Sharing Research Data for Fun and Profit
Advancing Innovation and Scholarship

Prof. Matthew Zook
Department of Geography
University of Kentucky

Open Access Week
University of Kentucky
October 23, 2014
** My research **

How are geotagged, user-produced data and social media interacting with and mediating urban space and places?
http://dx.doi.org/10.1016/j.geoforum.2014.01.006


The Floatingsheep blog is an ongoing effort to explore some of the less serious sides of big geosocial media data. Joint effort with the FloatingSheep Collective - Dr. Mark Graham (Oxford Internet Institute), Ate Poorthuis (University of Kentucky), Taylor Shelton (Clark University) and Dr. Monica Stevens (University of Buffalo).

www.floatingsheep.org
Church versus Beer

Moran's I spatial autocorrelation
Geotagged tweets for the words "church" and "beer" are highly spatially clustered based on an IDW function with a z-score of 14.34 (p=0.000), implying this highly-clustered pattern is not a result of random chance.
FloatingSheep.org
DOLLY = (Data On Local Life and You)

- One of the key challenges we spent time on was optimizing systems to collect and store geotagged social media.
- Using funding the VP for Research, the College of Arts and Sciences and the Department of Geography we developed DOLLY
- DOLLY = every geotagged tweet in the world since July 2012 (9+ billion tweets)
Stream Workers

Built on top of Node.JS

Process Workers

DOLLY = (Data On Local Life and You)

Cassandra

Open Source Enterprise Messaging

RabbitMQ

Streams Workers

elasticsearch

Front End

Built on top of Node.JS

Post-processing
Arc, R, Illustrator, Google API, TileMill

Or other geosocial media

Ate Poorthuis
DOLLY Interface (alpha version)

http://www.floatingsheep.org/p/dolly.html
Can search by location, time, keyword, hashtag, username, etc.
Sharing Data, Collaborating on Research
tweets consisting of words related to displacement [Starbucks, hipster] were... located in census tracts of Los Angeles that are going through or just finished revitalization projects did contain such tweets. ... In other words, the VGI Method detected a signal for potential displacement
To explore the potential for using Twitter data to track health-related behaviors, we calculated the rate of geo-located tweets mentioning well-known fast food chains for each state in the U.S., then compared these rates to the proportion overweight, and the average number of fruit and vegetable servings consumed per day, by respondents to the Behavioral Risk Factor Surveillance System (BRFSS). We also conducted a parallel intra-state analysis using 41 regions of Kentucky developed specifically for aggregating BRFSS data. The results of this study demonstrate that residents of U.S. states and regions of Kentucky with higher rates of fast food tweets consumed fewer fruit and vegetable servings per day on average ($r=-0.52$, $p<0.0001$; $\rho=-0.59$, $p=0.0001$; respectively). These results suggest Twitter data might be useful for tracking health behaviors related to chronic disease.
Dr. Jay Christian, Assistant Professor, Dept. of Epidemiology, University of Kentucky
Using Twitter to Aid Outdoor Scene Understanding

Goal is to use geo-tagged tweets as a weak form of supervision to train automated algorithms to describe changes in outdoor scenes, from changing weather conditions to human activity. Aim is to use fully automated machine learning-based methods to find connections between the images from a static outdoor camera and the material in the twitterverse.

Temporal distribution of each topics in Twitter

Before and after the snow
Dr. Jin-Kyu Jung. School of Interdisciplinary Arts and Sciences, University of Washington-Bothell

Code clouds: Qualitative geovisualization of geotweets

Using lessons taken from qualitative GIS and geographic visualization to develop “code clouds” as a technique of visualizing spatial information. Code clouds can depict and visualize analytic codes, or codes identifying key ideas and themes, that are generated through digital qualitative research rather than transforming qualitative forms of data into categories or numbers.
Dr. Jin-Kyu Jung. School of Interdisciplinary Arts and Sciences, University of Washington-Bothell
Haiyun Ye and Dr. Keith Clarke, Department of Geography, University of Santa Barbara

Does Volunteered Geographic Information (VGI) after disasters share the time-space structure of other user-contributed Web content?

Studying the case of Boston Marathon Bombings in April 2013 to examine whether Twitter postings follow patterns (the donut hole and the long tail) seen in other user-contributed content relating to disasters. Some possible implications are how social media feeds can act as sensors for disasters and assist in emergency response.
Right after the bombing (~15 minutes)
Within 15 minutes after the bombing, people started twitting about the “breaking” news, which happened at the “finish” “line”.
Two days after the bombing: Tweeting about the “arrest” the “suspect”, and “media” have more “news” and “report.”
Tweets within 15 min of the event
Other Projects

• Dr. Yuko Aoyama, Clark University: Disasters and Urban Transit in Tokyo
• Mobility with the Louisville Metropolitan Area
• Access to Ponds and Lakes in New England
The New Mappings Collaboratory is an initiative of the College of Arts & Sciences at the University of Kentucky to promote creativity, excellence, and interaction around emergent mapping technologies. Working on developing an online graduate program in digital mapping.

http://newmaps.as.uky.edu/
Thank you.

zook@uky.edu
@mattzook

floatingsheep.org
@floating_sheep