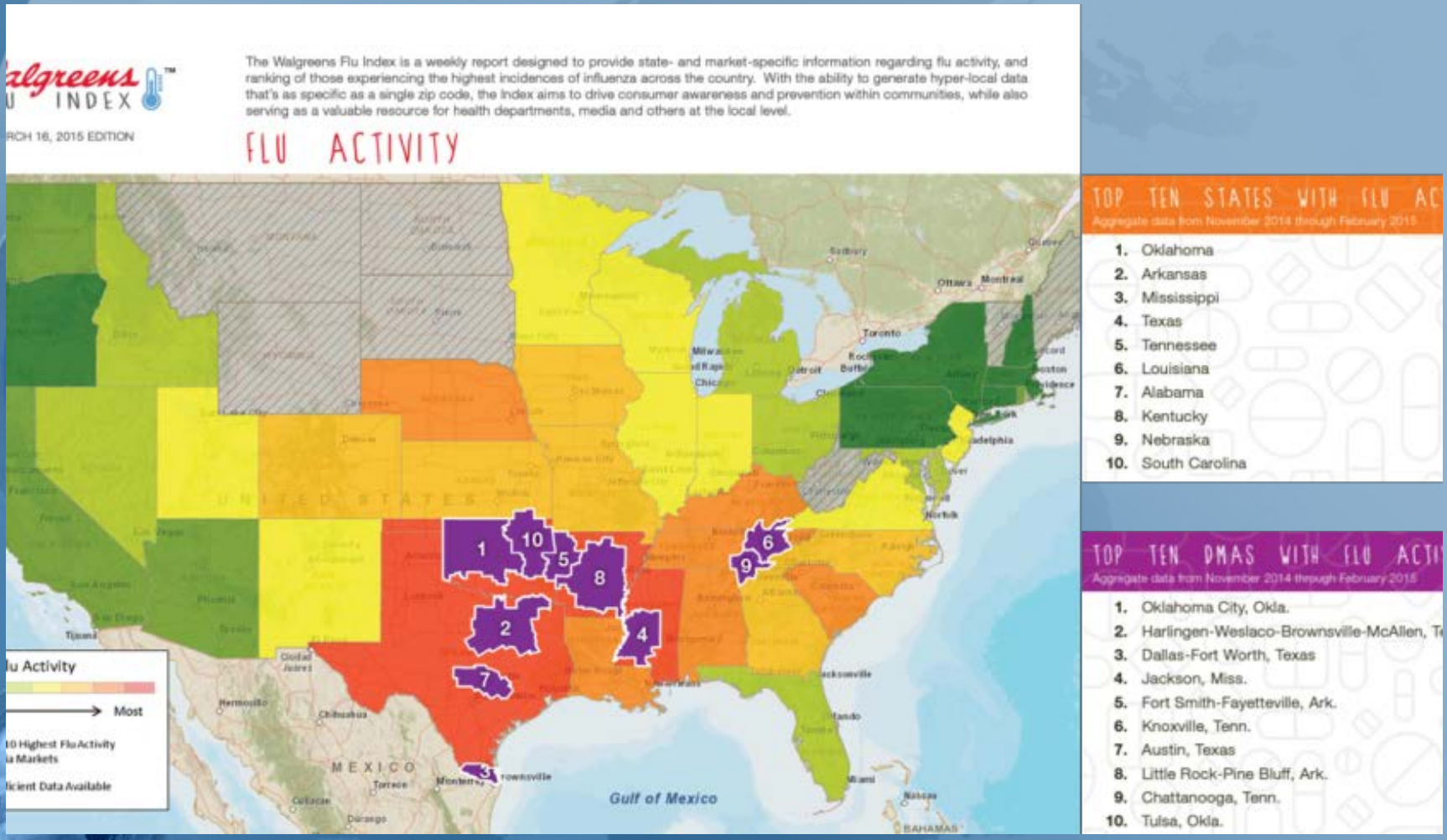
Towards data driven decision-making using crowd-sourced geographic information

Linna Li (linna.li@csulb.edu)
University of Redlands
December 2, 2015

WalMap: Walgreens store location map



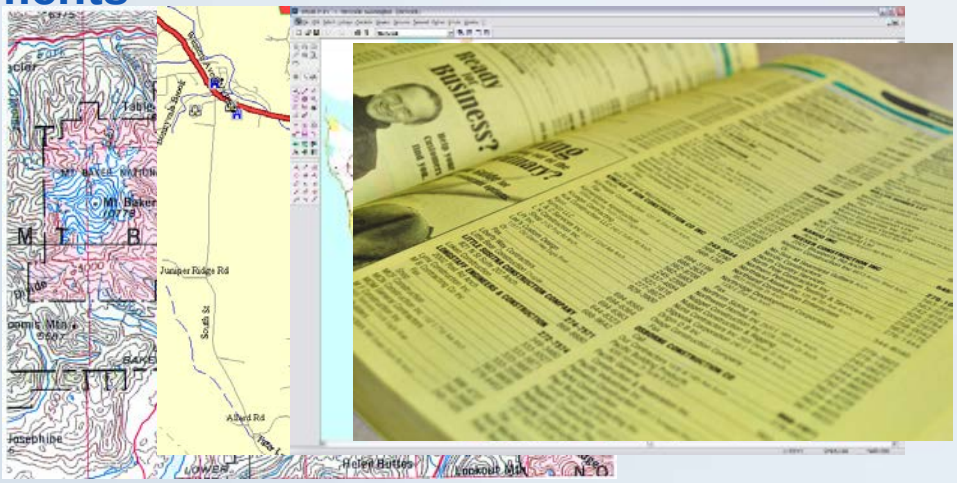
Geographic information

- **information about specific characteristics of places on or near the Earth's surface**
- **<x, z> where x is a location in space-time, z is a set of general properties associated with x**

- **information about pairs of locations**
- **$<x^1,x^2,z>$ where z is distance, interaction, travel cost, travel time, connectivity, direction, *etc*.**

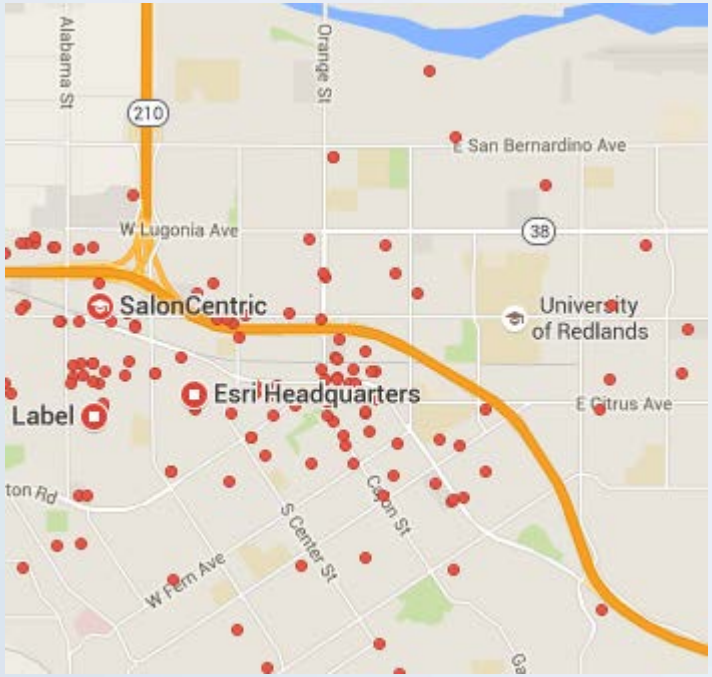
Traditional sources of geographic data

- **By authorities and their experts**
  - **USGS**
  - **U.S. Census Bureau TIGER files**
    (**T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing)
  - **NGA**
  - **military in many countries**
  - **state and local governments**
  - **data companies, e.g.**
    street networks,
    business locations



Geospatial technologies

- **Positioning on the Earth's surface**
  - **GPS, RFID**
- **Acquisition systems**
  - **remote sensing**
  - **field measurement**
- **Web GIS and mapping**
  - **Google Maps, Google Earth**
  - **OpenStreetMap**



Business locations in Google Maps

New sources of geographical data

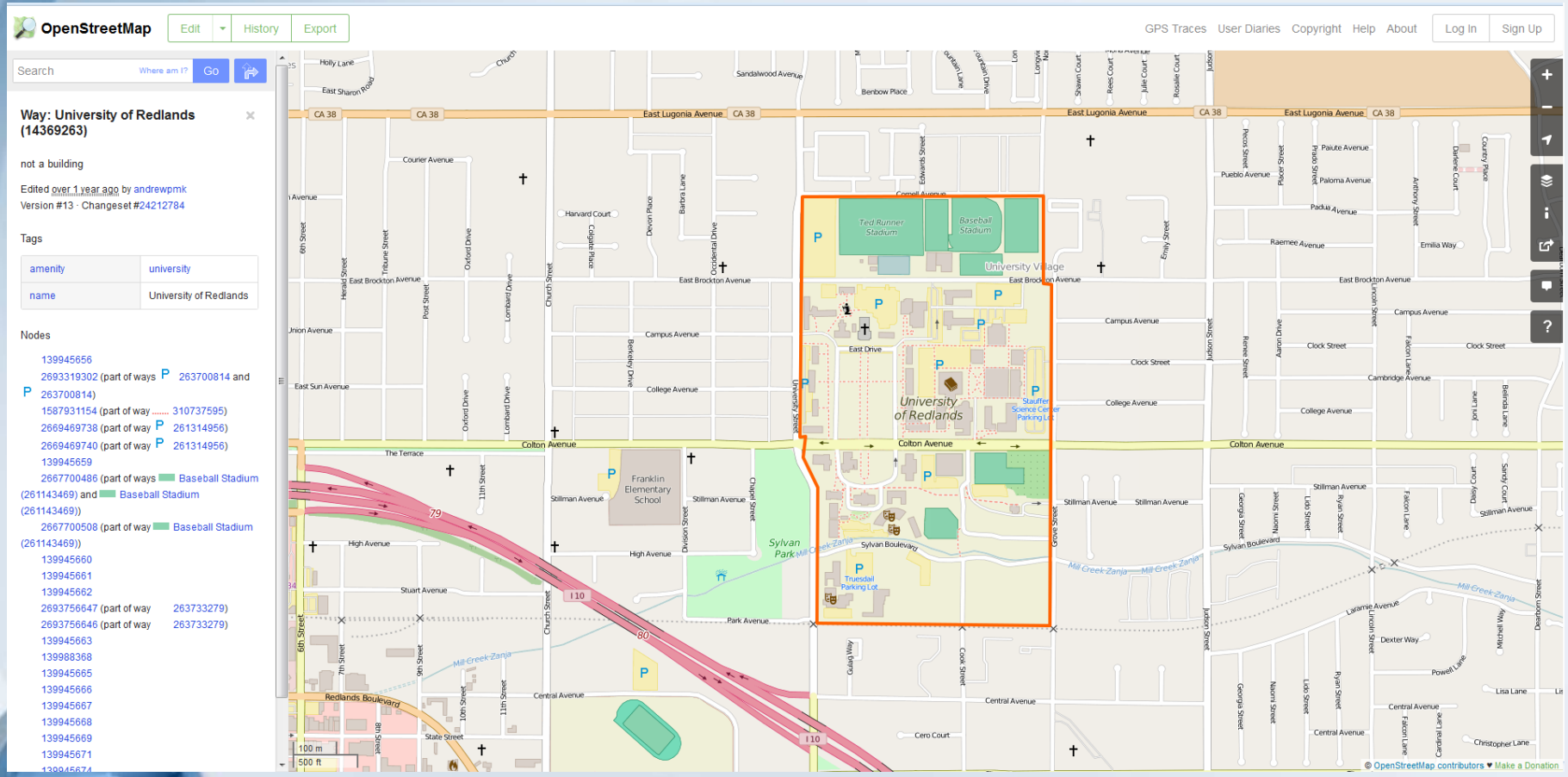
- **Volunteered geographic information (VGI; Goodchild 2007)**
  - **a phenomenon of the 21st Century**
  - **user-generated content**
  - **asserted information**
  - **the empowerment of millions of private citizens**
    - **largely untrained**
    - **no obvious reward**
    - **no guarantee of truth**
    - **no authority**
- **Spatio-temporal footprints generated by social media**
  - **Twitter, Flickr, *etc.***

Big data

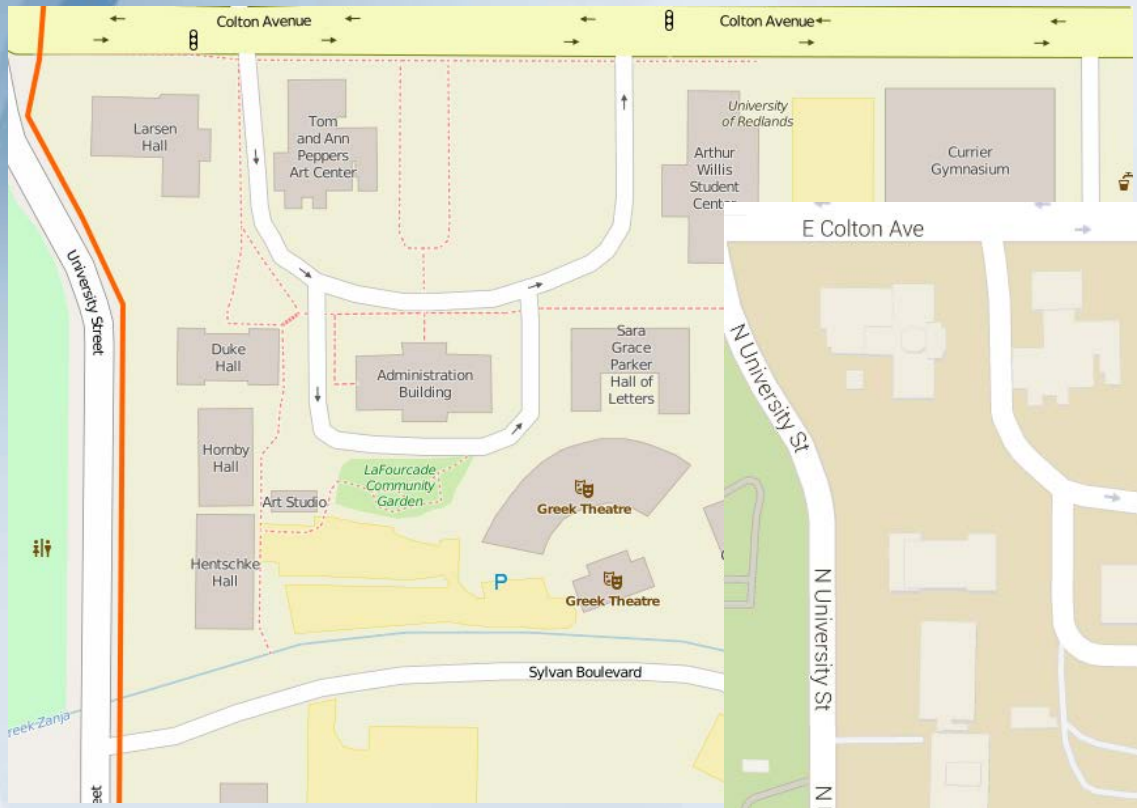
- Big data: emails, blogs, photos, videos, geographic datasets
- Widely disseminated over the Internet: Twitter, Flickr, Youtube, Openstreetmap

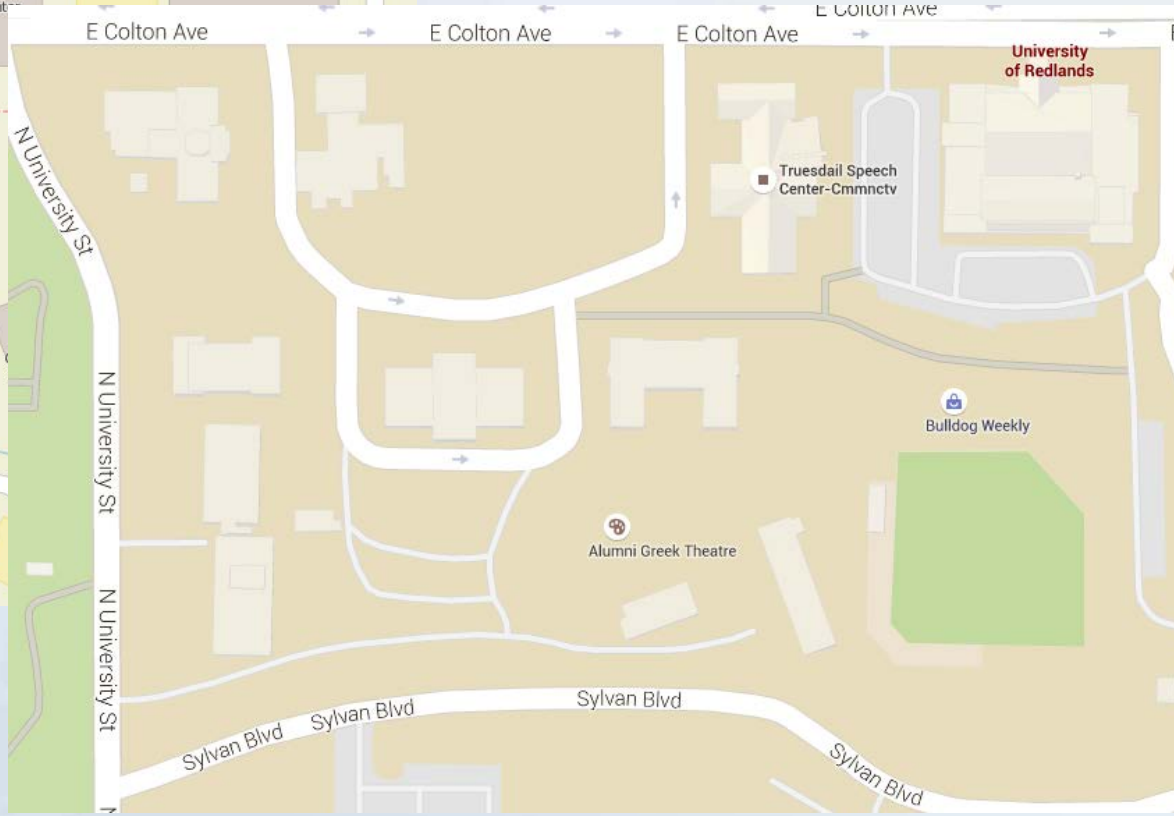
- 92% of babies in the US have online presence before two years old (AVG).

www.openstreetmap.org



OpenStreetMap vs. Google Maps





Geographic data from online communities

- Twitter: a social networking site
- Flickr: online photo management service
- Georeferenced tweet messages in Twitter and georeferenced photos in Flickr

Tweets and photo metadata in the database

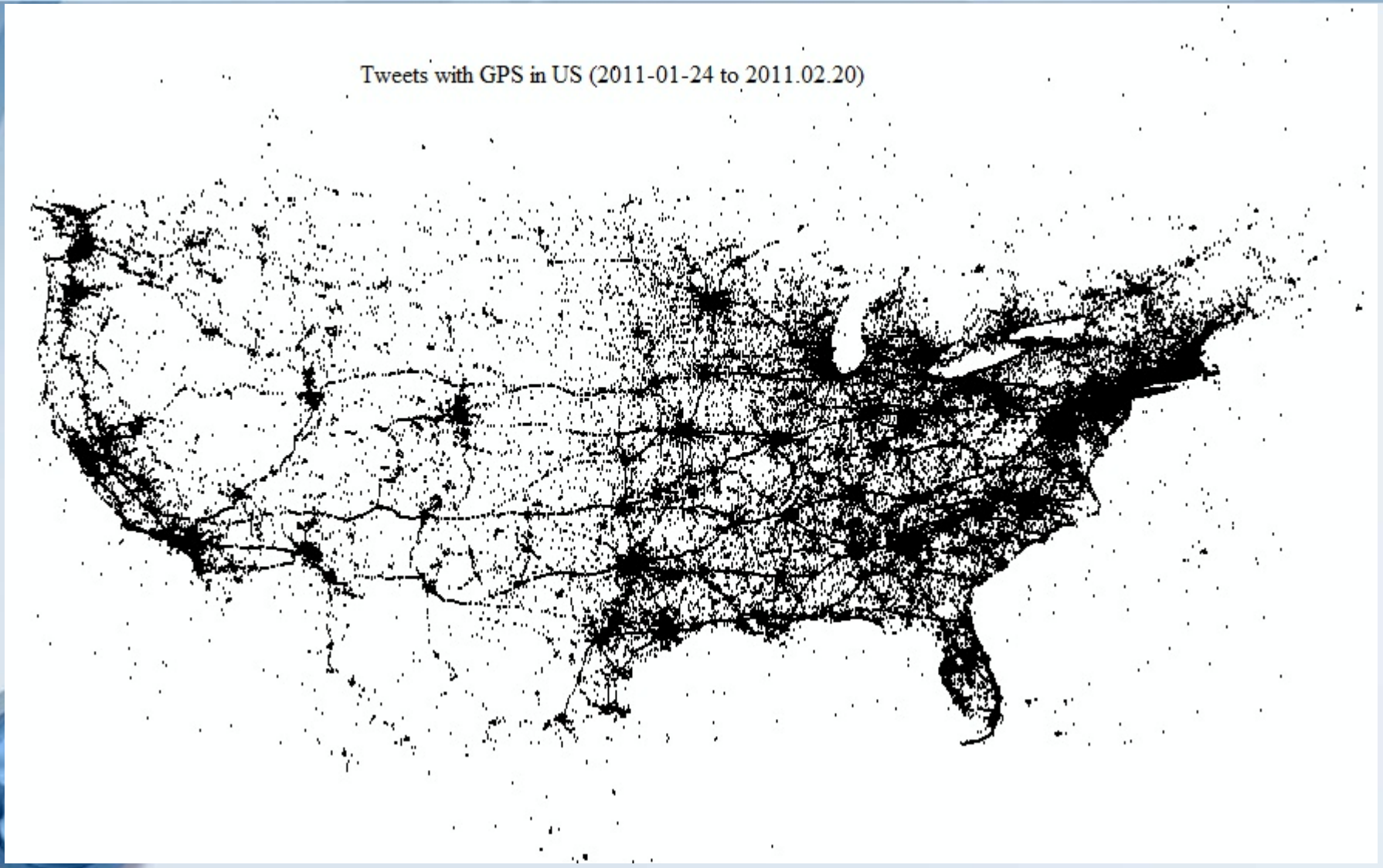
January 23rd to February 12th , 2011, more than 4 million tweets

tweet_text	created_at	geo_lat	geo_long
@Rocknrealty volunteering for Alejandro's high sch...	2011-01-20 17:49:24	30.49307	-97.77580
I'm at El Kartel (1025 Robson St., at Burrard St.,...	2011-01-20 17:49:24	49.28649	-123.12775
I'm at Phillips Seafood (900 Water St SW, Washingt...	2011-01-20 17:49:24	38.88053	-77.02669
@ThenAndreaSaid 'Just keep swimming, just keep swi...	2011-01-20 17:49:24	26.21436	-98.13900
Bear Grylls is the funniest person ever !!!! Haha	2011-01-20 17:49:24	35.78715	-78.59540

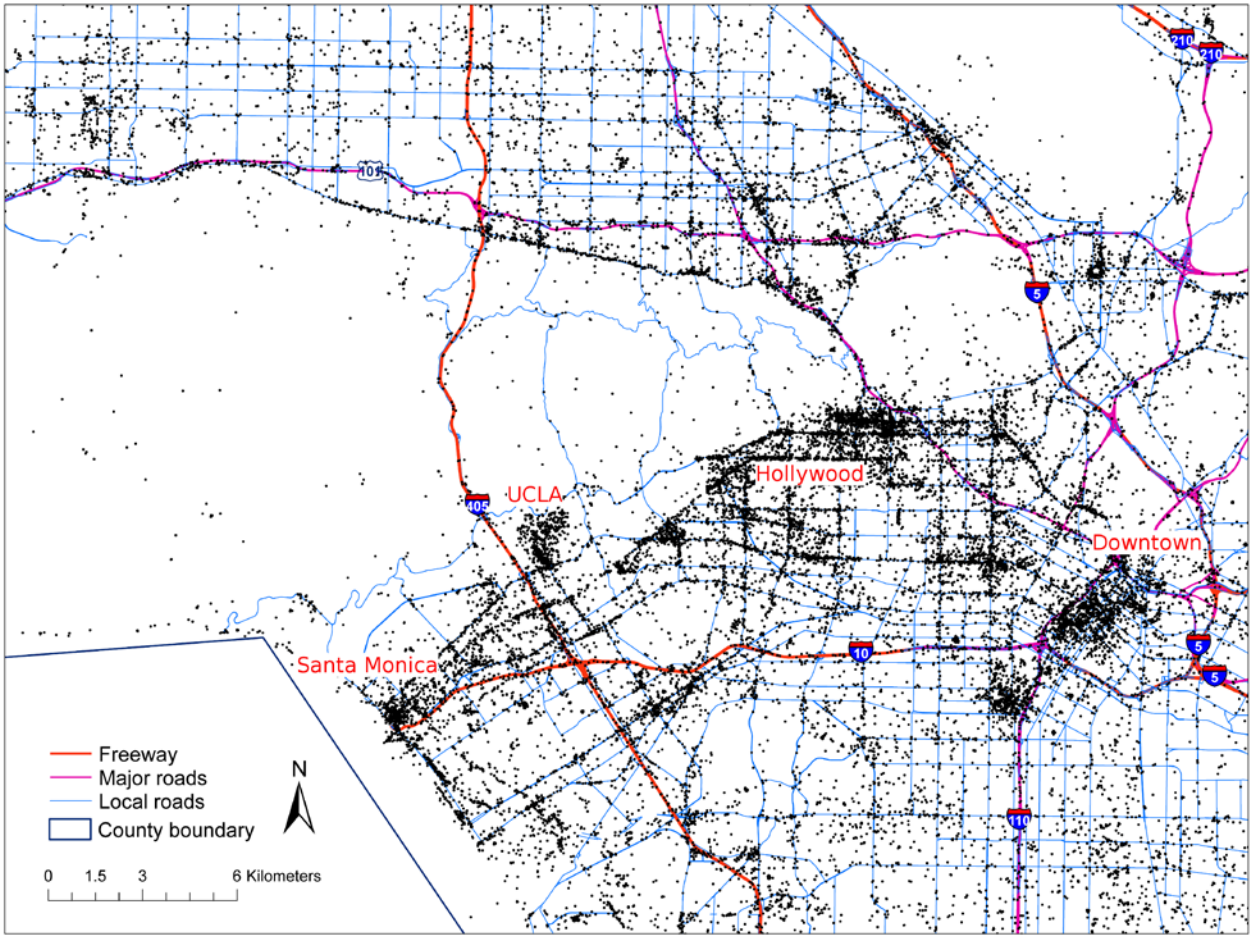
January, 1st to February, 25th, 2011, about 800,000 photos

photo_title	description	photo_tags	date_taken	geo_lat	geo_long
Seattle, WA	Pike Place Corner Market		2008-08-21 12:12:22	47.60892	-122.34058
Happy New Year! 2011 will rock.		square squareformat iphoneography instagramapp upl...	2011-01-01 00:01:10	45.53898	-122.63066
Olivia for President 2012	A paraody sure to please a child in the next elect...	olivia president 2012	2010-12-31 23:56:46	47.62035	-122.34900

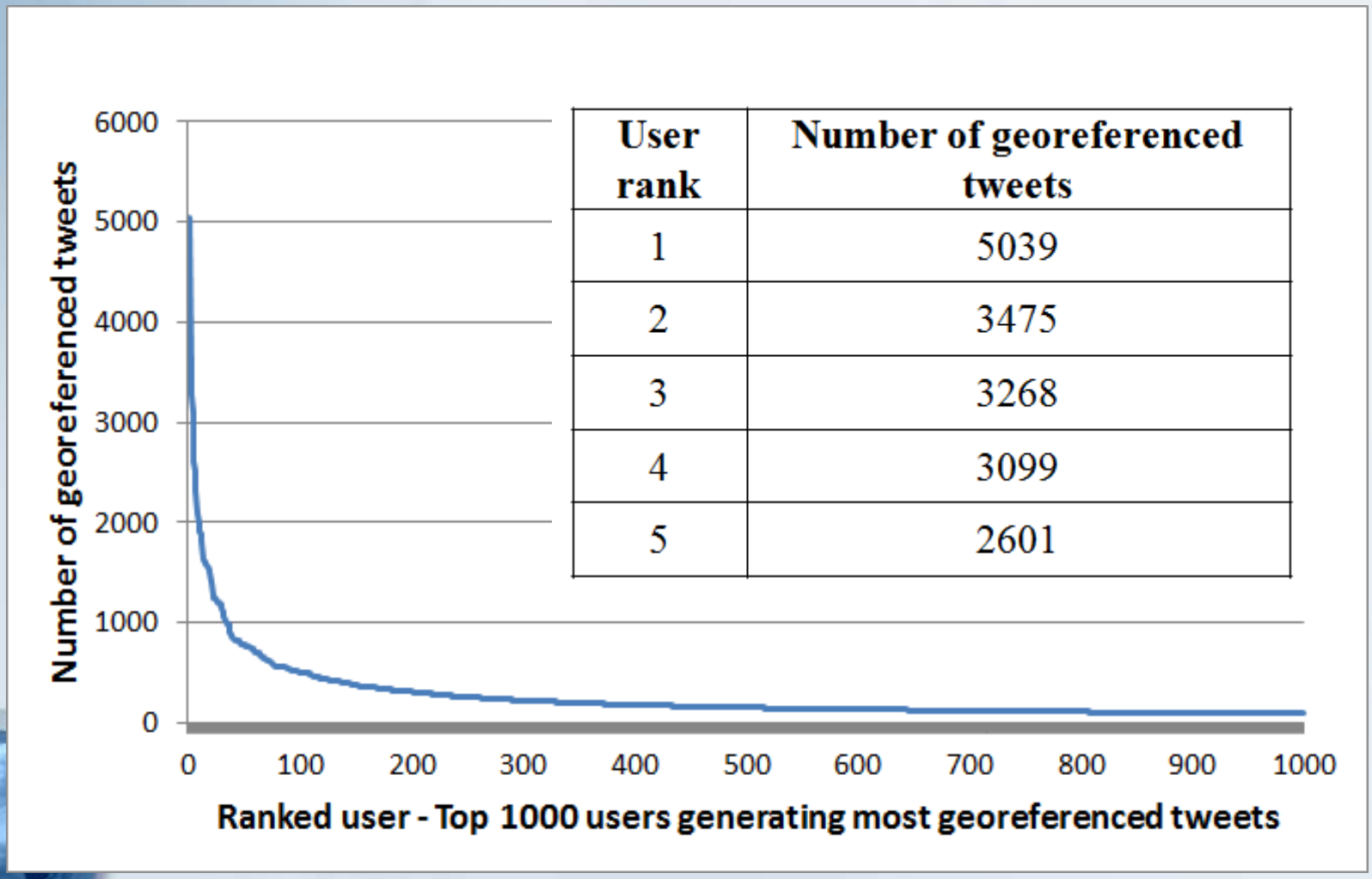
Georeferenced tweets within three weeks



A close-up of the georeferenced tweets in Los Angeles county

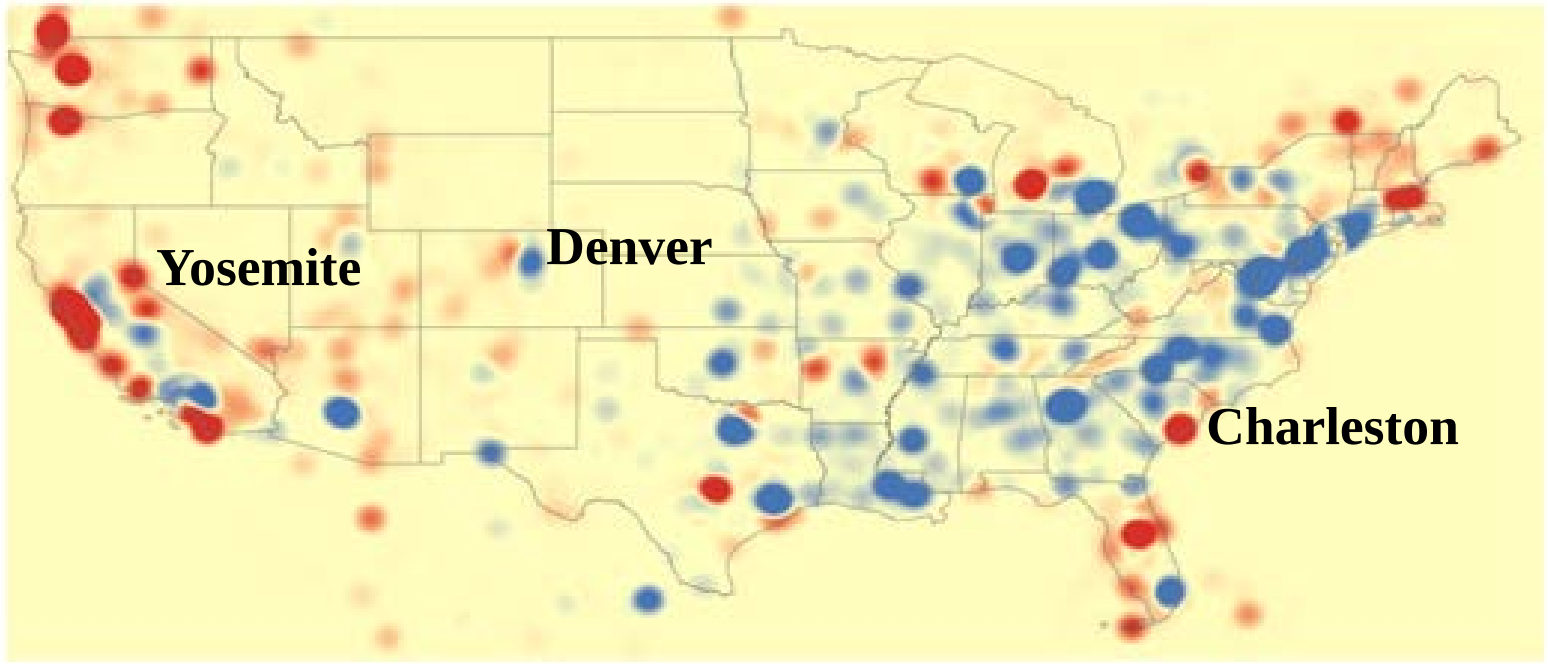


Number of georeferenced tweets generated by the top 1000 users



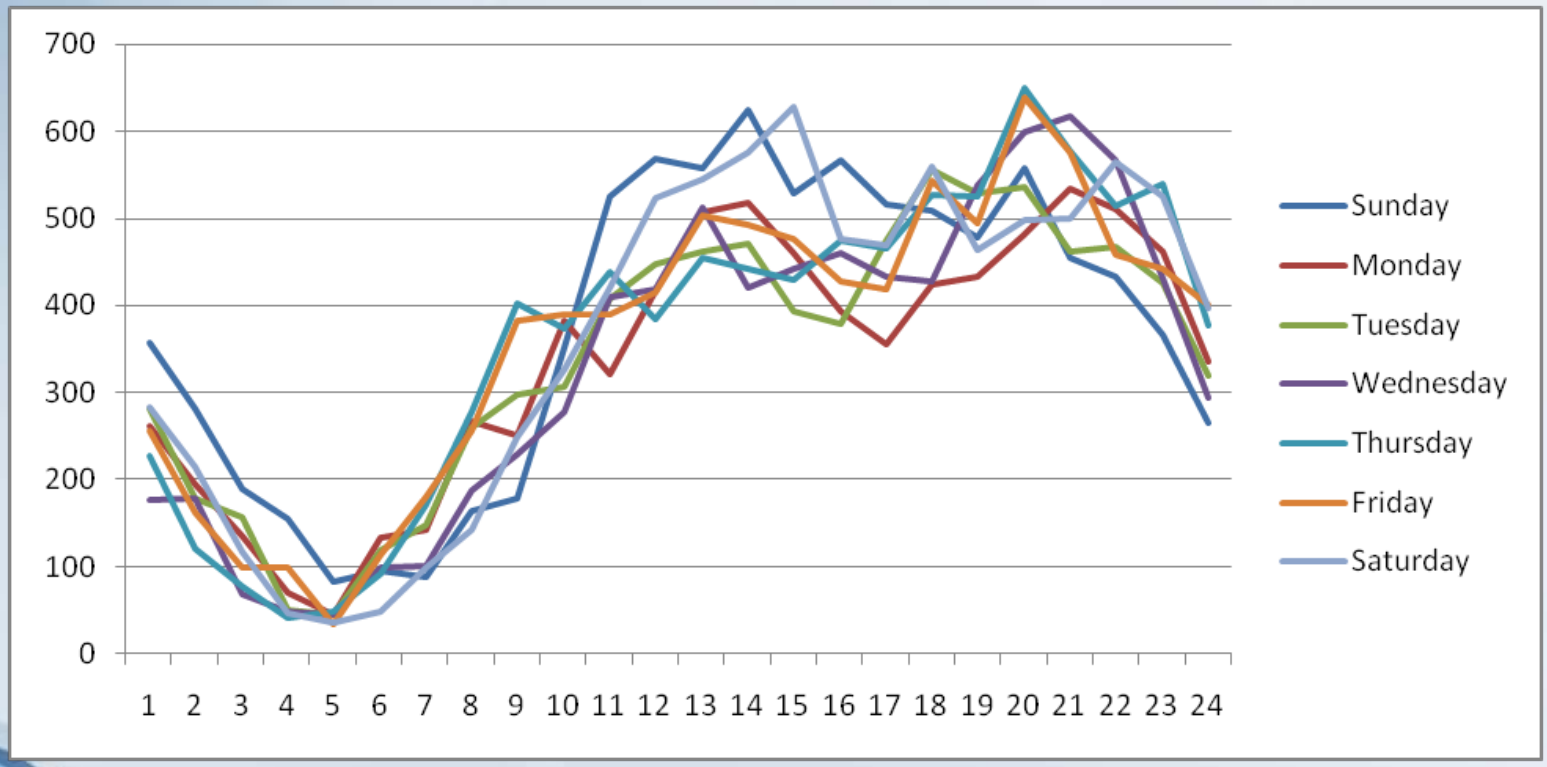
User rank	Number of georeferenced tweets
1	5039
2	3475
3	3268
4	3099
5	2601

Density differences between Flickr photos and tweets



(Li, Goodchild, Xu, 2013)

Temporal variation of tweets



Place and footprints

- **Place**
  - **Fundamental concept in everyday life.**
  - **Place names are omnipresent in conversations, documents, and news stories.**
  - **Place referencing : meeting locations (e.g., University of Redlands), navigation (e.g., from CSULB to University of Redlands), trip planning**
- **Footprints**
  - **Traces of people interacting with their surroundings**
  - **Coordinates: latitude, longitude**

A location on the Earth's surface

- **Two ways:**
  - **Place names: adopted by human beings**
  - **Coordinates: adopted by computers**

- **Defining places from footprints:**
  - **What are the significant places that are identified in space?**
  - **Where are they?**

Location, place, gazetteers

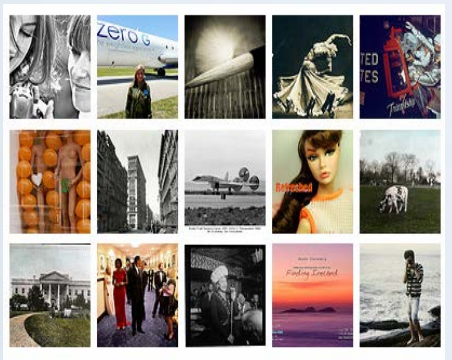
- Location is a key concept in everyday life.
- Coordinates
  - Point, polyline, polygon
  - One universal spatial coordinate system: latitude, longitude
- Place names
  - Easier to comprehend and remember
  - Frequently used in human discourse
  - Place denotes the significance of a location
- Gazetteer: a bridge between place names and spatial coordinates
  - Three core elements in a gazetteer: place name, place type, and spatial location
  - E.g., GeoNames

Representation of vague places

- **Participants identify and draw downtown Santa Barbara (Montello et al., 2003)**
- **User needs for place data based on an empirical study, as well as the implications for modeling place in GIS (Davies et al., 2009)**
- **The use of place names harvested from the Web to find the spatial extents of vague places (Jones et al., 2008)**
- **The use of geotags to build bottom-up gazetteers and compared extracted place location with the location in a gazetteer (Keßler et al., 2009)**

Data

- Footprints associated with photos in Flickr
- Collected photo metadata using the Flickr public API.
- Approximately 150,000 photographs georeferenced within the bounding box of the Île-de-France region.
- Metadata include photo ID, title, description, tags, time when a photo was taken and uploaded, latitude and longitude, as well as information about the user who uploaded the picture.



Preprocessing

- **Flickr data are a type of VGI, so they lack mechanism for quality control.**

- **Variations of place names are present.**
- **Eiffel Tower: eiffeltower, eiffel, theeiffeltower, towereiffel, eifeltower, toureiffel, torreeiffel, latoureiffel, paristorreeiffel, toureiffelparis, eiffeltoren, eiffelturm, eiffeltã, eiffeltour, eifelturm, eifel, toureifel, and torreeifell.**

- **Errors are present in location**
- **Users mass upload to the same point location with the same group of tags a set of pictures that were taken during a trip that involves several places.**
- **In such as case, only one photograph is retained.**

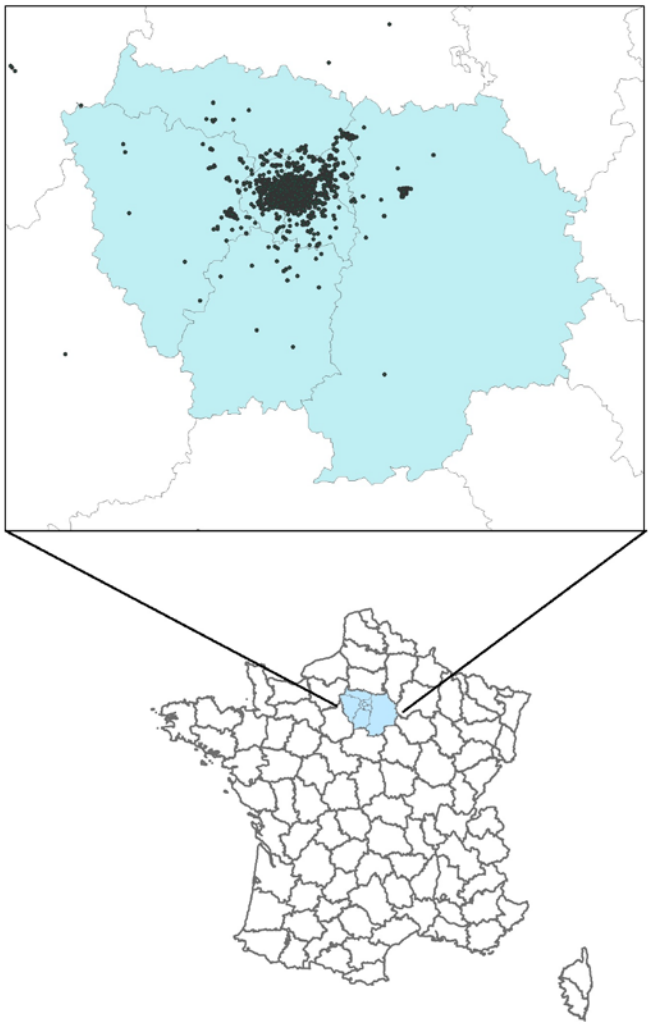
Spatial extent of place

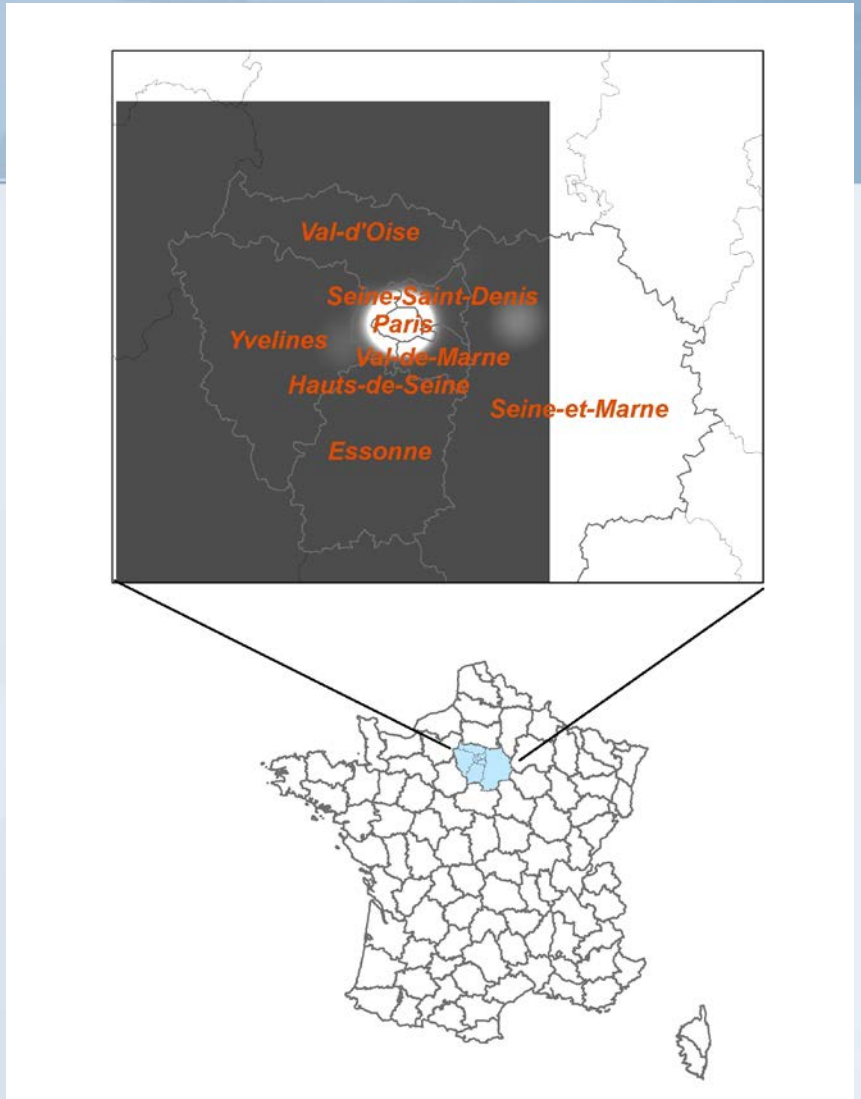
- **Places under study: Paris, Eiffel Tower, Louvre, Montmartre, Notre Dame, Palace of Versailles, and Disneyland.**
- **Method: kernel density estimation**
- **The kernel estimator**

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (1)$$

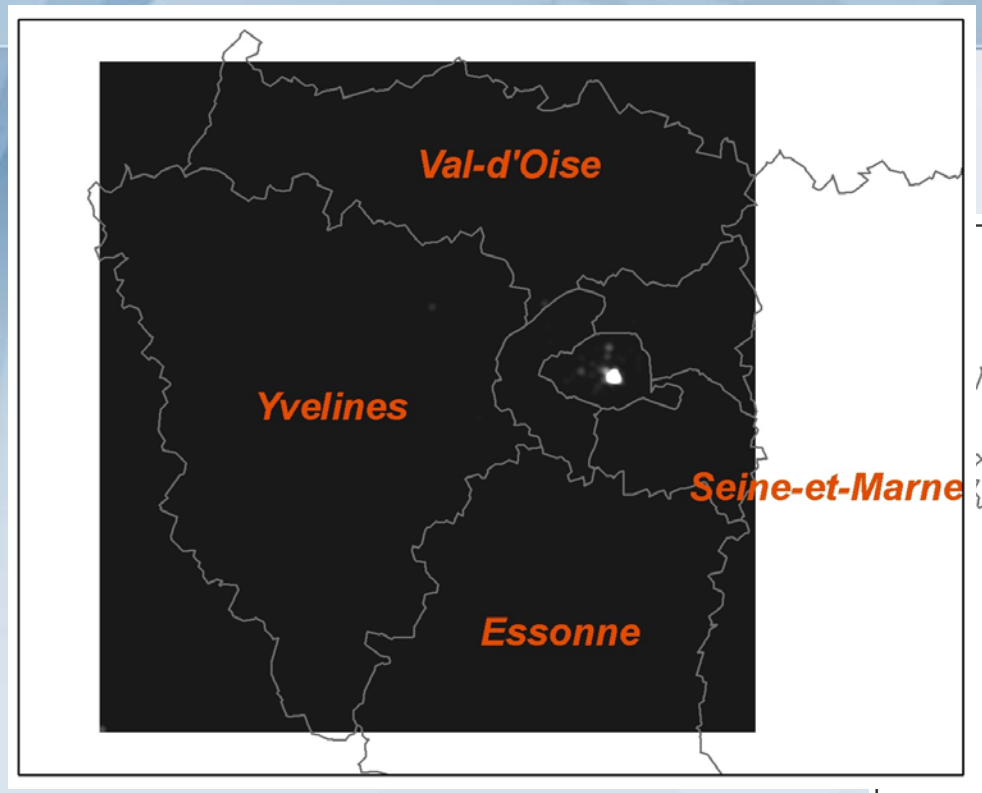
- **where *n* is the number of points, *h* is the bandwidth that determines the amount of smoothing, *K* is the kernel function, *x* is the location to be estimated, and x_i is a point location.**

Footprints of photographs with the geotag "Paris" over the administrative boundaries of France



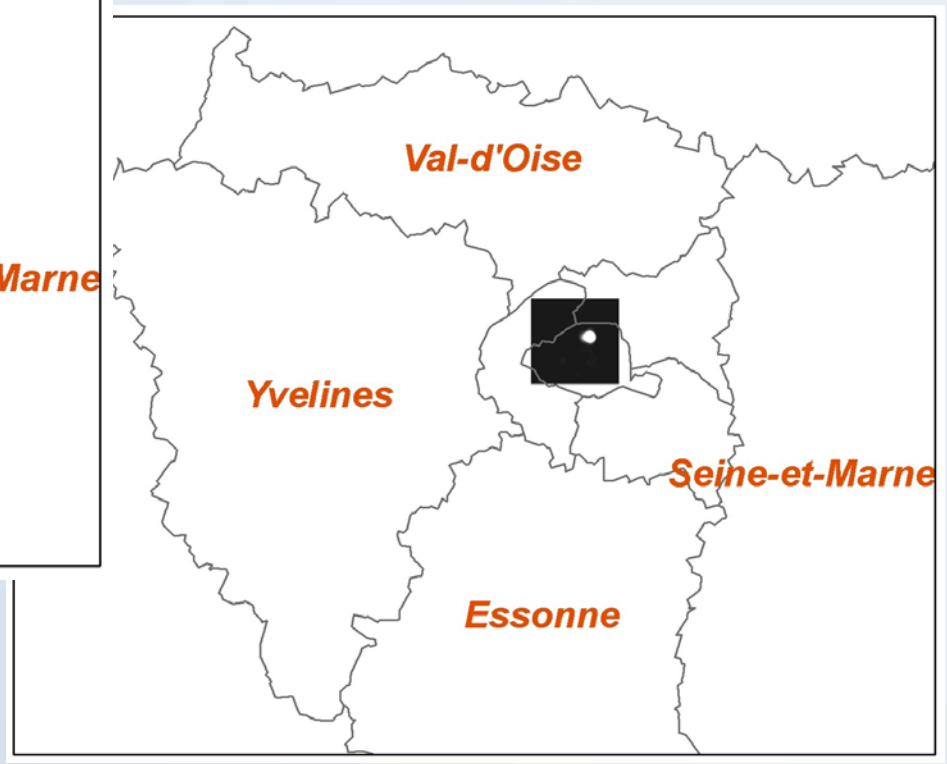


The spatial extent of Paris generated from geotagged photos



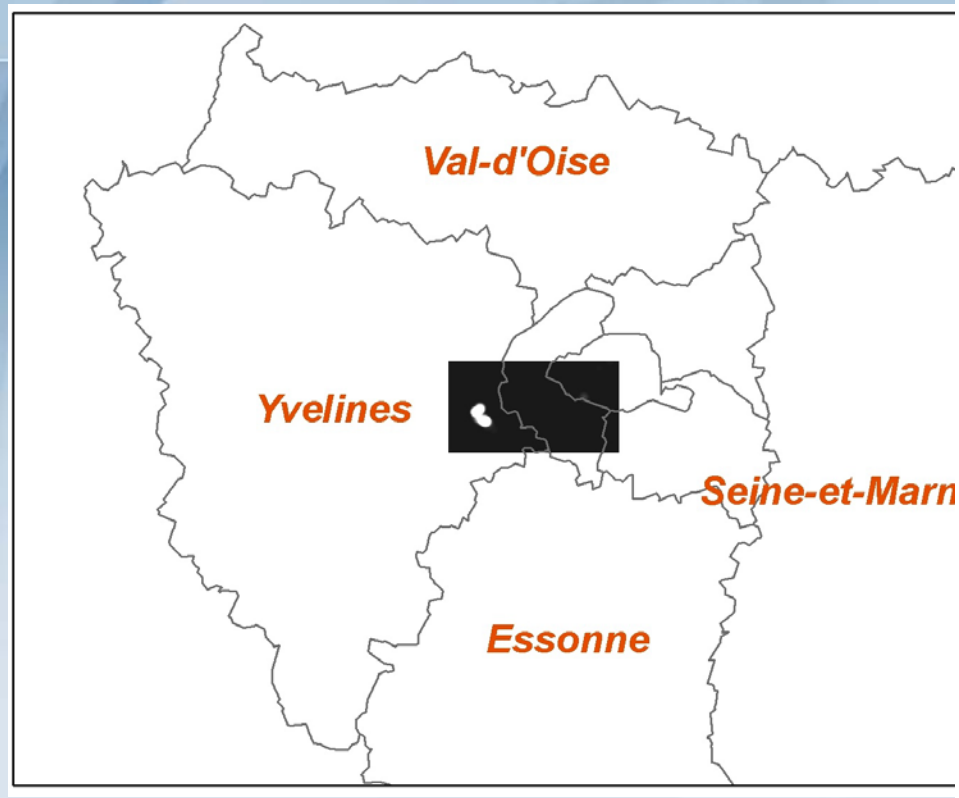
Notre Dame

Uncertainty: 78,200m

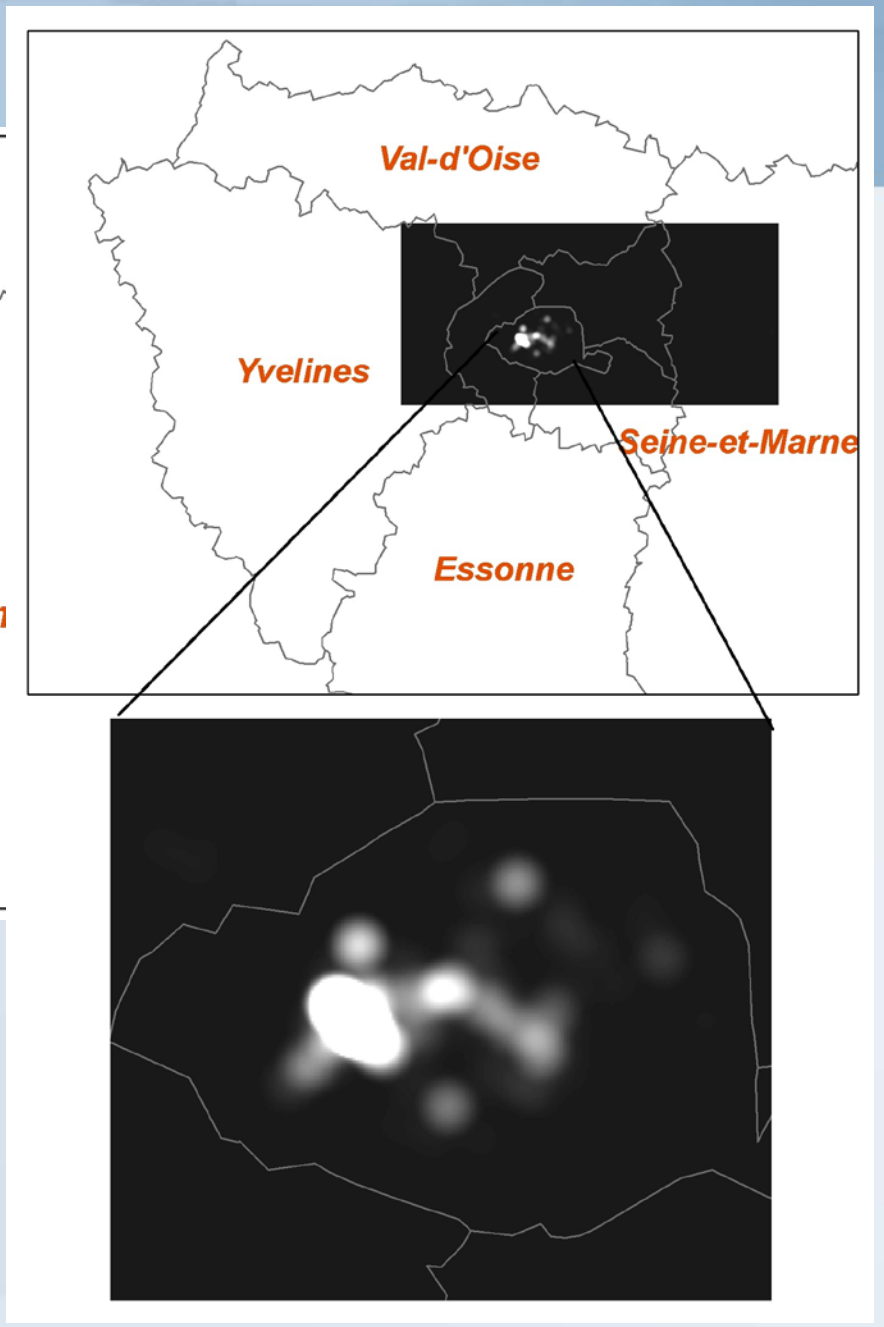


Montmartre

Uncertainty: 9,600m



Palace of Versailles

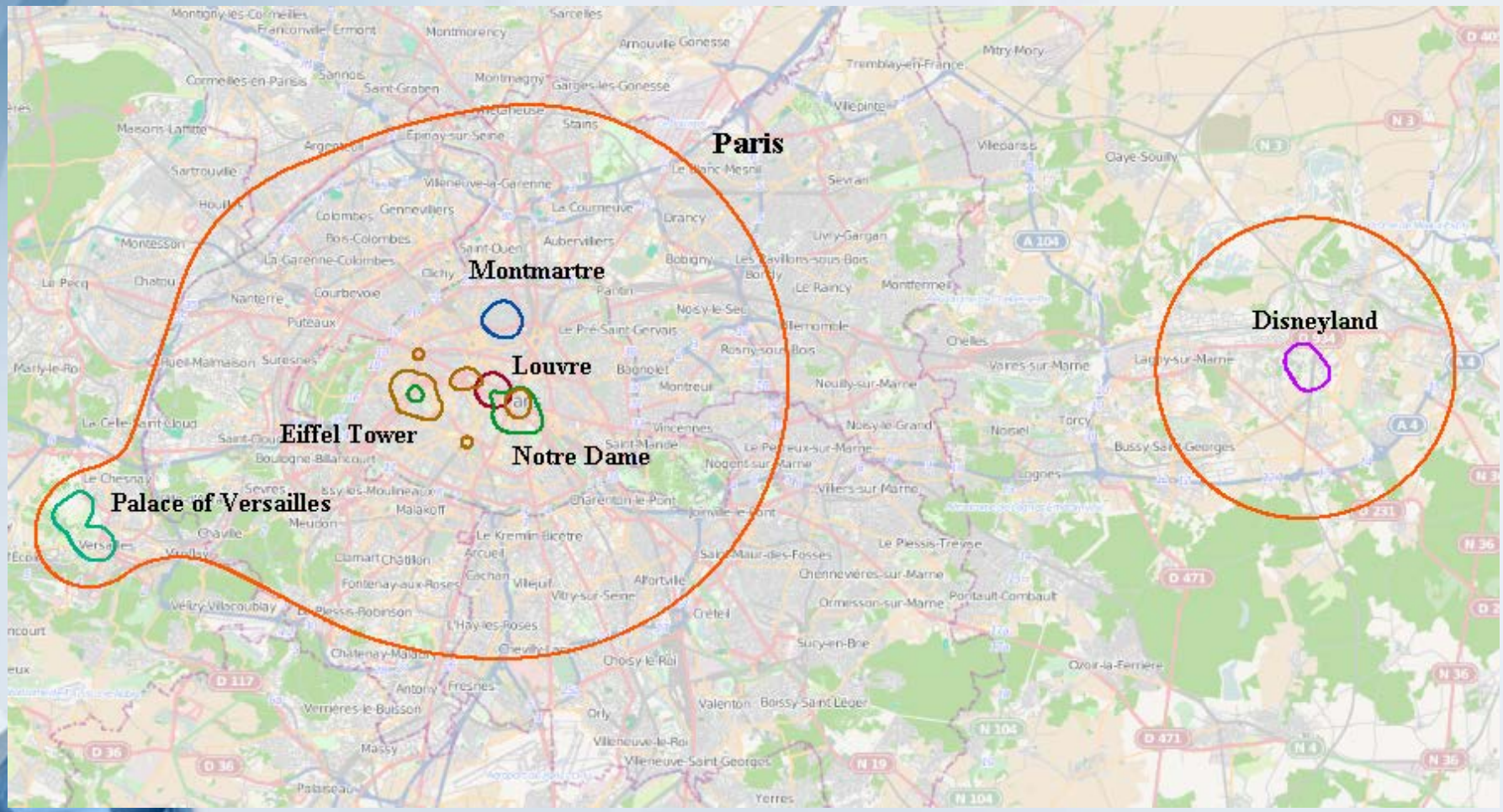


Eiffel Tower

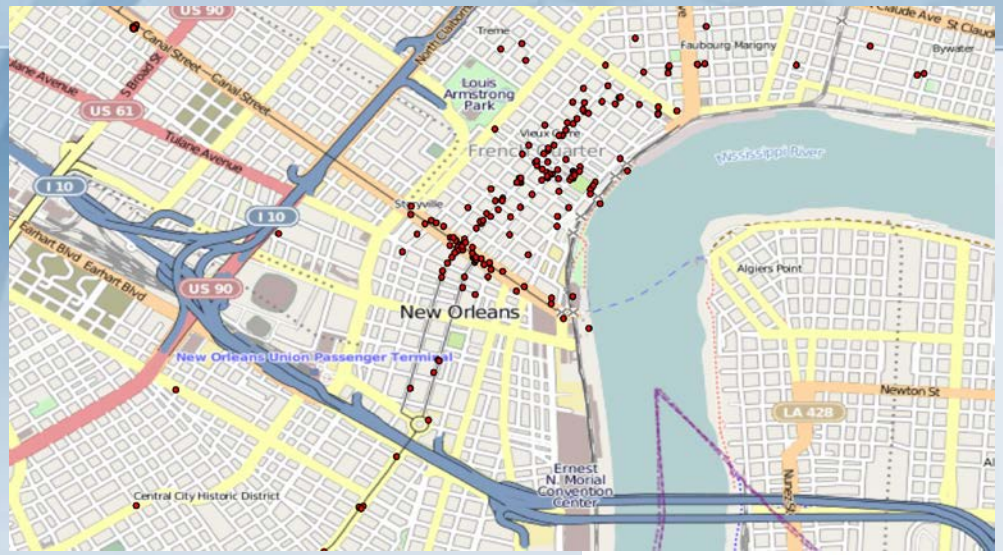
Relation between places

- **Relationships between a larger place and smaller places that compose the larger place can also be inferred from spatial footprints.**
- **Constituent places are not exclusive of each other.**
- **They also do not cover all locations in a larger place.**
- **Smaller places are distributed over different parts of a larger place.**

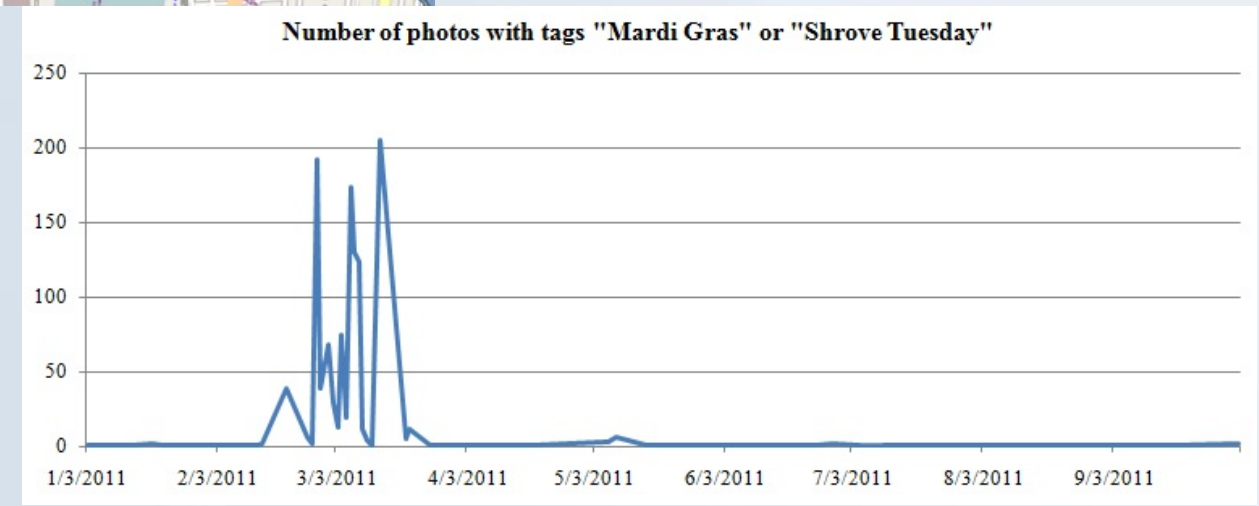
Hierarchical relationship between Paris and its constituent places



Events



Spatial footprints of Mardi Gras



Temporal footprints of photos with tags "Mardi Gras" or "Shrove Tuesday"

Place and footprints

- **Spatial footprints that are harvested from geotagged Flickr photos can be used to investigate the spatial extent of individual places and associated spatial uncertainty in footprints.**
- **Case study: Paris and its six constituent places.**
- **In the location density surfaces, peaks on footprint density surfaces suggest a high degree of agreement on that location.**
- **Hierarchical relationships and events can also be inferred from spatial footprints.**

Place and footprints

Implications:

- **It is a non-intrusive method of data collection**
- **Spatial knowledge of places may complement and improve existing gazetteer**
- **It can not only answer questions regarding location of a single place, but also questions about relations between places.**

Limitations:

- **User-generated spatial data**
- **No sampling strategy is adopted**
- **Only places with a sufficient amount of data can be studied**

Emergency response

- **Disasters:**
  - **Hurricane Katrina, 2005**
  - **The Wenchuan Earthquake, China, 2008**
  - **The Haiti and Chile Earthquakes, 2010**

- **Time is critical**

- **What information would you want to know**
  - **if you are a victim who needs help?**
  - **if you are an emergency manager?**

- **Location!!**
  of resources, of victims, etc.

Geographic information for emergency response

- **Data needs: reliable, accurate, timely geographic data**
- **Data problems:**
  - **Data provided by government: inadequate**
  - **Data from remote sensing: only phenomena visible from above**
  - **Not enough GIS data experts**
  - **Not enough professional emergency workers**
- **The response efforts during Katrina "were significantly hampered by a lack of information from the ground" (a report to the US Congress on Hurricane Katrina, by Secretary of Homeland Security Michael Chertoff, 2005)**

Information sharing in emergencies

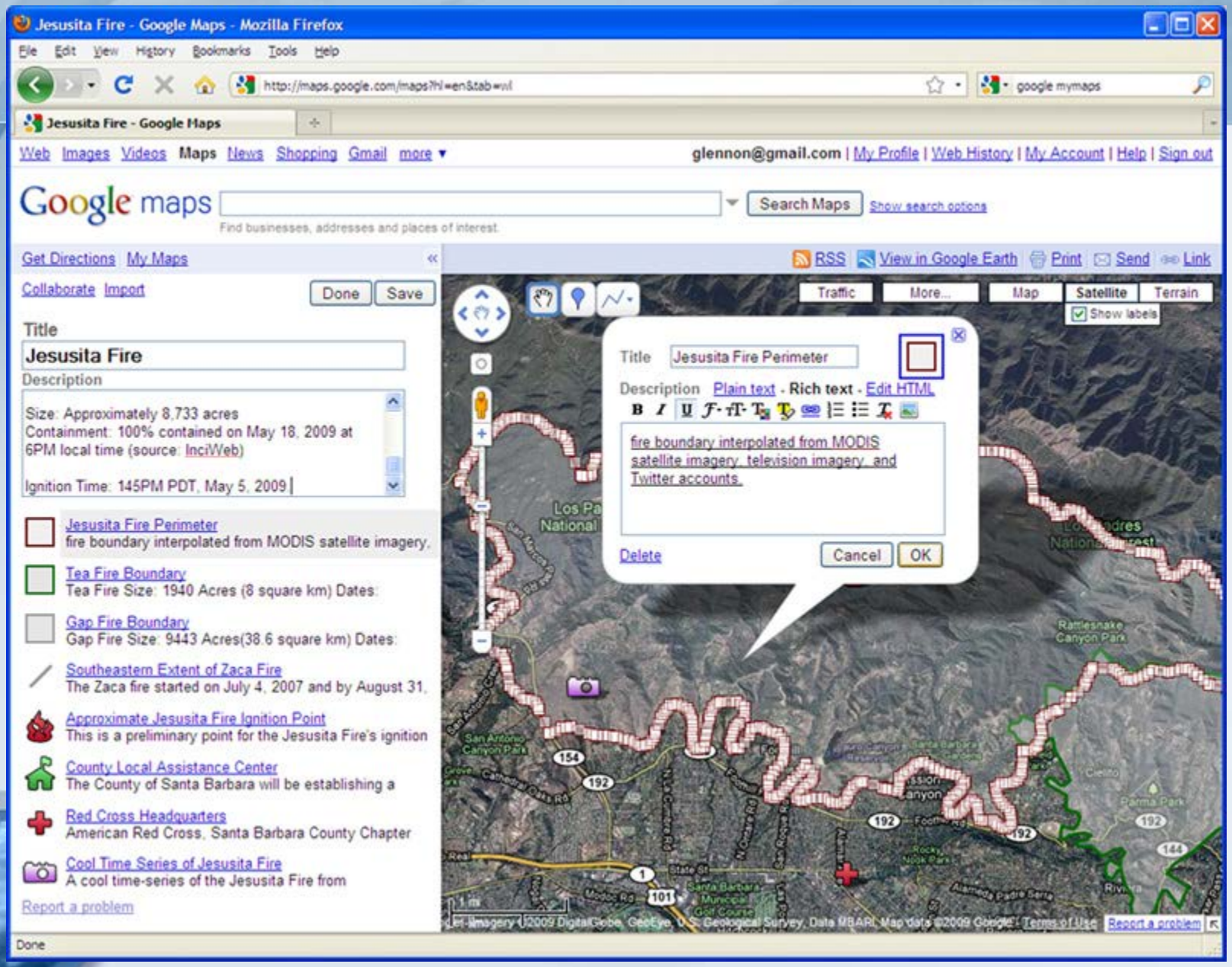
- **Important to facilitate coordination**
- **Important between emergency managers and local residents**
- **Important for local residents for emotional reasons**
- **Existing communication channels: neither sufficient nor adequately used**
- **Alternative: online and mobile social networking services**

Jesusita Fire

- May 5-18, 2009
- 8,733 acres burned
- 80 homes destroyed; 15 homes damaged
- Estimated costs $17 million

Data about Jesusita Fire

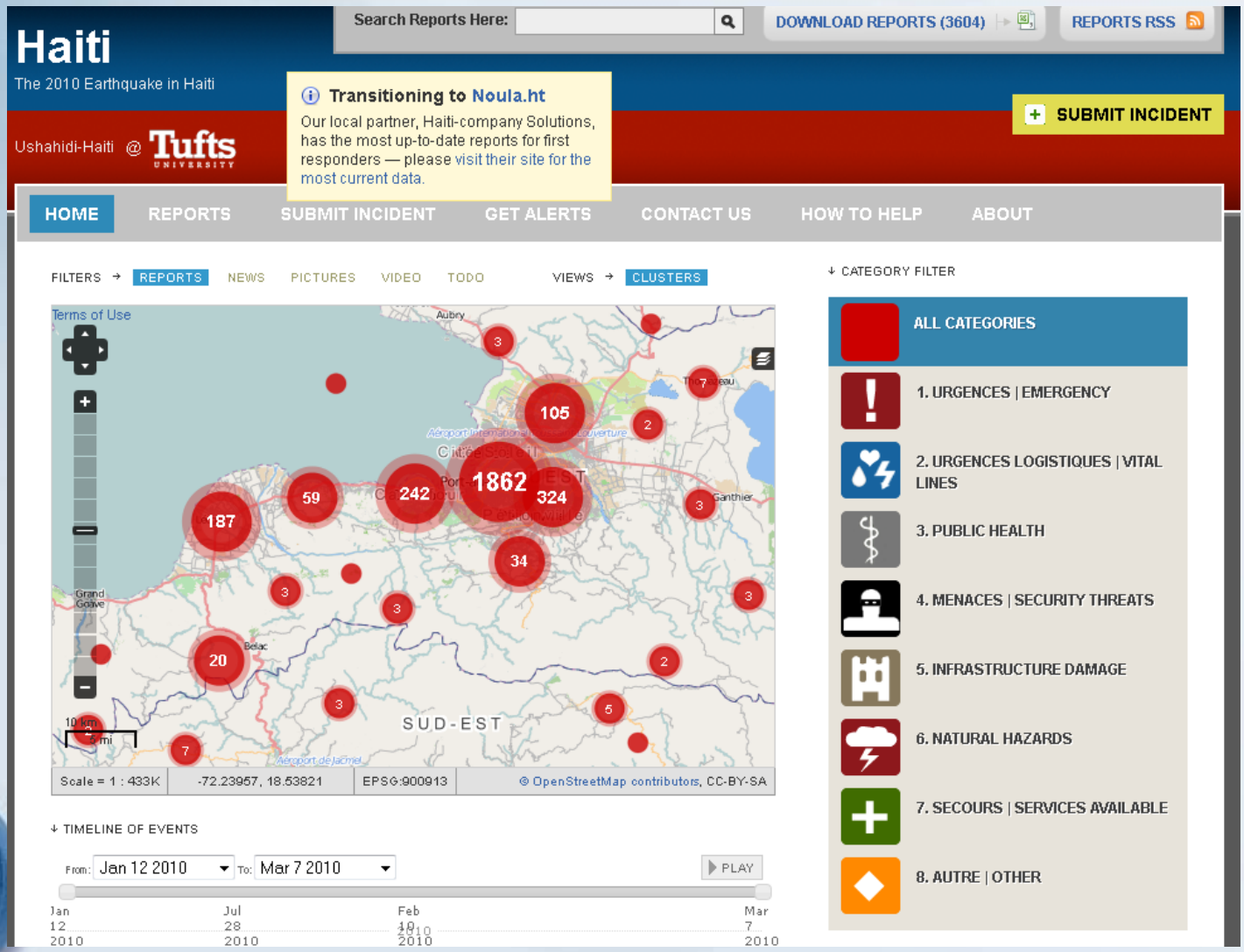
- **Community web: Independent**
- **Social networks: Twitter, Flickr, Facebook**
- **Television and radio: both traditional and online**
- **Government and official outlets**
- **Satellite imagery**
- **Newspaper stories**
- **Conversations with first and secondhand eyewitnesses**


Jesusita Fire - Google Maps - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://maps.google.com/maps?hl=en&tab=wl
google mymaps
Jesusita Fire - Google Maps
Web Images Videos Maps News Shopping Gmail more ▾
glennon@gmail.com | My Profile | Web History | My Account | Help | Sign out
Google maps
Search Maps
Show search options
Find businesses, addresses and places of interest.
Get Directions | My Maps
Collaborate Import
Done Save
Title
Jesusita Fire
Description
Size: Approximately 8,733 acres
Containment: 100% contained on May 18, 2009 at 6PM local time (source: InciWeb)
Ignition Time: 145PM PDT, May 5, 2009
Jesusita Fire Perimeter
fire boundary interpolated from MODIS satellite imagery,
Tea Fire Boundary
Tea Fire Size: 1940 Acres (8 square km) Dates:
Gap Fire Boundary
Gap Fire Size: 9443 Acres(38.6 square km) Dates:
Southeastern Extent of Zaca Fire
The Zaca fire started on July 4, 2007 and by August 31,
Approximate Jesusita Fire Ignition Point
This is a preliminary point for the Jesusita Fire's ignition
County Local Assistance Center
The County of Santa Barbara will be establishing a
Red Cross Headquarters
American Red Cross, Santa Barbara County Chapter
Cool Time Series of Jesusita Fire
A cool time-series of the Jesusita Fire from
Report a problem
RSS View in Google Earth Print Send Link
Traffic More... Map Satellite Terrain
Show labels
Title Jesusita Fire Perimeter
Description Plain text - Rich text - Edit HTML
fire boundary interpolated from MODIS satellite imagery, television imagery, and Twitter accounts.
Delete
Cancel OK
Rattlesnake Canyon Park
San Antonio Canyon Park
Parma Park
Santa Barbara Municipal Golf Course
Terms of Use
Report a problem
Done

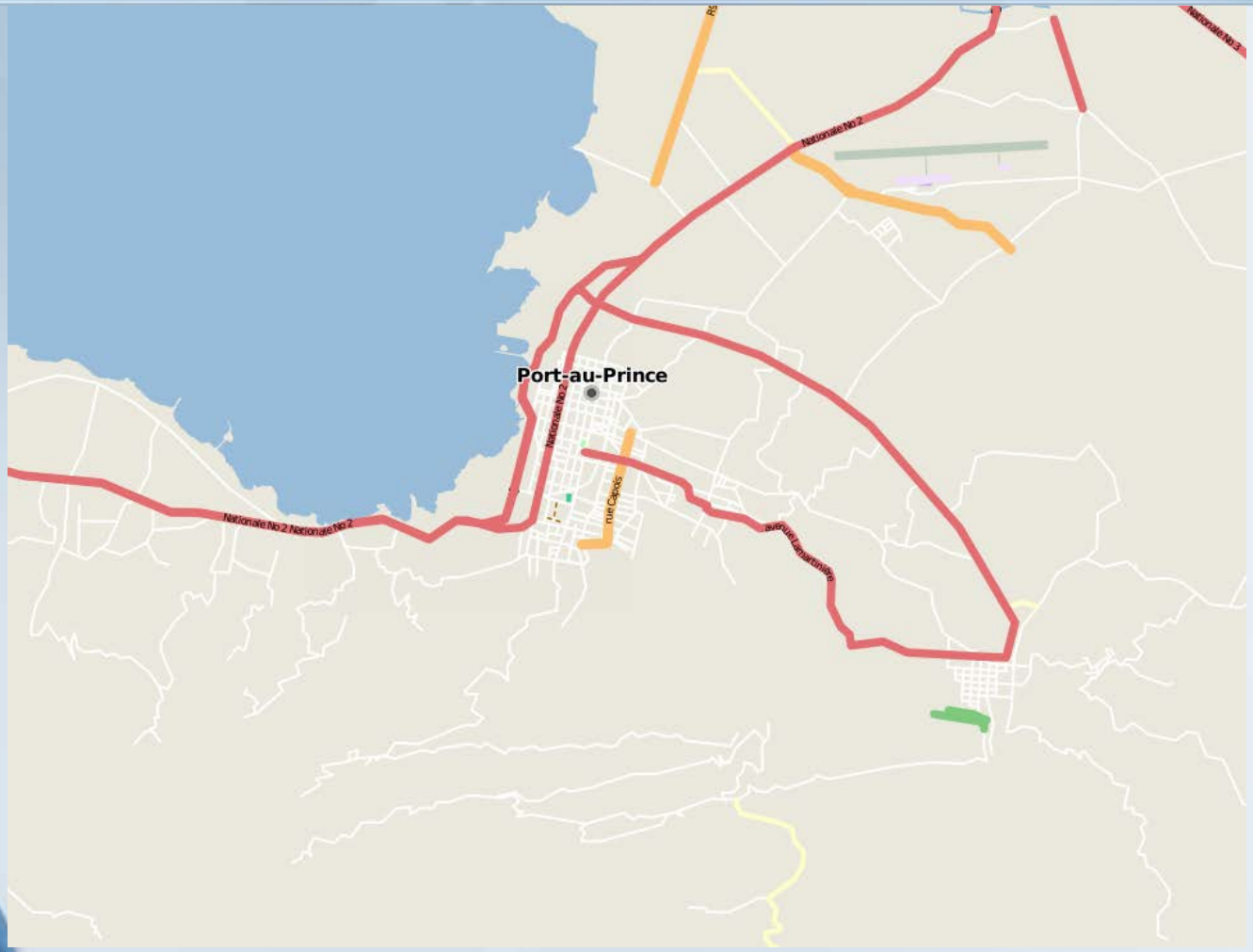
2010 Haiti 7.0 M_W earthquake

- 16:53 local time on Tuesday, 12 January 2010
- three million people were affected
- an estimated 316,000 people died
- 300,000 were injured and 1,000,000 made homeless
- 250,000 residences and 30,000 commercial buildings collapsed or severely damaged.

Ushahidi


Haiti
The 2010 Earthquake in Haiti
Search Reports Here:
DOWNLOAD REPORTS (3604)
REPORTS RSS
Transitioning to Noula.ht
Our local partner, Haiti-company Solutions, has the most up-to-date reports for first responders — please visit their site for the most current data.
SUBMIT INCIDENT
Ushahidi-Haiti @ Tufts University
HOME
REPORTS
SUBMIT INCIDENT
GET ALERTS
CONTACT US
HOW TO HELP
ABOUT
FILTERS → REPORTS NEWS PICTURES VIDEO TODO
VIEWS → CLUSTERS
Terms of Use
105
1862
324
242
59
187
34
20
SUD-EST
Scale = 1 : 433K
-72.23957, 18.53821
EPSG:900913
© OpenStreetMap contributors, CC-BY-SA
CATEGORY FILTER
ALL CATEGORIES
1. URGENCES | EMERGENCY
2. URGENCES LOGISTIQUES | VITAL LINES
3. PUBLIC HEALTH
4. MENACES | SECURITY THREATS
5. INFRASTRUCTURE DAMAGE
6. NATURAL HAZARDS
7. SECOURS | SERVICES AVAILABLE
8. AUTRE | OTHER
TIMELINE OF EVENTS
From: Jan 12 2010
To: Mar 7 2010
PLAY
Jan 12 2010
Jul 28 2010
Feb 10 2010
Mar 7 2010

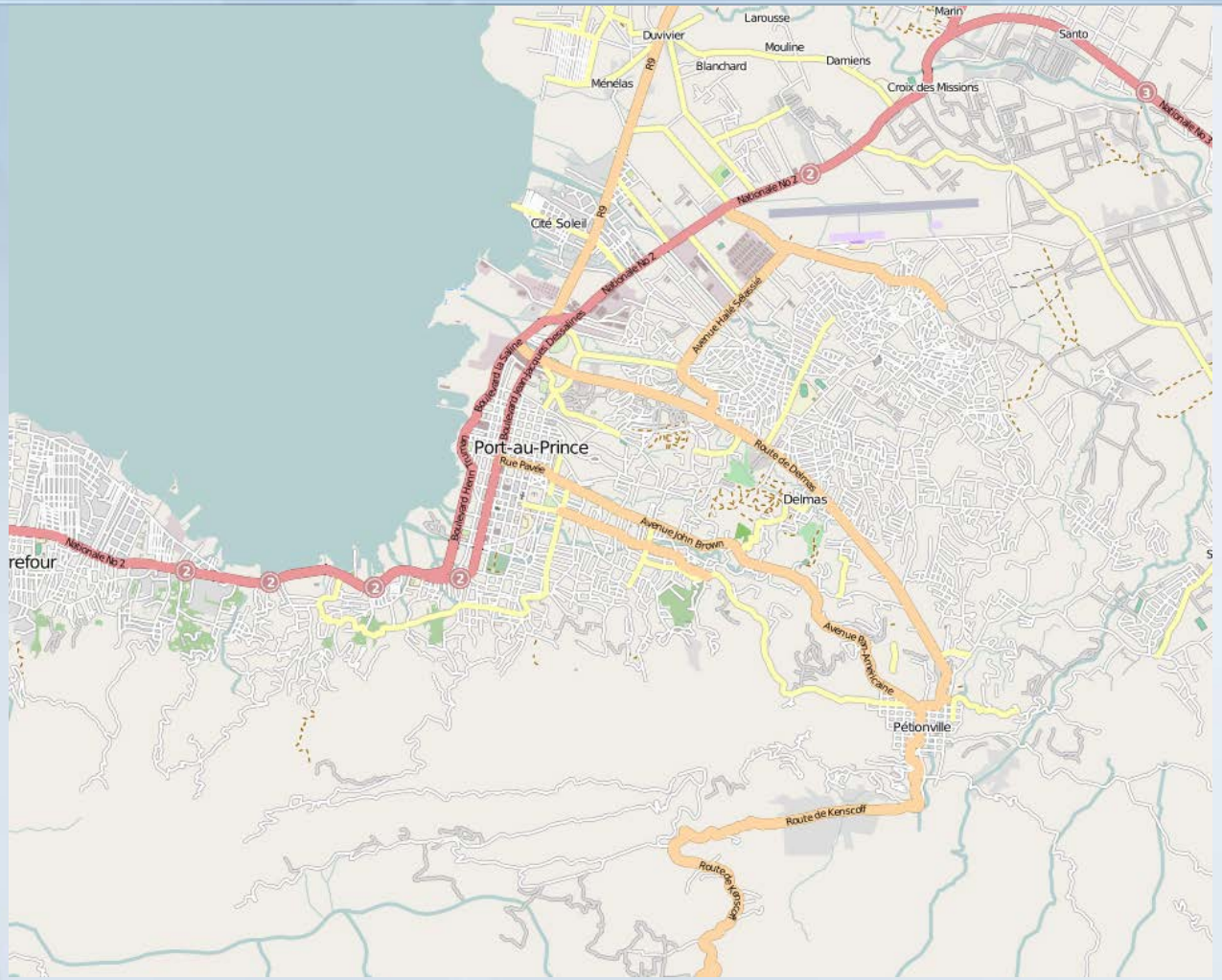
OpenStreetMap, Haiti



Before the 2010 Earthquake

http://www.openstreetmap.org/

OpenStreetMap, Haiti



January 14, 2010, within only 48 hours

http://www.openstreetmap.org/

Outcomes

- http://vimeo.com/9182869
- **Complementary technologies (open geographic data and reports from victims) integrated**
- **Fast cycle, open maps and data saved lives**
- **Partnerships and community formed**

Can social media rescue street children?

- **Street Children are often excluded from the mainstream society.**
- **Child beggars are prevailing in China. (1 to 1.5 million homeless with 20,000 children abducted every year)**

By Xining Yang, California State University, Chico

Geography
Ohio State University

THE OHIO STATE UNIVERSITY

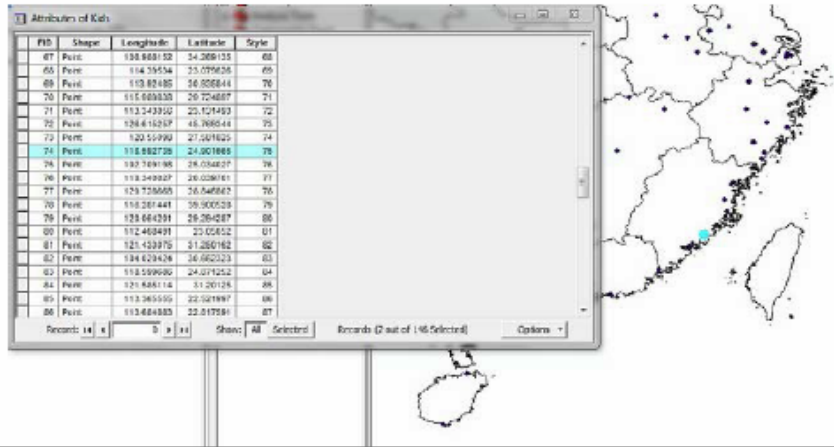
Data

- Dataset 1: “Street Photos to Rescue Child Beggars”, Sina Weibo **Twitter** (http://www.weibo.com/jiejiuqier)
- Dataset 2: “BaoBeiHuiJia”, web community ran by a NGO. (http://www.baobeihuijia.com)

Baby come home: missing children database

VGI Data

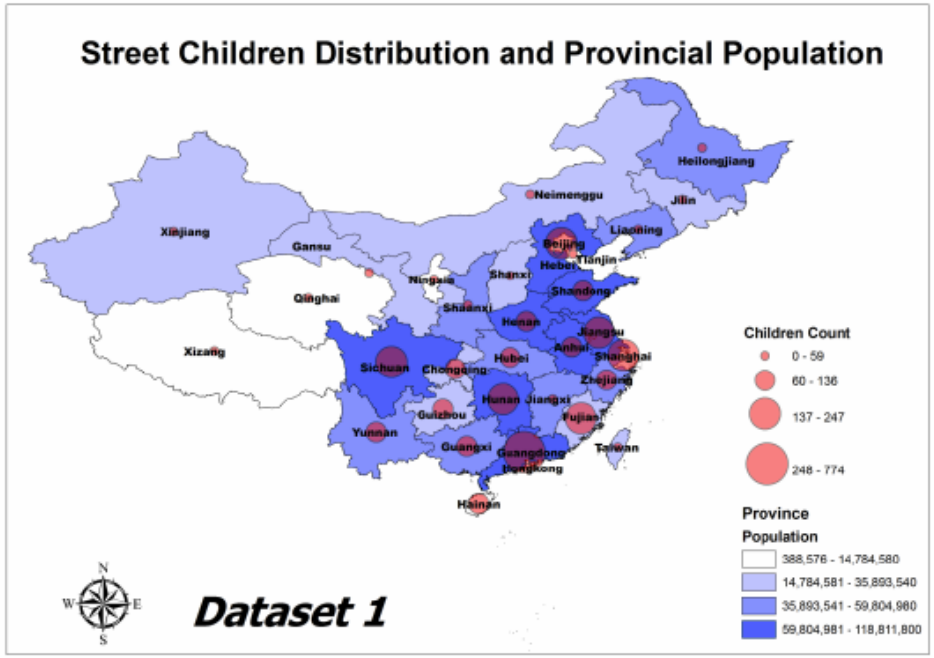
- The raw data is collected and interpreted via a java application.
- The data is then geocoded into shapefile.

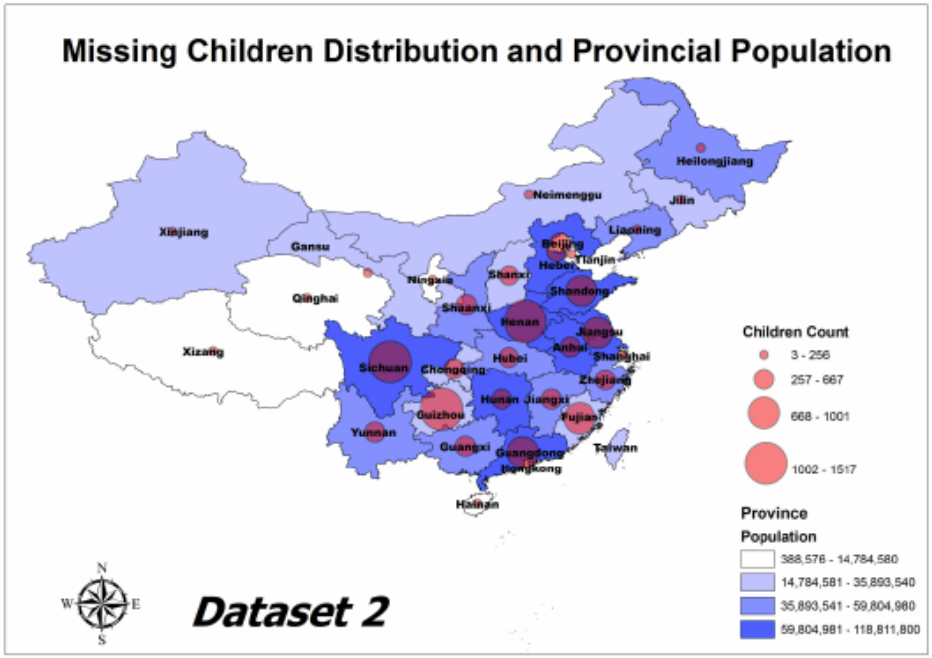


Geography
Ohio State University

THE OHIO STATE UNIVERSITY

Spatial Distribution





By Xining Yang

Challenges (Li and Goodchild, 2010, Goodchild and Li, 2012)

- data access and data quality
  - How to discover relevant information?
  - How to validate VGI?
- How to synthesize geographic data
  - Uncertainty
  - Data conflation
- How to use social networks to support crowdsourcing mapping?
- Privacy

Inferred home and work locations from tweets (Li and Goodchild, 2013)

- **Linked the georeferenced tweets with land use data provided by Southern California Association of Governments (SCAG).**
- **Land use data in LA contain use code at the parcel level.**
- **Major categories of land use: vacant land, single family residence, multiple dwelling, commercial, industrial, agricultural, exempt property, natural resources, and others.**
- **The parcel with the largest number of tweets was regarded as a main location of user activities.**

Percentage of georeferenced tweets and land use in the parcel with most tweets for the top 5 users

User rank	Percentage of georeferenced tweets in the parcel	Land use of the parcel
1	73.07%	Single family residence
2	35.39%	Apartment complex
3	50.70%	Combination Stores and Offices
4	32.83%	Apartment complex
5	54.52%	Apartment complex

Georeferenced tweets for the first user

Changed perception of privacy?

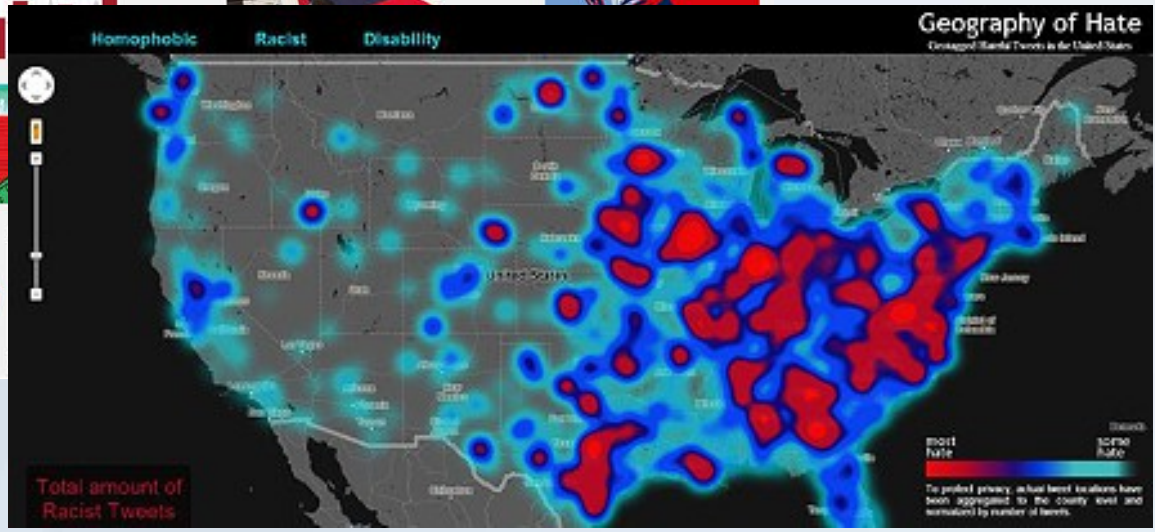
- **Inferred home and work locations of individuals from georeferenced tweets**

- **Two possible reasons:**
  - **First, they are not aware of the disclosure of geopersonal information**
  - **Second, they are willing to share spatial footprints due to low concern of geographic privacy**
    - **Less worried about geopersonal information compared to other sensitive personal information such as bank accounts**
    - **Give out personal information to take advantage of convenience and efficiency, e.g., cell phones, credit cards, location search, etc.**

- Home location may no longer be a secret.
- User contributed locations used to predict real-time traffic
- Sometimes difficult or even disadvantageous not to share geopersonal information
- Many people are willing to share geographic information when using location-based services
- Geoprivacy may be redefined
- Location sharing is a choice, not privacy violation

Geovisualizing emotional response to policy

- A study of the patient protection and affordable care act



By Brent Dell, CSU Long Beach

Research Questions

- **Patient Protection and Affordable Care Act**
  - **Is there a geographic trend in the distribution of emotional response to the PPACA?**
  - **Is there a correlation between the emotional responses and demographic, economic, or political factors?**

- **Twitter**
  - **Do PPACA consumers respond to their experiences on the internet, or more specifically twitter?**
  - **Is twitter an accurate or reliable sampling of the PPACA consumer base?**

Future trends

- **Location-based services in mobile devices: marketing analysis, vacation spots**
- **Know the position of almost everything**
- **Let geospatial technologies do the geo part, we people are responsible for the social part**
- **Volunteered spatial analysis and modeling in various aspects of life**
- **Decision-making based on crowd-sourced spatial data and analytics**