



Our Local Surroundings and Sentiments on Twitter

Therese Norman

Wesley DeWitt

Research project: two parts

1. **Analysis: Identifying significant relationships between an individual's health and spatial variables**
2. **Visualization: Disseminating results through an interactive web map using open source software**

Data and Method

- Geographic Extent
 - ∅ Census tracts in LA County
- Data
 - ∅ Tweets through Twitter API
 - ∅ Static socioeconomic and environmental variables from CCEHST
 - ∅ Reference layers(roads, cities, etc.) from various sources
- Method
 - ∅ Collecting/streaming Tweets through Twitter API using Python
 - ∅ Quantifying sentiment of incoming messages
 - ∅ Interactive web map through QGIS leaflet
 - ∅ Statistical Analysis (correlations, hot spots and GWR) using QGIS, ArcGIS, R, Stata

Method – Tweets

- Downloaded between 03/17/2015 – 04/09/2015
- Total tweets in LA County: 90,076
- Keyword examples:

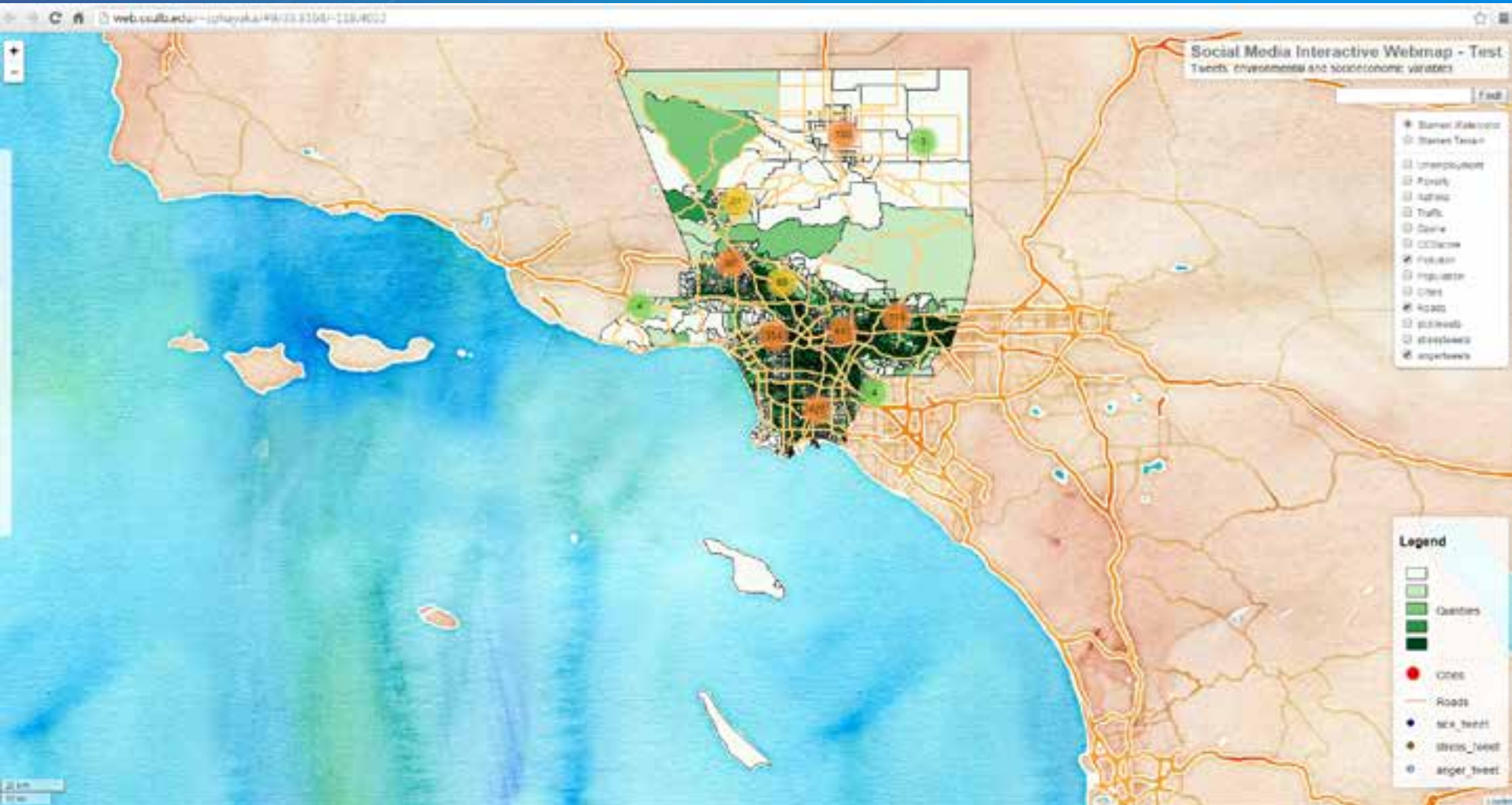
Tweet	Keywords
“sick”	am sick, am ill, feeling sick, feeling ill, cough, a cold, runny nose, the flue, a fever, influenza, sore throat, throw up, puke, under the weather, etc.
“stress”	stress, panic, anxiety, miserable, nervous, sleepless, restless, uneasy, etc.
“anger”	angry, anger, pissed, annoyed, hate, hating, bitter, enraged, furious, irritated, awful, sucks, horrible, loath, resentment, dread, excruciating, distasteful, painful, terrible, disturbing, gruesome, appalling, unpleasant, etc.

- Anger tweet example: *“Everything is somehow my fucking fault. I hate this shit.”*

Bias of Big Geosocial Data

- **Generated by a small fraction of social media participants**
- **Geocoded tweets about 2% of all tweets**
- **Skewed demographic use Twitter**
- **Geotagged data might be produced in or be about another location**
- **Non-human generated content (ads)**

Visualization: Interactive web map



Analysis: Share of Tweets

Standardized by all tweets. Randomly scattered pattern

Share of tweets about being sick, stressed or angry compared to all tweets per tract

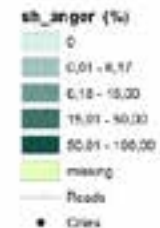
share "sick"



share "stress"

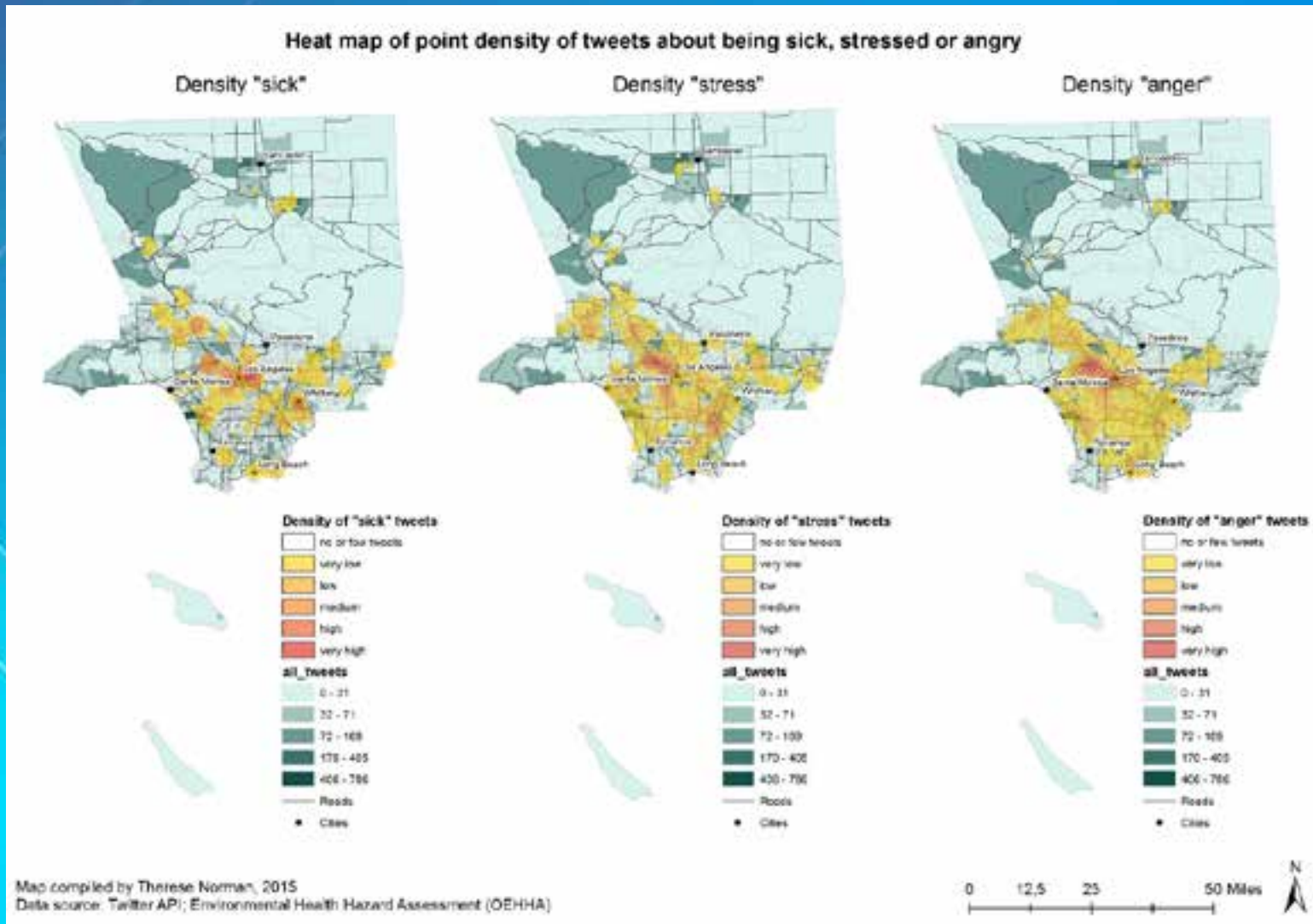


share "anger"



Analysis: Heat maps of Tweets

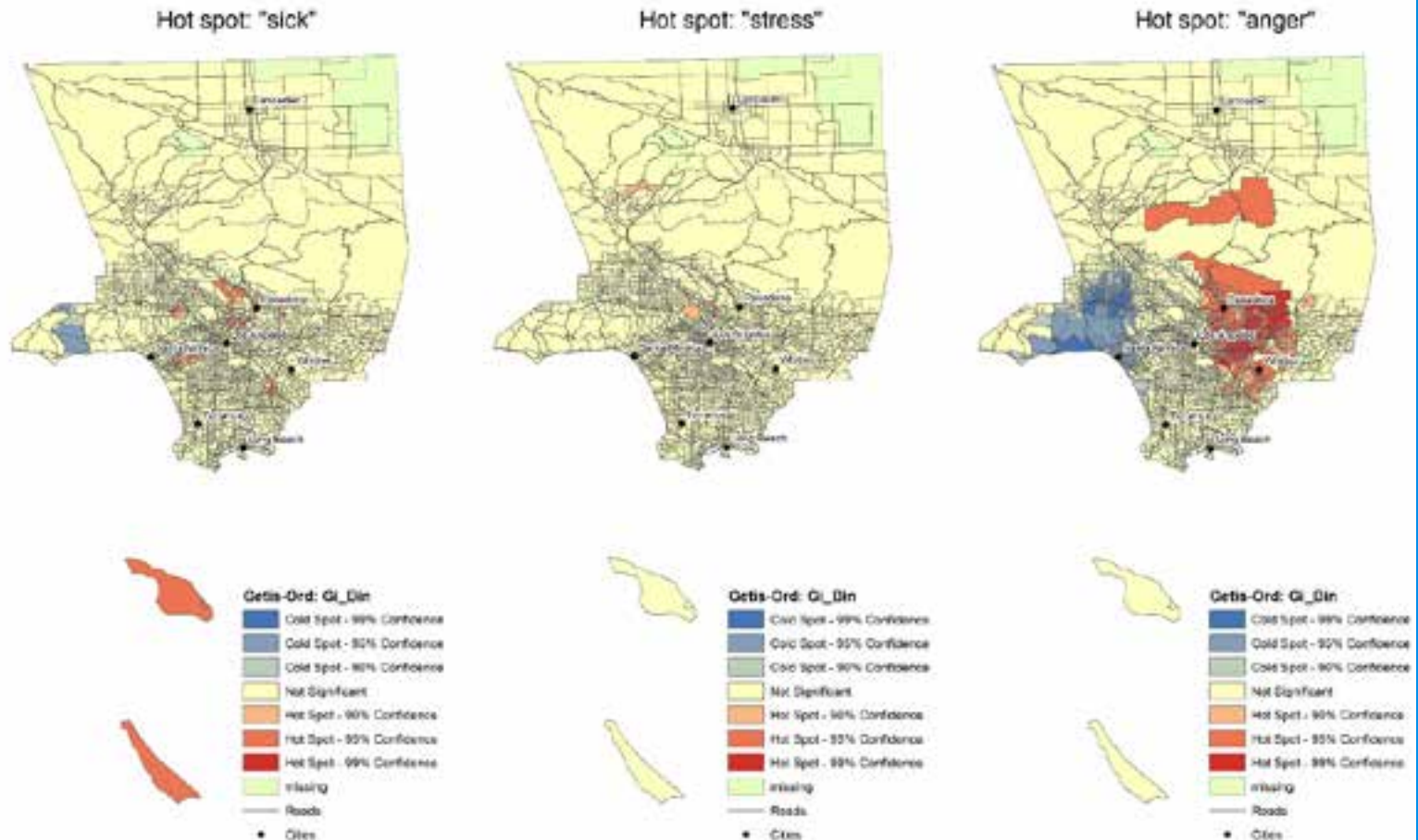
Unstandardized. Higher density near downtown LA. No "sick" along coast



Analysis: Hot spots (Getis-Ord)

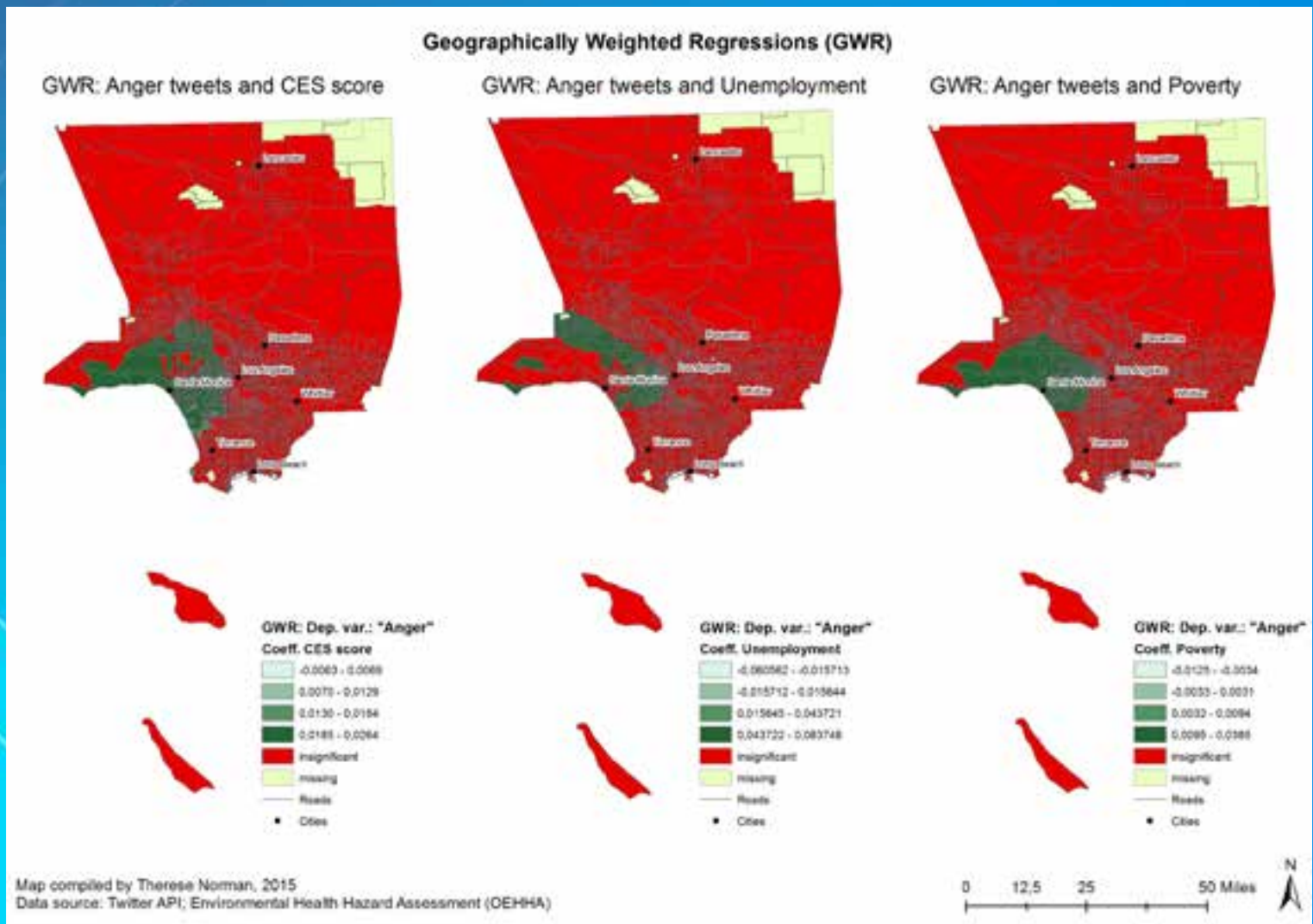
Standardized. Low anger around Malibu/Santa Monica. High anger east of LA

Getis-Ord Hot spots of share of tweets about being sick, stressed or angry



GWR: Anger explained by socioeconomic conditions

Standardized. Low anger – low CES score, unemployment and poverty



Conclusion

- **Preliminary results show a positive relationship between poor environmental and socioeconomic conditions and prevalence of especially anger tweets**
- **Road ahead: Incorporate a live Twitter feed, sentiment analysis and dynamic statistical analysis into the website**