertise

- What does it mean to be a geographical expert?
 - Skills and training in geographic methods and tools?
 - Understanding of spatial concepts, processes and patterns?
 - *Familiarity or local knowledge for a specific areas?*
- Citizen science and VGI studies recognize
 - Non-experts can produce GI of similar quality to experts
 - Non-experts often have local, experiential or place-based knowledge experts lack
- Can the expert-amateur divide be recast to capture continua of geographic knowledge?

Untapped expertise in Oregon?

- Mats Jarlstrom fined for *'talking freely about language, universal language, which is mathematics and physics'*.
- Independently confirmed with 1959 authors of traffic light algorithm that yellow light timing did not account for cars turning

CBC As It Happens - <http://bit.ly/2pUVbw9>

Wednesday May 03, 2017

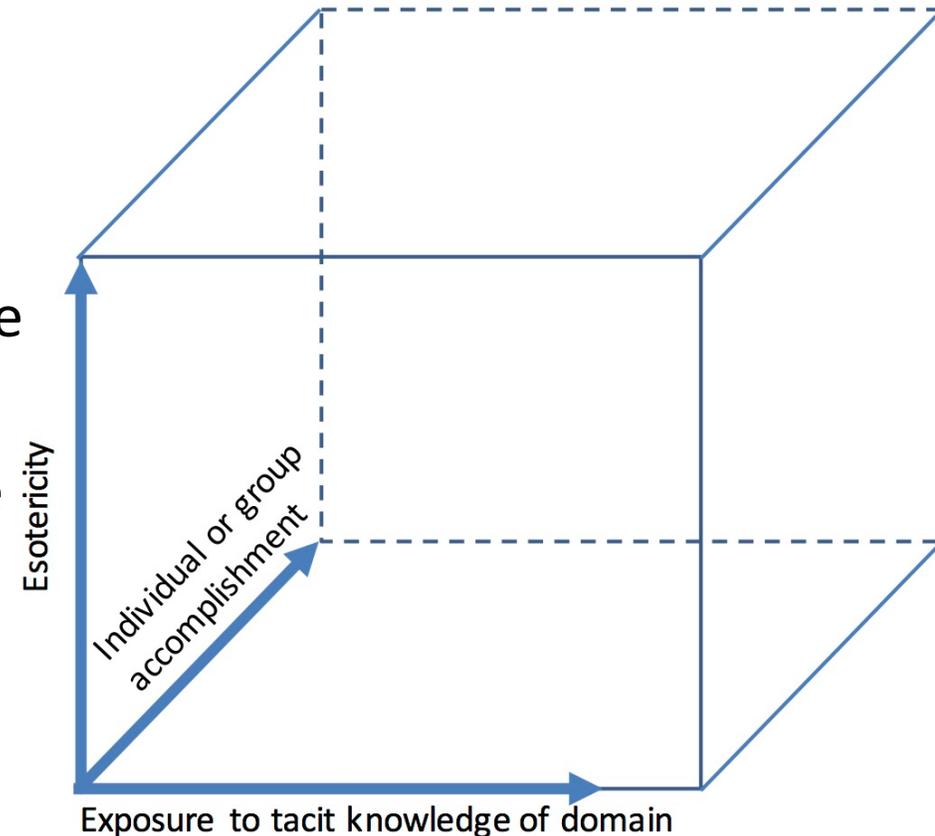
'I felt violated,' says U.S. man fined \$500 for talking about traffic lights



Mats Jarlstrom was fined \$500 for billing himself as an engineer while promoting his research on traffic lights. (Institute For Justice)

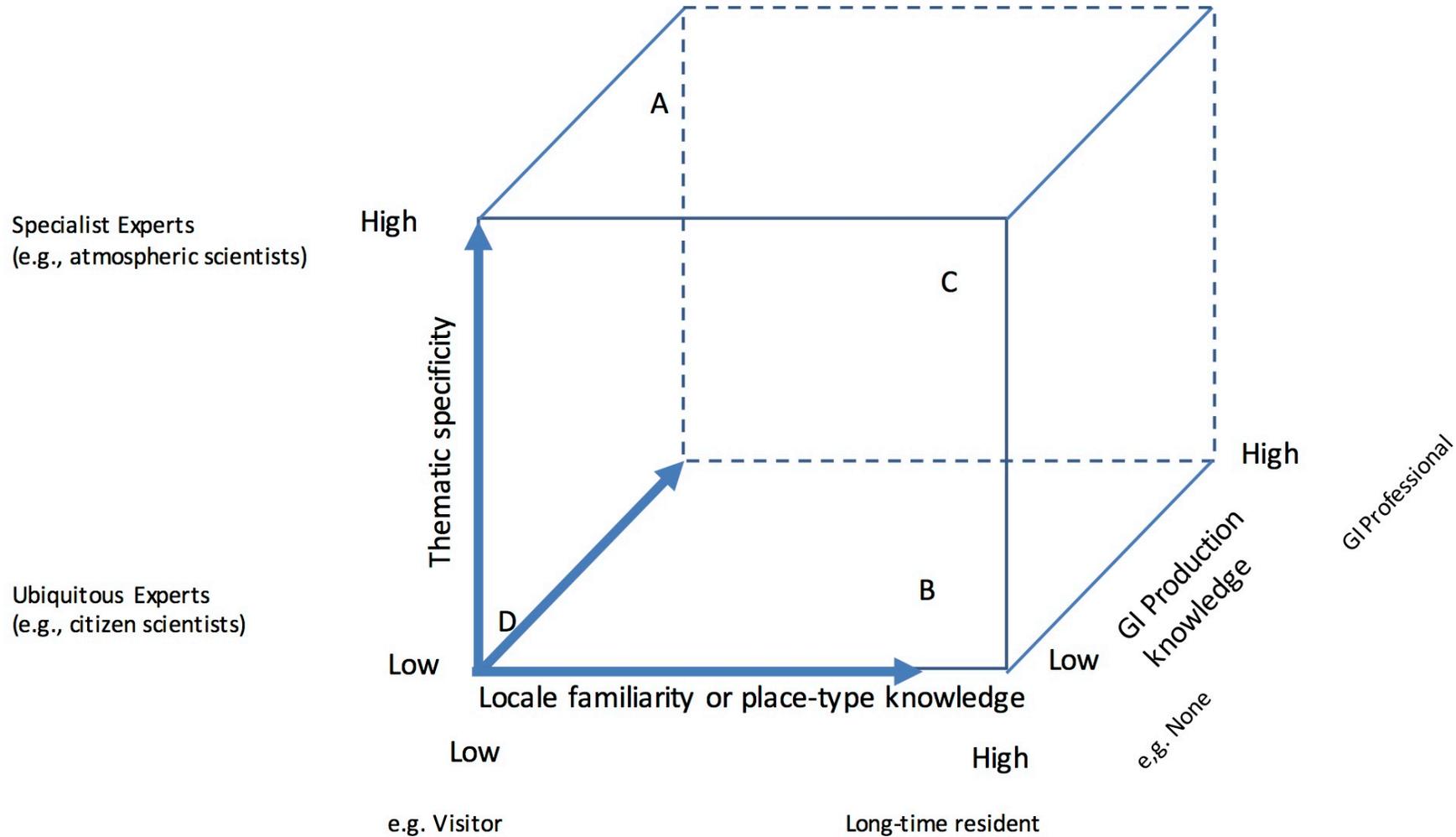
Studies of expertise and experience

- Collins and Evans (2002) – 3 dimensions of expertise
 - Contributory expertise
 - via formal training within a domain
 - Interactional expertise
 - via socialization & exposure to tacit knowledge
 - Esotericity
 - degree that expertise is widespread or unique
- Expertise for any one person in a domain can be measured across these dimensions

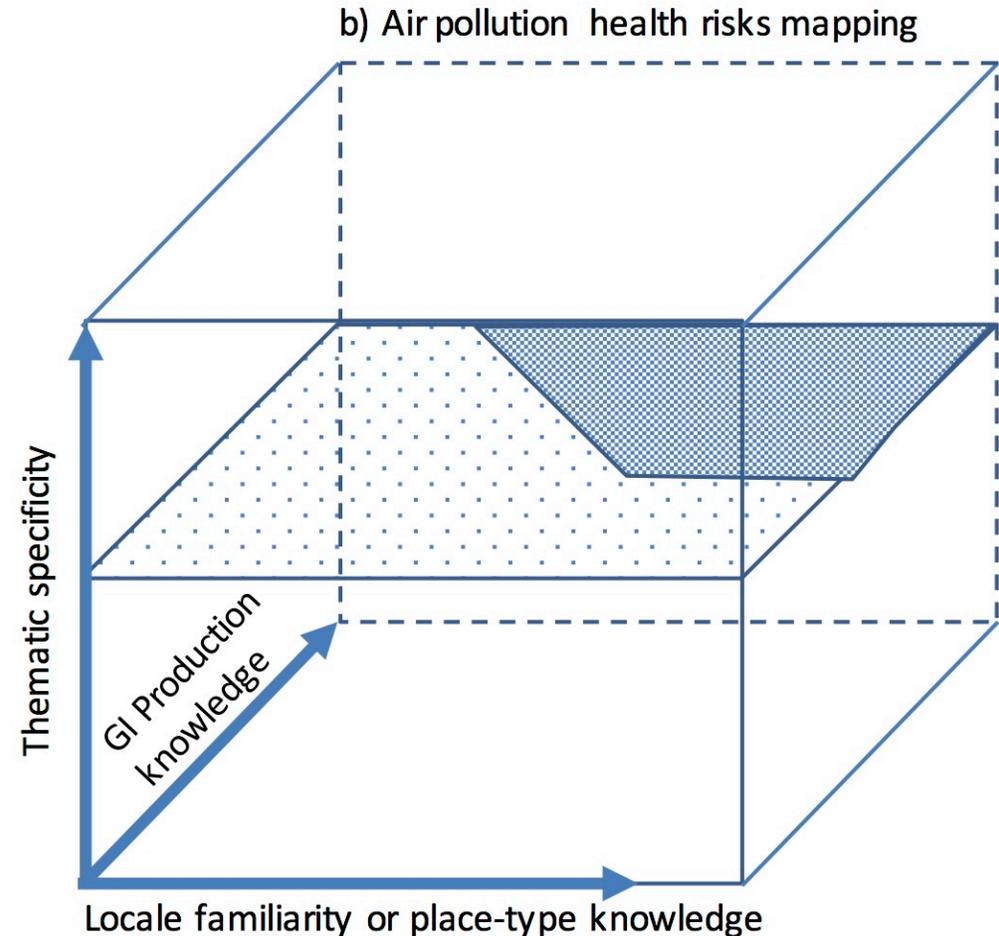
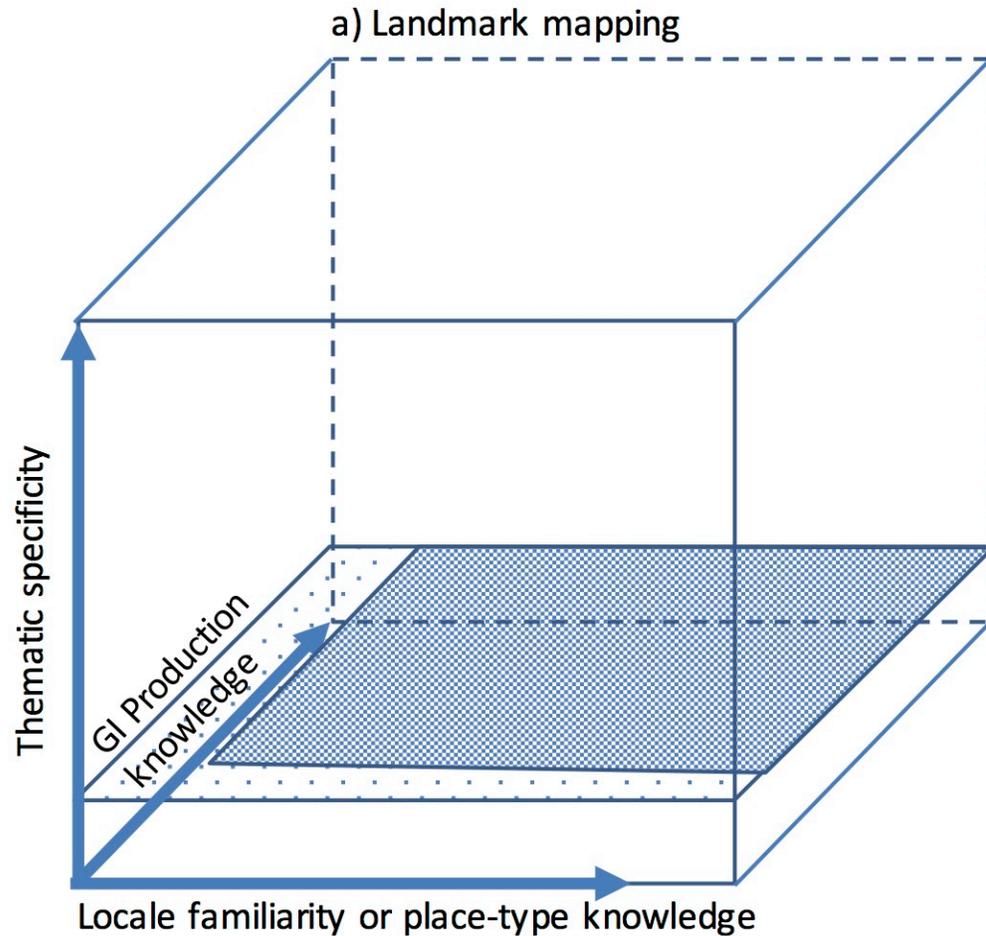


People

Conceptualizing Geographic Expertise

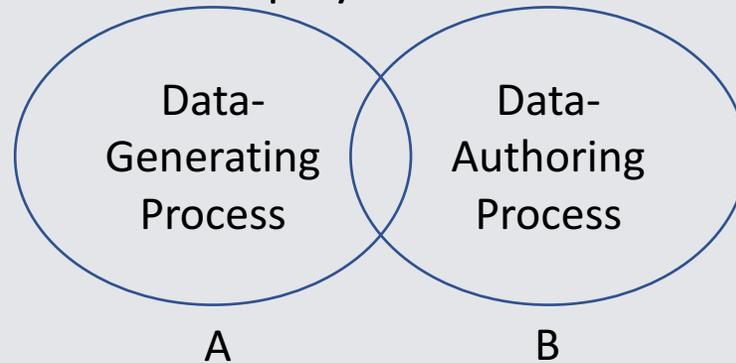


GI authoring – expertise requirements

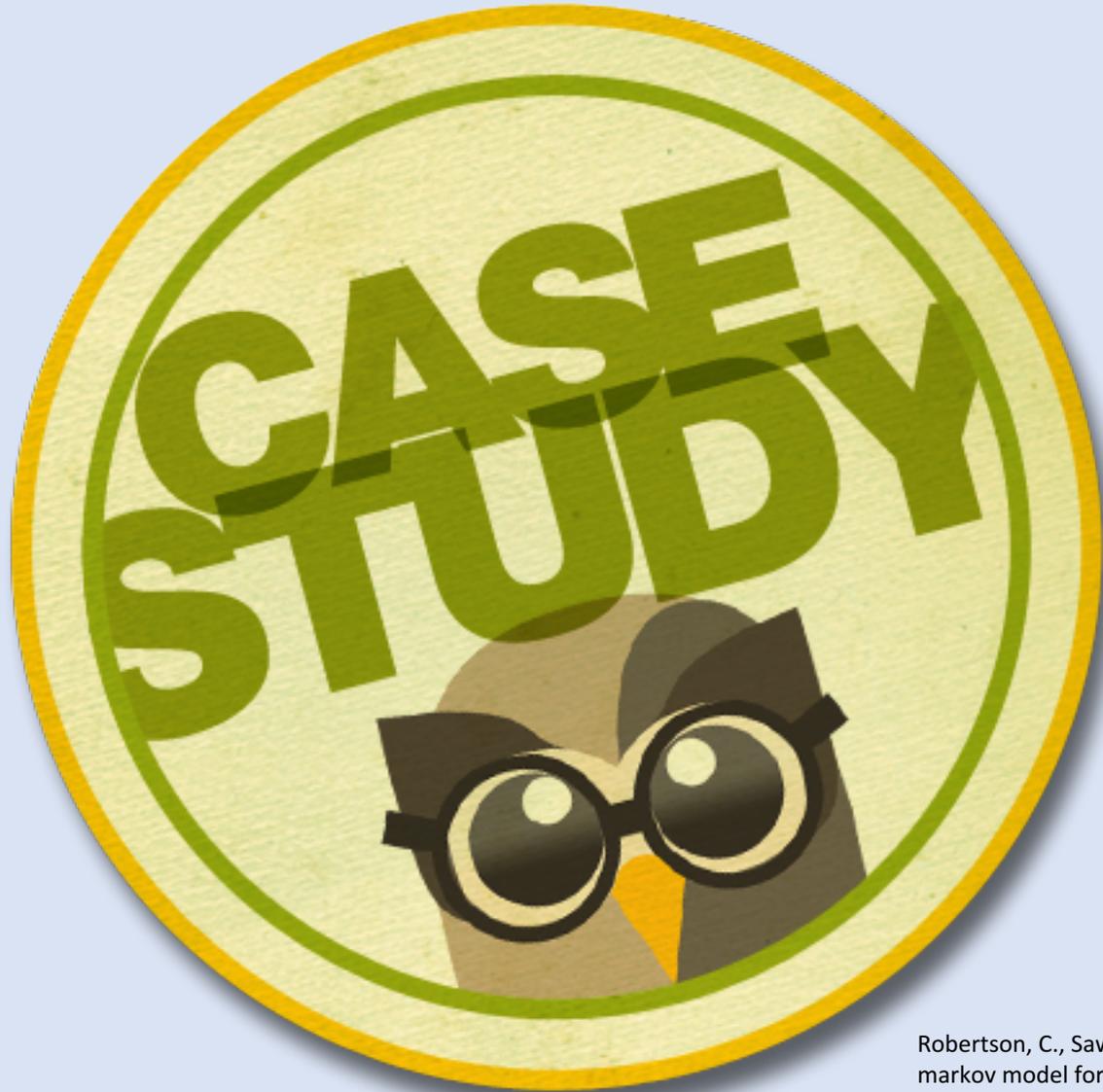


VGI User, Spatial, and Temporal Patterns

- Analytics in Big Data is often a 'search for patterns' to identify interesting spatial/temporal regularities in data
- Question for big data is – what can we do with this data? Same for much VGI sources...
- In VGI, we have two types of processes at play



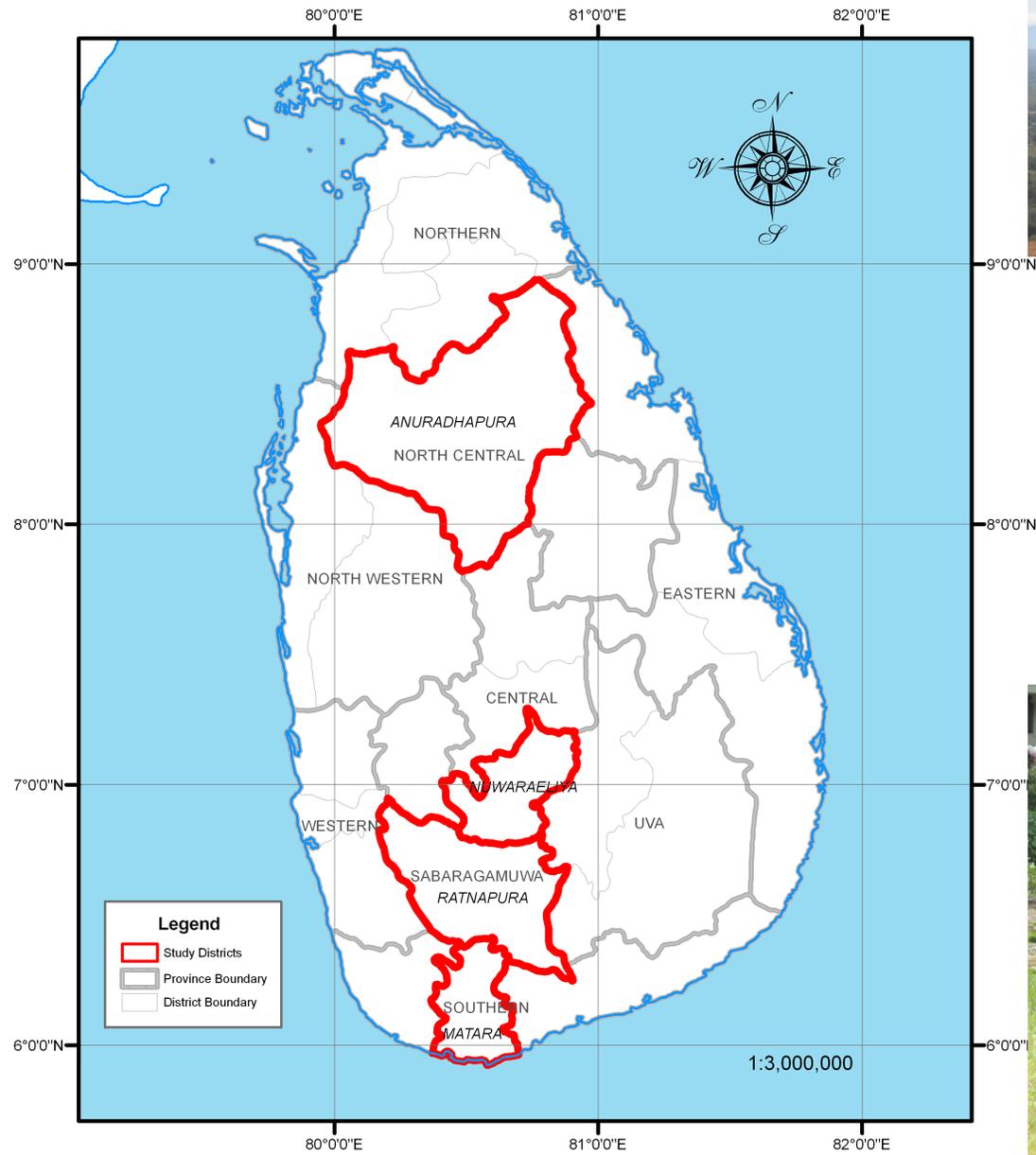
- Patterns of interest can be
 - 'A NOT B'
 - 'B NOT A'
 - Rarely is it 'A OR B'
- Nature of GE inherent in VGI (i.e., research context) dictates appropriate approaches for data filtering and analysis



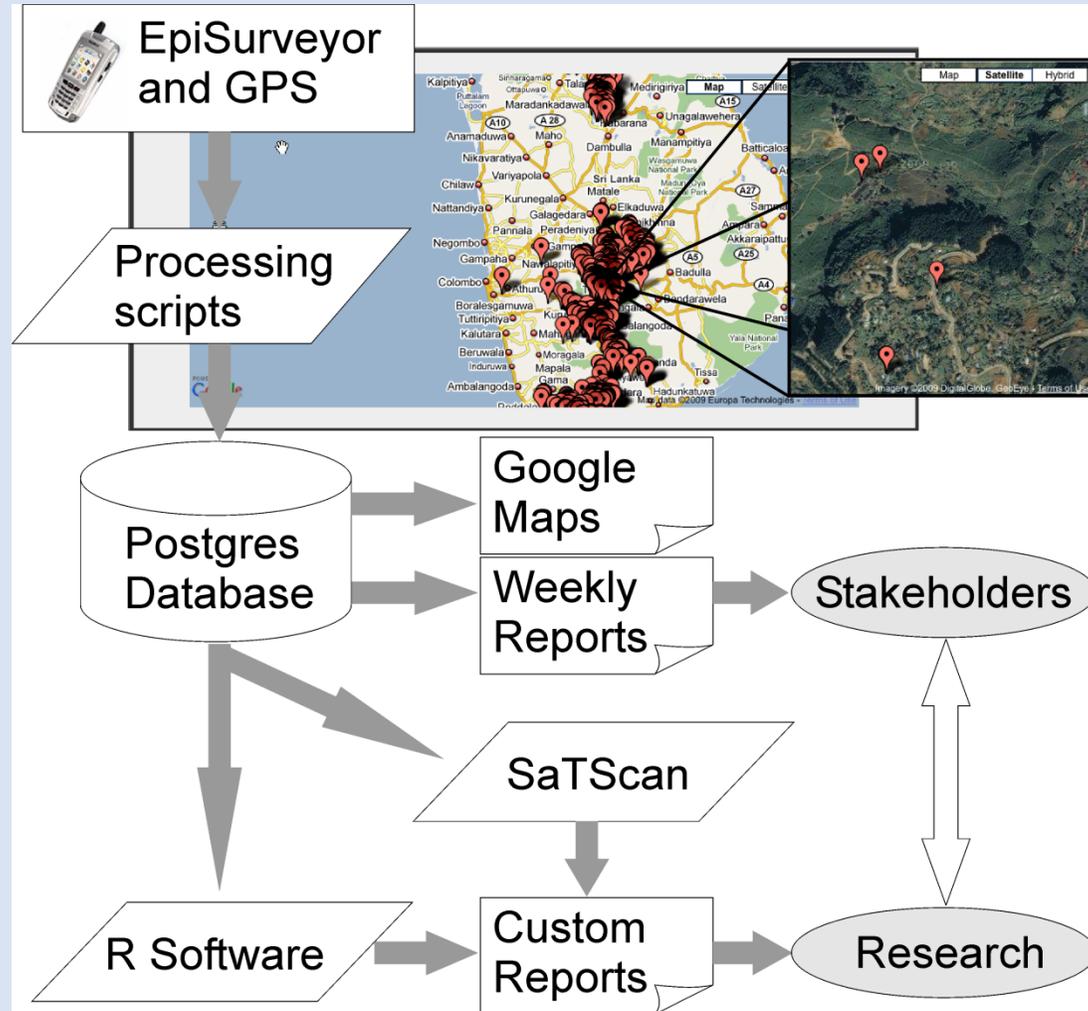
#1

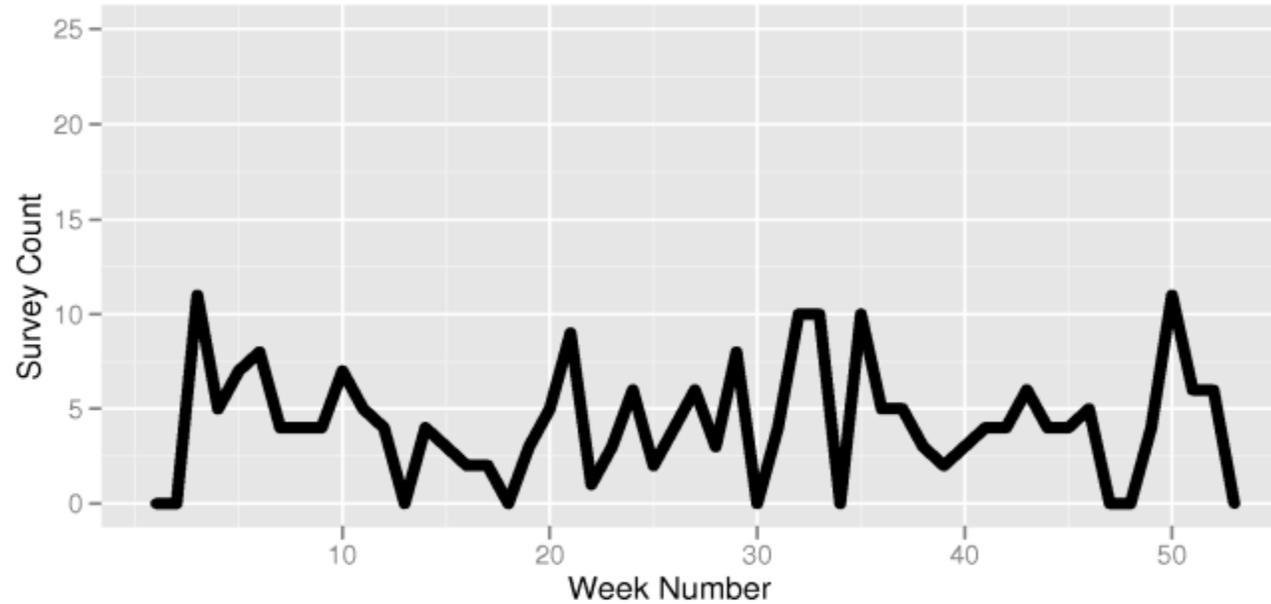
Infectious Disease Surveillance in Sri Lanka (IDSAS-SL)

- Animal health syndrome/diagnoses near real time surveillance system in cattle, poultry, and buffalo – data includes location
- Local government veterinarians engaged as participant data collectors, reporting location and characteristics of animal health events encountered in field
- Varying levels of
 - ability and expertise in using mobile phone and touch screen interfaces
 - Engagement / interest in the project

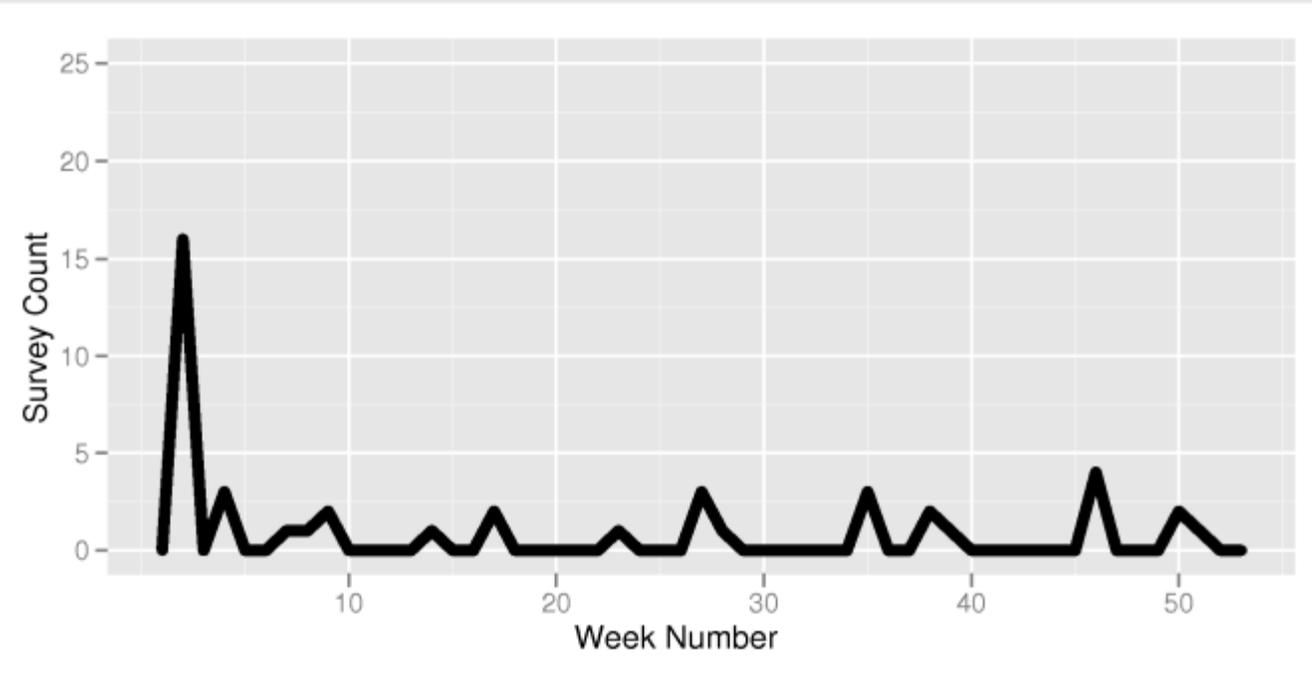


IDSAS System





Survey Date	Species	Reported Symptoms	Suspected Diagnoses
01/12/09	cattle	lameness	Footrot
01/12/09	cattle	decreased feed intake/milk production	OTHER diagnosis
01/12/09	cattle	decreased feed intake/milk production	Bloat
01/12/09	cattle	decreased feed intake/milk production	Bacteremia/Septicemia
01/13/09	cattle	respiratory signs	Pneumonia
01/13/09	cattle	decreased feed intake/milk production	OTHER diagnosis
01/13/09	cattle	decreased feed intake/milk production	Mastitis
01/16/09	cattle	decreased feed intake/milk production	OTHER diagnosis
01/16/09	cattle	decreased feed intake/milk production	Milk fever/hypocalcemia
01/16/09	cattle	decreased feed intake/milk production	Mastitis
01/16/09	cattle	decreased feed intake/milk production	Bacteremia/Septicemia
01/19/09	cattle	decreased feed intake/milk production	OTHER diagnosis
01/19/09	cattle	decreased feed intake/milk production	Milk fever/hypocalcemia
01/19/09	cattle	decreased feed intake/milk production	Mastitis
01/22/09	cattle	decreased feed intake/milk production	Mastitis
01/22/09	cattle	decreased feed intake/milk production	Ketosis/fatty liver
01/26/09	cattle	decreased feed intake/milk production	Ketosis/fatty liver
01/27/09	cattle	decreased feed intake/milk production	Ketosis/fatty liver
01/27/09	cattle	decreased feed intake/milk production	Babesia
01/28/09	cattle	lameness	Muscle inflammation

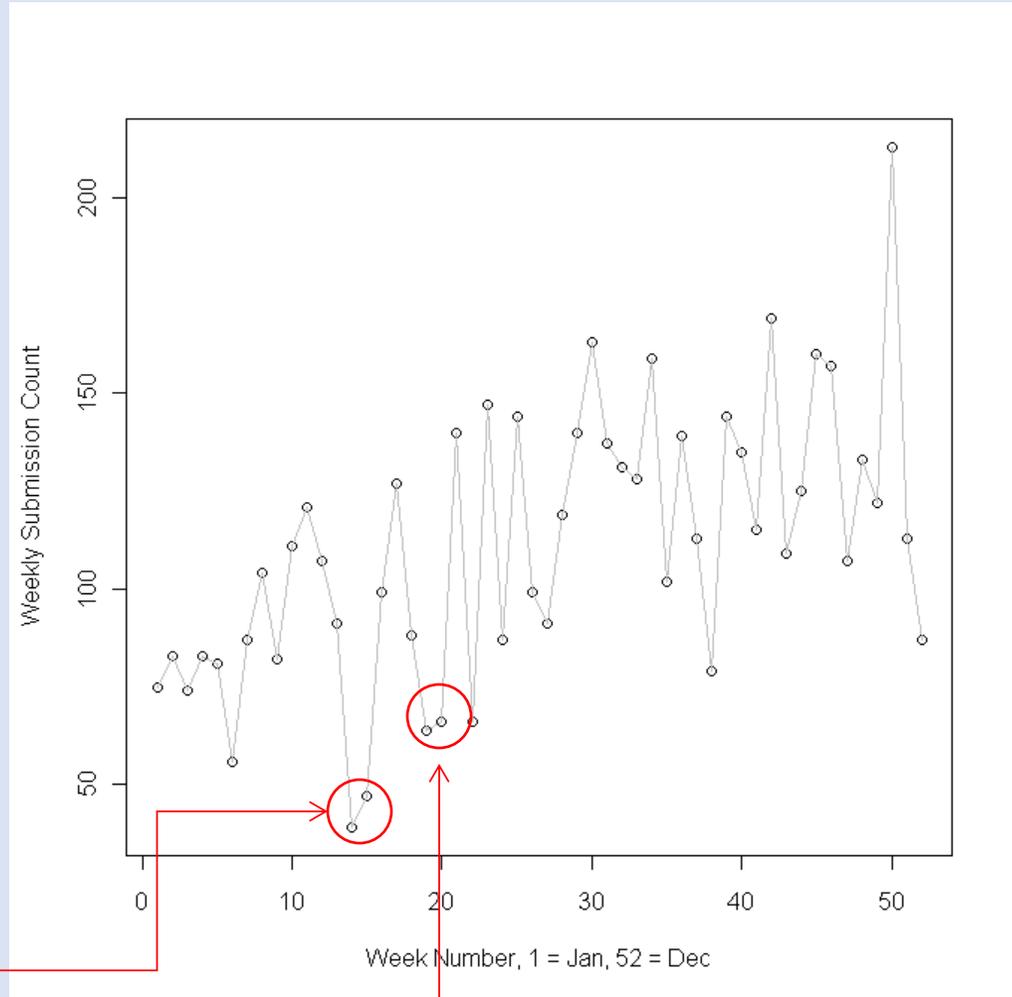


Survey Date	Species	Reported Symptoms	Suspected Diagnoses
01/08/09	buffalo	decreased feed intake/milk production	OTHER diagnosis
01/08/09	buffalo	decreased feed intake/milk production	Malnutrition
01/08/09	cattle	decreased feed intake/milk production	Mastitis
01/08/09	cattle	decreased feed intake/milk production	Bloat
01/08/09	cattle	decreased feed intake/milk production	Bloat
01/08/09	poultry	decreased egg production/weight gain/ appetite	IBD/gumboro
01/08/09	poultry	decreased egg production/weight gain/ appetite	Heat stress
01/08/09	poultry	decreased egg production/weight gain/ appetite	Coccidiosis
01/09/09	buffalo	decreased feed intake/milk production	OTHER diagnosis
01/09/09	cattle	decreased feed intake/milk production	Mastitis
01/09/09	cattle	decreased feed intake/milk production	Malnutrition
01/09/09	cattle	decreased feed intake/milk production	Bloat
01/10/09	buffalo	lameness	Trauma/fractures
01/10/09	buffalo	decreased feed intake/milk production	Metritis
01/10/09	cattle	decreased feed intake/milk production	Mastitis
01/10/09	poultry	decreased egg production/weight gain/ appetite	Coccidiosis

Patterns

IDSAS Data by Week

Total Submissions	5758
Avg / Week	111
Min / Week	39
Max / Week	213
Median / Week	110



National Examinations for Veterinarians

End of 25 year civil war

Sources of Variability

Data Generating Process

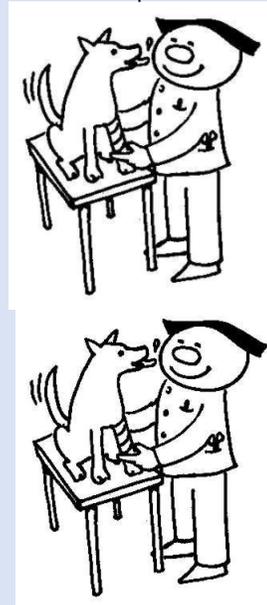
normal abnormal



$\Pr(X1) = ?$ $\Pr(X1) = ?$
 $\Pr(X2) = ?$ $\Pr(X2) = ?$

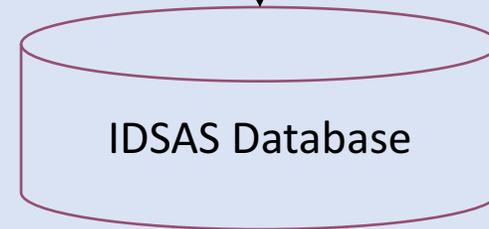
↑ ↑
Transition probability matrix

Data Authoring Process



Daily Electronic Submissions

Observed Data



Markov Model Development

$$\begin{array}{l}
 \text{HMM}_3 \left\{ \begin{array}{l}
 \text{Covariates} \\
 \lambda_{it} = \exp(\mu_{c_{it}} + \beta_1 \text{Male} + \beta_2 \text{Training} + \beta_3 \text{Years} \\
 + \beta_4 t + \beta_5 \text{Precip} + \beta_6 \text{Temp})
 \end{array} \right.
 \end{array}$$

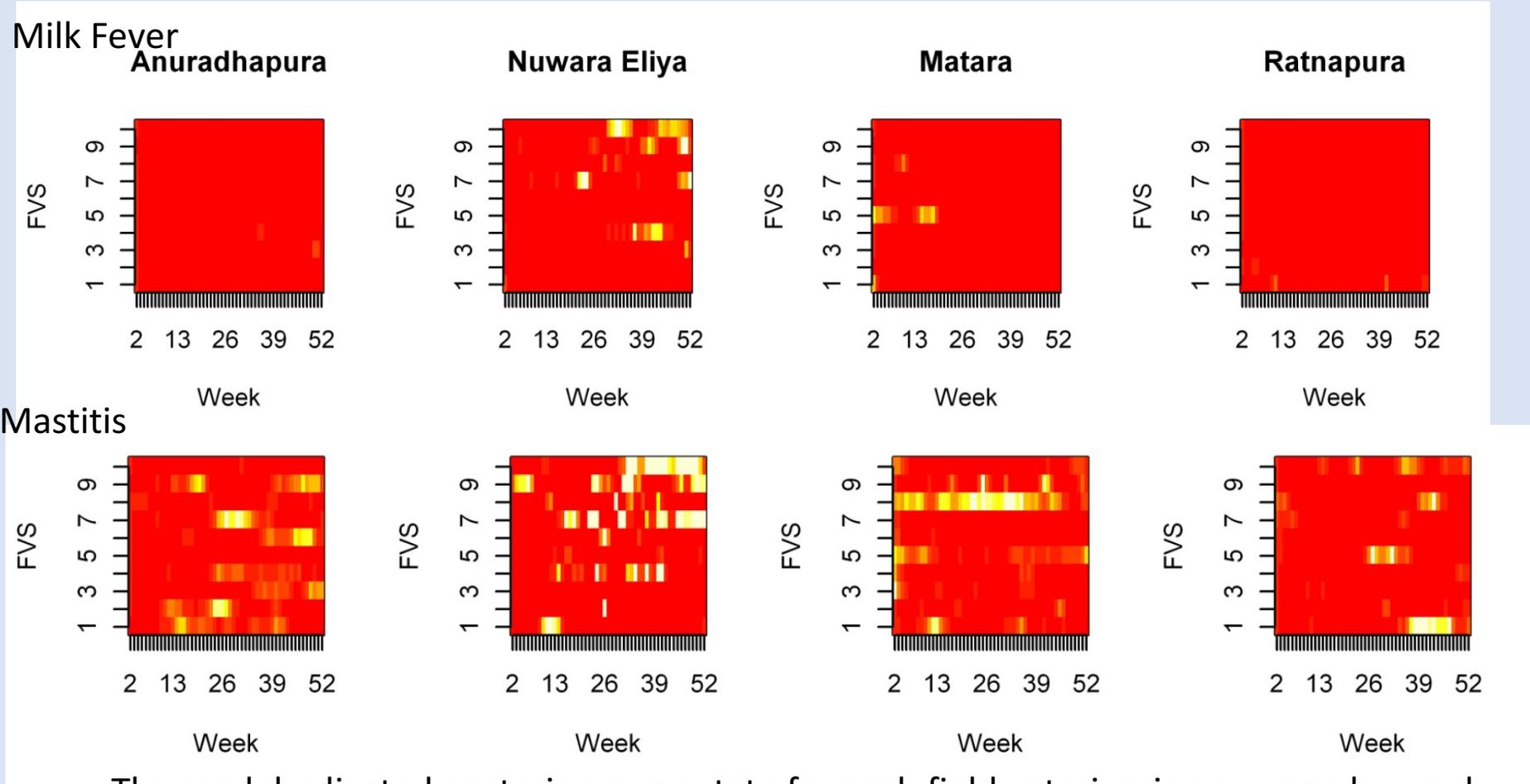
Data Authoring Terms

Data Generating Terms

Model Selection

- Bayesian Information Criterion (BIC)
 - Relative measure of model fit and complexity
 - Lower score = more parsimonious model

Visualizing 'Unusualness' At the User-Level



The model-adjusted posterior mean state for each field veterinarian surgeon by week

From Patterns to Populations

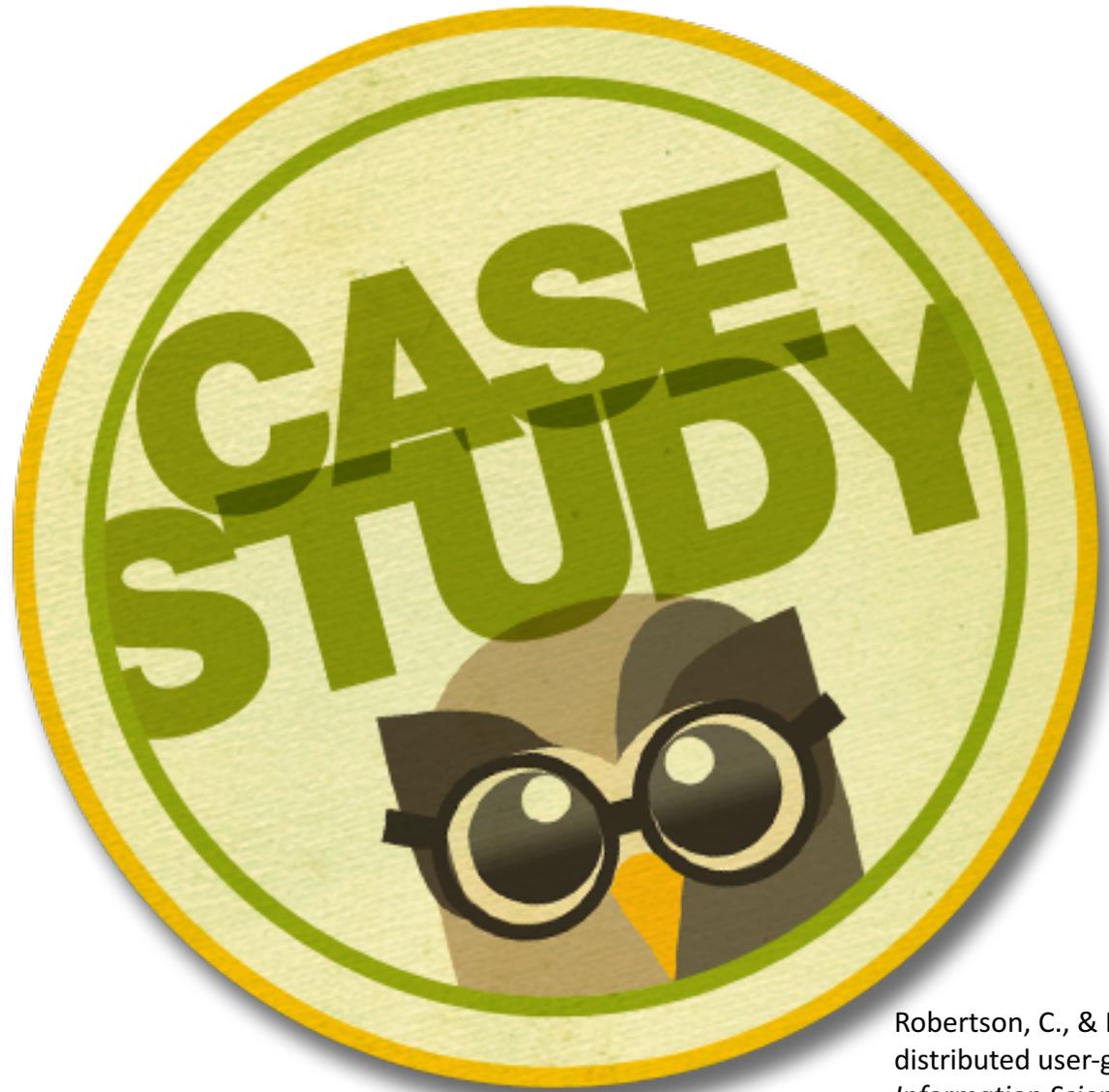
Key Concept - Representativeness

- VGI is often inherently uncertain, noisy and ‘collected’ for other purposes
- Research has focused on understanding dimensions of VGI data quality and increasingly probing issues of who is represented
 - socioeconomic class
 - demographic factors
 - motivational factors
- Digital Divides and Differentials...
 - empowering coders, engineers and data providers
 - disempowering those in “digital shadows”
 - reinforcing, amplifying biases

From Patterns to Populations

Key Concept - Representativeness

- Implications for analytics
 - Does the sample reflect the population?
 - sampling design
 - sampling frames
 - replication and validation
 - Do we even need to think in terms of populations and samples?



#2

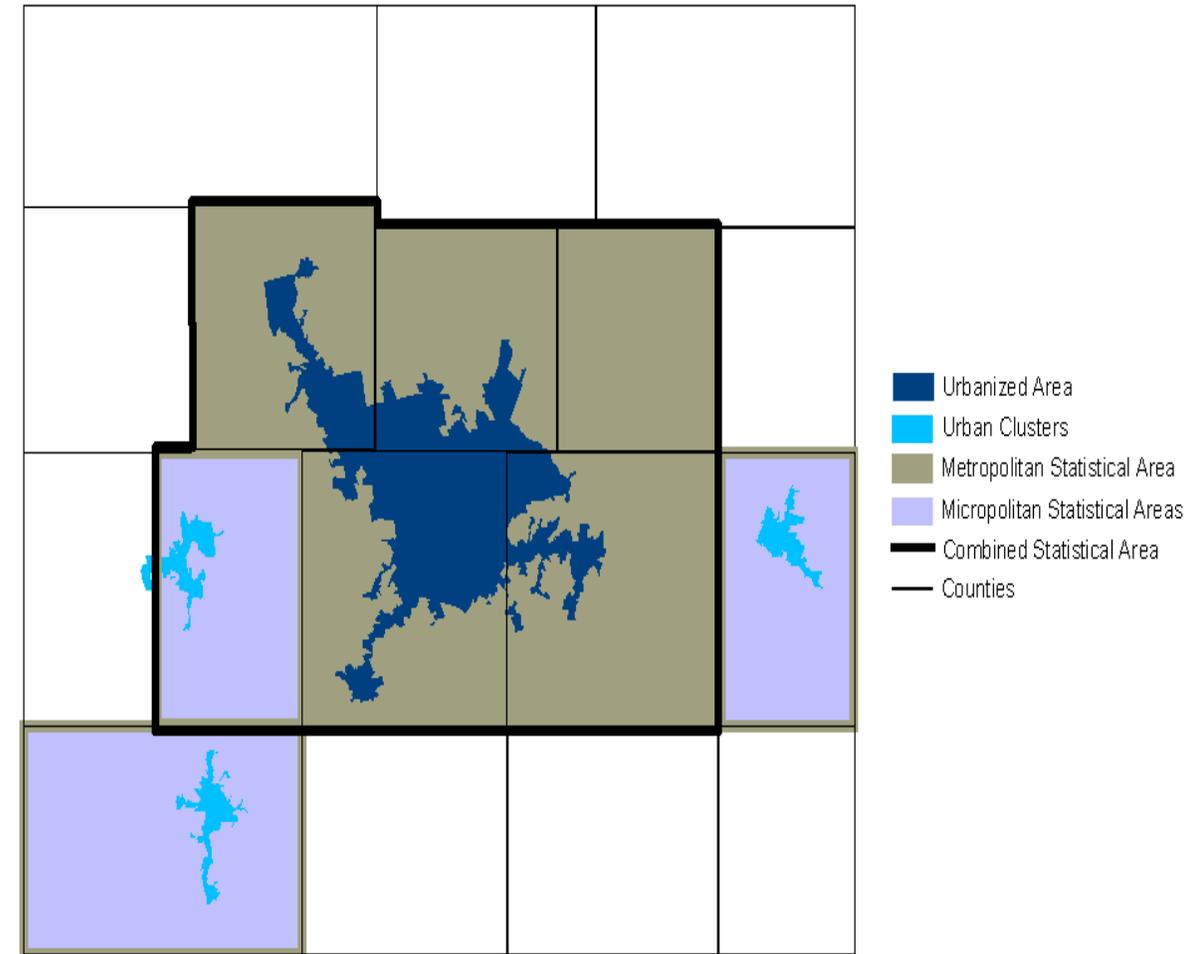
Robertson, C., & Feick, R. (2015). Bumps and bruises in the digital skins of cities: unevenly distributed user-generated content across US urban areas. *Cartography and Geographic Information Science*, 0(0), 1–18.

Representativeness Analytics Example

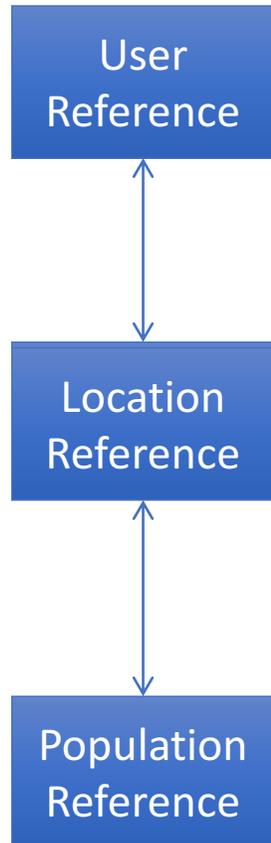
- Represent densely developed territory
 - Residential
 - Commercial
 - other non-residential urban land uses in which social and economic interactions occur
- Represent the actual “Urban Footprint”
- 481 UAs - continental US

Data Acquisition Strategy

- Obtained Flickr point data (metadata) from Flickr public API
- For each Urban Area (UA), laid a grid of points as search areas over the entire UA – 1.5 km spacing
- Repeated search queries to Flickr API over all search points to obtain GTP records
- Total number of points: 2.7 million points



Defining Local Populations



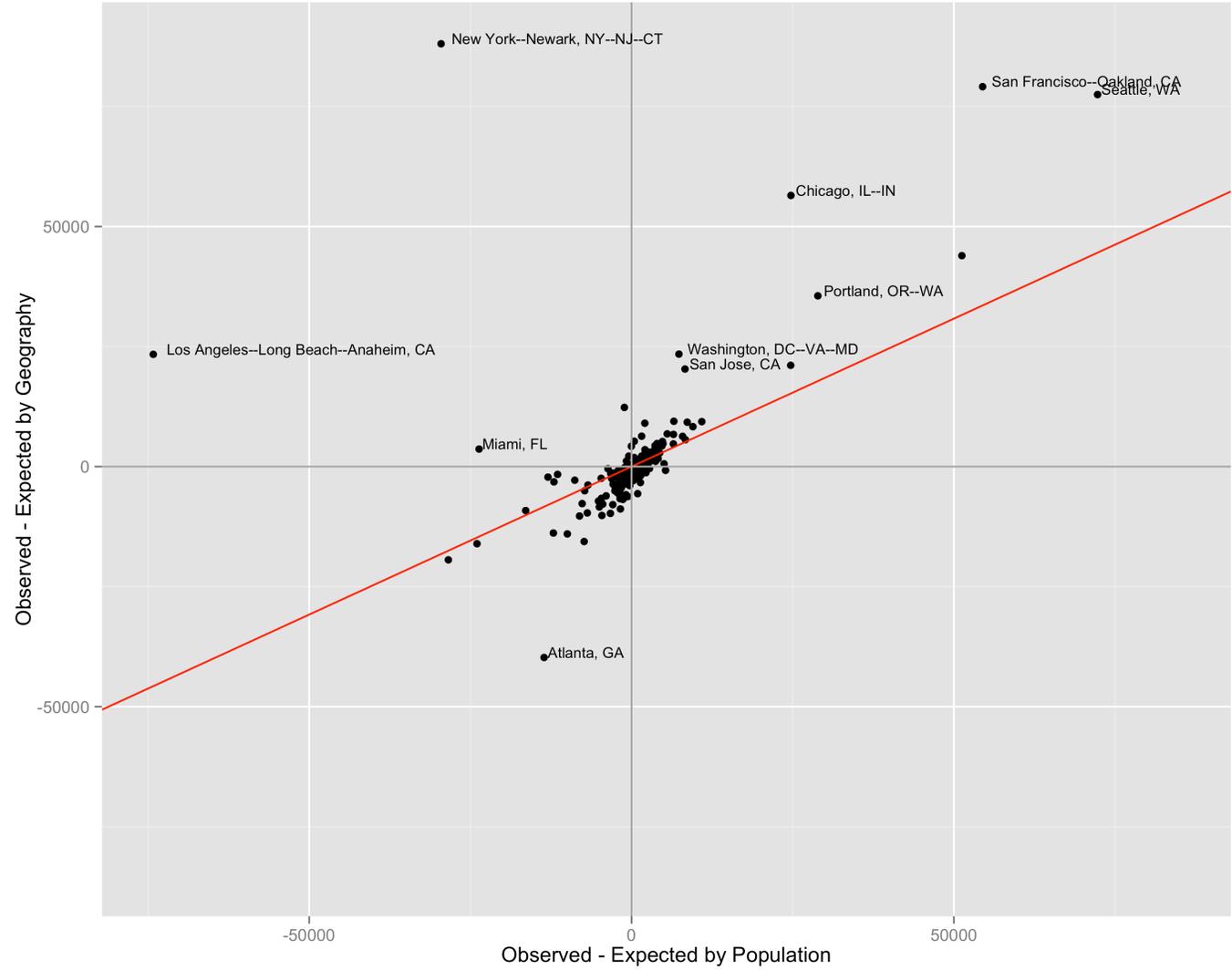
- Making connection from GTP locations to underlying populations through location reference can be facilitated if we can distinguish between
 - local residents
 - non-local residents
- Used 10-day time-location threshold to distinguish local and non-locals in each UA
 - Li, Goodchild and Xu (2013); Jankowski et al (2010)

Multi-Scale Analysis

- Within each Urban Area – stratify GTPs by census tract
 - GTP Population of each census tract
 - Income inequality within each census tract (Gini)
- Fit hierarchical linear model with UA-level random effect and identify effects of local scale factors
 - income inequality
 - % tourists
 - Population

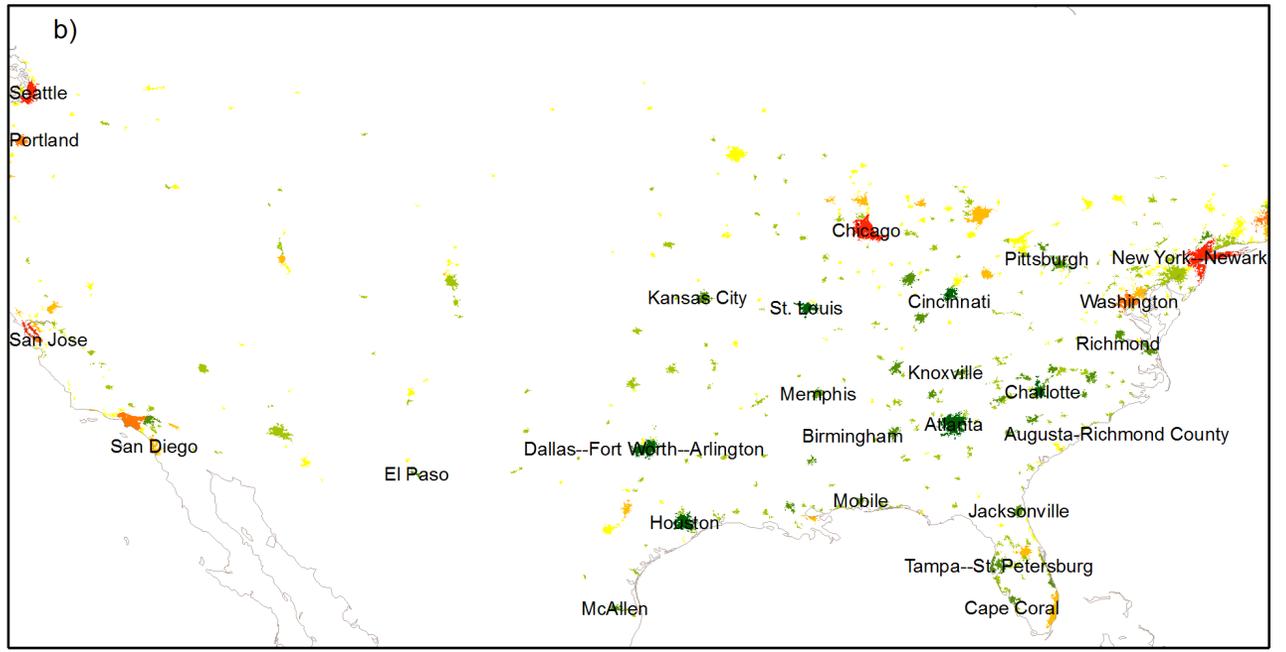
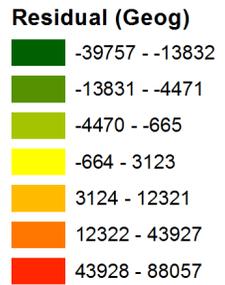
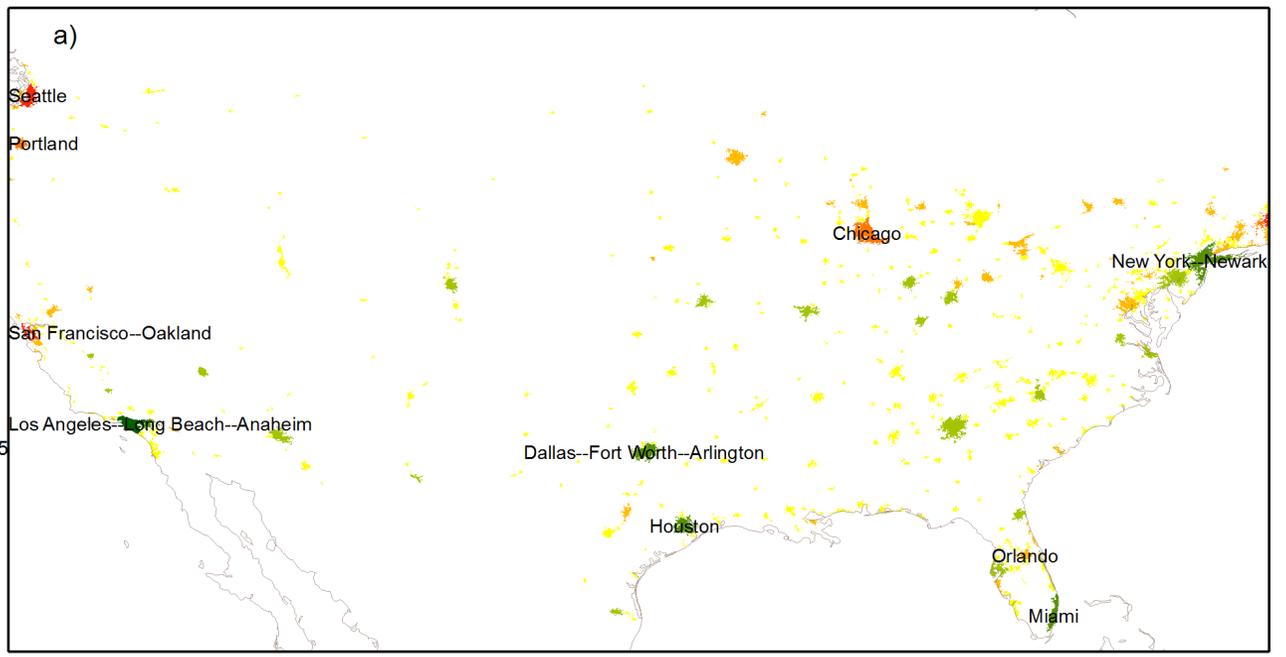
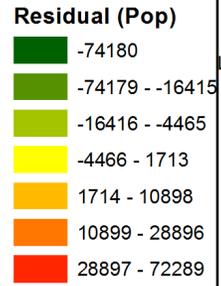
GTPs vs. Unique User Counts

- Widely varying baseline rates of GTPs relative to # of unique users in each area
- Different geographic forms associated with GTP production



Populations

Population vs Geography as basis for Representativeness



Local vs Non-Local Populations

Model A -

Coefficient	Estimate	t-value
Intercept	-15.88	-3.85
GINI	1.54	18.66
Population	< 0.00	1.46
Model AIC: 589198		

Model B -

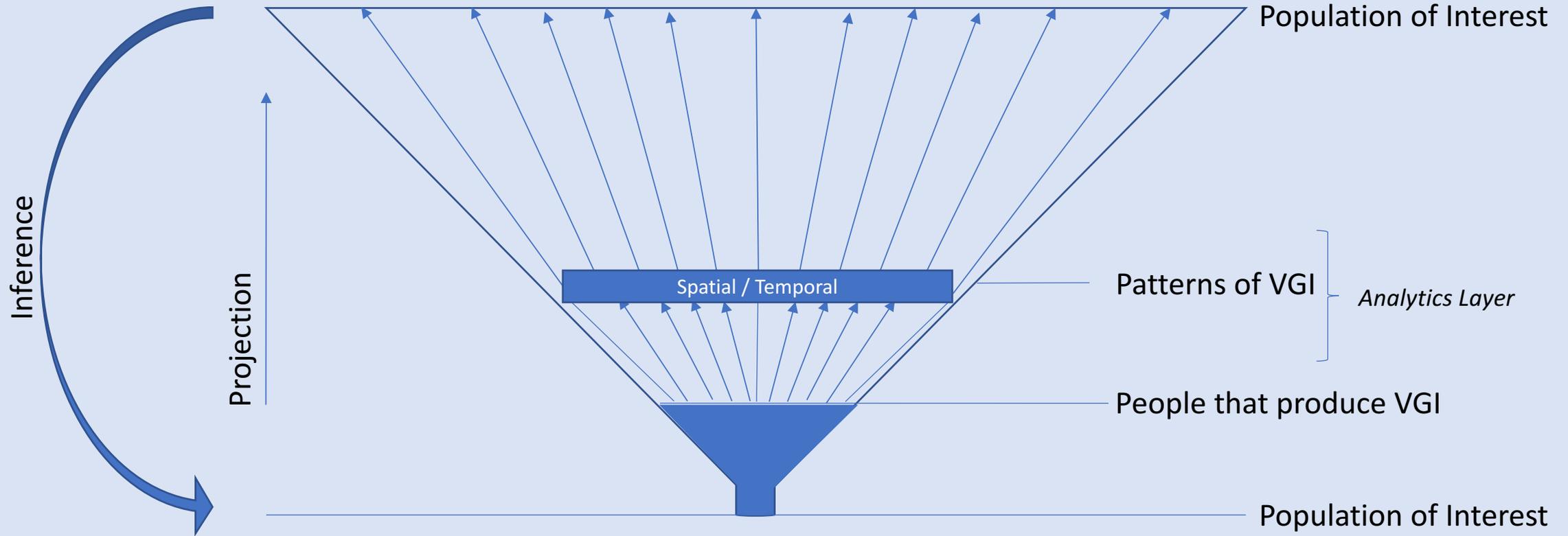
Coefficient	Estimate	t-value
Intercept	-29.65	-4.77
GINI	1.92	13.93
Population	< 0.00	0.017
% Tourists	48.05	3.73
Population x % Tourists	-0.92	-2.94
Model AIC: 541585		

Person-Place Linkages in VGI

Table 4 – Poisson modeling results for selected covariates and geotagged photos at the urban area scale across the United States (* indicate significant at $\alpha = 0.05$).

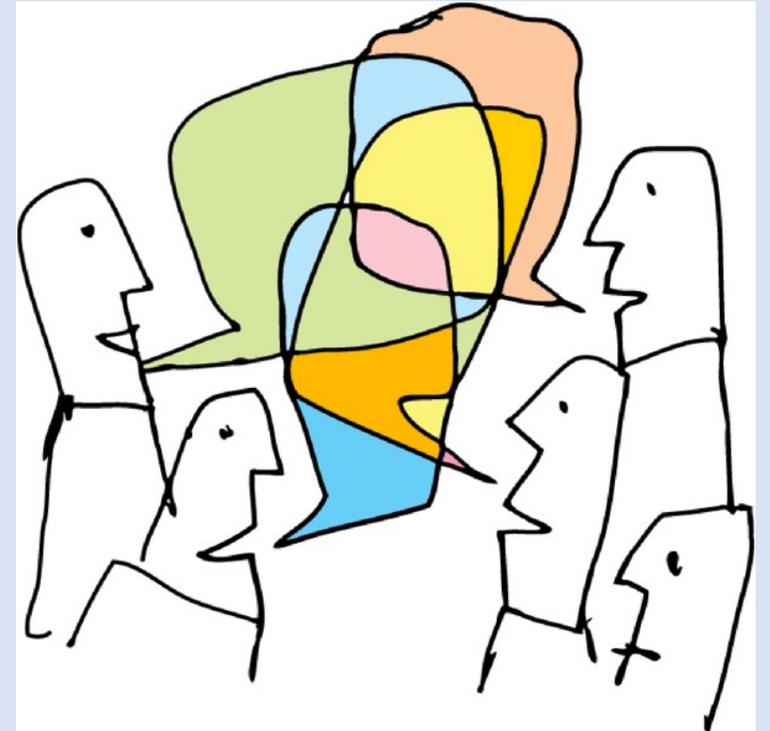
Coefficient	Estimate	p-value
Intercept	-3.572	<0.001*
GINI coefficient	0.396	0.748
% Under poverty line	-0.113	<0.001*
% Unemployed	0.046	0.010*
% Vacancy	0.009	0.117
% 1 unit detached housing	0.001	0.694
Estimated housing value	<0.001	0.436
% Walk to work	0.097	<0.001*
Estimated travel time to work	-0.031	<0.001*

People, Patterns and Populations Revisited



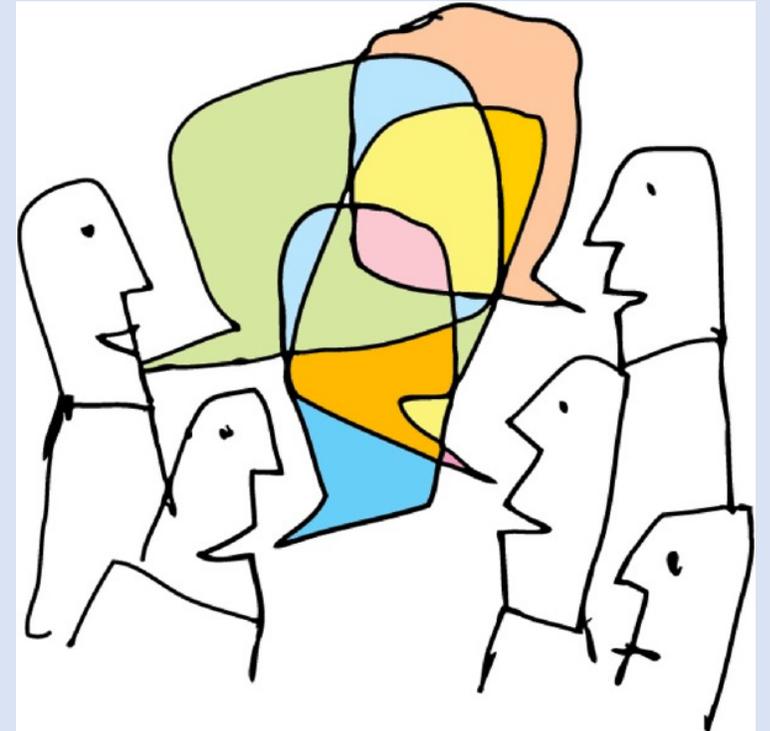
Discussion

- Can we define new categories of users, new metrics, and new analytics assemblages that support production of new geographic knowledge?
- Citizen science example demonstrates the fluid nature of expert-amateur binaries
 - Estimating both the data-generating and data-authoring processes separately?
 - Need to take user-heterogeneity seriously – could be informed by cross-platform data
- Flickr modelling take multi-scale modelling approach to VGI pattern detection
 - Where are unusual areas within and across cities?



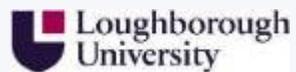
Discussion

- Key questions raised by these examples
 - Does representativeness matter in VGI analytics? Is the sample enough? Person-place linkages?
 - Can we develop new ways to characterize the experience and expertise of users
 - Across platforms or objects?
 - Privacy considerations?
 - Is distinguishing between VGI and GI still relevant? How do we move beyond binary classifications to inform method and tool development; research design considerations; etc.



Acknowledgements

- Canadian Institute of Health Research (Canada)
- National Sciences and Engineering Research Council (Canada)
- Social Sciences and Humanities Research Council (Canada)
- Numerous study participants – both passive and active!
- Colleagues and collaborators on several research projects...



Thanks for your attention!

People, Patterns, and Populations

Considering multiple perspectives in the spatial analysis of VGI

Colin Robertson crobertson@wlu.ca, [@colinr23](https://twitter.com/colinr23), www.thespatiallab.org

Rob Feick robert.feick@uwaterloo.ca, athabasca.uwaterloo.ca/sdslab