How ACS Data Can Make Smart Cities Even Smarter:
A method for combining bike sensor data with ACS demographic data
Peter Viechnicki, Deloitte Center for Government Insights, 12 May 2017
pviechnicki@deloitte.com, @pviechnicki
As Cities get ‘Smarter’, Sensors Proliferate

“The Instrumented Environment”

Solving the Mobility Problem: means helping citizens get around cheaply, efficiently, safely

Cities are installing sensors to monitor flows of many kinds of traffic:

• Automobile
• Transit
• Bikers / Pedestrian

BUT: Having a flood of new data doesn’t always clarify things...

The solution? Connecting demographic baseline data to sensor data can help derive meaning from this welter of information.

The Bottom Line: In this presentation we’ll show how we can use dimensionality reduction techniques and spatial overlays to connect ACS data to bike counter data from the Washington DC region.

Our method provides a clearer story from the bike counter data than is possible without ACS.

Context: Bike Commuting and Bike Count Sensors
Washington, DC Region

• Tops nation in traffic congestion*
• Small but growing cadre of bike commuters**
  • 2010: 5.0%
  • 2011: 5.41%
  • 2012: 6.22%
  • 2013: 6.91%
  • 2014: 7.39%
  • 2015: 7.87%
• Growing Network of Bike Trails
• Increasing use of Sensors
• Bike trail usage data provided by regional partners (API)
• But how to make sense of all the data generated by these sensors?
• Build on prior research:
  • Weekend vs. weekday patterns (Jake VanderPlas, Seattle)
  • Commuter vs. Recreational Counters (Fraser McLaughlin, Eco-Counter)

Sources:
* Texas Transportation Institute Urban Mobility Scorecard
** ACS 5-year estimates, B08301 Means of Transportation to Work, DC MSA
**Data**
- Bike counters: 45 sensors collecting observations every hour for 3+ years
- (34,265 counter X days)
- ACS Demographic variables (2015 ACS 5-year estimates, census tract geography)

**Methods**
- Dimensionality reduction through Principal Component Analysis (PCA)
- Spatial Correlation of PCA factors to ACS variables

**Overview**
- Pull data from bikearlington.com api
- 1.4 million observations (hours X counters X directions)
- Principal Component Analysis
- Keep top three factors
- Spatial correlation with ACS
- Find Meaning!
Methods Part 1: Dimensionality Reduction of Sensor Data
Principal Component Analysis and Visualization

Follow method of VandenPlas for Seattle biking patterns:

- Convert each sensor-day to a 24-dimensional vector, 1 dim per hour

- Principal Component analysis reveals 3 significant factors explain
  - Factor 1: Weekday commuting usage pattern
  - Factor 2: Recreational (afternoon) usage
  - Factor 3: High late-night (night shift?) usage
  - Factor 4: 3-6 AM peak

- Visualize sensors in factor1 X factor2 space
Methods Part 2: Spatial overlay of sensors with ACS data

**ACS data selection:** 2015 5-year census tract data for:

- Bike commuter modal share
- Commute characteristics
- Vehicle ownership
- Household size
- Age
- Education level
- Income

**Select tracts within 1-mile radius of counter**

**Average ACS variables over list of tracts for each sensor**

**Calculate Pearson product-moment correlations between 3 PCA factors and ACS averages**
Methods Part 3: Analysis and Results

Identify significant* correlates

Commuter Segment correlates with:
- Areas with high bike commute share and transit use
- Areas with fewer vehicles

Recreational User Segment correlates with:
- Transit use
- Longer commutes
- Younger residents
- Smaller households

Late-night User Segment correlates with:
- Younger residents
- Smaller households
- Lower incomes
- Longer commutes

<table>
<thead>
<tr>
<th>ACS Correlate</th>
<th>Factor 1: Commuters</th>
<th>Factor 2: Recreational</th>
<th>Factor 3: Late Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biking Share</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transit Share</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Vehicles Available</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel Time to Work</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HH Size</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>High Income Share</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Work in County Share</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Correlations significant at the level of p < .05
Discussion:
Results and Future Research

Generalizable Uses:
• Make outlier detection of sensors more robust
• Predict usage patterns of points on trails
• Improve trail signage and lighting based on 3 user segments

Future research:
• Refine spatial correlation methods
• Rationalize choice of buffer distance
• Weight demographic variable averages using distance decay
• Use CTPP origin-destination data in addition to ACS
• Extend analysis to more complex traffic flows