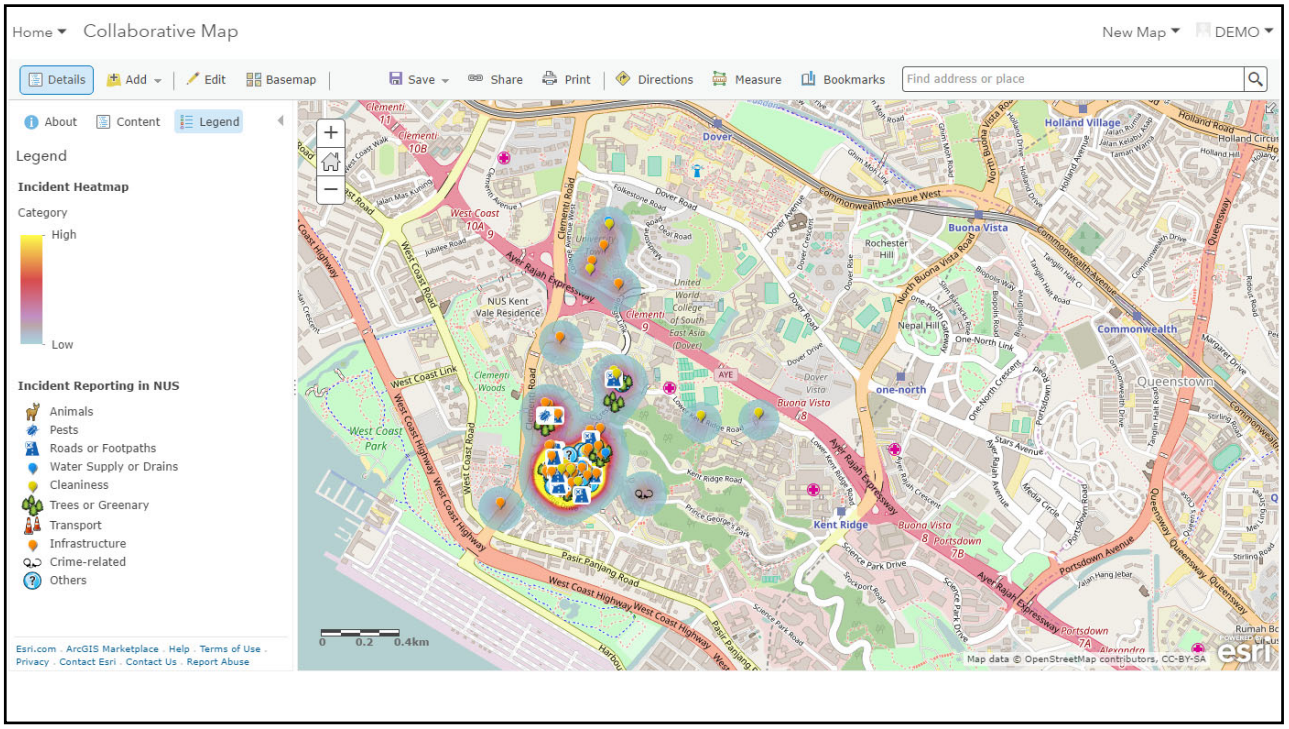
05. Volunteered Geographic Information (VGI)

GE3238 GIS Design and Practices

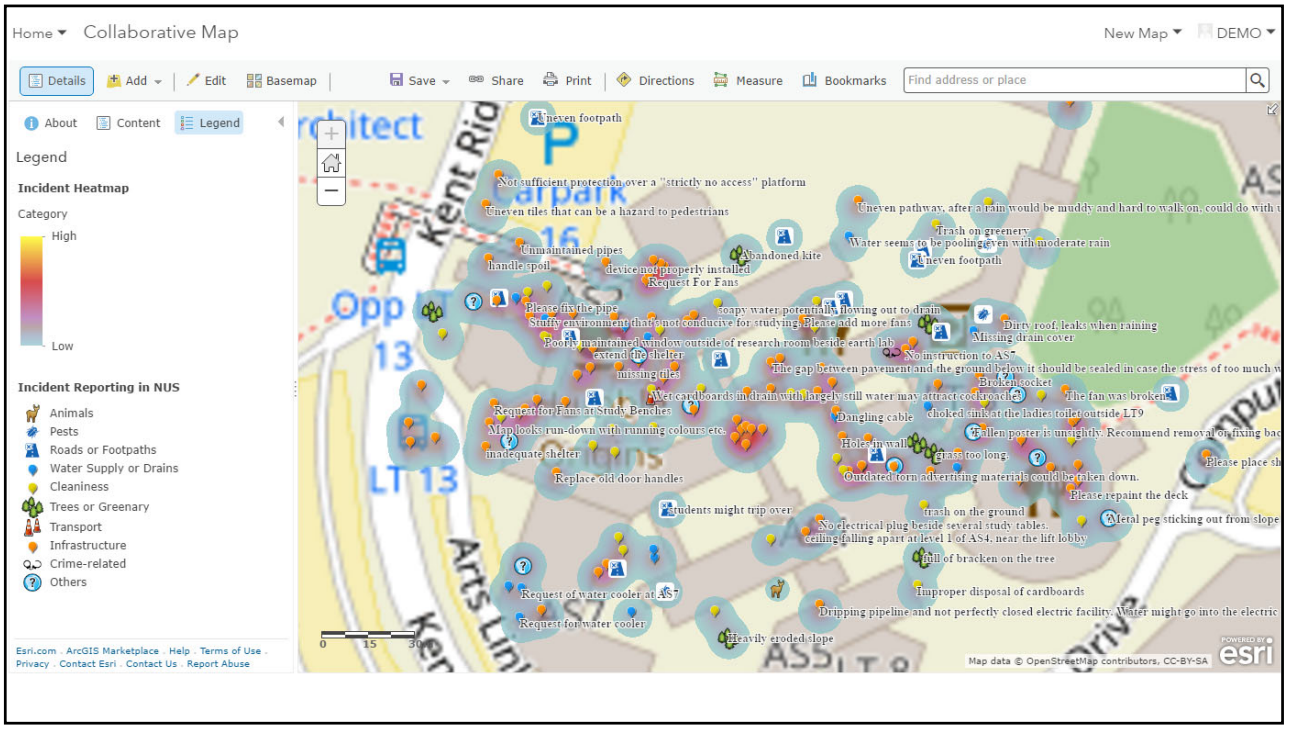
Geography, NUS

Chen-Chieh FENG

1



2



3

VGI is not new!

- Citizen scientists: communities or networks of citizens who act as observers in some domain of science, e.g.,
  - Christmas Bird Count (since 1900)
    - http://www.audubon.org/conservation/science/christmas-bird-count
  - [Butterfly | Dragonfly | Garden Bird] Watch
    - https://www.nparks.gov.sg/nature/community-in-nature#citizen-science
- Mail surveys
- For Geography: land use, land cover, soil classes, tree species

Key terms

VGI
Citizen scientists
Mail survey

4

Goals

- VGI and its characteristics
  - OpenStreetMap (OSM) as an example
  - (VGI is not new) Implications for questions involving geospatial?
- Getting insights from two other VGI sources?
  - Flickr (images): Post-disaster recovery detection
  - Twitter (micro-blogging): Information diffusion during public health event

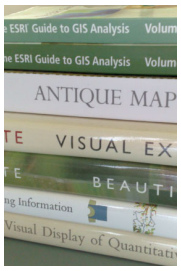
5

VGI and Some of Its Characteristics

- The engagement of large number of citizens often with little formal qualification in the creation of geographic information
  - Their actions are almost always been voluntary and nonmonetary
  - Their results may not be accurate (Goodchild, 2007, Haklay 2010)
- They collectively present a dramatic innovation with impact on GIS, Geography, and its relation to the general public (Goodchild, 2007)

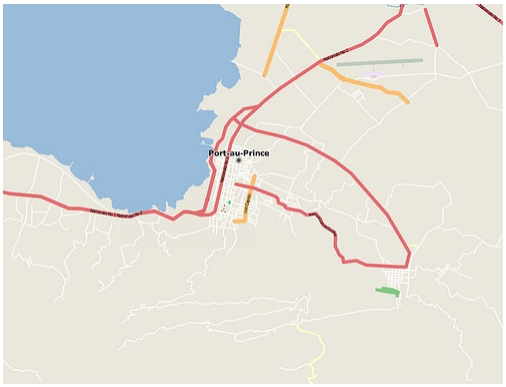
Key terms

VGI
Voluntarism
Accuracy
Collaboration

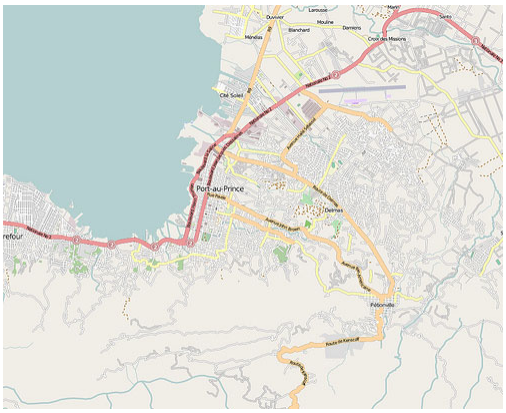


6

Haiti Disaster and Collaborative Mapping



The earthquake occurred at 16:53 local time (21:53 UTC) on Tuesday, 12 January 2010 (Wikipedia, 2010 Haiti Earthquake).



Haiti OpenStreetMap Response

January 14, 2010 at 9:45 pm · Filed under disaster, openstreetmap

7

Data Quality of VGI?

NCGIA Research Initiatives

- **Initiative 1: Accuracy of Spatial Databases**

[Leader: Michael Goodchild (Santa Barbara) - begun December 1988, closed October 1990]

http://www.ncgia.ucsb.edu/research/initiatives.html

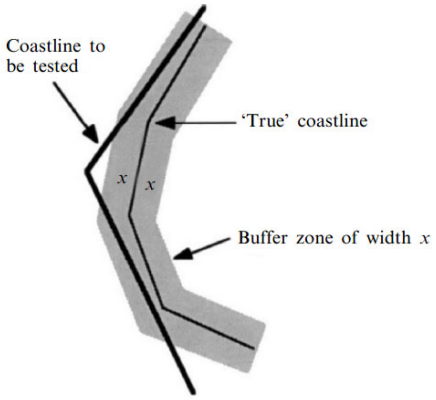
Key terms

Data quality
Measurement

- How do you verify the quality of data from volunteers or the crowds?
  - Eyeballing?
  - Confidence?
  - Crowd voting?
  - Benchmark against a data source with higher data quality?

8

Positional Accuracy of OSM, London



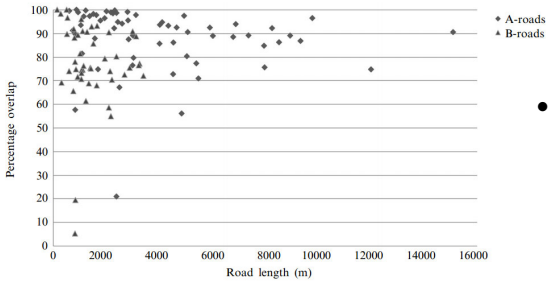
- Motorway
  - Average overlap of 80%, with variability of 60% – 89%
- Smaller roads (i.e., A- and B-roads)

Key terms

Data quality
Positional accuracy
OSM
Buffer
Overlay

9

Positional Accuracy of OSM, London



- Motorway
  - Average overlap of 80%, with variability of 60% – 89%
- Smaller roads
  - A-road: average overlap of 88%
  - B-road: average overlap of 77%

Key terms

Data quality
Positional accuracy
OSM

10

Completeness

- Measured by the different between the total of **OSM road lengths** and **authoritative data road lengths**

Key terms

Data quality
Completeness

11

Implications?

Data quality indicators

- Positional accuracy
- Attribute accuracy
- Logical consistency
- Lineage
- Completeness

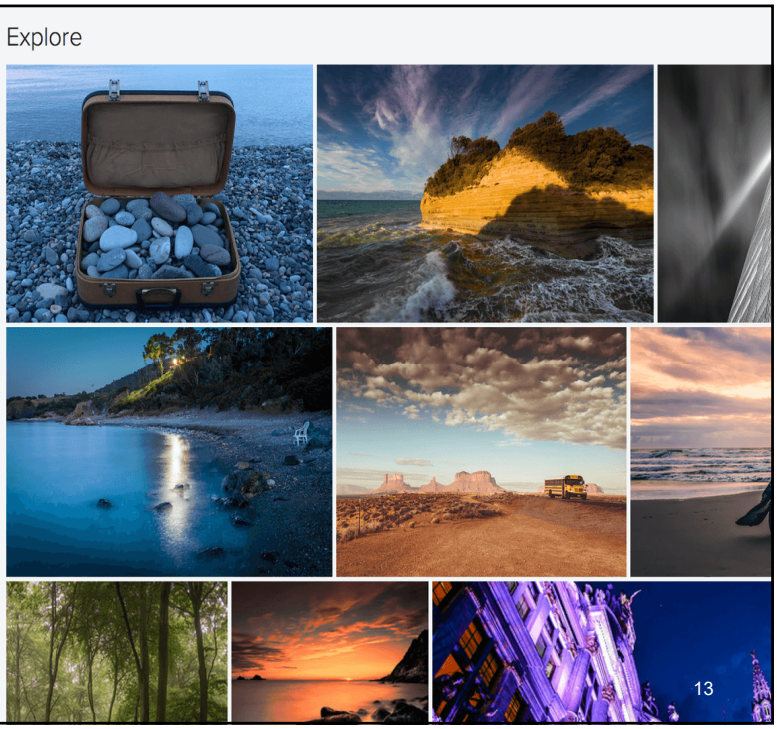
Usage of VGI in geospatial applications

- ?

12

Images as the format of VGI input

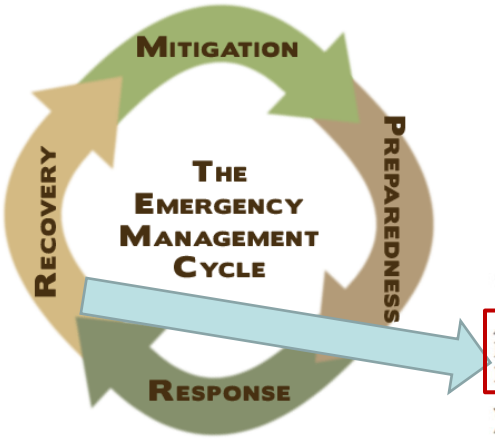
Flickr is a media platform used for uploading, organizing, and sharing digital media such as photos and videos

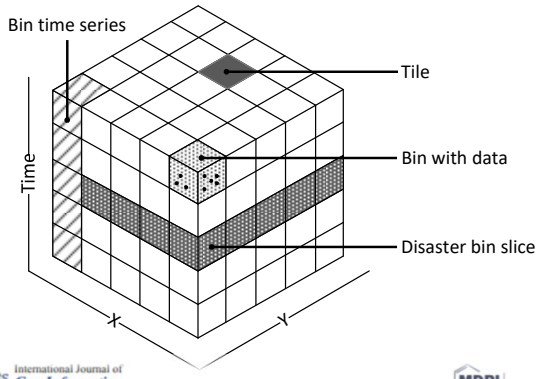


13

- VGI can be spatial big data
- Hum, some photos are available. For GIS design and practices, so what?
  - What insights can I get from these photos?
  - How do I get these photos and what structural information is available that I can take advantage of?

14





isprs International Journal of Geo-Information
MDPI

Article

Monitoring and Assessing Post-Disaster Tourism Recovery Using Geotagged Social Media Data

Yingwei Yan *, Melanie Eckle, Chiao-Ling Kuo, Benjamin Herfort, Hongchao Fan and Alexander Zipf

Institute of Geography, Heidelberg University, 69120 Heidelberg, Germany; eckle@uni-heidelberg.de (M.E.); kuo@chiaoling.com (C.-L.K.); herfort@uni-heidelberg.de (B.H.); hongchao.fan@uni-heidelberg.de (H.F.); zipf@uni-heidelberg.de (A.Z.)

* Correspondence: Yingwei.Yan@uni-heidelberg.de or yanyingwei@u.nus.edu; Tel.: +49-6221-54-5509

Academic Editors: Marguerite Madden and Wolfgang Kainz

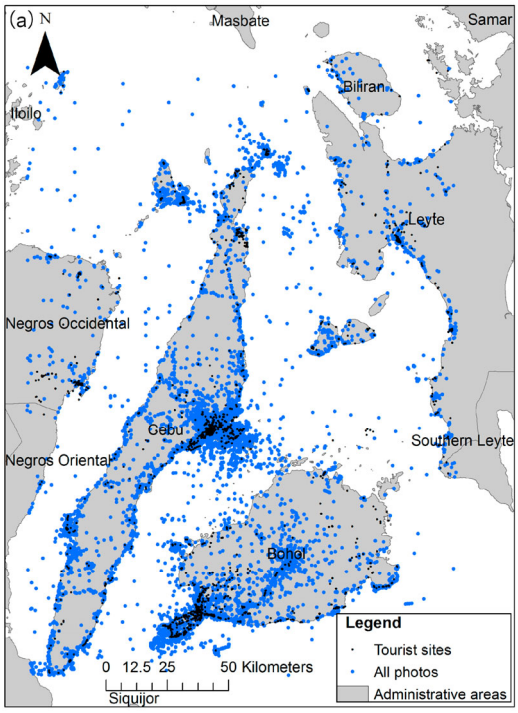
Received: 27 January 2017; Accepted: 27 April 2017; Published: 3 May 2017

15

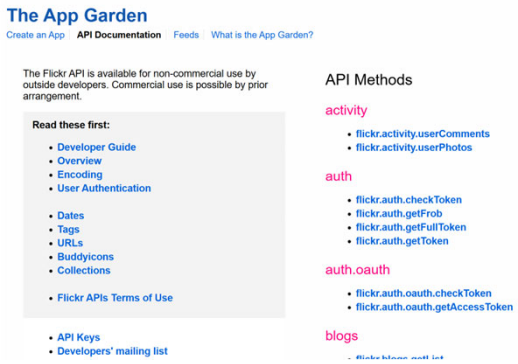
The Big Deal (of VGI)

- Can be a cost-effective source of geographic data with big volume and big spatiotemporal coverage
- Can be a source of timely geographic data during emergencies
- Non-experts can be engaged in and benefit from the spatial knowledge discovery, and academic research can benefit from the engagement of non-experts

16

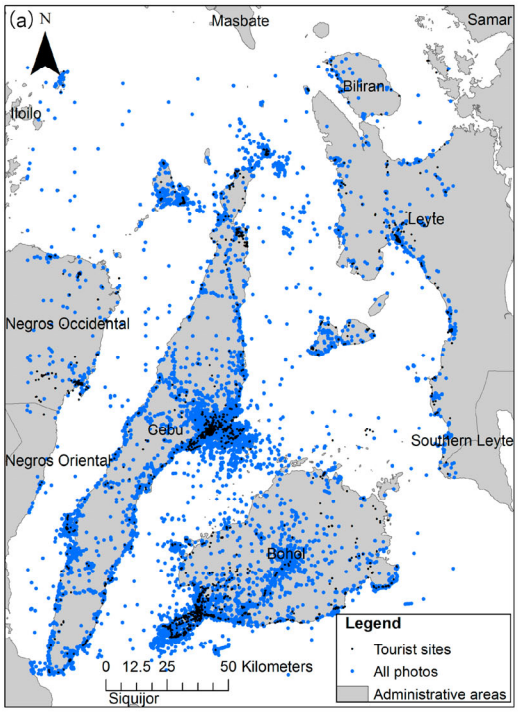


- Data accessible via Flickr API

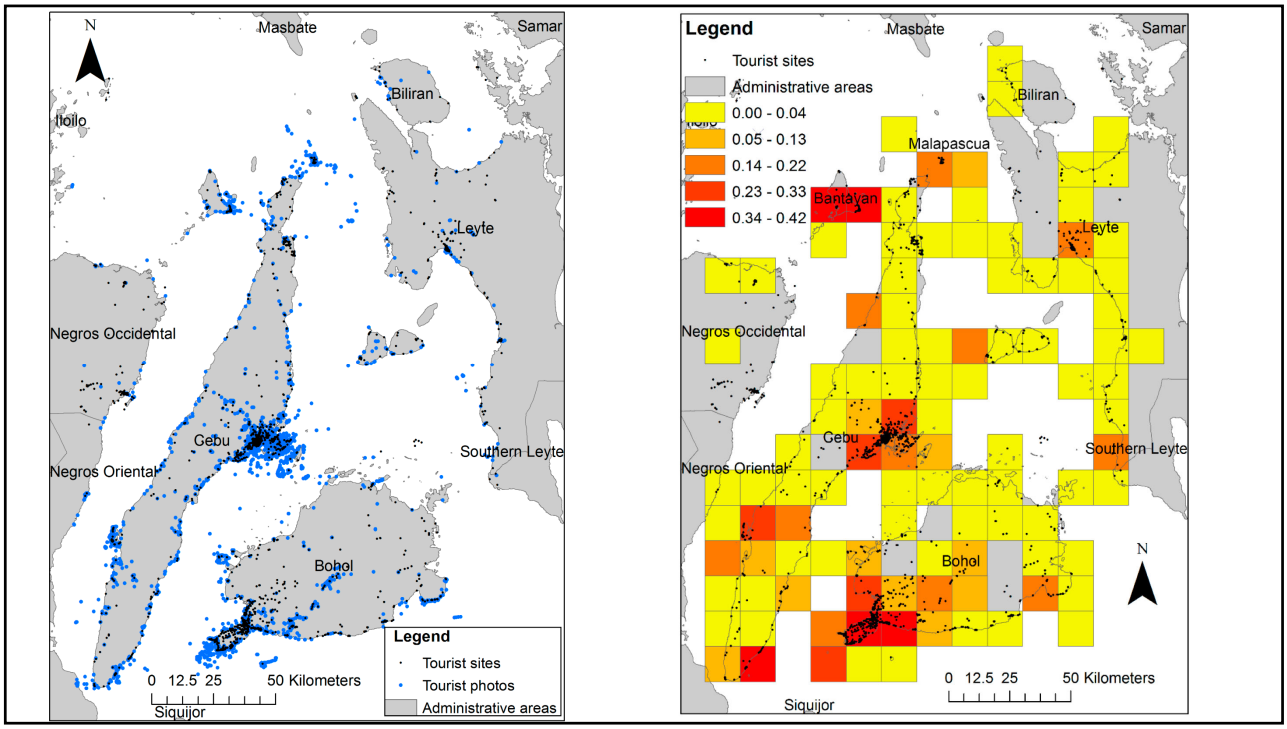


https://www.flickr.com/services/api/

17



18



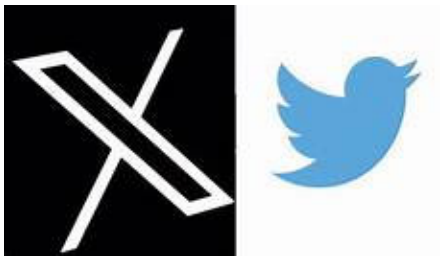
19

Food for thoughts

- Similar to the understanding of population (or demographic) distribution
  - Daytime versus nighttime
  - Static versus dynamic
  - Generic versus specific (population)
- Can we predict population distribution and predict it at a finer spatiotemporal resolution?

20

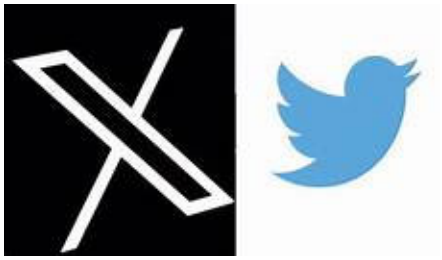
Short texts as the form of VGI input



21

21

How can the diffusion of rumors about [something] be captured?



22

22

Background

- The epidemic event Middle East respiratory syndrome (MERS) infection in Korea in 2015
  - Information delivery by CDC Korea is not real-time
  - Incomplete or no information by mass media
- New media (as opposed to mass media)
  - User generated content; human dynamics on the ground
  - Rumors and their propagation is unprecedented

23

Methods and Data

- Twitter data with keyword *MERS* in Korean (*메르스*)
  - Data parsed with KoNLPy (Korean NLP Python toolkit)

https://konlpy.org/en/latest/

24

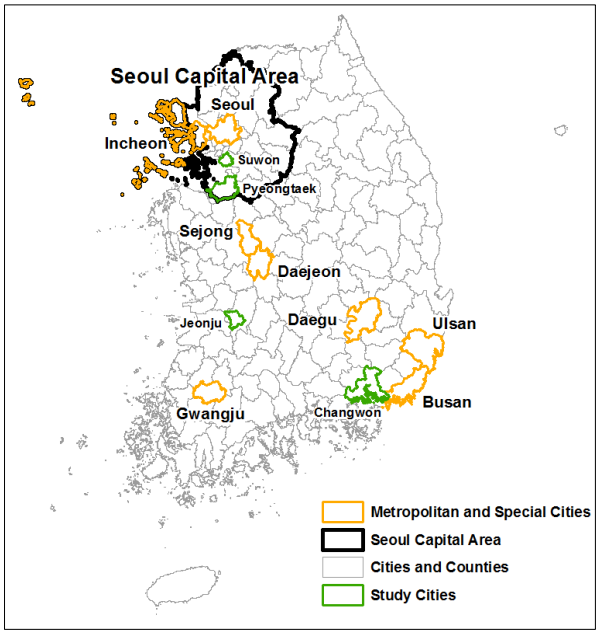
Methods and Data

- Twitter data with keyword *MERS* in Korean (*메르스*)
  - Data parsed with KoNLPy (Korean NLP Python toolkit)
- Events (all in 2015)
  - 20 May: first diagnosed case reported
  - 1 June: first death case reported
    - 4 June: Seoul mayor released partial information
    - 7 June: CDC South Korea released all information
  - 23 Dec: MERS outbreak officially over
- Data collected from 25 May 2015 – 23 Dec 2015
  - About 3,500K tweets before 31 July were analyzed
  - Twitter users: about 250K

25

Spatial Pattern and Interaction

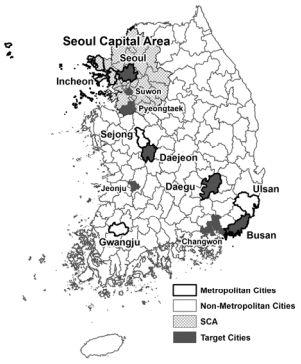
- Eight target cities that are spatially dispersed
  - Metropolitan cities: Seoul, Daejeon, Daegu, and Busan
  - Nonmetropolitan cities: Pyeongtaek, Suwon, Jeonju, and Changwon
- Locations
  - Twitter users and the locations mentioned
  - Weighted mean center was used

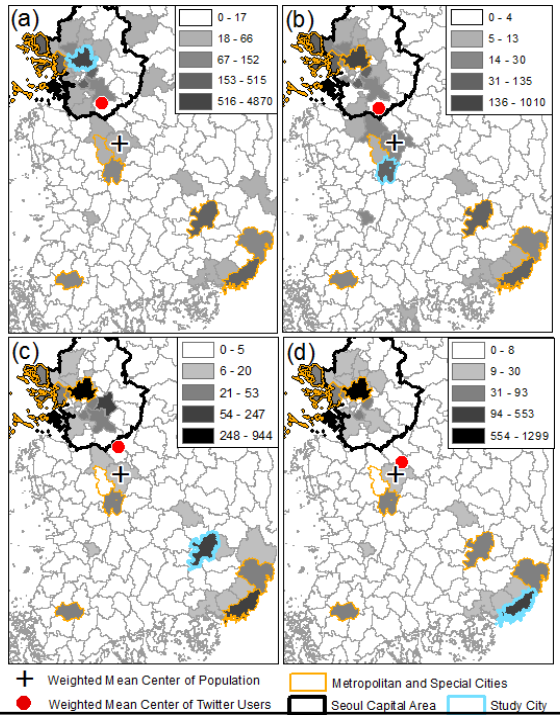


26

Spatial Pattern & Interaction

- Spatial distribution of Twitter users who mentioned the target metropolitan city of:
  (a) Seoul;
  (b) Daejeon;
  (c) Daegu; and
  (d) Busan
- Darker grays indicate higher numbers of Twitter users

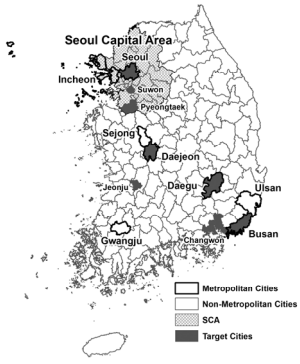


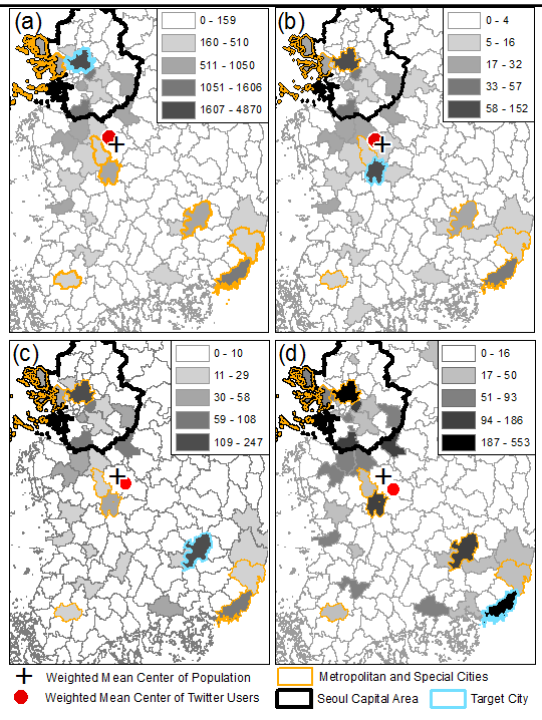


27

Spatial Pattern & Interaction

- Spatial distribution of cities mentioned by Twitter users from the target metropolitan city of:
  (A) Seoul;
  (B) Daejeon;
  (C) Daegu; and
  (D) Busan.





28

Summary

- VGI and its role in GIS design and practices
- Technical
  - May provide information not possible from authoritative data sources, but can be messy
  - ArcGIS Pro has tools to convert these formats, but you need to know their existence
- Conceptual
  - Data quality issue is even more critical and challenging; It is relative, contingent on the application involved

29