Location Data Evolution & Overview

- **Mobile Device Data:**
  - Wireless Carrier Data *(service pings, CDR, etc.)*
    - Received directly from wireless carriers
  - “Sensors” *(bluetooth, WiFi, cameras, etc.)*
    - Requires hardware to be installed in field at each study location
  - GPS Data *(from mobile apps)*
    - Received from App data aggregators

- **Vehicle Data:**
  - GPS
    - Received from data aggregators
      - Connected Vehicles *(CVs)* *(built-in)*
      - Aftermarket devices *(external/added)*
# Data Source Comparison

<table>
<thead>
<tr>
<th>Attribute</th>
<th>LBS / Smartphones</th>
<th>Connected Vehicles</th>
<th>Carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Source</td>
<td>Usually GPS</td>
<td>GPS</td>
<td>Carrier location</td>
</tr>
<tr>
<td>Location Accuracy</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Sampling Rate</td>
<td>Variable</td>
<td>Very-High</td>
<td>Medium- High</td>
</tr>
<tr>
<td>“Who” / “What”</td>
<td>People</td>
<td>Vehicles</td>
<td>People</td>
</tr>
<tr>
<td>Representativeness</td>
<td>Medium-High</td>
<td>Slightly-Skewed</td>
<td>High</td>
</tr>
<tr>
<td>Other Data Features</td>
<td>Inferred individual demographics</td>
<td>Actual speeds, headings, and vehicle types</td>
<td>Inferred individual demographics</td>
</tr>
</tbody>
</table>
Level of Accuracy of Data Sources

- **LBS/Smartphones:**
  - Building or building cluster location
  - Transportation network location

- **Connected Vehicles (CVs):**
  - Transportation network location

- **Carriers:**
  - Neighborhood location
Methodologies

- **Raw data sourcing**
  - Duplicated records
  - Accuracy of actual versus reported location
  - Representative of population vs. unique groups
  - For LBS, don’t want too many apps of same type

- **Data vetting and cleansing**
  - Not all devices are equal
    - High versus low visibility
    - Want only devices with meaningful insight

- **Data representation**
  - Extrapolation/expansion is everything
Data Options and Considerations

- **Study area**
  - Consider majority of “influencers” in study area
    - All trips must have an Origin (“O”) and a Destination (“D”) in study area for an O-D trip matrix
    - “Halo zones”

- **Big data tells a story; but how it’s told is more important**
  - Devices don’t have a “string” following them
  - Sample size is critical; don’t make it too small
    - Typical weekdays through a month vs. Tuesday - Thursday, 4/12 - 4/15
    - Wider range is more representative

- **Fleet data is not the same as heavy vehicle data**
Use-Cases

- **Travel Demand Modeling**
  - Input for O-D data
  - Model calibration and validation

- **Transportation Demand Management**
  - Highlight common O-D pairs for:
    - Low-hanging fruit
    - Express bus service
    - Carpooling and vanpooling outreach
    - Peak spreading

- **Understanding of transportation network users**
  - Where do they live?
  - Where do they work?
  - How often do they make a trip?
Final Thoughts about Location Based Data

- Data is like most things - you get what you pay for

- Having the best information to make a better informed decision matters:
  - Cost of a turn-lane = ~$200k-$350k
  - Widening 2-lane to 4-lane roadway = $Millions

- Ask questions of your data providers
  - Understanding the output is key to work with it for your projects
    - Raw counts (i.e., people or vehicles)
    - Estimated trips (i.e., people or vehicles)
    - Modeled output (i.e., estimated ADTs and Turning Movement Volumes)
    - Impacts of desired study period and granularity of output
Big Data & Tech in Transportation

Q & A

Robert Kohler, PTP
RKohler@AirSage.com
AirSage