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
  Ÿ Census tracts in LA County

• Data
  Ÿ Tweets through Twitter API
  Ÿ Static socioeconomic and environmental variables from CCEHST
  Ÿ Reference layers (roads, cities, etc.) from various sources

• Method
  Ÿ Collecting/streaming Tweets through Twitter API using Python
  Ÿ Quantifying sentiment of incoming messages
  Ÿ Interactive web map through QGIS leaflet
  Ÿ 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:

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>“sick”</td>
<td>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.</td>
</tr>
<tr>
<td>“stress”</td>
<td>stress, panic, anxiety, miserable, nervous, sleepless, restless, uneasy, etc.</td>
</tr>
<tr>
<td>“anger”</td>
<td>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.</td>
</tr>
</tbody>
</table>

- 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

http://web.csulb.edu/~sphayaka/
Analysis: Share of Tweets

Standardized by all tweets. Randomly scattered pattern
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.
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