Project Report

Assembling Integrated Data Sets for Analyzing Connections Between Travel Behavior, Attitudes, and The Built Environment

Prepared for **Teaching Old Models New Tricks (TOMNET) Transportation Center**

TOMNET Transportation Center

Teaching Old Models New Tricks

Arizona State University

Georgia Tech

UNIVERSITY of WASHINGTON

USF UNIVERSITY OF SOUTH FLORIDA

By,

Deborah Salon
Email: dsalon@asu.edu

Matthew Wigginton Conway
Email: mwbc@unc.edu

School of Geographical Sciences and Urban Planning
Arizona State University
Coor Hall, 5th floor
975 S. Myrtle Ave.
Tempe, AZ 85287

August 2018

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
N/A</td><td>2. Government Accession No.
N/A</td><td colspan="2">3. Recipient's Catalog No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Assembling Integrated Data Sets for Analyzing Connections between Travel Behavior, Attitudes, and the Built Environment</td><td colspan="2">5. Report Date
September 2018</td></tr>
<tr><td colspan="2">6. Performing Organization Code
N/A</td></tr>
<tr><td colspan="2">7. Author(s)
Deborah Salon, https://orcid.org/0000-0002-2240-8408
Matthew W. Conway, https://orcid.org/0000-0002-1210-2982</td><td colspan="2">8. Performing Organization Report No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
School of Sustainable Engineering and the Built Environment
Arizona State University
Tempe, AZ, 85281</td><td colspan="2">10. Work Unit No. (TRAIS)
N/A</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
69A3551747116</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
U.S. Department of Transportation,
University Transportation Centers Program,
1200 New Jersey Ave, SE, Washington, DC 20590</td><td colspan="2">13. Type of Report and Period Covered
Research Report
(2017 – 2018)</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
USDOT OST-R</td></tr>
<tr><td colspan="4">15. Supplementary Notes
N/A</td></tr>
<tr><td colspan="4">16. Abstract
The project described in this report is a data fusion exercise in which multiple datasets were fused together with travel survey data in two states: California and Arizona. As yet, there is only preliminary analysis and no real findings.</td></tr>
<tr><td colspan="2">17. Key Words
Attitudes; Travel Behavior; Data Fusion</td><td colspan="2">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif.(of this report)
Unclassified</td><td>20. Security Classif.(of this page)
Unclassified</td><td>21. No. of Pages
18</td><td>22. Price
N/A</td></tr>
</table>

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This study was funded by a grant from A USDOT Tier 1 University Transportation Center, supported by USDOT through the University Transportation Centers program. The authors would like to thank the TOMNET and USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project.

TABLE OF CONTENTS

DISCLAIMER 3
ACKNOWLEDGMENTS 3
INTRODUCTION 6
DATA 7
  Arizona Data Assembly Project 8
    Access Across America: Transit 2014 9
    Decennial Census and LODES data 10
    Maricopa County 2012 Assessor Data 10
    Maricopa County Land Use Dataset 10
    TIGER/Line Data 11
    USGS Protected Area Database 11
  Preliminary Data Exploration 11
  California Data Assembly Project 13
  Preliminary Data Exploration 15
NEXT STEPS 18
REFERENCES 18

LIST OF TABLES

Table 1: Datasets to be fused with ASU Travel Survey 9

LIST OF FIGURES

Figure 1: Simplified Conceptual Model of the Determinants of Trip Distance and Mode Choices 6
Figure 2: Simplified Conceptual Model of the Determinants of Trip Frequency 7
Figure 3: Phoenix-area Transportation Analysis Zones (TAZs) 8
Figure 4: Relationships between attitudes, residential choice, and travel, Arizona dataset 13
Figure 5: Relationship between attitudes, residential choice, and travel, California dataset 18

INTRODUCTION

Important travel behavior outcomes include distance, mode, route, and time of day for a particular trip, as well as number of trips taken overall by an individual or a household. Observed travel behavior depends on some combination of attitudes and preferences about these factors; attitudes and preferences about vehicle ownership, vehicle type, and residential location; and physical (both individual capability and the built environment) and budget constraints (Figures 1 & 2).

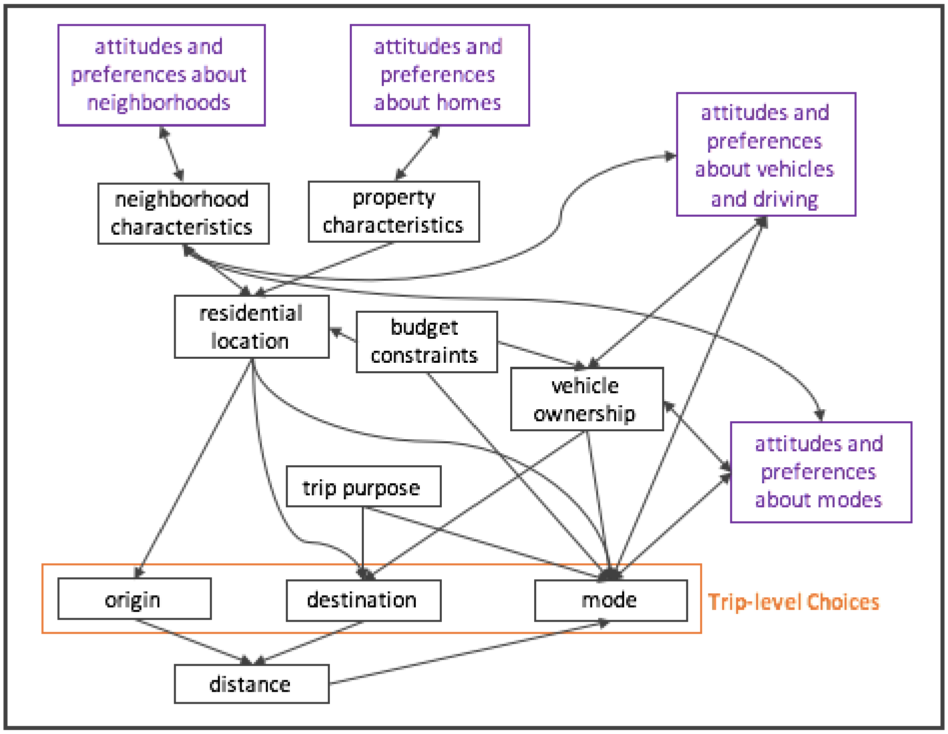
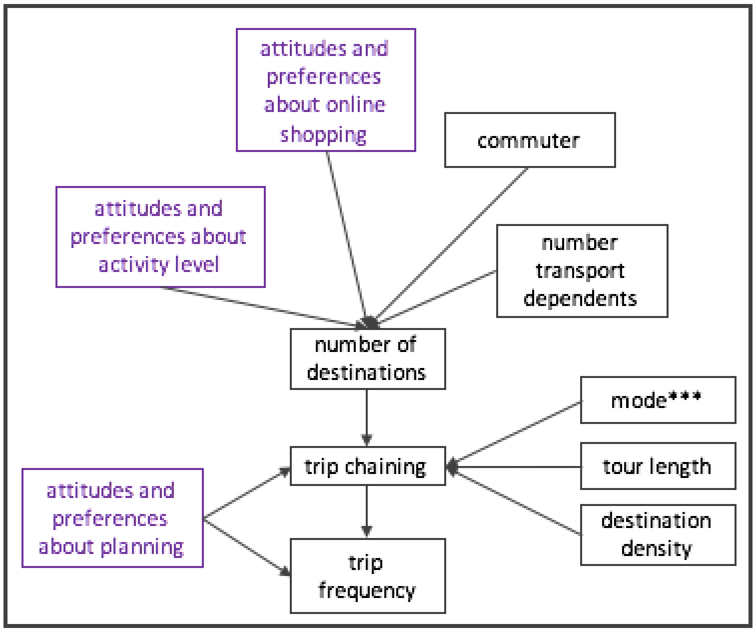


Figure 1: Simplified Conceptual Model of the Determinants of Trip Distance and Mode Choices



*** Mode is both affected by and affects multiple other factors - see Figure 1

Figure 2: Simplified Conceptual Model of the Determinants of Trip Frequency

Travel surveys -- that are used to quantify these relationships and form the basis for regional travel demand models -- contain only a portion of the relevant variables. Most travel diary-style datasets include details of trip origins, destinations, modes, time of day, purpose, frequency, and vehicle ownership. Many also include residential neighborhood information, and socioeconomic and demographic data about the households. Few include attitudes and preferences at all. If included, attitudes and preferences about modes are the most common. We are not aware of any travel diary-style datasets that include detailed home property characteristics and attitudes and preferences about neighborhoods and homes.

Separate surveys do exist that include questions regarding attitudes and preferences about modes, vehicles, and neighborhoods. However, these detailed attitudinal surveys do not collect detailed travel behavior information, chiefly based upon the (probably correct) concern that the respondent burden would be too great.

This project addresses this data gap by fusing existing datasets that are rich in attitudinal information with vehicle ownership and use, home neighborhood, and individual property characteristics datasets. These combined datasets will allow us to investigate questions about how attitudes influence residential location choice, and how the combination of attitudes and location choice influences travel behavior. The development of fused datasets is the deliverable from Year 1 of this project, and the subject of this report.

The rich datasets of attitudes, location attributes, and travel behavior produced by this project will be made available to other TOMNET researchers for use in other projects as well.

DATA

This project fused observations in two attitudinal surveys previously collected by TOMNET team members: the ASU Travel Survey (2012) and the California Millennials Dataset (2015). We used home location information provided in each survey to merge household observations with neighborhood and property characteristics information from multiple sources. The details are documented in the following two subsections.

Arizona Data Assembly Project

The Arizona portion of the data assembly project fused data from the 2012 Arizona State University Travel Survey with a number of other datasets to produce a combined dataset of attitudes and residential locations. Since the ASU travel survey did not record addresses, there are no attributes of individual properties; all attributes are aggregated, mostly at the TAZ level. TAZs in the Phoenix area are relatively small, as shown in Figure 3.

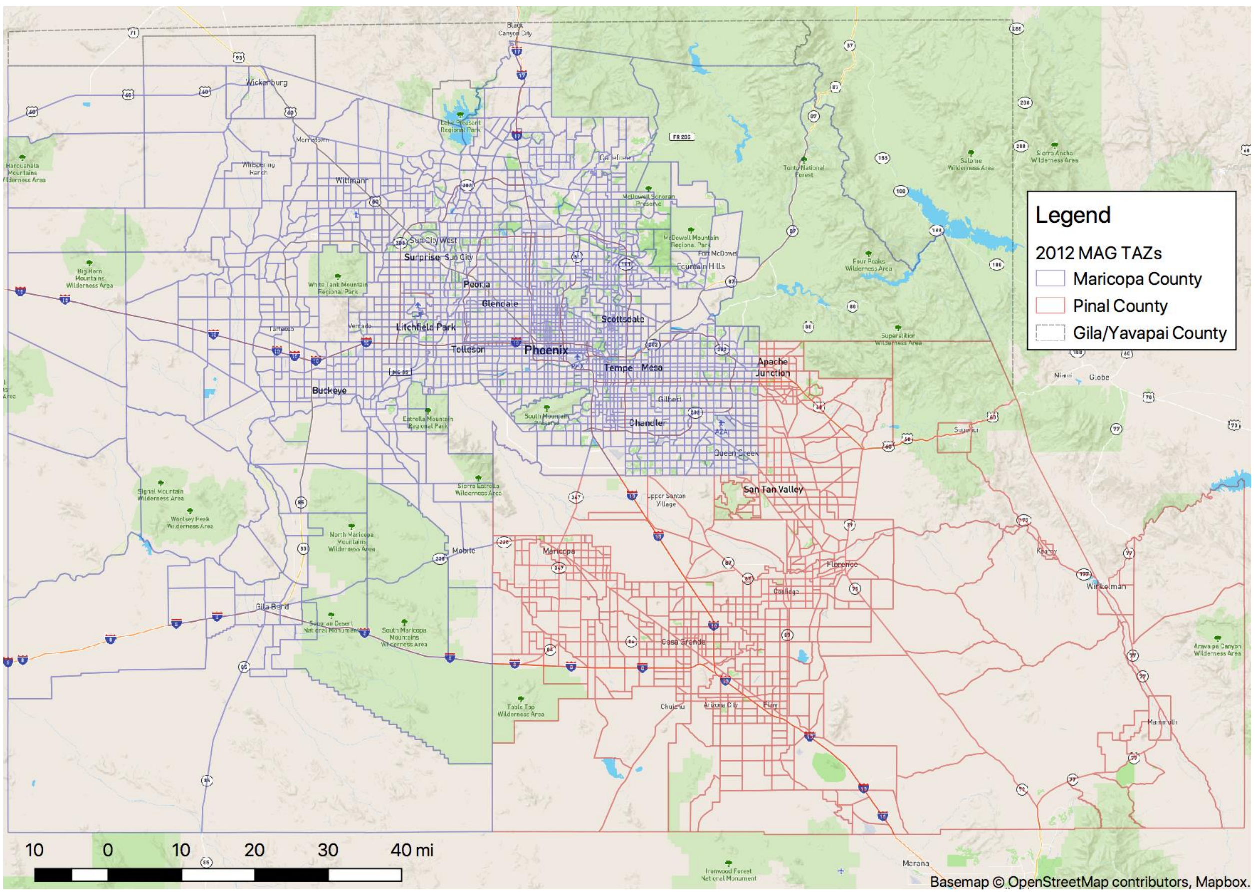


Figure 3: Phoenix-area Transportation Analysis Zones (TAZs)

The following datasets were compiled for this project and associated with the TAZs of the respondents to the ASU travel survey. Datasets marked with an asterisk (*) were only available for Maricopa County; approximately 2% of respondents live outside Maricopa County. Datasets marked with a † are open and publicly-available.

Table 1: Datasets to be fused with ASU Travel Survey

Dataset	Features extracted	Citation and URL
Access Across America: Transit 2014†	Job accessibility via transit	Owen and Levinson, 2014 http://dx.doi.org/10.13020/D6MW2Q
2010 Decennial Census, Maricopa and Pinal Counties†	Population density	https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html U.S. Census Bureau, 2010 and 2017
Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LEHD LODES)†	Employment density	https://lehd.ces.census.gov/ U.S. Census Bureau, 2012a and 2017a
Maricopa County 2012 Assessor Data*	Home characteristics for neighborhood	https://mcassessor.maricopa.gov/index.php Maricopa County Assessor, 2012
Maricopa County 2012 Land Use Data*	Land use	Maricopa County Planning and Development Department, 2012
TIGER/Line 2012†	Intersection Density	https://www.census.gov/geo/maps-data/data/tiger-line.html U.S. Census Bureau, 2012b
USGS Protected Areas Database, version 1.4†	Park access	https://gapanalysis.usgs.gov/padus/data/download/ U.S. Geological Survey, 2016

There is a data dictionary for the combined dataset included with this report, and the Stata version of the output contains extensive variable labels. The data processing steps for each dataset incorporated are presented below. All data processing was performed in Python version 3.6, using the Python Data Analysis Library (pandas), GeoPandas and Numpy. Jupyter Notebooks containing the code used are provided with the data.

Access Across America: Transit 2014

code file: Access Across America.ipynb

The Access Across America dataset (Owen and Levinson 2014) contains block-level job accessibility by transit for major metropolitan areas. The data presents the number of jobs accessible by transit, on average, from the centroid of each block during the morning commute (7:00 to 8:59). The data are presented at the block level; they are summarized to TAZs through a weighted averaging process. For the accessibility to jobs from the Home TAZ, the result is the average job accessibility from each block in the TAZ, weighted by the population of that block. If a block is partially covered by a TAZ, the weight is multiplied by the fraction of the block that is in the TAZ, assuming that the weighting variable (population in this case) is evenly distributed within the block. The accessibility to jobs from the work TAZ is calculated identically, except that the weighting variable is total employment in each block (from the LODES dataset). With these weights, the TAZ-level average accessibility reported represents the average accessibility experienced by a resident living in that TAZ, or a worker working in that TAZ, respectively. The accessibility variables are only presented for Maricopa and Pinal Counties (one respondent resided in Yavapai county, and no workers worked outside Maricopa or Pinal Counties).

Decennial Census and LODES data

code file: LODES.ipynb

Population density and employment density for each TAZ in Maricopa and Pinal counties were extracted from the 2010 Decennial Census table P001 and the 2012 LODES dataset. These data are presented at the block level. To summarize these data to the TAZ level, a similar averaging process was undertaken, where block-level population and employment density variables are averaged weighted by the area of the block. This is mathematically equivalent to the density in the entire TAZ, as long as one assumes that population density is constant across the TAZ. There is a proof of this in the source code.

Maricopa County 2012 Assessor Data

code files: Join Parcel Data.ipynb, Summarize Parcel Data.ipynb

Many property characteristics were extracted from the 2012 Maricopa County Assessor Rolls. Since the ASU Travel Survey does not have information on exact addresses, we instead summarize property characteristics at the TAZ level. Since many neighborhoods are relatively homogenous, this provides a potentially useful proxy for people's preferences. In order to allow data users to construct, as much as possible, the measures most appropriate for their studies, we have provided marginal distributions for all of the variables of interest from the assessor's data, as well as medians and the sample size of properties in that TAZ. Only properties of type Residential were considered.

There are 14 categories of assessed home full cash value, in $50,000 increments from $100,000 to $700,000. Each category contains the proportion of residential properties in the TAZ with an assessed full cash value in the relevant range. There is also a median value, and a variable for the count of properties that had a non-null assessed value.

Livable area, lot size, and year built are all summarized in the same way—as a marginal distribution, a median, and a count of non-null values. Lot size excludes condominiums and apartments (Maricopa County Property Use Codes 07 and 03, respectively).

There are also columns for Refrigeration and Pool, which represent, respectively, the proportion of residential properties in the relevant TAZ which have refrigeration (air conditioning) as a cooling type (alone or in combination with evaporative cooling), and the proportion which have a pool. As before, there is a variable for the number of properties that had this information—although in the case of pools, it is not possible to distinguish missing data from the lack of a pool.

Maricopa County Land Use Dataset

code file: Land Use.ipynb

The 2012 Maricopa County land use dataset was used to assess neighborhood land use. As with the assessor data, the land use data are summarized in a relatively disaggregate form, in order to facilitate the creation of many different land use metrics. There are 27 different land use categories in the source dataset. The source dataset is not aggregated; it is simply a Shapefile with features showing where particular land uses exist. These are aggregated to home and work TAZs, with one variable for each land use type containing the proportion of the TAZ that is of that land use—these can be aggregated and metrics such as land use entropy (Kockelman 1997) can be calculated. Note that the transportation land use includes railroads, arterials, etc., but does not contain local streets or parking lots.

TIGER/Line Data

code file: TIGER.ipynb

TIGER/Line data from the US Census Bureau is used to extract intersection density, which is a common metric for assessing the friendliness of the street network to pedestrians and bicyclists. TIGER/Line 2012 All Edge data for Maricopa and Pinal counties were retrieved, and all edges with MAF/TIGER Feature Class Codes of S1100, S1200, S1400, and S1640 (primary, secondary, local, and service roads) were extracted. Highway on-ramps and alleys were omitted. The TIGER data contains a node ID for each end of each line; any nodes with more than 3 (bidirectional) edges remaining after the filtering described previously were considered to be intersections for the purposes of this section.

Each TAZ was then buffered by 50 meters, and the density of intersections in this area was computed. Since TAZs are often defined by roads, this buffer was intended to avoid issues with intersections falling exactly on the edge of a TAZ being counted or not counted based on small rounding errors, thus potentially biasing the results.

USGS Protected Area Database

code file: Protected Areas Database.ipynb

The USGS Protected Areas Database (US Geological Survey 2016) identifies areas in the US that are protected from development through municipal ownership or otherwise. It is used to identify local parks and wilderness areas. There are variables in the output for the proportion of land within a 1-mile radius that is dedicated to local parks, and the proportion within a 5-mile radius that is wilderness area (under the assumption that people are willing to travel further to wilderness areas than to local parks).

The Protected Areas Database classifies protected areas into 15 categories within the MAG region based on the managing agency. “Local Park,” “State Historic or Cultural Area,” and “State Other”[1] were classified as local parks, while “National Public Lands,” “National Monument or Landmark,” “National Forest,” “Inventoried Roadless Area,” “Area of Critical Environmental Concern,” “Recreation Management Area,” “State Park,” “Conservation Area,” “State Conservation Area,” “Local Recreation Area,” “National Scenic or Historic Trail,” and “State Recreation Area” were classified as wilderness areas. Additionally, *all* areas larger than three square kilometers were classified as wilderness areas, to capture areas such as Phoenix’s South Mountain Park. Areas that were not classified as open access were excluded.

Preliminary Data Exploration

Figure 4 shows the results of an initial analysis of the data, and demonstrates some relationships between the attitudinal and travel-behavior information in the merged datasets. Figures 4(a) and (b) use the full sample of 12,011 respondents, while Figure 4(c) uses a smaller sample of 10,475 commuters.

---

[1] State Other had only 0.03 km of area within the MAG area, so misclassification within this catchall category is not a significant concern.

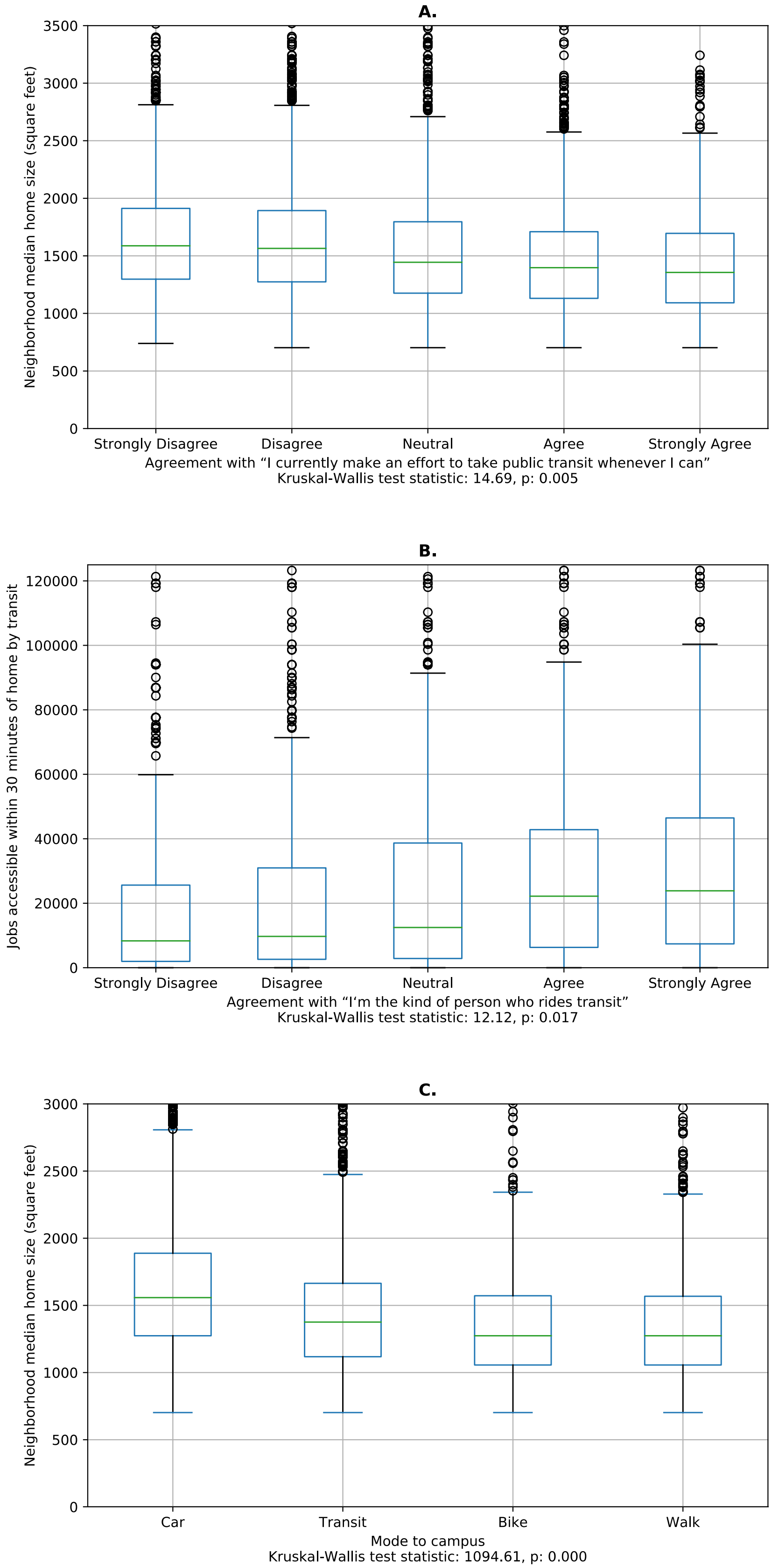

A.
3500
3000
2500
2000
1500
1000
500
0
Neighborhood median home size (square feet)
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree
Agreement with “I currently make an effort to take public transit whenever I can”
Kruskal-Wallis test statistic: 14.69, p: 0.005
B.
120000
100000
80000
60000
40000
20000
0
Jobs accessible within 30 minutes of home by transit
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree
Agreement with “I‘m the kind of person who rides transit”
Kruskal-Wallis test statistic: 12.12, p: 0.017
C.
3000
2500
2000
1500
1000
500
0
Neighborhood median home size (square feet)
Car
Transit
Bike
Walk
Mode to campus
Kruskal-Wallis test statistic: 1094.61, p: 0.000

Figure 4: Relationships between attitudes, residential choice, and travel, Arizona dataset

Figure 4(a) shows the relationship between residential location choices and intentions of traveling by public transportation. People who agree with the statement "I currently make an effort to ride transit whenever I can" are more likely to live in neighborhoods (TAZs) with lower median home sizes. While the distributions overlap, indicating a large amount of noise in this relationship, the Kruskal-Wallis test that at least one of the groups has a different median is significant.[2] This indicates a connection between neighborhoods and behavior. People who live in areas with smaller homes are likely to make more of an effort to ride transit. Attitudes can both affect and be affected by behavior (Kroesen, Handy, and Chorus 2017). Thus, this could be an effect of more convenient transit access in these neighborhoods (because they are likely older and more central), or it could be an effect of people predisposed to use transit self-selecting into neighborhoods more conducive to transit use.

Figure 4(b) shows the relationship between residential location choices and one's attitude towards public transportation as a mode of travel for oneself. The x axis shows agreement with the statement "I'm the kind of person who rides public transit," while the y axis shows the number of jobs accessible within 30 minutes of public transit travel. People who can see themselves riding transit are more likely to live in areas where more opportunities are accessible via transit. This relationship may be because people in high-accessibility areas experience better transit service, and can therefore see themselves using it; alternately, these people may be predisposed to choose to live in areas with better transit service. The Kruskal-Wallis test again indicates significant differences between the groups.

Finally, Figure 4(c) shows how these residential location choices can affect travel outcomes. It shows that people who drive to campus are more likely to live in neighborhoods with larger homes, transit users are slightly less likely to live in areas with larger homes, and people who bike and walk to campus are least likely to live in such areas (presumably because they live near campus). Thus, if a household had a preference for a larger home, they might find themselves needing to drive to access employment or education.

These results show that attitudes and residential location choices are related (Figures 4(a) and (b)), as are residential location choices and travel (Figure 4(c)). While both of these relationships have been discussed before, this combined dataset will allow us to consider them simultaneously. The models we will build in year 2 of the project will explore these relationships in more detail, and attempt to tease out how attitudes affect residential location, and how that in turn affects travel.

California Data Assembly Project

The 2015 California Millennials Survey has a rich set of attitudinal and travel behavior questions, making it a very valuable dataset for the analysis of the impact of attitudes on travel behavior. However, we believe that many travel behavior choices are moderated through residential location, which is itself affected by attitudes. In order to better understand the three-way relationship between attitudes, residential location choice, and travel behavior, we need a dataset that contains information about all three. The Millennials survey contains two of the three.

For a large subset of the Millennials dataset, respondents reported complete, valid addresses (the remainder reported only their cross streets, or had invalid addresses). Using these addresses, we can merge the millennials dataset with other datasets containing information about the respondent's precise residential location. Zillow's ZTRAX database (Zillow, 2018) is an ideal

[2] The Kruskal-Wallis test is used rather than the more common ANOVA to avoid the assumption that population standard deviations are constant across groups (SciPy Community, 2018).

dataset; it is a unified database of assessment records for every home in the United States. By joining the subset of the Millennials dataset with complete addresses to the ZTRAX dataset, we can supplement the attitudinal and travel behavior information in the Millennials dataset with information about the homes occupied by those respondents.

The only common key between the Millennials dataset and the ZTRAX dataset is the address itself. The addresses in the Millennials survey are relatively clean, as self-reported addresses go, but heuristic techniques still had to be used to allow them to match to the ZTRAX addresses.

In order to match the addresses, all addresses from both the ZTRAX and the Millennials dataset were expanded and tagged. Address expansion was done using the *libpostal* library; the *libpostal* library's expansion engine parses an address and returns many possible alternate spellings. For example, for the address 121 St. Paul St., Sacramento, CA 93116, *libpostal's* expansion engine returns:

- 121 SAINT PAUL SAINT SACRAMENTO CALIFORNIA 93116
- 121 SAINT PAUL SAINT SACRAMENTO CA 93116
- 121 SAINT PAUL STREET SACRAMENTO CALIFORNIA 93116
- 121 SAINT PAUL STREET SACRAMENTO CA 93116
- 121 STREET PAUL SAINT SACRAMENTO CALIFORNIA 93116
- 121 STREET PAUL SAINT SACRAMENTO CA 93116
- 121 STREET PAUL STREET SACRAMENTO CALIFORNIA 93116
- 121 STREET PAUL STREET SACRAMENTO CA 93116

While clearly these are not all valid alternate spellings, they contain many reasonable alternate spellings.

Each possible alternate spelling is then parsed using the Python library *usaddress*. This library identifies components of the address such as street number, street name, city, and ZIP code.

Prior to tagging, any expansion containing the letters 'san' when the original did not contain those letters is filtered out. The reason for this is that, in some cases, an S before a road name (e.g. S California Ave) is expanded to 'san' (e.g. San California Ave), and in some cases the tagging will then identify the name of the street as simply 'San.' This creates problems in the matching if this happens to multiple streets. For example, 111 S California Ave might match to 111 S Forest Ave, because both were expanded and tagged to 111 San. Excluding these incorrect expansions prevents the issue. (If the original, un-expanded name contained the letters 'san', for instance 203 San Antonio Road, expansions containing 'san' are not discarded).

Once all addresses have been expanded and parsed, they are matched using the parsed fields. In order to be considered a match, the ZIP code, street number, street name, and city name of two addresses must match exactly. To break ties, a weighting approach is used; matching unit numbers are weighted 100 points, and matching street types (e.g. avenue or road) and street directionals (e.g. N or South) are each weighted 50 points. The match with the highest weight is considered the unique match for that location.

Remember that for both a source address and a possible candidate match address, there are many expanded addresses. Two addresses are considered a match if any pair of their expansions meets the criteria defined above, and the weight of the match is defined as the maximum of the weight of any pair of expanded address. These are computed independently; the pair of expanded addresses where all required fields match exactly may not be the same as the pair of expanded addresses that yielded the highest weight.

The Millennials dataset has two address fields, one a cleaned version of the user input, and another a version that has been parsed by Google's geocoder. Both are used to create the expansions of addresses used in the matching, in order to maximize the match rate.

This process successfully matched approximately 80% of the addresses. The remaining addresses were matched by hand. These were predominantly addresses in multi-family developments (ranging from duplexes to large apartment complexes). Since the ZTRAX data is tabulated at the parcel level, the addresses of individual units within these developments do not necessarily match the address recorded for the parcel. By using a combination of Google Maps, county assessor maps, and human intuition, the vast majority of these records were matched to the correct ZTRAX record. Note that in the case of apartment complexes, there frequently is no unit-level information in the ZTRAX data, but the parcel record for the complex yields important information such as the number of units, parcel area (and thus density), and building type (e.g. high-rise, mid-rise, or low-rise).

Some addresses were automatically matched to multiple possible properties in the ZTRAX data. These were primarily condominiums; the correct match was chosen by the research team. In some cases, the respondent did not include the unit number for their condominium, or the unit number they reported was not in the ZTRAX data; in these cases, a random potential match was chosen. These cases are flagged in the final data so users can ensure they do not affect their analysis. Additionally, some properties matched two records for other, unknown reasons (possibly due to duplicate ZTRAX records); these are flagged as well.

As a final step, given the manageable sample size, each address match was given a "sanity check," in which each pair of matched addresses was displayed and the research team confirmed that the match appeared to be correct. The few addresses matched incorrectly by the computer were manually corrected.

The survey was also matched with a number of other neighborhood-level variables from other datasets. Respondents were matched to 2010 Census tracts (U.S. Census Bureau, 2018) and information on commute modes and the percentage of households with children was added from the 2017 5-year American Community Survey (U.S. Census Bureau 2017b, 2017c). Transit accessibility information was added from the 2015 Access Across America dataset (Owen, Levinson, and Murphy, 2017), and other neighborhood attributes were added from the EPA Smart Location Database (U.S. Environmental Protection Agency, 2013).

Preliminary Data Exploration

Figure 5 shows selected relationships from the California dataset. The data are similar to the Arizona data presented in Figure 4, with a few key differences. There are more attitudinal statements in the California survey. More importantly, since we have exact addresses in the California dataset, we have properties of the exact home from the Zillow dataset. However, the sample size is smaller; Figure 5 presents data from the 387 respondents who reported addresses that were successfully matched to Zillow records, and who lived in single family homes. Since the Zillow data often does not include detailed information about multifamily homes, they are excluded from this preliminary analysis; future analyses will include them when possible. Figure 5(c) includes only 211 commuters.

The panels of Figure 5 show the value of merging the attitudinal survey information with attributes about residential location. Panels (a) and (b) show how the attitudinal information in this data is related to residential choices, both at the home level and the neighborhood level. Figure

5(c) shows how these residential choices are related to travel choices.

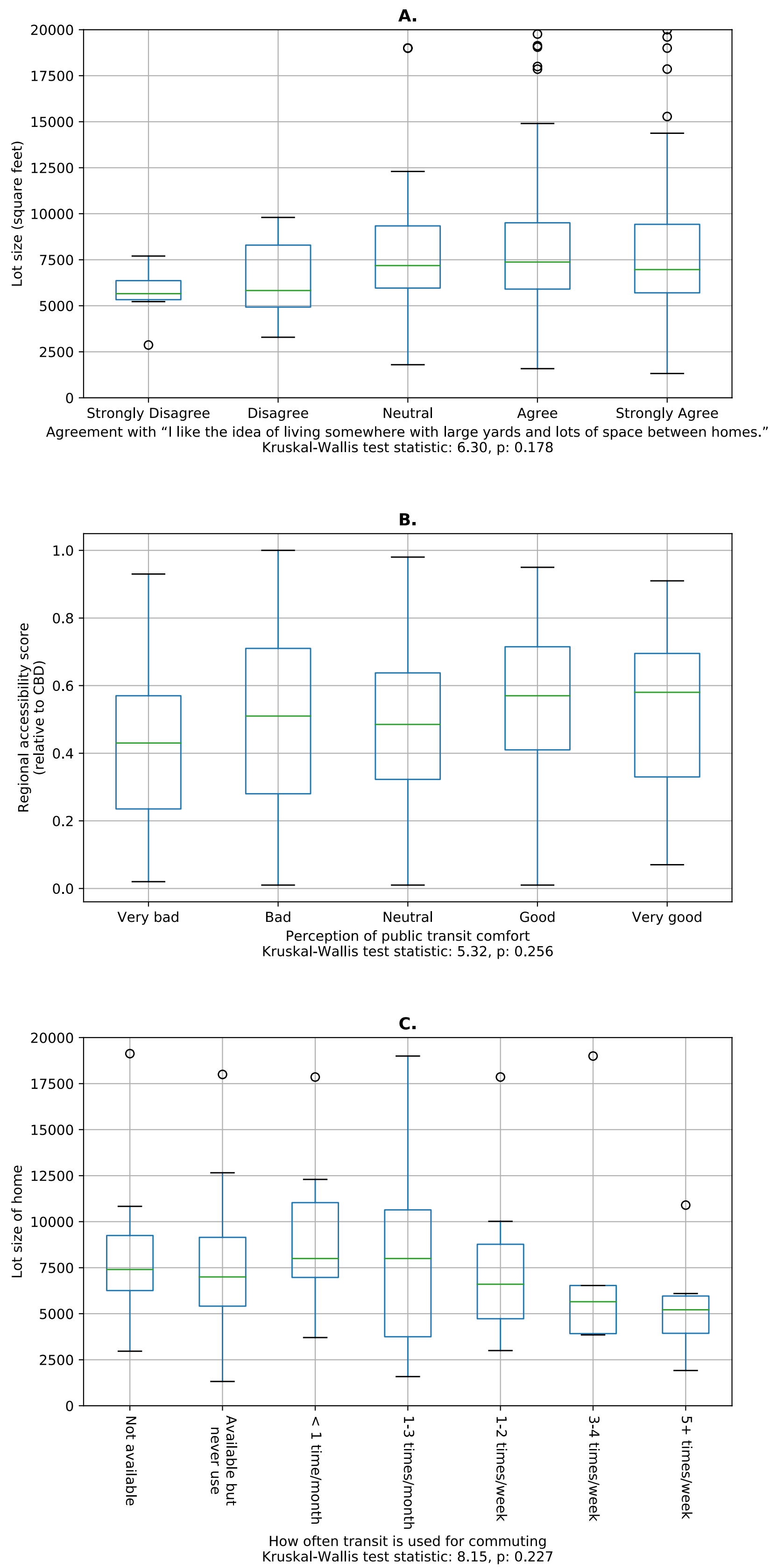

A.
Lot size (square feet)
20000
17500
15000
12500
10000
7500
5000
2500
0
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree
Agreement with "I like the idea of living somewhere with large yards and lots of space between homes."
Kruskal-Wallis test statistic: 6.30, p: 0.178
B.
Regional accessibility score (relative to CBD)
1.0
0.8
0.6
0.4
0.2
0.0
Very bad
Bad
Neutral
Good
Very good
Perception of public transit comfort
Kruskal-Wallis test statistic: 5.32, p: 0.256
C.
Lot size of home
20000
17500
15000
12500
10000
7500
5000
2500
0
Not available
Available but never use
< 1 time/month
1-3 times/month
1-2 times/week
3-4 times/week
5+ times/week
How often transit is used for commuting
Kruskal-Wallis test statistic: 8.15, p: 0.227

Figure 5: Relationship between attitudes, residential choice, and travel, California dataset

Figure 5(a) suggests that attitudes about residential location choice are related to residential choices (although the direction of the causality is not known). People who agree with the statement that they “like the idea of living somewhere with large yards and lots of space between homes” are, unsurprisingly, more likely to live in homes on larger lots. This is an analysis only possible with the California data, since the Arizona data does not have addresses that can be matched to exact property characteristics. The Kruskal-Wallis test for this difference is not significant, however, meaning we cannot conclude that this difference is not due to sampling error. This is potentially due to the small sample size.

Figure 5(b) suggests that people who find public transit more comfortable may be more likely to live in more central areas, as measured by accessibility to jobs from their home location relative to the maximum accessibility to jobs from any location in the metro area. Accessibility to jobs via auto was used since accessibility to jobs via transit is not available in the EPA Smart Location Database for many of the areas in the sample. Again, this could be due to people predisposed to liking public transport choosing to live in more central areas, or people who live in more central areas becoming accustomed to higher-quality public transport service. As before, the Kruskal-Wallis test is not significant, possibly due to small sample size.

Finally, Figure 5(c) suggests that the relationship between attitudes towards large lots and residing in homes on large lots may have travel behavior implications. It shows that people who use transit for commuting more often are less likely to live in homes with large lots. Again, though, the Kruskal-Wallis test is not significant.

NEXT STEPS

This project has succeeded in fusing multiple datasets at different spatial scales together with attitudinal travel survey data in two states. The next step is to analyze these data using multivariate statistical models to test the extent to which attitudes relate to residential location choices, travel choices, or both. Concurrently with this data fusion project, the leads on this work have been participating in the development of a new TOMNET-funded survey that will allow them to go even further with this research approach.

REFERENCES

Kockelman, K. (1997). “Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area.” *Transportation Research Record 1607*:116–125. doi:10.3141/1607-16.

Kroesen, M., Handy, S.L., & Chorus, C. (2017). Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. *Transportation Research Part A, 101*, 190–202. https://doi.org/10.1016/j.tra.2017.05.013

Maricopa County Assessor. 2012. Tax assessment database.

Maricopa County Planning and Development Department. 2012. Land Use Database.

Owen, A., & Levinson, D.M. (2014). Access Across America: Transit 2014 Data. Retrieved from the Data Repository for the University of Minnesota, http://dx.doi.org/10.13020/D6MW2Q.

Owen, A., Levinson, D.M., & Murphy, B. (2017). Access Across America: Transit 2015 Data. Retrieved from the Data Repository for the University of Minnesota, https://doi.org/10.13020/D63G6F.

SciPy Community. (2019). scipy.stats.f_oneway. In: *Scipy v1.2.1 Reference Guide,*

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html. Accessed 16 Apr 2019.
U.S. Census Bureau. 2010. Decennial Census, table P001, block level. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
U.S. Census Bureau. 2012a. Longitudinal Origin-Destination Employment Statistics, Arizona. https://lehd.ces.census.gov/
U.S. Census Bureau. 2012b. All lines, TIGER/Line shapefiles. https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2012&layergroup=All+Lines
U.S. Census Bureau. 2017a. 2010 Census blocks, TIGER/Line shapefile. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
U.S. Census Bureau. 2017b. 2013-2017 5-year American Community Survey, table B08006 (commute mode). https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
U.S. Census Bureau. 2017c. 2013-2017 5-year American Community Survey, table B11005 (presence of children). https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
U.S. Census Bureau. 2018. 2010 Census tracts, TIGER/Line shapefile. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
U.S. Environmental Protection Agency. (2013). Smart location database, version 2.0. https://www.epa.gov/smartgrowth/smart-location-mapping#SLD
U.S. Geological Survey, Gap Analysis Program (GAP). May 2016. Protected Areas Database of the United States (PAD-US), version 1.4 Combined Feature Class.
Zillow. (2018). Zillow Transaction and Assessment database (ZTRAX). https://www.zillow.com/research/ztrax/