The purpose of this study is to continue a comparison of large and continuous quantities of volunteered geographic information about trip planning to methodically collected, but significantly smaller travel survey data. Prior work compared the origin-destination trip planning searches done in The Transit App - a popular smartphone application used to plan transit journeys - to the Regional Household Transportation Survey (RHTS) from the New York Metropolitan Transportation Council. The work proposed here would make similar comparisons with the Transit App to Regional Household Transportation Surveys from other regions to assess the repeatability of the original findings and refine the methodology.

The first paper found that while the Transit App data has significantly fewer dimensions than the New York RHTS, it is counterbalanced by its significantly larger volume of records - showing many OD pairs where the RHTS had no data. Correspondingly, the Transit App data can give seasonality information due to its continuous collection unlike the RHTS which simulates only a ‘typical’ weekday when school is in session. When the comparison was limited to trip-pairs where only both sets contained data it was found that the Transit App produced origin-destination travel patterns in similar ratios to the RHTS. Taking this model a step further, a simple regression found that just origin and destination variables could explain 50% of the variation between data sets. We expect that including sociodemographic variables into the regression analysis can raise its explanatory power significantly.

If this pattern holds in other spatiotemporal dimensions, it suggests that internet based trip queries occur at rates translatable to actual travel. This can ultimately aid in transport modeling and planning functions. There are also potential cost savings: the New York RHTS is a $4.5 million endeavor that takes 2 years to produce and is conducted approximately every 10 years; the Transit App trip data is a continuous byproduct of its user-centric operational service.
Selected Results from Prior Research

Table 1: Geographic & Data Descriptives from NYC Study

<table>
<thead>
<tr>
<th></th>
<th>Transit App</th>
<th>RHTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained records collected</td>
<td>2,064,921*</td>
<td>143,925**</td>
</tr>
<tr>
<td>Constrained records selected</td>
<td>1,146,977</td>
<td>10,351</td>
</tr>
<tr>
<td>public transit mode; weekday, school year timestamp; in NYC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYC Community Boards</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Populated</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Unpopulated</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Community Board Origin-Destination Pairs</td>
<td>5,041</td>
<td></td>
</tr>
<tr>
<td>O-D combinations &gt;0 records</td>
<td>4624</td>
<td>1,897</td>
</tr>
<tr>
<td>OD combinations &gt;30 records</td>
<td>1,885</td>
<td>42</td>
</tr>
<tr>
<td>Original O-D pairs in which both data sets have &gt;0 records</td>
<td>1,896</td>
<td></td>
</tr>
<tr>
<td>Count of those records</td>
<td>975,939</td>
<td>10,350</td>
</tr>
</tbody>
</table>

*Only reflects records in NYC limits, Transit App data exists globally in 109 different regions.

**RHTS draws from the 28 county NYC region in 3 states.

Figure 1: The percent of Transit App queries to RHTS Trip Estimates by NYC Community Board

\[ y = 0.9226x + 4E-05 \]
\[ R^2 = 0.4958 \]

Reference: