Applying public data to make mobility more efficient and equitable

Don MacKenzie
University of Washington
We define sustainability broadly

Sustainable Transportation System

Technology
Policy
Behavior
Economy
Environment
Equity
Context:
Revolution in Transportation Data
Traditionally, transportation data were pretty sparse, but everyone had access.
New mobility services are concentrating data in private companies

Who has access?

Level of Access: How detailed are the data?

- Companies: Rich
- Government: Sparse
- Everyone: Sparse

Who has access?
Governments are pressing for disclosure, but data are still not widely available.

<table>
<thead>
<tr>
<th>Who has access?</th>
<th>Level of Access: How detailed are the data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Companies</td>
<td>Sparse</td>
</tr>
<tr>
<td>Government</td>
<td>Rich</td>
</tr>
<tr>
<td>Everyone</td>
<td>Sparse</td>
</tr>
</tbody>
</table>
My opinion: governments should press for less data, but more widely available.

Level of Access: How detailed are the data?

- Sparse
- Rich

Who has access?

- Companies
- Government
- Everyone
Two applications of harvesting API data to understand transportation system performance:

• Is car2go competing with transit?
• Do UberX drivers avoid low-income or minority neighborhoods?
Is car2go augmenting or competing with transit?

Xiasen Wang
PhD student

Zhiyong Cui
PhD student
car2go is being used in Seattle for some late-night trips where transit is infeasible
Lots of ways to assess complement vs competitor question

- In this study:

*Is car2go used for trips:*
  - that are poorly served by transit
  - where car2go offers disproportionately large time savings?
We scraped the car2go API for available vehicles every 30 seconds

- Vehicle ID
- Location (lat/lon)
- Vehicle condition
- Fuel level

Compare list of available vehicles to identify trip starts & ends

Identify origins & destinations

Screen out maintenance trips
We are interested in direct, one-way trips

Seattle, Jan – May 2016
329,478 total trips identified

Available vehicle locations from API every 30 seconds

Total time > worst case + 30 mins
37,286 trips removed

Estimated time < 2 mins
23,356 trips removed

268,836 trip in cleaned data set
Ability to book car2go vehicles 30 minutes ahead complicates travel time analysis

\[ T_{total,i} = \sum_j \beta_{1j} O_{ij} + \beta_2 \cdot t_{Google,i} + \sum_k \beta_{1k} D_{ik} + \varepsilon_i \]
To estimate walking time, assume everyone chooses closest available car.

$$E(d) = \frac{\iint_{A} (|x - x_{c2g}| + |y - y_{c2g}|) dx \, dy}{\iint_{A} dxdy}$$
Pre-booking + walking time averages about 8 minutes
Parking time averages about 3 minutes

Histogram of destination tract coefficients

seconds

Frequency
car2go does **not** appear to be used where it offers extra travel time savings

<table>
<thead>
<tr>
<th></th>
<th>Actual car2go trips</th>
<th>PSRC Travel Survey Transit Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>car2go</td>
<td>car2go</td>
</tr>
<tr>
<td>Walking time</td>
<td>6.2</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>9.2</td>
<td>6.2</td>
</tr>
<tr>
<td>In-vehicle + transfer</td>
<td>17.8</td>
<td>33.7</td>
</tr>
<tr>
<td></td>
<td>33.7</td>
<td>16.1</td>
</tr>
<tr>
<td>Pre-waiting</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total time</td>
<td>24</td>
<td>42.9</td>
</tr>
<tr>
<td></td>
<td>66.9</td>
<td>41.6</td>
</tr>
<tr>
<td>car2go savings</td>
<td>44%</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>46%</td>
<td>68%</td>
</tr>
</tbody>
</table>
Summary

- car2go does not appear to be used primarily on routes poorly served by transit
- For trips taken by car2go, car2go travel time averaged 44 – 64% less than taking the same trip by transit
- For trips taken by transit, car2go travel time would have been 46 – 68% less than taking the same time by transit
- Based on average difference in fare of $7.19 and average time savings of 19 minutes, car2go users are paying about $23 per hour saved
Does UberX provide equitable service?

Ryan Hughes
MS, 2015
Do Uber / Lyft drivers discriminate against passengers based on race and/or gender?
There are multiple opportunities for discrimination to occur in ride-sourcing

**Before Ride**
- **Motives**: Maximize long-term revenue potential
- **Information**: Surge Prices, Neighborhood Demographics, Hearsay
- **Action**: Where do I drive?

**Requests**
- **Motives**: Maximize short-term revenue; minimize per-trip risk
- **Information**: Location, Star Rating, First Name
- **Action**: Reject or Cancel

**Fares**
- **Motives**: Maximize short-term revenue; obtain good rating
- **Information**: First Name, Interaction with Passenger
- **Action**: Route & Service

**After Ride**
- **Motives**: Advise other drivers of experience
- **Information**: First Name, Interaction with Passenger, Gratuity
- **Action**: Star Rating
Waiting times for UberX observed every 4 seconds for 2 months

~ 1 million observations of UberX waiting times from Uber API, May – July 2015
How do waiting times correlate with neighborhood characteristics?

(a) Population Density, Thousands / Square Mile

(b) Employment Density, Thousands / Square Mile
How do waiting times correlate with neighborhood characteristics?

(c) Minority Share of Population

(d) Average Income
We used a spatial error regression model to test effects of density, income, minorities

\[
\text{LogWaitTime}_i = \beta_0 + \beta_1 \times [\text{Populationdensity}]_i + \beta_2 \times [\text{Employmentdensity}]_i \\
+ \beta_3 \times [\text{Averageincome}]_i + \beta_4 \times [\text{Minoritypopulationfraction}]_i + \epsilon_i.
\]

\[
\epsilon = \lambda W \epsilon + \zeta
\]
Judged by waiting time, UberX is not just for "white & wealthy" areas

\[
\text{LogWaitTime}_i = \beta_0 + \beta_1 \times [\text{Populationdensity}]_i + \beta_2 \times [\text{Employmentdensity}]_i + \beta_3 \times [\text{Averageincome}]_i + \beta_4 \times [\text{Minoritypopulationfraction}]_i + \epsilon_i.
\]

\[
\epsilon = \lambda W\epsilon + \zeta.
\]
There are multiple opportunities for discrimination to occur in ride-sourcing (in Seattle, 2015).
Summary
For many important questions, API data are a poor substitute for more detailed system data

• But API data can democratize research and oversight of transportation markets

• Access to existing APIs is nearly costless

• We do not (necessarily) need "all" the data in order to make sound policy choices
A major challenge is that API data are subject to terms of use set by private companies

- Access is a challenge and subject to withdrawal at any time

- Common restrictions on aggregating and saving data are barriers to research

- My opinion: more generous access to APIs should be required as a condition of market access, to legitimize and democratize oversight
Thank you!

Don MacKenzie: dwhm@uw.edu  @donmackenzie9
Regina Clewlow: hello@populus.ai  @populus_ai