Open Source Web-based Transit Demand Modeling (Using GTFS)
A Data Science Approach to Transit Planning

Dr. Catherine T. Lawson
Albany Visualization And Informatics Lab (AVAIL)
Lewis Mumford Center
University at Albany, SUNY
Shifting Transportation Landscape
The Rise of Transit On-Demand

• Built by an army of Data Scientists
• Ridership is On Demand and In Your Hand
• Low Network Overhead
• Efficient Business Model
• Spreading Fast

UBER

Vs

TRANSIT

• Dense Urban Networks of Transit Infrastructure
• A History of Serving Communities
• Employs Planners to Anticipate / Respond to Ridership Demand
• Building a Cache of Underutilized Operations Data

How will Public Transit compete with the data-driven revolution in transportation?
We could utilize data science
Some of us already do.
What if all of us did?

What would it look like?
Data Scientists start by breaking a complicated concept into data-strands.

In the case of Transit, we start by breaking transit ridership into:

- Route Geographies
- Market Area / Census Tract Geographies
- Service Costs
- Service Levels
  - Stops/Sq Mile
  - Scheduling
- Fare structure
- Ridership Count and Survey Data
- Market Area Demographics
  - American Community Survey
  - Census Transportation Planning Products
  - Other Economic Census Datasets
Data Scientists then re-weave the data-strands into user-friendly analytics dashboards by first building tools for each data-strand that are designed to be integrated back together.
Application Programming Interface (APIs)

APIs are the data “pipes.” APIs allow data scientists to build user-friendly software to analyze and visualize complex data.

Data Scientists can turn on different faucets for different users with different levels of access.
Open Data

Allows the assembly of data elements using a consensus process like building blocks

Data Specifications + APIs = Universal Applicability

E.G., General Transit Feed Specification (GTFS)
APIs and GTFS have already improved the transit experience for riders by allowing software designers to develop real-time transit service smartphone apps.
But What About Us?
Can transit planners use these tools to plan more effective and efficient transit service?
• What kinds of data-driven tools would help us to plan our transit services more effectively and efficiently?

• What kinds of data specifications could we collectively deploy to help all of us analyze our respective markets and service?

• GTFS was revolutionary, even transformative, but what about creating a standard for Farebox Data?

• Can we pool together to develop data-driven transit analytics tools?
About Our Data Science Approach to Transit Planning
The General Transit Feed Specification (GTFS) is a data standard for encoding transit schedule and operations information.
Mapping GTFS

AVAIL has developed data science tools for easily mapping GTFS files.
Using GeoJSON
We make geo-coded data web-friendly

GeoJSON holds the same data as a shapefile. It is a text only version of a shapefile.

Many if not all of the open source mapping libraries for JavaScript (Leaflet and MapBox most notably) support GeoJSON and allow GeoJSON layers to be added to their maps.

QGIS, the open source desktop mapping application, also allows the user to view and export data in the GeoJSON format. This has made it extremely popular as a geodata format.

GeoJSON was born out of a format called JSON, which stands for JavaScript Object Notation.

JSON is like a table turned on its side; each row gets its own section, and inside of that section is a series of properties and values unique to that row.
Our GTFS editor allows users to draw and drag GTFS, set schedules, and export GTFS files.

Uploading and Modifying GTFS Files

Uploading new GTFS is easy: A user simply drags and drops a GTFS zip file to the drop zone on the screen.
Census Demographics

Mapping and visualizing your market area’s demographics

- The American Communities Survey (ACS) API comes out every year, allowing planners to view up-to-date market area demographics.

- The CTPP Dataset allows planners to see number of work trips by origin/destination and by census tract.
Regression Models

• Census tracts in designated market area (GTFS adjacent)
  • ACS from any 5-year series
    • Dependent Variable – “bus to work”
    • Independent variables explored using correlation matrix and simple linear regression models
    • Multiple linear regression models specified using SPSS
  • 2010 CTPP modified by ACS regression forecast
  • Ability to develop a suite of regression models to test

Bus to work = f(0 car households, age groups, industry, etc.)
Filter your survey and farebox visualizations based on route or ridership demographic to locate trends.
Multimodal trip planning & analysis

OpenTripPlanner provides a range of passenger information and transportation network analysis features using our infrastructure for finding itineraries combining transit, pedestrian, bike, and car segments.

OpenTripPlanner (OTP) is an open source platform for multi-modal and multi-agency journey planning. It follows a client-server model, providing several map-based web interfaces as well as a REST API for use by third-party applications. OTP Analyst goes beyond passenger information, applying OpenTripPlanner's routing engine to problems in transportation planning, public policy, and the social sciences. The transit system model and optimization logic originally
### Trip Tables

AM Peak generates trips from household location to employment location at a census tract level

<table>
<thead>
<tr>
<th>Depart Time</th>
<th>From Tract</th>
<th>From Coords</th>
<th>To Tract</th>
<th>To Coords</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:02am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:23am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:28am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:27am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:10am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:28am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:08am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:01am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:07am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:29am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:00am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:16am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:14am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>7:24am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>8:09am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
<tr>
<td>7:18am</td>
<td>34001000100</td>
<td>39.358709,-74.460683</td>
<td>34001002400</td>
<td>39.36314,-74.417092</td>
</tr>
</tbody>
</table>
Multimodal trip planning & analysis

OpenTripPlanner provides a range of passenger information and transportation network analysis features using our infrastructure for finding itineraries combining transit, pedestrian, bike, and car segments.

OpenTripPlanner (OTP) is an open source platform for multi-modal and multi-agency journey planning. It follows a client-server model, providing several map-based web interfaces as well as a REST API for use by third-party applications. OTP goes beyond passenger information, applying OpenTripPlanner's routing engine to problems in transportation planning, public policy, and the social sciences. The transit system model and optimization logic originally
Model Analysis Tools:
Analyzing Model Output and Comparing Models

• Bus-Route Ridership on Route Level and on Trip Level
• Bus-stop Level Boarding and Alighting
• Trip Duration
• Wait Times
• Walking Times

Data generated by Open Trip Planner allows users to analyze the results of the microsimulated models.
Model Validation Against Farebox Data
The Market Area Analysis Tool allows a user to manipulate demographic inputs to reflect assumptions about the future. Changes in:

- Land-Use Development
- Population and Employment Changes
- Socio-Economic Trends
Bringing Data Science into Transit Planning

The Problem:
- There simply aren’t enough data scientists.
- Programmers from Computer Science Departments come equipped with math, statistics and coding skills but aren’t typically trained on how to apply computer science to domain specific problems, or how to design user interfaces.
- Web-developers typically don’t focus on data.
- Domain specific researchers don’t typically have computer science or design skills.

One Data Scientist =
Computer Programming Skills + Domain Specific Knowledge + User Interface Web-Design Skills
The Solution:

The Data Science Team Approach: building data science teams out of domain experts, computer programmers and graphic designers.

A national effort to share open source resources, beginning with GTFS: the Spatial Backbone for transit planning

Following the principles and practices of data science would create tremendous benefits and make it possible to compete effectively for market share in the future.
Web-based Open Source Data Science Research Tools

Open Sourcing the research products allows multiple teams to customize the software interface, develop new tools, and build upon legacy research.