

Social Media on Attitudes: Mapping Strategies when Location is Optional

Mary Sue Passe-Smith
Department of Geography
University of Central Arkansas

and

John Passe-Smith
Department of Political Science
University of Central Arkansas



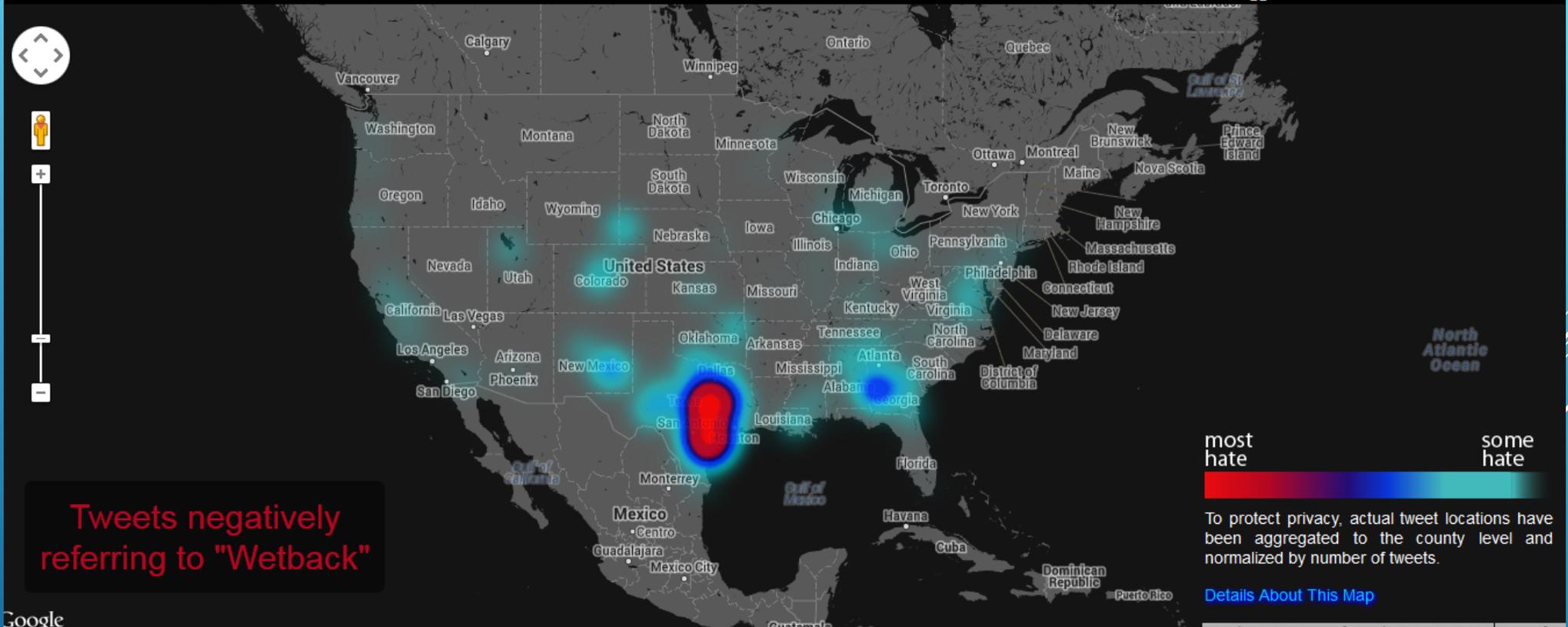
Geography of Hate

Geotagged Hateful Tweets in the United States

Homophobic

Racist

Disability



Source: http://users.humboldt.edu/mstephens/hate/hate_map.html# Site last visited May 31, 2016.

Group Competition Theory		Contact Theory
Labor market: jobs, promotion, pay (Mayda, 2004)		Labor market: having outgroup coworkers (Voci and Hewstone, 2003)
General Resources: welfare, health care, education, etc. (Bobo and Hutchings, 1996)		Cultural Abstraction diminished: having outgroup neighbors, friends, dating, marriages lessening anxiety (Wagner et al. 2006)
Cultural control/advantage: maintaining language, religious institutions, “racial” purity (Branton et al. 2007)		Distinguishing countrywide vs. direct contact: emphasis on direct contact in order to diminish use of abstract threat of “other” by (Schlueter and Scheepers, 2009)

Complicating the Simple Hypothesis...

- Exaggerate outgroup population (Alba et al. 2006)
- Perceived population creates positive/negative attitudes (Semyonov et al.)
- Agitators can cause greater miscalculation of population (Shleuter and Scheepers (2010)
- Host more easily misled when negative image of outgroup held (Shleuter and Scheepers (2010)
- Humans are abstract thinkers. Feared consequences of competition for an entire group does not have to apply to the individual.
- So...is the actual map of outgroup concentrations important at all?

Simple Hypotheses

Group Competition Theory	Contact Theory
Greater Concentration Of outgroup → hostility	Greater Concentration Of outgroup → No hostility (positive reaction)

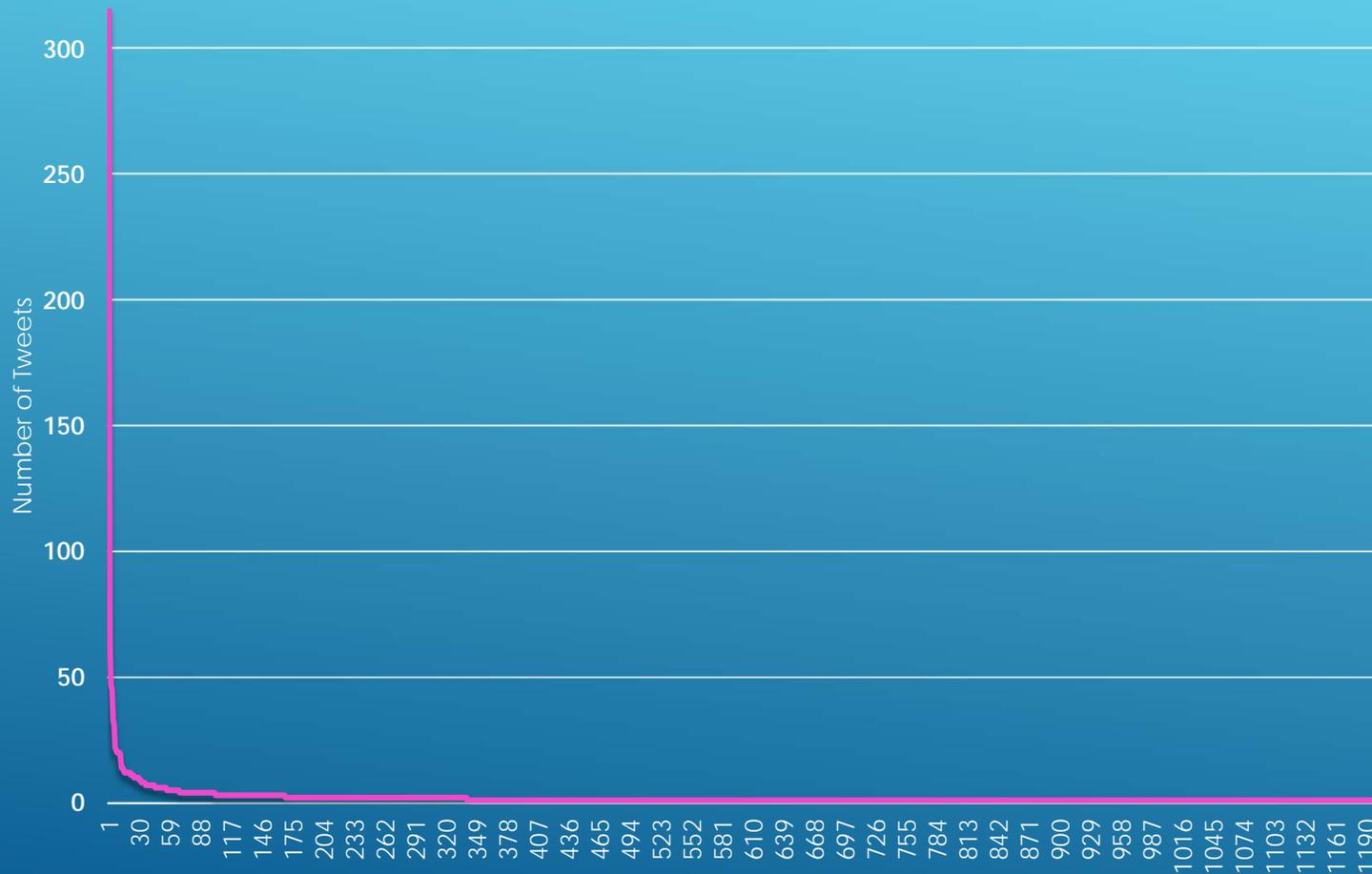
- ▶ Only 26% have listed location as specific as city (Cheng *et al.* 2010)
 - ▶ In this study: _____
- ▶ Far fewer (.42% Cheng, .77% Hecht) are geotagged (lat/long)
 - ▶ In this study: _____
- ▶ Hecht *et al.* (2011) report only 42% of Twitter users report VALID locations in profile
 - ▶ Sample using April 2016 AND those who provide lat/longs, 83.4% were actually where their profile stated – caveat!
 - ▶ Who gives permission to geotag? 38 percent of those with lat/longs were commercial in nature: hiring across US (also “I’m at...” plants) – not ‘personal’ tweets

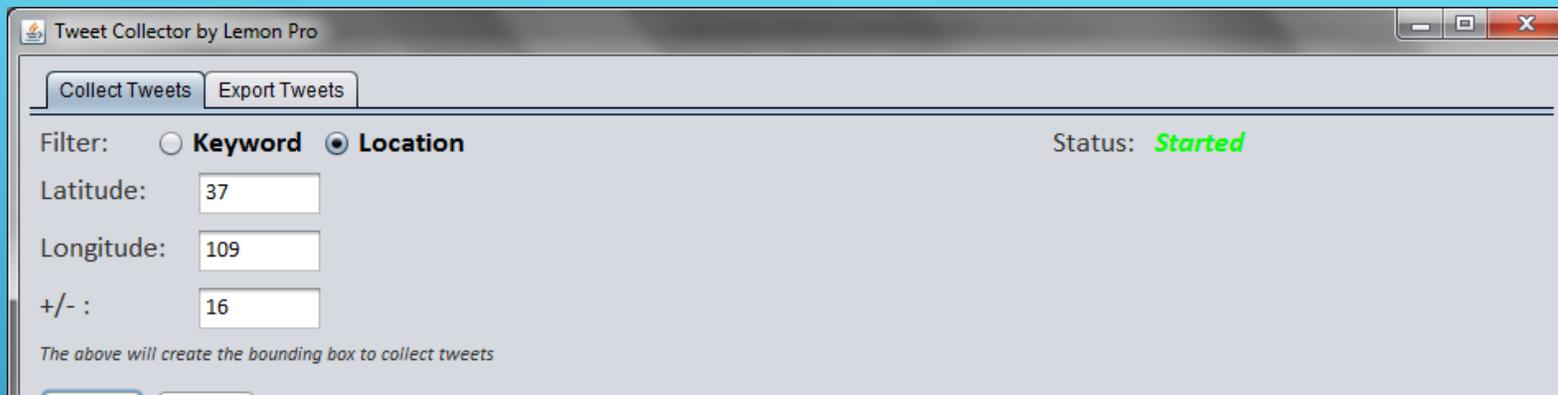
REALLY COMPLICATING THE METHOD:
MYRIAD LOCATIONAL ISSUES WITH TWEETS

- ▶ *Li et al.* note difficulty estimating REPRESENTATIVENESS of Tweets in studying human attitudes/behaviors...

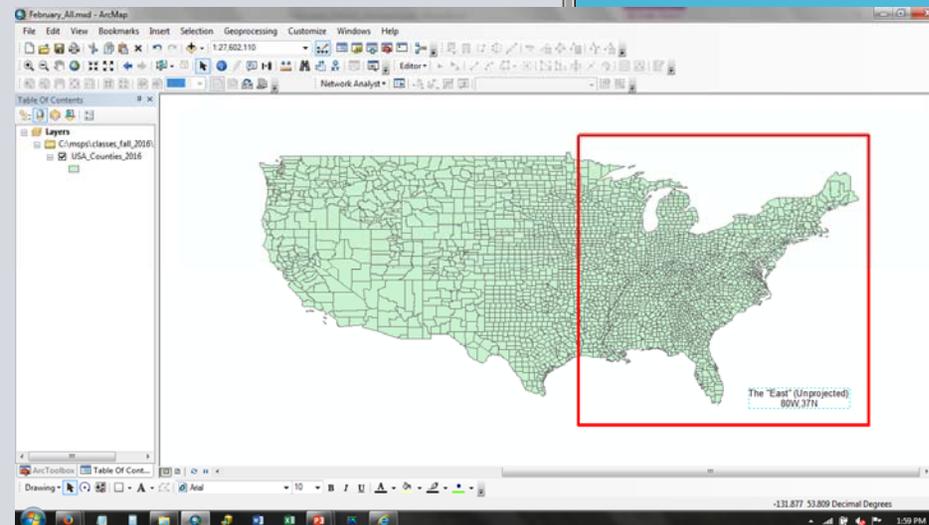
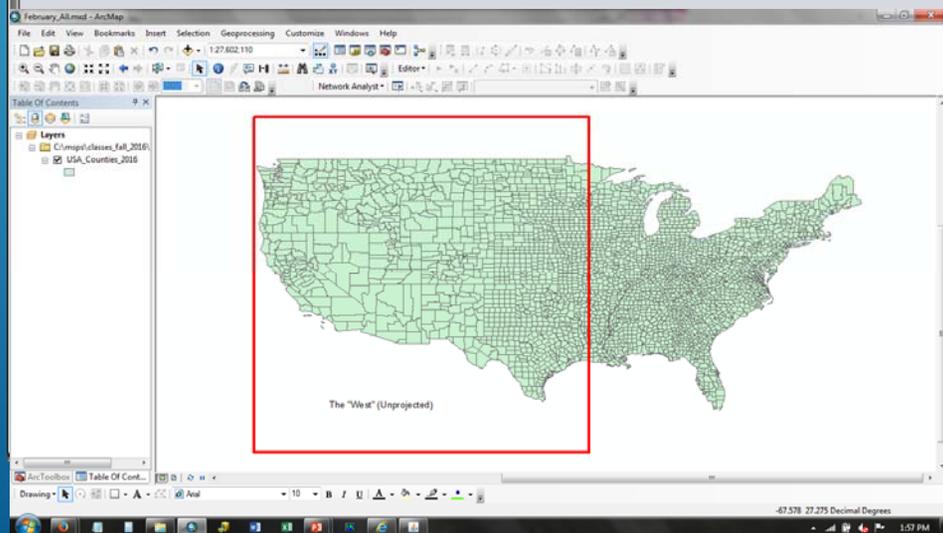
HOW REPRESENTATIVE...?

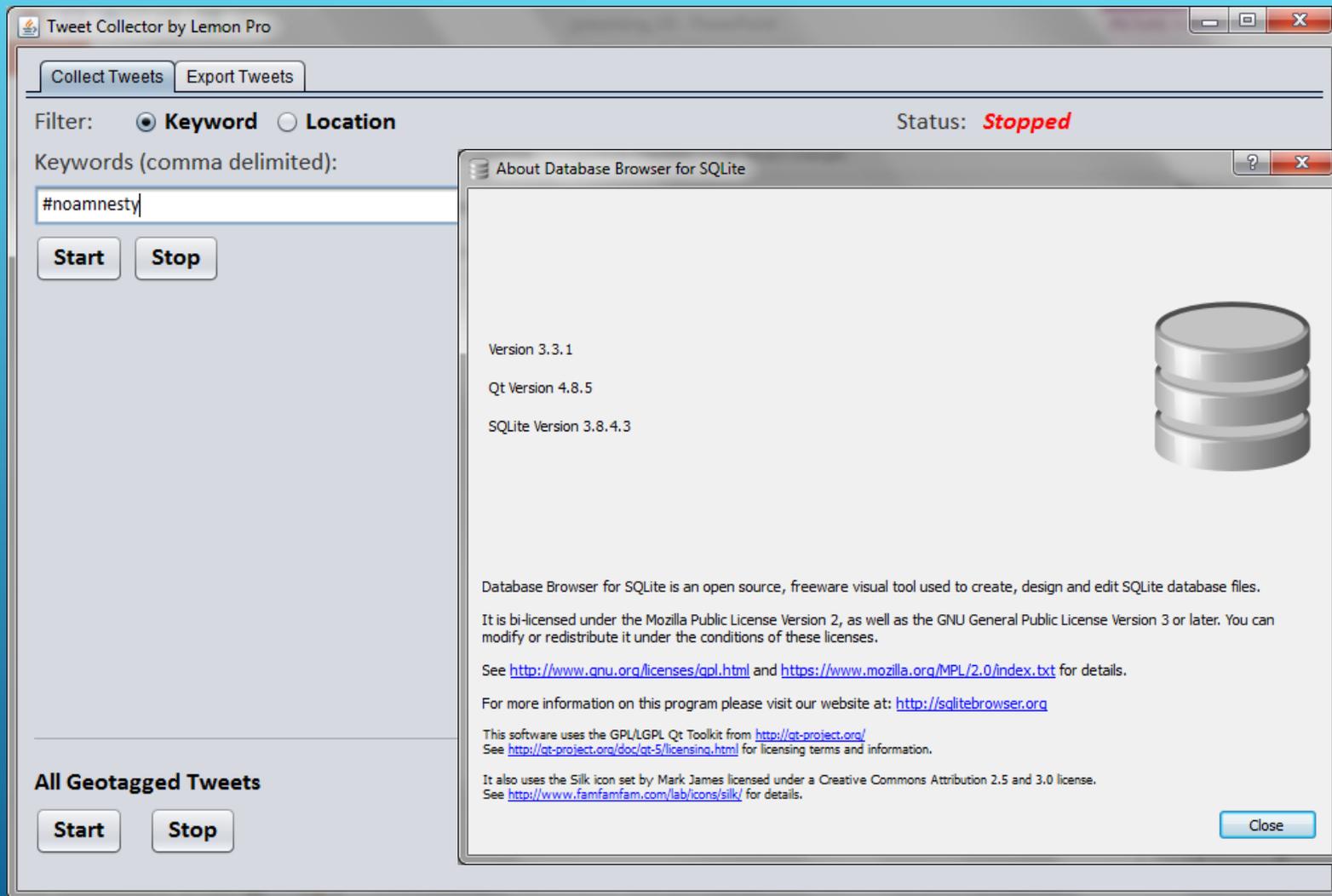
Anti-Immigrant Tweet Count by Username: Very Preliminary Study (2015)





FEB 1-
APRIL 30 2016
(capture many
primaries)





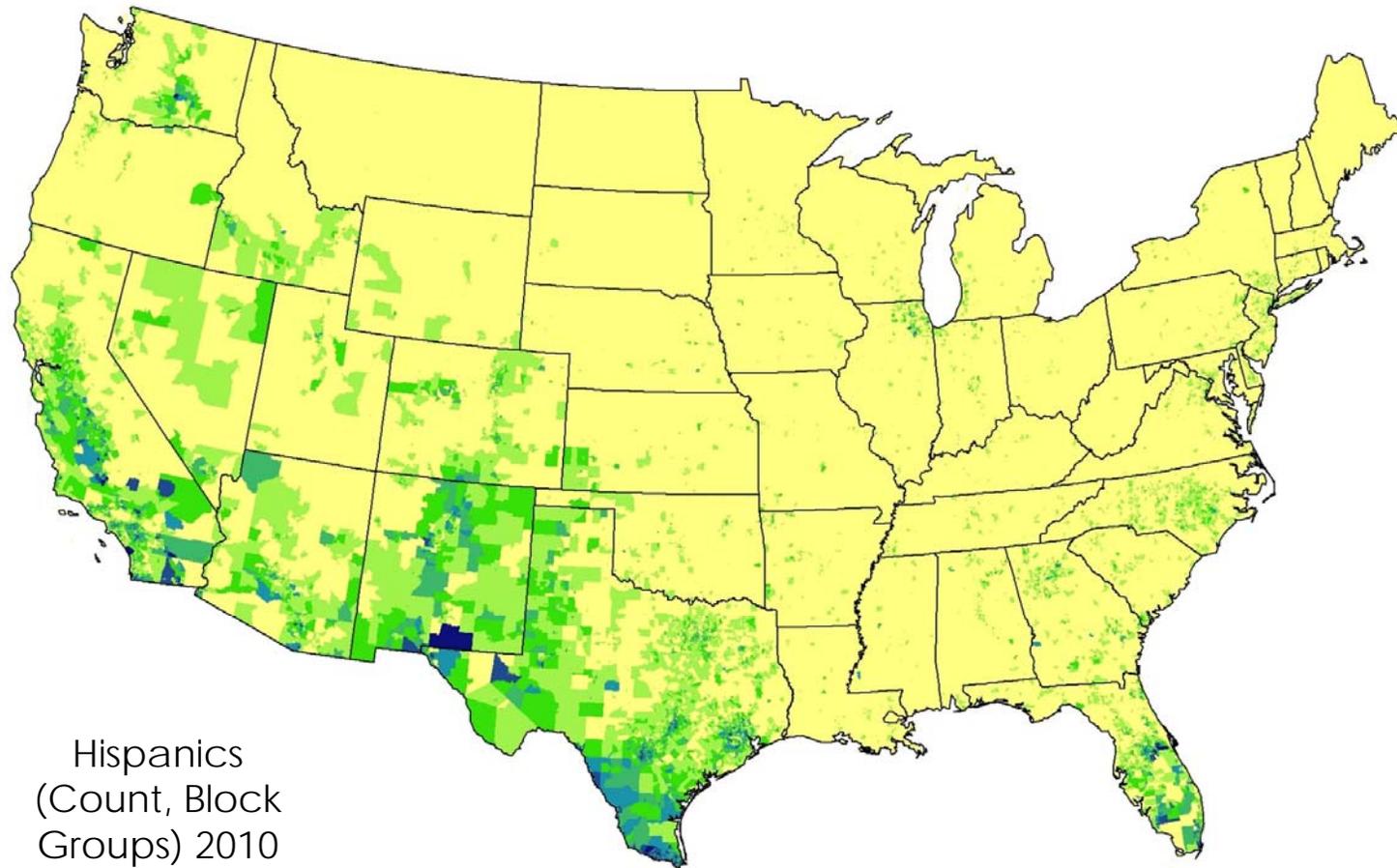
Open source software from <http://sqlitebrowser.org/>

...and now, let the mapping begin!

Three white diagonal lines of varying lengths and positions, extending from the bottom right towards the top right, crossing the text.

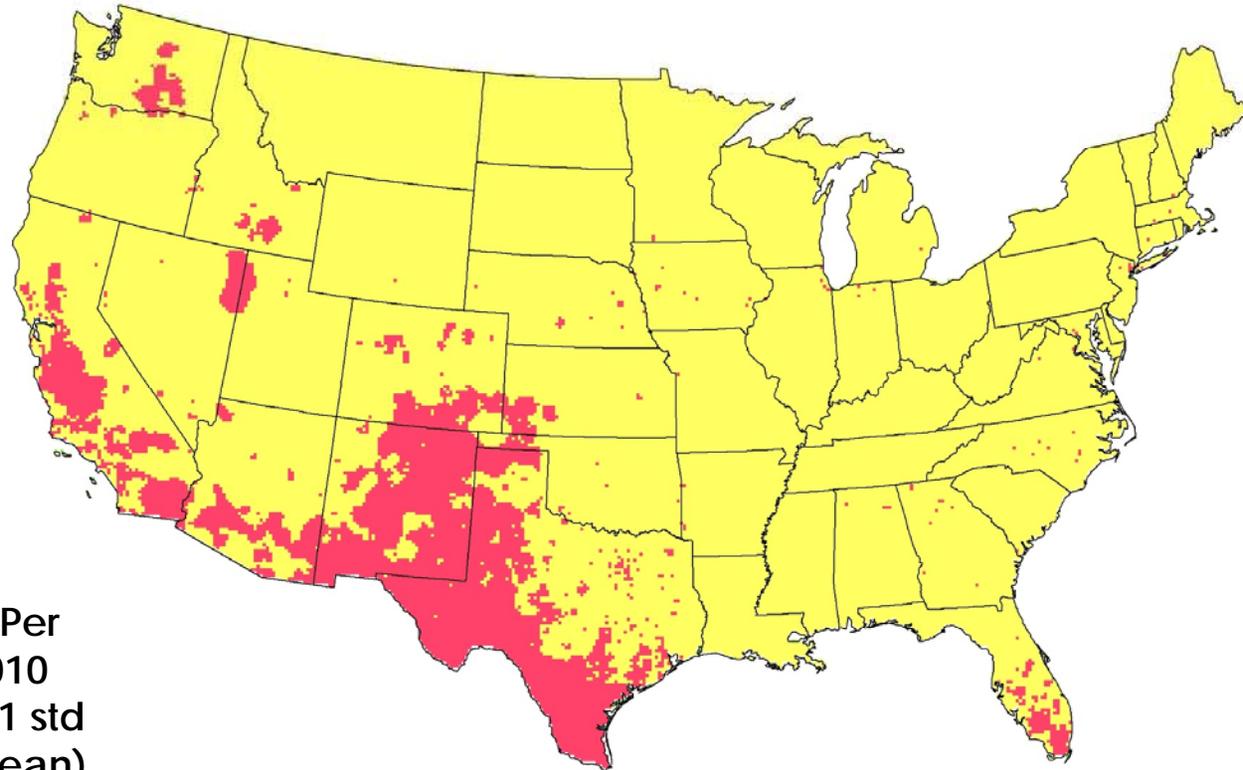
- ▶ All BG in the US; created points (including all demographics, Hispanic)
- ▶ Calculated Hispanics per capita (CAVEAT: not all Hispanics are immigrants/Mexicans!)
- ▶ Interpolate Percentage Population that is Hispanic to show distribution
- ▶ Found mean/standard deviation
- ▶ Found all areas over mean 1 std dev over mean (Boolean 1,0)
- ▶ **GOAL IS OVERLAY HIGH % HISPANIC WITH HIGH % 'BAD TWEETS' to see if they are similarly distributed, or opposite, or mixed**

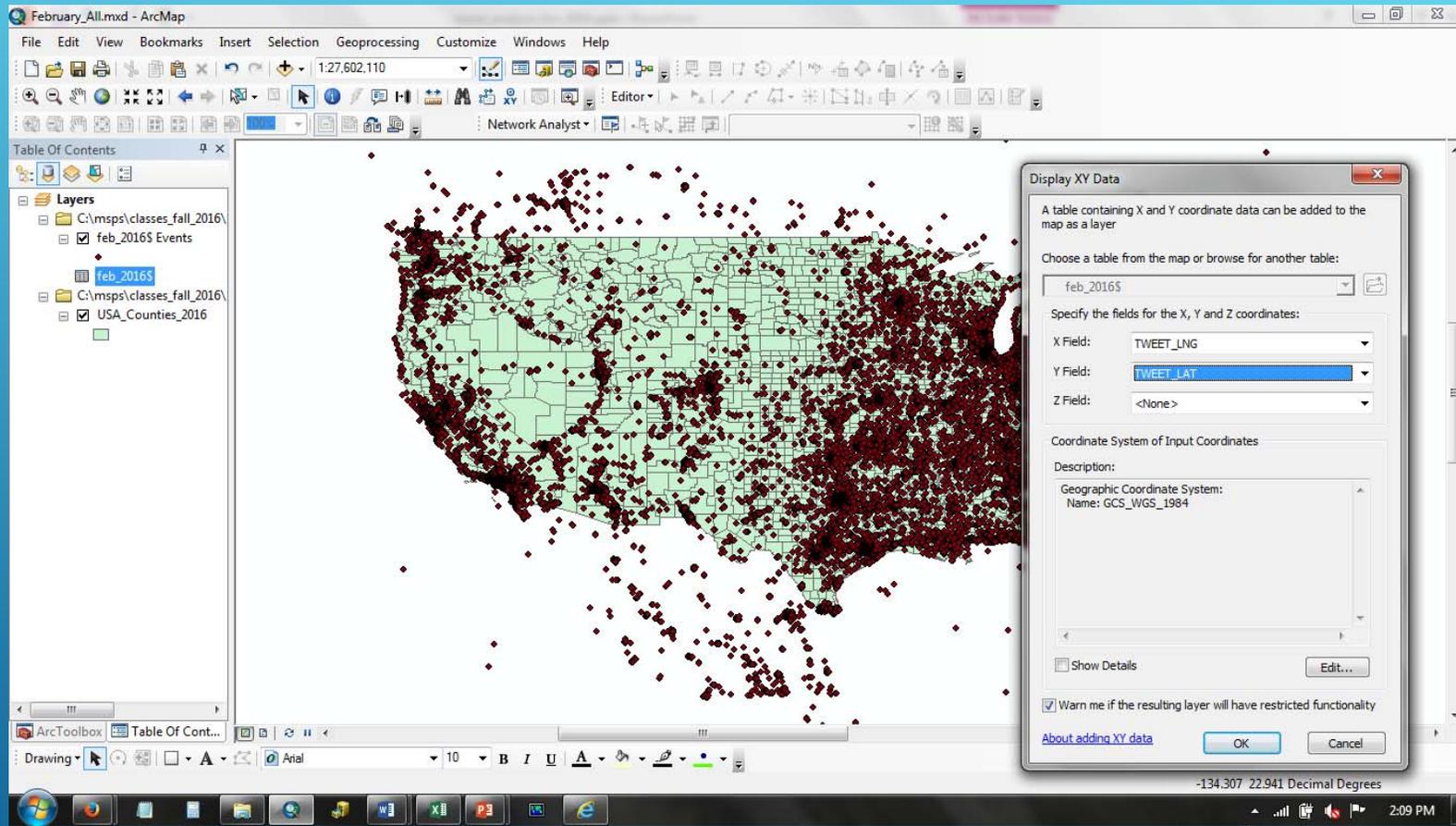
CREATING THE COMPARISON

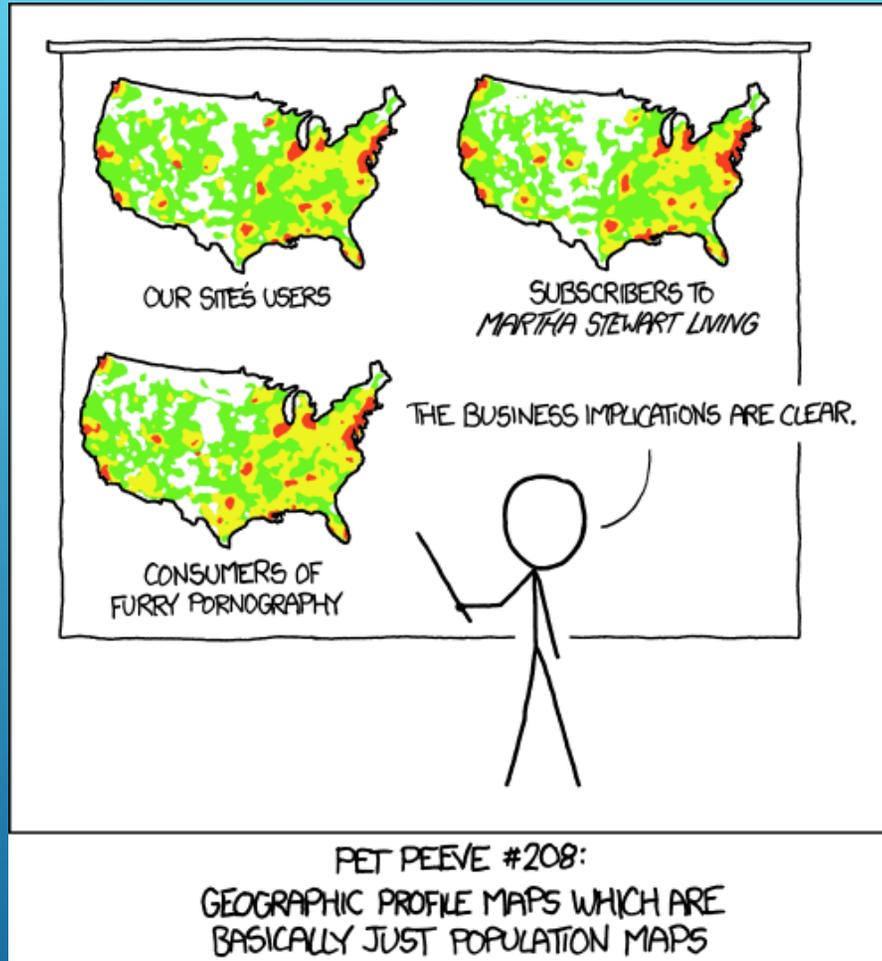


Hispanics
(Count, Block
Groups) 2010

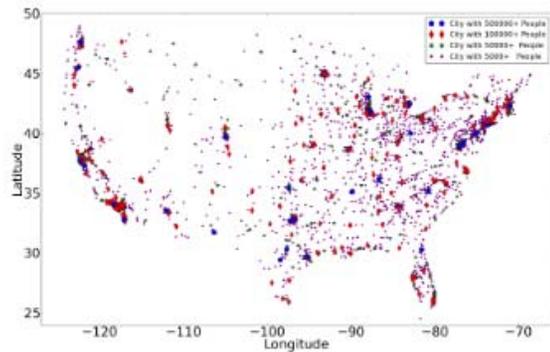
Hispanics Per
Capita 2010
HOTSPOTS (1 std
dev over mean)



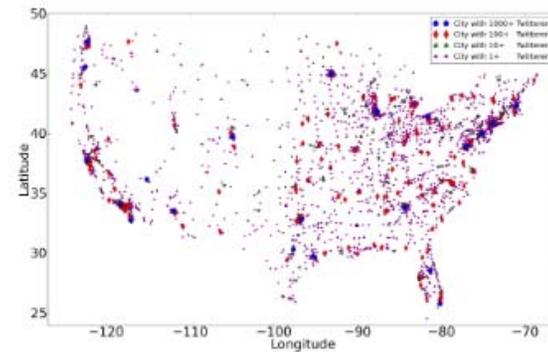




<http://imgs.xkcd.com/comics/heatmap.png>



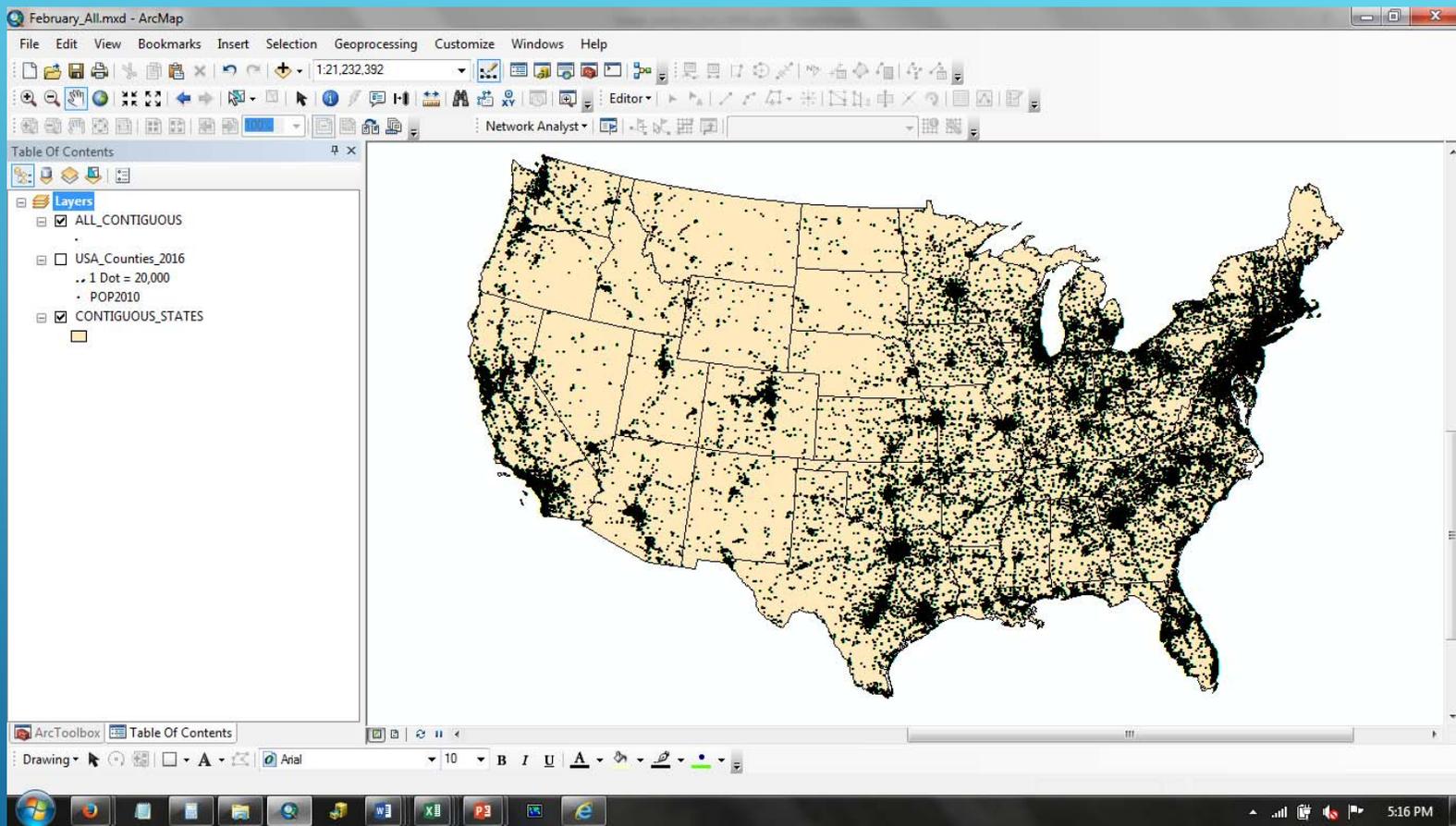
(a) Population Distribution of the Continental United States



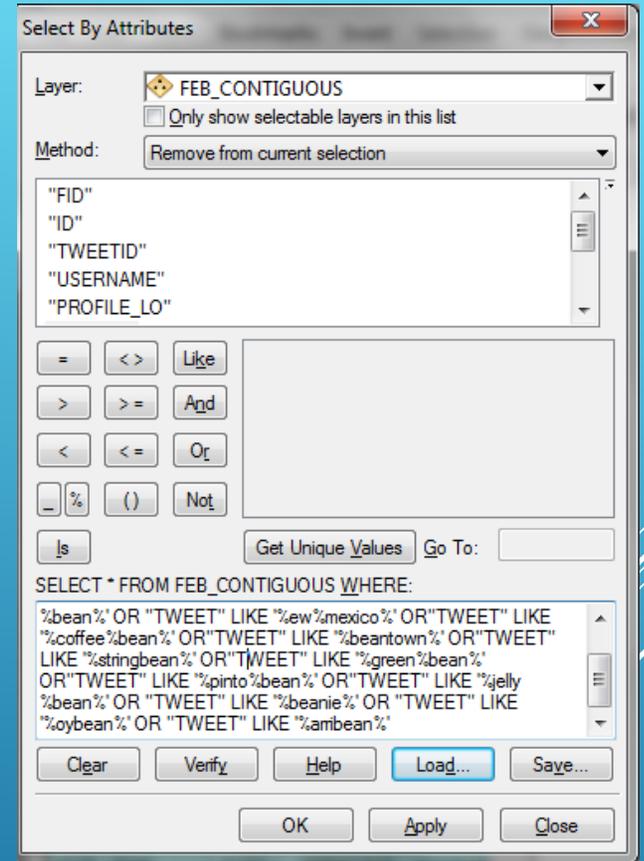
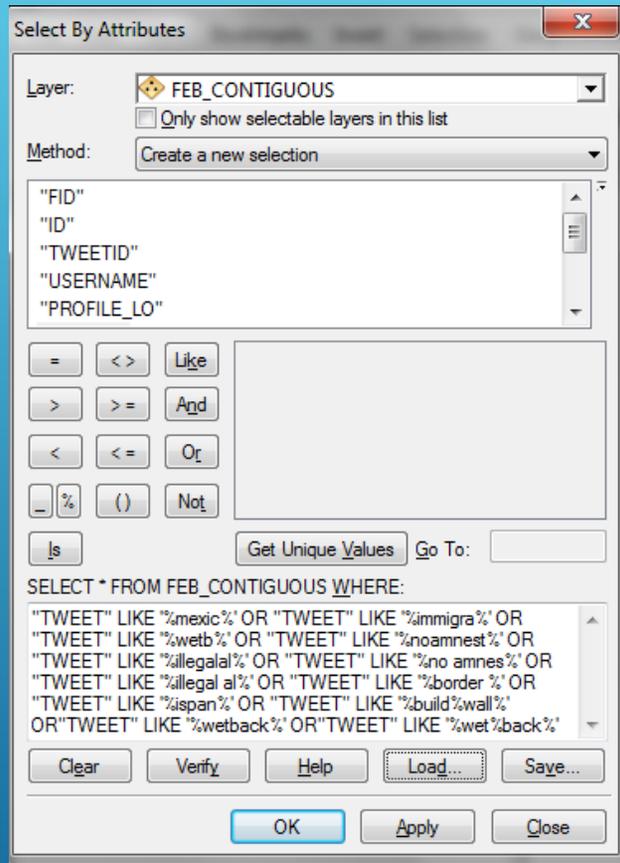
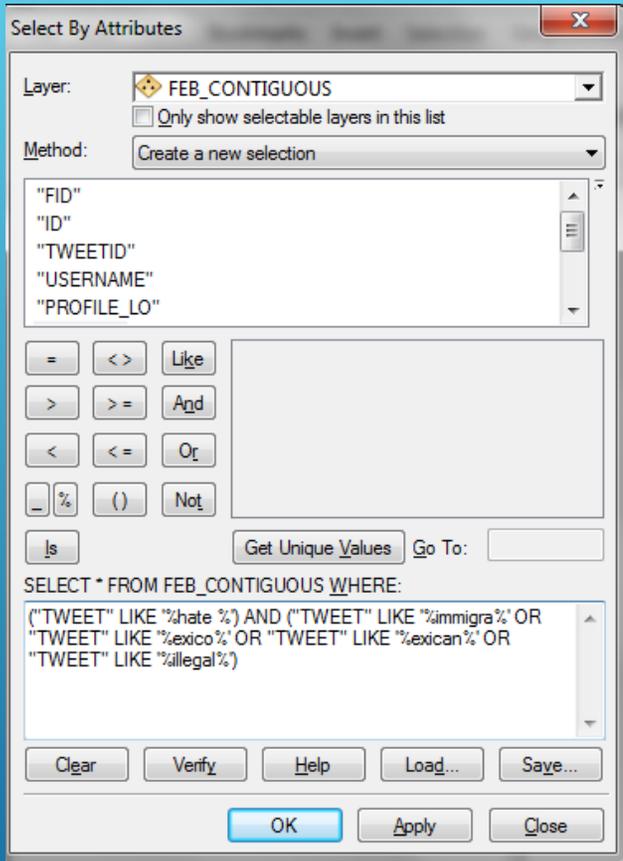
(b) User Distribution of Sampled Twitter Dataset

Figure 1: Comparison Between the Actual US Population and the Sample Twitter User Population

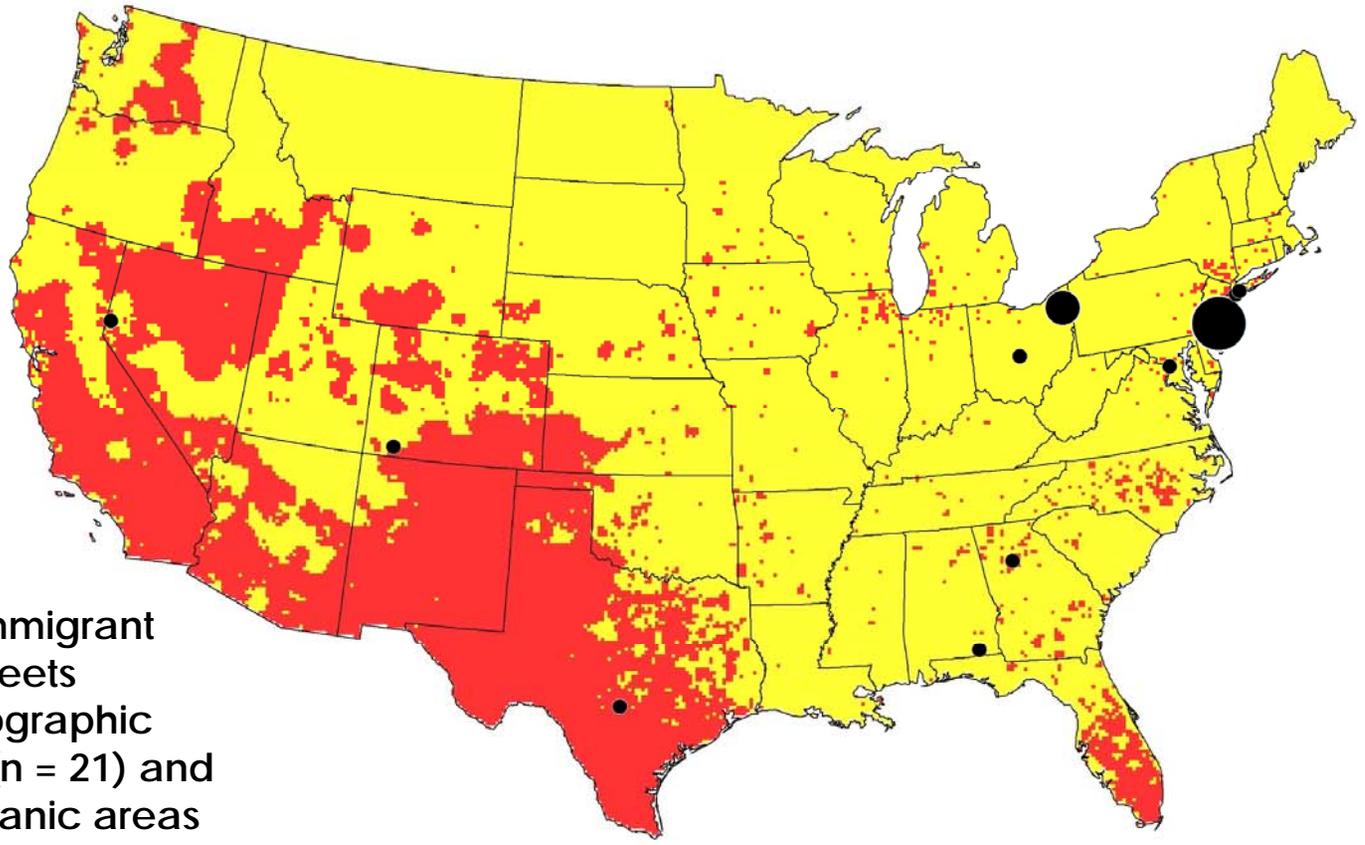
Cheng *et al.*'s 2010 comparison of their 4 million+
 Tweet locations and population distribution
 "...sample...representative of the actual population of
 the United States"



692,000 Tweets with Longitude/Latitudes Included...



Anti-Immigrant
Tweets
w/Geographic
Location (n = 21) and
high Hispanic areas



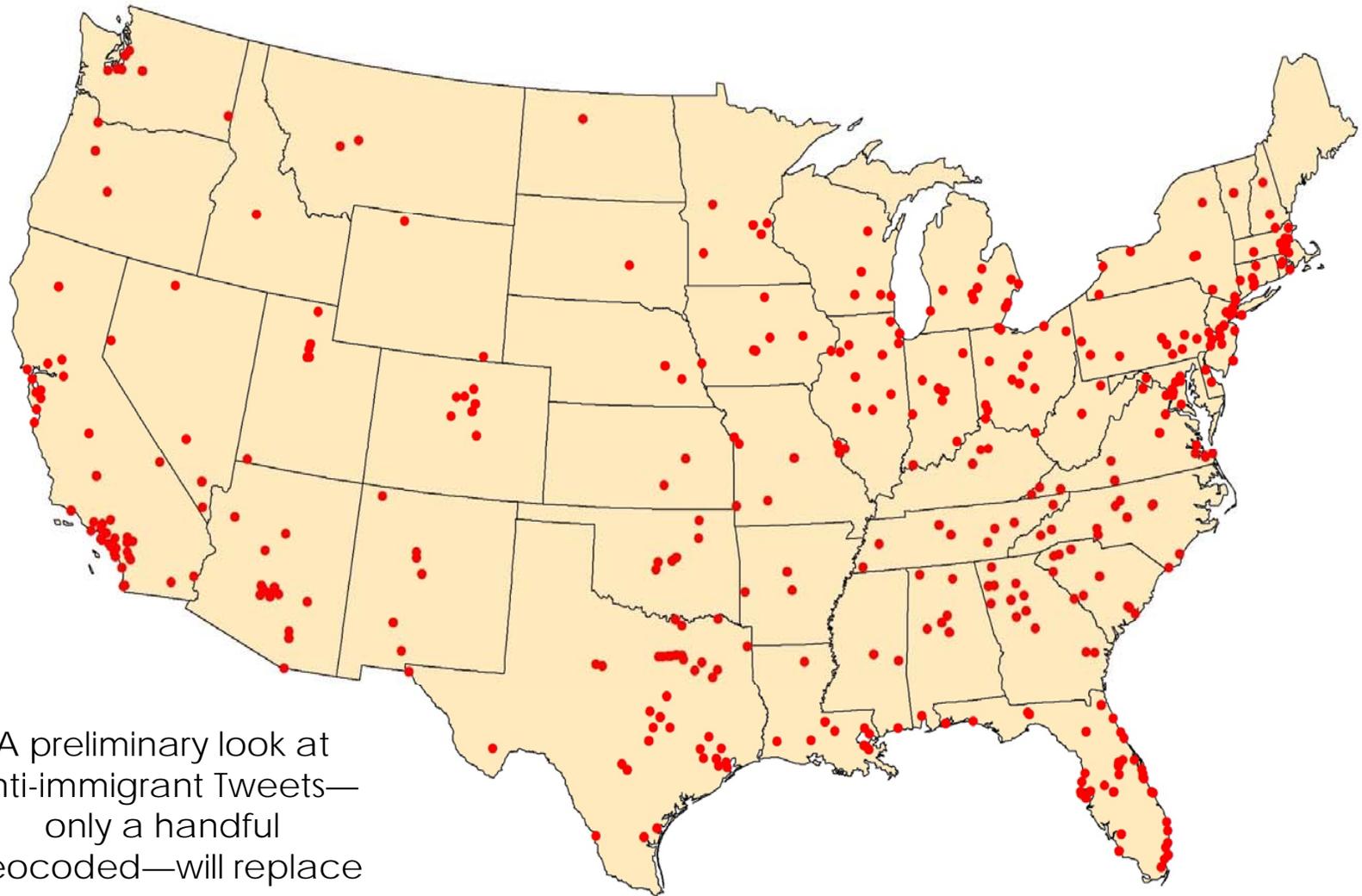
The #HASHTAG Search Sample

Geographic Information Provided in Tweet	N	% of Sample
No Geographic Information		
Compass Direction and/or Geographic Feature (no state, county, or coordinates)		
State Only		
State plus Compass Direction		
State plus Geographic Feature		
State plus County		
State plus City		
Coordinates		
Total		

February 1-April 30, 2016
 Key search terms: noamnesty, no_amnesty, build a wall, build wall

- ▶ New address locators needed!
 - ▶ Use Esri city dataset from ArcOnline
 - ▶ County centroids (geographic)
 - ▶ State POPULATION centroids (not GEOGRAPHIC)
 - ▶ Census Bureau Gazetteer (mountains, small towns, lakes, etc)
- ▶ Some tough decisions
 - ▶ Where IS “by my husband’s side”???
 - ▶ Border Gal in Arizona
 - ▶ Our own table of XY (look up places not in the above data sets)

FEW (N = ???) WITH XY COORDINATES SO...

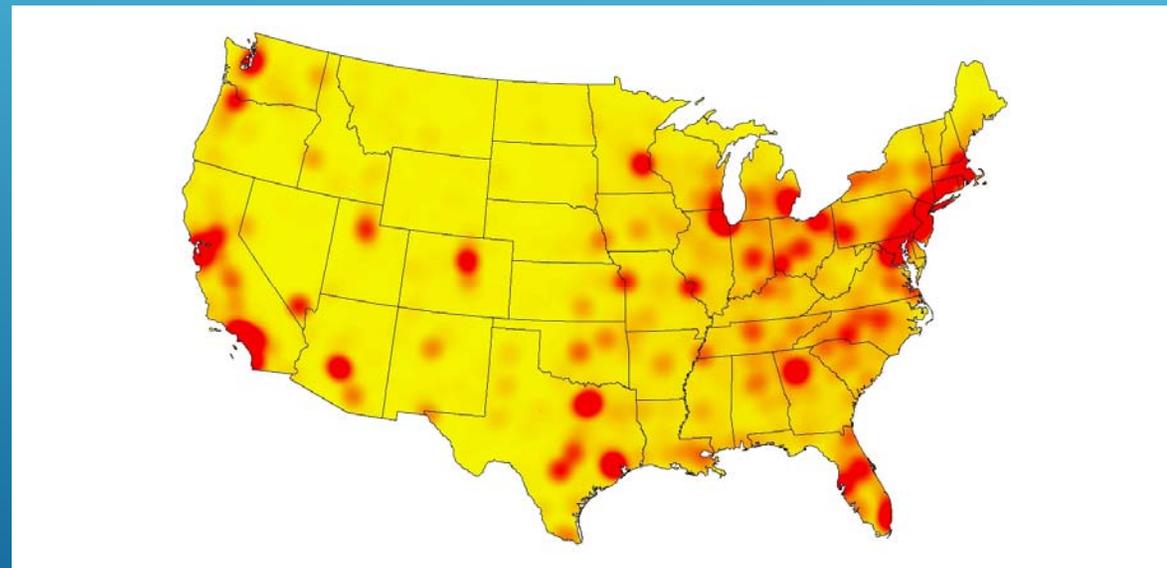


A preliminary look at anti-immigrant Tweets—only a handful geocoded—will replace

INSERT MAP OF DENSITY OF 'HATE' TWEETS (not per capita)

US Population

'Hate' Tweets

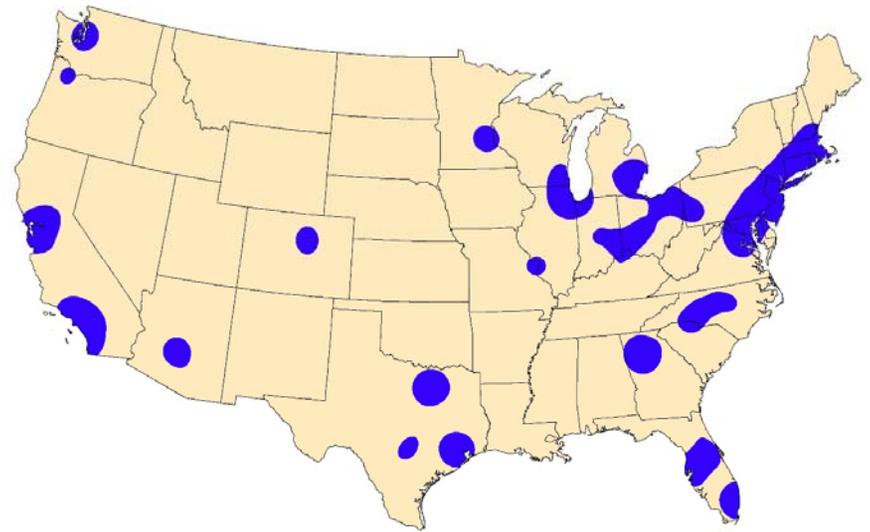
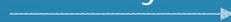


INSERT MAP OF PER-CAPITA DENSITY OF HATE TWEETS

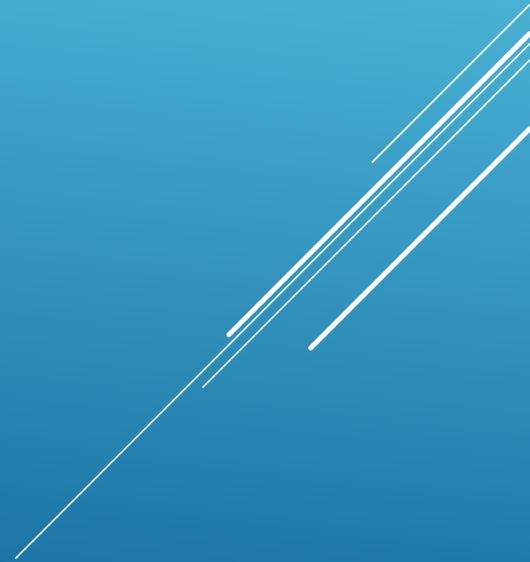
High Tweets Per
Capita

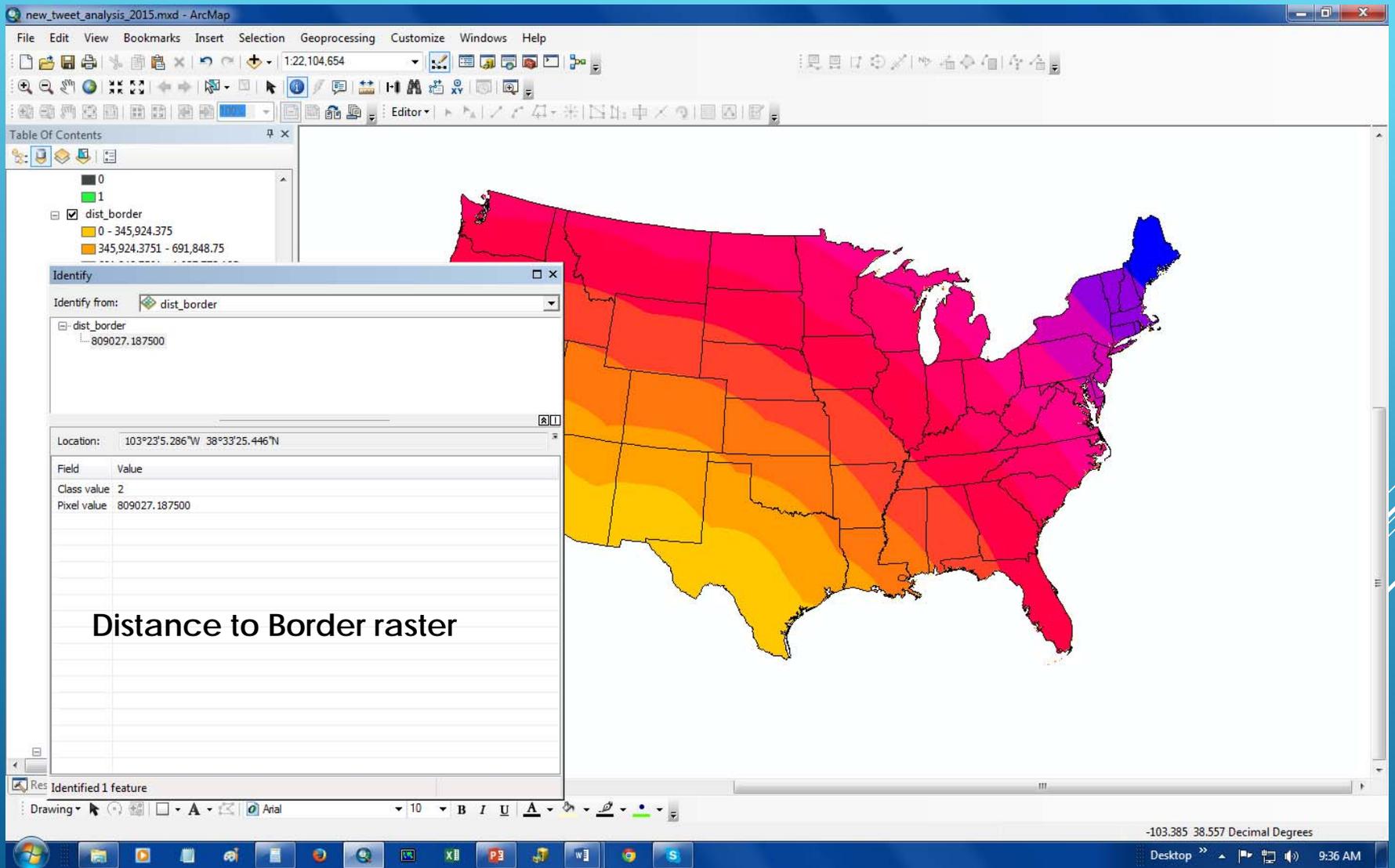


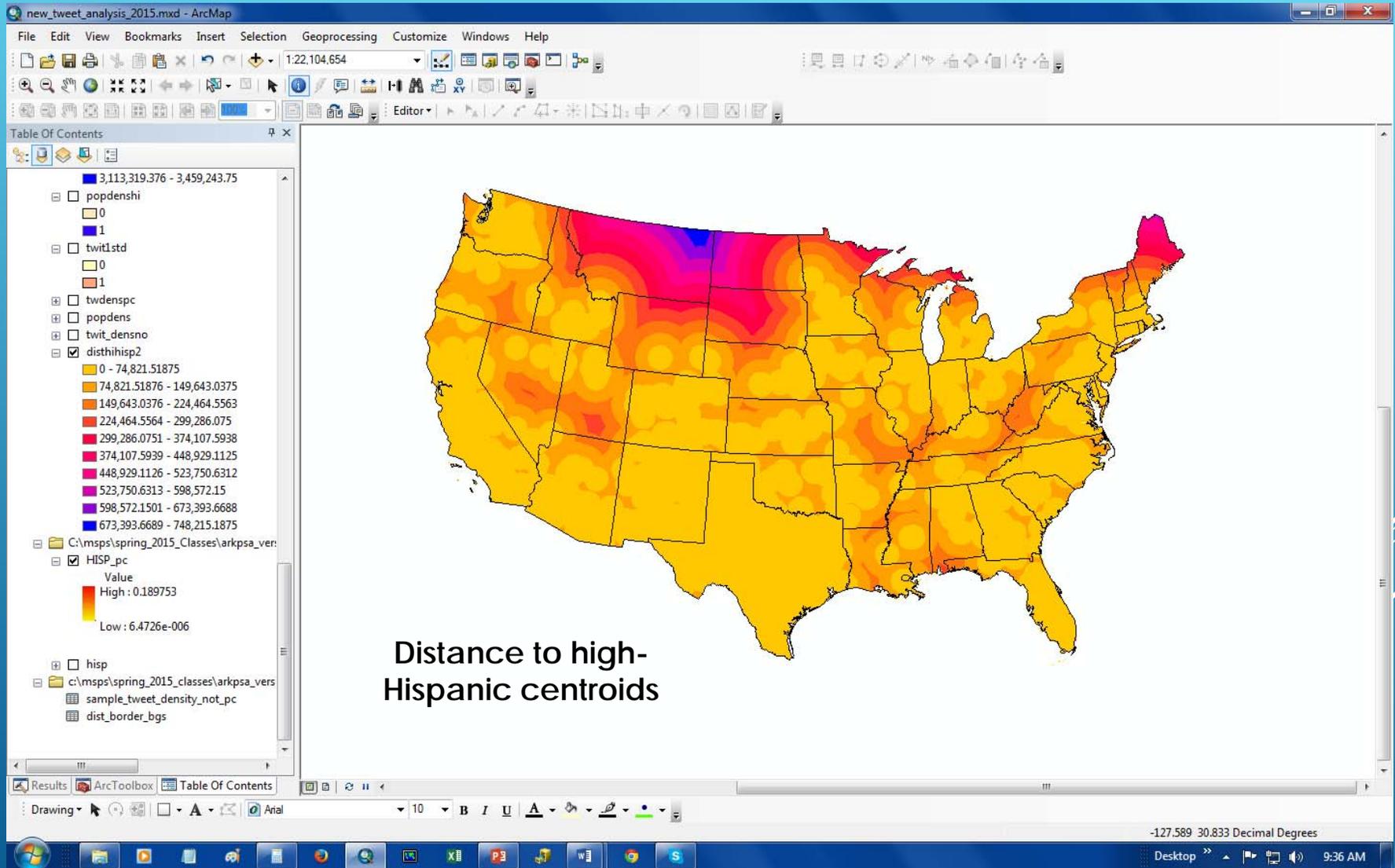
High Population
Density



MAP or MAPS
OF WHERE THE
TWO
OVERLAP...if
they do!







Distance to high-Hispanic centroids

GOAL: will create and insert graphs comparing distance to border and to high Hispanic areas with density of 'hate' Tweets to see if the anti-immigrant feelings ease or increase as distance increases (regression, perhaps—not enough data to work with yet)



- ▶ *Twitter users who allow their geographic location to be seen are generally very unlikely to express anti-immigrant (hateful) feelings; those that do tend to not have a profile location*
 - ▶ *Question: do they even know they are geotagged?*
- ▶ *TO ANSWER: Does hate* overlap areas of extremely large percentage of Hispanics?
- ▶ Does distance from the border/high Hispanic areas coincide with higher or lower anti-immigrant Tweet numbers (per capita) or density
- ▶ IMPORTANT CAVEAT LAST THOUGHT! Some heavily Hispanic areas are notably *not Mexican; but most of the Tweets seem directed at Mexicans*

CAUTIOUS CONCLUSIONS (TENTATIVE!!!!)

And then there's the problems

1. **Not** a random sample
2. **Not** a representative sample
3. **Not** a survey
4. Individuals can retweet same message numerous times & retweet others'
5. Individuals can sign up for multiple identities
6. People can lie about and change profile location (their 'address')
7. People must choose to release coordinates
8. Coordinates do not indicate that the person is from that area (may be driving through)

CANNOT OVERANALYZE, BUT TOO PROMISING TO
IGNORE IT (SUPPLEMENT OTHER RESEARCH)

- ▶ Attempt to find specifically Mexican areas (Census)
- ▶ Isolate specific subsets: areas heavily Mexican, areas with very few/no Mexicans; compare with anti-Mexican Tweet activity
- ▶ Tweak the Tweet finder: to specific areas (and no 0,0), look for CONTENT rather than HASHTAGS
- ▶ Do statistical analysis (regress # Tweets with % Mexican, etc.)
- ▶ MORE TWEETS!

FUTURE

- ▶ USA City, Gazetteer: ESRI ArcOnline maps
- ▶ Block Groups: US Census Bureau (data and TIGER line files)
- ▶ Edward Briggler at LemonPro wrote the Javascript (<http://tc.lemonprogis.com/>)
- ▶ Database Browser for SQLite: Sourceforge freeware (<http://sourceforge.net/projects/sqlitebrowser/>)
- ▶ References available on request

DATA/SOFTWARE SOURCES