Transportation Research Board Conference on Applying Census Data for Transportation, Kansas City, Missouri

Understanding the Role and Relevance of the Census in a Changing Transportation Data Landscape

Gregory D. Erhardt[1] & Adam Dennett[2]

[1] Department of Civil Engineering, University of Kentucky
261 Oliver H. Raymond Bldg., Lexington, KY 40506, USA, greg.erhardt@uky.edu

[2] Centre for Advanced Spatial Analysis (CASA), University College London (UCL)
90 Tottenham Court Road, London, W1T 4TJ, United Kingdom, a.dennett@ucl.ac.uk

Submitted: August 21, 2017

Revised and Resubmitted: October 6, 2017

Abstract

The data landscape is changing in terms of both data availability and the demands for new and more types of data. New data sources such as mobile devices, GPS and sensor data expand the possibilities of data collection and analysis. Using a review of recent literature as a starting point, this paper explores how Census data relate to these emerging and evolving data sets for transportation planning applications. It identifies areas where one or the other is used more commonly, and areas where they are complimentary, and finds that the Census data remain relevant, especially for the demographic and socioeconomic context they provide and for their universal availability.

The paper goes on to consider the prospects for keeping the Census data relevant to transportation planning, in the face of challenges such as the changing nature of mobility and of work, as well as opportunities to expand the role and relevance of Census data. It considers the results of a recent evaluation of the future of the United Kingdom (UK) Census and the overlap of the issues faced by the US Census. The paper considers strategies to be considered for keeping the Census relevant, which are offered as a range of visions that the Census could take. The authors provide a clear recommendation on one of the strategies considered, recommending against the "Give up and go home" strategy, and urging the Census Bureau, transportation planning organizations, and universities to continue their historic role of providing data as a public resource.

Keywords: Census, American Community Survey (ACS), Census Transportation Planning Package (CTPP), Longitudinal Employment-Household Dynamics (LEHD), Journey to Work, Big Data, Mobile Phone Data, Passively Collected Data, Transportation Planning

1. Introduction

The US Census has long been an important data source for transportation planning and forecasting. The population and housing data provide the basis for populating traffic analysis zones (TAZs); demographic and socioeconomic data are used to understand the effects of transportation projects on different populations; journey-to-work (JTW) data provide insight into commute patterns, mode shares and the demand for transportation; and the longitudinal employment-household dynamics (LEHD) data provide consistent estimates of employment throughout the US.

Transportation planning also has a long history of leveraging other data as a complement to the Census, including household travel surveys, traffic counts, transit ridership counts, state employment records and local land use data. More recently, a new generation of data have come online, and transportation planners have started developing methods to capture and use these so-called Big Data.

Big Data include a range of sources that are typically passively collected, meaning that they emanate from sensors, transactions or administrative records without the need for an active response on the part of the participant. In transportation, these include data such as transit automated vehicle location and automated passenger count data; transit fare card transactions; electronic toll transponder transactions; GPS traces from commercial vehicle movements; and trip tables derived from mobile phone data. These data offer several advantages over traditional travel surveys and Census Data, including potentially much larger sample sizes, potential cost savings, and the ability to better measure changes due to their continuous nature. Big Data, however, brings its own set of challenges and limitations. Of note are the fact that the biases inherent in the data are often unknown, and that the data often excludes contextual information, such as demographics and socio-economics, that can be included in an active data collection scheme. For these reasons, and due to the relative immaturity of the Big Data field, Smith (2013) argues for a hybrid approach that draws from the best aspects of each, while Johnson and Smith (2017) suggest that Big Data is best viewed as a supplement to, not a substitute for traditional surveys.

This paper examines the relationship between Census Data and emerging Big Data sources in the context of transportation planning, and considers the ways in which they serve as substitutes versus complements. It does this through a semi-structured literature review that identifies recent transportation planning papers and articles that reference either the Census or Big Data. The search reveals both overlapping and non-overlapping topic areas, indicating some potential for competition versus complementarity in those topic areas. A subset of the literature is reviewed in more detail to better understand the uses and limitations of each type of data.

The US Census is not unique in facing the emergence of new data and technology—other nations are faced with similar issues and opportunities. This paper reports the recommendations of a recent effort to modernize the United Kingdom Census, and considers the relevance of those recommendations to the US.

The paper goes on to consider some key policy questions of the future, and how the existing Census data structure fits or does not fit with those questions.

Given this three-tiered foundation, a menu of options is offered for keeping the Census relevant to transportation planning. These options are segmented into a competition track and a complementarity track. With a single exception, the authors refrain from recommending a path forward, and instead offer the options with the hope of stimulating a debate about the future of the Census.

2. Emerging Data Sources and Their Relationship to the Census

To identify areas of overlap and non-overlap between the uses of Census data and Big Data, we conducted a semi-structured review to identify relevant literature.

The TRID database was used as the search engine. TRID combines the records from the Transportation Research Board's (TRB's) Transportation Research Information Services (TRIS) Database and the OECD's Joint Transport Research Centre's International Transport Research Documentation (ITRD) Database, providing an extensive database focused specifically on transportation research. The search was limited to articles and papers, published in English, within the Planning and Forecasting subject area. The date range was from 2008 through August 2017. Papers focusing on research conducted outside the United States are included in an effort to learn from the international experience.

Two separate searches were conducted, one for the keyword "census", and one for the key word "big data". The Census search returned 513 articles and the Big Data search returned 232 articles. A third search, for "census" and "big data" returned only five articles, constituting a subset of both. While it would be possible to expand the results by searching for specific types of data—such as "mobile phone" or "GPS"—the 232 articles retrieved provides a sufficient basis for identifying the themes discussed in this paper.

2.1 Keyword Analysis

To get a sense of the topic areas that are prominent in the research, the key words from each of the 740 (513 + 232 - 5) articles returned from either search were tabulated. Supplemental Table 1 shows the frequency of each keyword in the Census search and in the Big Data search. Only the 253 keywords (out of 1,727 total keywords) used by more than five articles are shown. Each keyword is categorized as high frequency or low frequency for each search, with high frequency defined as being used by more than five papers in that set of search results. This grouping allows us to identify which keywords have a high frequency in both searches, in just the Census search, in just the Big Data search, or in neither search. Those without a high frequency in either search are of little interest and are not examined further.

Table 1 shows the number of articles returned for each year in the searches. It is clear that Big Data is a recent trend, with few articles published prior to 2014, but the numbers growing to be on-par with the number of Census articles by 2015. The number of Census articles also grows during this period, indicating that Big Data research is not necessarily detracting from research that uses Census data.

To get a sense of the topic areas that are prominent in the research, the key words from each of the 740 (513 + 232 - 5) articles returned from either search were tabulated. Supplemental Table 1 shows the frequency of each keyword in the Census search and in the Big Data search. Only the 253 keywords (out of 1,727 total keywords) used by more than five articles are shown. Each keyword is categorized as high frequency or low frequency for each search, with high frequency defined as being used by more than five papers in that set of search results. This grouping allows us to identify which keywords have a high frequency in both searches, in just the Census search, in just the Big Data search, or in neither search. Those without a high frequency in either search are of little interest and are not examined further.

Table 1: Articles by year for "Census" and "Big Data" search terms

Year	"Census"	"Big Data"
2017	35	33
2016	68	57
2015	57	59
2014	60	35
2013	60	11
2012	55	11
2011	52	4
2010	42	9
2009	46	8
2008	38	5
Total	513	232

Table 2 shows the keywords that occur with high frequency in both searches, sorted by the total frequency. The keywords show a number of terms indicating a range of applications relevant to transportation planning, travel forecasting and travel behavior analysis, traffic and transit. These are areas where there is potential for Big Data to serve as a substitute for Census data, although the mere presence of the terms in both searches does not necessarily indicate that it is a substitute. It could also be that each is used for different specific applications, or each is used in a complementary way.

Table 3 shows the Census dominant keywords. These are keywords that occur frequently in articles within the Census search, but infrequently in articles within the Big Data Search. For parsimony, only the top 40 are shown. The top keyword in this group is "traffic counts". An inspection of the papers using this keyword reveals that they are traffic-related, but not obviously Census related. It appears that either there is an anomaly in the coding, or that these articles use the term in a different context.

The remaining keywords in the Census dominant group are all more logical, and correspond to obvious applications of Census data. "Commuting", "work trips" and "commuters" all refer to analysis using the journey-to-work data. "Demographics", "Socioeconomic factors" and "Equity (Justice)" all use data that are available in the American Community Survey (ACS) or the Census long form. This is important because a characteristic of Big Data is that while they often provide detailed trajectory information, they usually lack characteristics of the individual or the household. Therefore, the Census remains the best source of this information. A number of terms that also show up relate to land use and the built environment ("Land use", "Neighborhoods", "Land use planning", "Residential location"), highlighting another area where the Census shines. A fourth theme that can be observed is several terms relate to non-motorized travel ("Bicycling", "Bicycles", "Walking", "Nonmotorized transportation"). This may be due to the limitations of Big Data in capturing non-motorized travel—people do not (yet!) have sensors built into their bodies that allow them to be directly tracked, and mode inference from GPS traces remains difficult, although inroads are starting to be made in this area (Bolbol et al. 2012), and more recently by technology start-ups such as TravalAi in the UK.

Table 2: Keywords with a high frequency in both searches

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Travel demand	84	21	105	High	High
2	Origin and destination	74	19	93	High	High
3	Data collection	46	39	85	High	High
4	Travel behavior	62	19	81	High	High
5	Public transit	57	19	76	High	High
6	Travel surveys	55	10	65	High	High
7	Mode choice	50	9	59	High	High
8	Case studies	34	22	56	High	High
9	Urban areas	44	11	55	High	High
10	Transportation planning	34	17	51	High	High
11	Travel time	29	16	45	High	High
12	Data analysis	15	30	45	High	High
13	Traffic data	24	20	44	High	High
14	Mobility	25	18	43	High	High
15	Geographic information systems	36	7	43	High	High
16	Travel patterns	26	16	42	High	High
17	Planning	36	6	42	High	High
18	Traffic flow	28	12	40	High	High
19	Traffic models	20	13	33	High	High
20	Traffic volume	27	6	33	High	High
21	Traffic congestion	16	14	30	High	High
22	Forecasting	22	8	30	High	High
23	Algorithms	13	15	28	High	High
24	Traffic forecasting	18	8	26	High	High
25	Global Positioning System	10	13	23	High	High
26	Choice models	15	8	23	High	High
27	Freight transportation	14	9	23	High	High
28	Vehicle sharing	15	7	22	High	High
29	Simulation	13	9	22	High	High
30	Ridership	12	9	21	High	High
31	Optimization	10	10	20	High	High
32	Decision making	8	12	20	High	High
33	Sustainable development	12	7	19	High	High
34	Infrastructure	11	6	17	High	High
35	Traffic simulation	9	7	16	High	High
36	Route choice	8	8	16	High	High
37	New York (New York)	8	6	14	High	High
38	Urban transportation	8	6	14	High	High
39	Sustainable transportation	8	6	14	High	High

Table 3: Census dominant keywords (top 40)

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Traffic counts	147	0	147	High	Low
2	Commuting	52	1	53	High	Low
3	Demographics	49	2	51	High	Low
4	Socioeconomic factors	47	2	49	High	Low
5	Spatial analysis	41	5	46	High	Low
6	Accessibility	36	4	40	High	Low
7	Land use	39	1	40	High	Low
8	Households	33	3	36	High	Low
9	Work trips	33	1	34	High	Low
10	Mathematical models	30	3	33	High	Low
11	Bicycling	27	4	31	High	Low
12	Traffic estimation	25	4	29	High	Low
13	Census	29	0	29	High	Low
14	Neighborhoods	27	1	28	High	Low
15	Commuters	23	4	27	High	Low
16	Automobile ownership	24	3	27	High	Low
17	United States	22	4	26	High	Low
18	City planning	20	5	25	High	Low
19	Walking	23	2	25	High	Low
20	Surveys	19	4	23	High	Low
21	Modal split	20	3	23	High	Low
22	Microsimulation	18	4	22	High	Low
23	Trip generation	20	2	22	High	Low
24	Canada	20	2	22	High	Low
25	Land use planning	19	2	21	High	Low
26	Nonmotorized transportation	21	0	21	High	Low
27	Activity choices	15	4	19	High	Low
28	Metropolitan areas	17	2	19	High	Low
29	Annual average daily traffic	18	1	19	High	Low
30	Demand	13	5	18	High	Low
31	Trip matrices	14	4	18	High	Low
32	Estimation theory	16	2	18	High	Low
33	Residential location	16	2	18	High	Low
34	Equity (Justice)	17	1	18	High	Low
35	Location	17	1	18	High	Low
36	Regression analysis	17	1	18	High	Low
37	Methodology	12	5	17	High	Low
38	Bicycles	15	2	17	High	Low
39	Statistical analysis	13	3	16	High	Low
40	Networks	10	5	15	High	Low

Table 4 shows the Big Data dominant keywords. There are a more limited number of these, and several are general terms ("Big Data", "Data mining", "Information processing", Technological innovations"). "Cellular telephones", "Smartphones" and "Smart cards" refer to specific types of data that are increasingly common. The applications in this group ("Intelligent transportation systems", "Real time information", "Logistics", "Supply chain management") are distinct from the other groups and are more operational or logistical in nature.

Finally, it is interesting to note that "China" and the "Netherlands" are in the Big Data dominant group, whereas the "United States" and "Canada" are in the Census dominant group. This may reflect clusters of research, but it also may relate to the quality and availability of Census data in those countries.

Table 4: Big Data dominant keywords

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Big data	2	42	44	Low	High
2	Intelligent transportation systems	2	26	28	Low	High
3	Data mining	5	14	19	Low	High
4	China	2	15	17	Low	High
5	Logistics	4	11	15	Low	High
6	Real time information	3	11	14	Low	High
7	Cellular telephones	5	8	13	Low	High
8	Information processing	5	6	11	Low	High
9	Smartphones	3	8	11	Low	High
10	Smart cards	3	7	10	Low	High
11	High speed rail	2	7	9	Low	High
12	Technological innovations	2	7	9	Low	High
13	Netherlands	2	6	8	Low	High
14	Supply chain management	0	6	6	Low	High

While the keywords provide an overview of the themes in each category, they provide little depth. To better understand the applications and uses of Census data and Big Data, the most frequent keywords within each group were examined in more detail. "Data collection" and "Big Data" were excluded from this exercise as not meaningful in this context, and "Socioeconomic factors" was excluded because it is similar to "Demographics", which was already included. For each keyword considered, a sub-search was conducted for articles using that keyword. The titles and abstracts of articles in the sub-search were examined, and a single paper was selected to illustrate a theme from that sub-search. Each of those papers is reviewed here in further detail.

2.2 Overlapping Topic Areas

Table 5 shows a summary of the articles reviewed for the top keywords with a high frequency in both the Census and Big Data searches. 10 articles are included—one for Census and one for Big Data with each keyword considered. The table shows the search terms, the author and year, the title, the full set of keywords used by that paper, the types of data used, and some brief notes.

Table 5: Summary of selected papers for top keywords with high frequency in both searches

Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Travel demand & Census	Yasmin, Morency, and Roorda 2017	Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model	Activity based models, Activity choices, Montreal (Canada), Origin and destination, Travel demand, Validation	OD survey, Canadian Census	Transfers TASHA from Toronto to Montreal. OD & census provide validation data.
Travel demand & Big Data	Huntsinger 2017	The Lure of Big Data: Evaluating the Efficacy of Mobile Phone Data for Travel Model Validation	Big data, Cost effectiveness, Data analysis, Data collection, Data quality, Households, Mobile telephones, Travel demand, Travel surveys, Validation	Mobile phone data (Airsage), HH travel survey	Airsage only available at district-level, but good for district-to-district flows. Proprietary nature makes it hard to evaluate.
Origin and destination & Census	Çolak, Alexander, Alvim, Mehndiratta, et al. 2015	Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities	Boston (Massachusetts), Cellular telephones, Origin and destination, Rio de Janeiro, Brazil, Traffic data, Travel behavior, Trip purpose	Mobile phone data (raw), Census, HH survey, OD survey.	Mobile phone data processed into OD matrices & expanded to Census, validated against surveys. Worked reasonably well.
Origin and destination & Big Data	Allos et al. 2014	New Data Sources and Data Fusion	Bluetooth technology, Data files, Data fusion, Global Positioning System, Origin and destination, Smartphones, Trip matrices	GPS data (Traffic Master), mobile phone data (Telefonica)	Passive data lacks segmentation and potentially biased, but big/complete sample size.
Travel behavior & Census	Jacques and El-Geneidy 2014	Does travel behavior matter in defining urban form? A quantitative analysis characterizing distinct areas within a region	Census tracts, Characterization, Factor-cluster analysis, Travel behavior, Urban form	Canadian Census, GIS land-use, OD survey, satellite images	Census provides housing & household measures.
Travel behavior & Big Data	Chen et al. 2016	The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis	Big data, Cooperation, Data files, Disciplines, Mobility, Transportation planning, Travel behavior	Mobile phone data (raw), Big Data in general	Scaling factors needed. Imputing modes is hard. Not clear what to validate against. Representativeness unclear. Longitudinal nature is an advantage.
Public transit & Census	T. Wang, Lu, and Reddy 2013	Maintaining Key Services While Retaining Core Values: NYC Transit's Environmental Justice Strategies	Census Transportation Planning Package, Costs, Environmental justice, Factor analysis, Impacts, Level of service, New York City Transit Authority, Public transit, Routes, Service changes, Social values, Transportation operations	2000 Census JTW, Census racial & income counts, trip planner (route schedules)	Evaluate equity of proposed service cuts.
Public transit & Big Data	Oort and Cats 2015	Improving Public Transport Decision Making, Planning and Operations by Using Big Data: Cases from Sweden and the Netherlands	Case studies, Data sources, Decision making, Netherlands, Planning, Public transit, Smart cards, Sweden, Transit vehicle operations, Vehicle positioning systems	Transit smartcard data, automated vehicle location (AVL) data, automated passenger count (APC) data	Illustrates range of applications: planning, operations, ridership prediction, real-time information. Promise in combining data sources.
Travel surveys & Census	Clark et al. 2014	Life Events and Travel Behavior	Aged, Bicycling, Commuting, Travel behavior, United Kingdom, Urban areas	UK Household Longitudinal Study, UK Census	Longitudinal data overcomes many estimation limitations.
Travel surveys & Big Data	Vij and Shankari 2015	When Is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis	Data files, Data quality, Errors, Global Positioning System, San Francisco (California), Statistical inference, Travel demand, Travel diaries, Travel surveys	HH travel survey, GPS based travel survey	Higher volume of GPS data is often offset by lower quality due to limits of inferring mode, purpose, etc.

Within the sub-search on travel demand and Census, a number of the papers focus on travel demand model validation, followed by population synthesis, cycling and origin-destination (OD) matrix estimation. Yasmin, Morency, and Roorda (2017) transfer the TASHA activity-based travel demand model from Toronto to Montreal, and use a combination of OD survey data and Canadian Census data to validate the transferred model. This aligns with our own experience using US Census journey-to-work data and auto ownership data to validate travel models.

The sub-search on travel demand and Big Data includes substantial topical overlap with the travel demand and Census search. Core topics include using Big Data to validate travel models and for OD matrix estimation, as well as one paper demonstrating the use of Big Data to estimate travel models. Huntsinger (2017) evaluates the effectiveness of Airsage mobile phone data for validating travel models. She compares the data to a household travel survey for the same region. The comparison is necessary because the proprietary (Black-Box) nature of Airsage makes it difficult to evaluate otherwise. The data comes in the form of district-to-district trip tables. It lacks the detailed travel characteristics and demographics of the survey, but due to the large sample size excels in the role of providing district-to-district flows.

Papers with an origin and destination keyword focus on generating OD trip matrices from a variety of sources, including mobile phone data, Bluetooth data, and transit farecard data. Many of the papers within the Census sub-search also use one of these Big Data sources. Çolak et al. (2015) present the methods for generating trip tables from mobile phone data, without purchasing a vendor's trip table directly. The method requires expanding the mobile phone data to the Census, and validating against surveys, showing that even with Big Data, the Census remains a crucial data product.

Allos et al. (2014) examine the process of creating OD matrices from GPS traces and mobile phone data in the UK. They report that the passive data provides a big/complete sample size, but lacks segmentation by purpose or income and is potentially biased. The potential for bias is important, with other research showing a transit smartcard data set to be biased against low-income and minority travelers, which can be problematic from an equity standpoint (Erhardt 2016b).

The travel behavior articles are more diverse. Within the Census sub-search, urban form and transit-oriented development are a common theme. Car sharing and activity patterns also come up repeatedly. The Big Data and travel behavior sub-search includes several conceptual papers on how Big Data can be used and some on tracing travel patterns. Jacques and El-Geneidy (2014) study the effects of different urban forms, using the Canadian Census, among other sources. Chen et al. (2016) offer a review of Big Data applications, arguing for stronger collaboration between traditional transportation planners and computer scientists and physicists doing Big Data research. Their review highlights several advantages and limitations of Big Data, noting that imputing modes is difficult, the representativeness of the data is unclear, and it is not clear what to validate against. On the other hand, the longitudinal nature of Big Data offers a clear advantage that is often not available in traditional data.

For the public transit and Census sub-search, commuting, accessibility and environmental justice emerged as core themes. Wang, Lu, and Reddy (2013) demonstrate a method of evaluating the equity of proposed service cuts using transit schedule data in combination with the Census journey-to-work.

The public transit and Big Data papers were split between the use of smart card data, conceptual papers on the value of Big Data, and approaches to imputing modes and walk distances. Oort and Cats (2015) illustrate a range of applications using smartcard data, automated vehicle location (AVL) data and automated passenger count (APC) data. They note that the greatest promises of Big Data lie in combining multiple data sources.

The travel surveys and Census sub-search largely includes methodology papers for how to conduct travel surveys and analysis of travel survey data. An interesting application considers the effect of life events on travel behavior, using the UK Census and the UK Household Longitudinal Study (Clark et al. 2014). This paper demonstrates how longitudinal data can be used to overcome some of the limitations of cross-sectional data, such as self-selection bias and co-linearity among certain variables.

The travel surveys and Big Data sub-search includes papers that discuss the strengths and limitations of Big Data and their value for travel model validation. Vij and Shankari (2015) examine GPS-only household travel surveys where mode, purpose and other attributes are imputed from the GPS traces, in comparison to travel surveys that ask for those attributes explicitly. They find:

> "In many cases, gains in the volume of data that can potentially be retrieved using GPS devices are found to be offset by the loss in quality caused by inaccuracies in inference. This study makes the argument that passively collected GPS-based surveys may never entirely replace surveys that require active interaction with study participants."

2.3 Census Dominant Topic Areas

Table 6 shows the papers reviewed within the Census dominant topic areas. In this table, only papers from the Census sub-search are included.

Traffic counts was the most frequent keyword in the Census search. The papers within the traffic counts sub-search include a number of bicycle related papers, as well as some about estimated annual average daily traffic (AADT) and others about OD matrix estimation. The relevance to the Census is not immediately obvious for many, indicating either a possible anomaly in the keyword coding or an alternative use of the term census. For example, the reviewed paper relates short term bicycle counts to continuous bicycle counts for the purpose of estimating annual average daily bicycle traffic. While it does not use Census data, it is relevant with respect to the expansion of Census bicycle commute mode shares to annual totals.

Commuting is a common application of Census data, split between an analysis of mode shares and commuter patterns. Wang (2017) presents an interesting example that considers cohort changes in commute mode shares using the integrated public-use microdata sample (IPUMS). The research demonstrates that it is valuable to be able to match data sets across time in a consistent format and with consistent data fields. Likewise, several national-level studies show up in this sub-search, highlighting that it is important to have consistent data across cities. This was also a theme to emerge from a recent workshop on the future of travel forecasting (Walker 2017): that in order to advance our knowledge as a field we need data and models that are developed across multiple cities.

Both demographics and socioeconomic factors are common keywords within the Census dominant group. The demographics keyword includes papers on aging populations, spatial distributions, equity and car sharing. Tyndall (2017) illustrates several of these by studying the equity of car sharing, with respect to the demographics of the neighborhoods where the cars are located. They use a Big Data source from the car share company to identify the car locations, but rely on Census data to understand the neighborhood demographics.

Table 6: Summary of selected papers for Census dominant keywords

Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Traffic counts & Census	El Esawey 2016	Toward a Better Estimation of Annual Average Daily Bicycle Traffic	Adjustment factors, Bicycle traffic, Bicycles, Traffic counts, Traffic estimation	Automated bicycle counters (inductive loops).	Does not use Census data. Relevant to expansion of JTW bike mode shares.
Commuting & Census	X. Wang 2017	Peak Car in the Car Capital? Double-Cohort Analysis for Commute Mode Choice in Los Angeles County, California, Using Census and ACS Microdata	American Community Survey, Carpools, Census, Cohort analysis, Commuting, Demographics, Forecasting, Los Angeles County (California), Microdata, Mode choice, Public Use Microdata Sample, Single occupant vehicles	Integrated PUMS from 2000 Census and 2009-2011 ACS.	Demographic data is important, as is the ability to match across multiple data sets for trend and cohort analysis.
Demographics & Census	Tyndall 2017	Where No Cars Go: Free-Floating Carshare and Inequality of Access	Demographics, Equity (Justice), Free-floating carsharing, Location, Mobility, Mode choice, Urban areas, Vehicle sharing	Carshare location data (Car2Go), ACS.	Big Data tells half the story, and is referenced to ACS demographics to understand equality considerations.
Spatial analysis & Census	Liu, Roberts, and Sioshansi 2017	Spatial Effects on Hybrid Electric Vehicle Adoption	Adoption models, Demographics, Hybrid vehicles, Neighborhoods, Peer groups, Spatial analysis, Spatial effects	Census, ACS, Ohio vehicle registration data.	Spatial distribution of demographic and socioeconomic factors is important.
Accessibility & Census	Owen and Levinson 2017	Developing a Comprehensive US Transit Accessibility Database	Accessibility, Alachua County (Florida), Geographic information systems, Methodology, Transportation disadvantaged persons	GTFS, LEHD	Accessibility is an increasingly important performance measure. Value in national consistency and availability of LEHD.

A number of papers also use Census data to study spatial effects. Often this applies to electric vehicles, urban form, or neighborhood characteristics. Liu, Roberts, and Sioshansi (2017) consider spatial effects on the adoption of hybrid electric vehicles, using a combination of Census, ACS and state vehicle registration data.

Accessibility is becoming an increasingly important performance metric. Accessibility measures the ease of access to destinations, as opposed mobility which measures the ease of movement. Owen and Levinson (2017) develop a comprehensive transit accessibility database. They use the general transit feed specification (GTFS) for transit schedules, and the LEHD as a spatially detailed measure of employment.

2.4 Big Data Dominant Topic Areas

Table 7 shows a summary of the papers reviewed within the Big Data dominant topic areas.

The most common keyword among the Big Data dominant topics is intelligent transportation systems. The papers in this area are largely focused on operational applications and on methods development. Xiao, Liu, and Wang (2015) develop a platform that combines a range of freeway-related data for performance management and operational analysis.

Table 7: Summary of selected papers for Big Data dominant keywords

Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Intelligent transportation systems & Big Data	Xiao, Liu, and Wang 2015	Data-Driven Geospatial-Enabled Transportation Platform for Freeway Performance Analysis	Data analysis, Data sharing, Freeways, Geospatial analysis, Performance measurement, Statistical analysis	Roadway geometric data, loop detector data, Bluetooth data, INRIX speed data, incident data, weather data, freeway travel time	Largely operational applications, and for performance management.
Data mining & Big Data	Zhang, Zhan, and Yu 2017	Car Sales Analysis Based on the Application of Big Data	Automobile industry, Automobile ownership, Big data, Data analysis, Information processing, Manufacturing, Sales	Scraped car sale data and reviews.	Aimed at providing insight to car makers.
China & Big Data	Hao, Zhu, and Zhong 2015	The Rise of Big Data on Urban Studies and Planning Practices in China: Review and Open Research Issues	Big Data, China, review, urban planning, urban studies	GPS, mobile phone data, smart card data, points of interests, volunteered geographic information, search engine data, digital land use data, parcel data, road networks.	Chinese language papers more likely to focus on plan making and management applications than English language papers.
Logistics & Big Data	Coyle, Ruamsook, and Symon 2016	Weatherproofing Supply Chains: Enable Intelligent Preparedness with Data Analytics	Data analysis, Logistics, Supply chain management, Weather conditions, Weatherproofing	50 year weather database, daily retail sales data by store	Ensure products are on shelves when storm hits. Applications from DOT or emergency management perspective are reasonable.
Real time information & Big Data	Fusco, Colombaroni, and Isaenko 2016	Short-Term Speed Predictions Exploiting Big Data on Large Urban Road Networks	Bayes' theorem, Floating car data, Mathematical prediction, Networks, Neural networks, Rome (Italy), Speed prediction models, Time series analysis, Traffic models, Urban highways	Floating car data (GPS), network.	Short-term operational focus.

Data mining shows up frequently as well, and the papers are often focused on mining a specific data set. One example uses scraped car sales data to provide insight to car makers (Zhang, Zhan, and Yu 2017).

China is among the top keywords in the Big Data dominant search, with the papers showing a range of applications including for transit, traffic, high speed rail and methods, as well as a wide range of data sets. Hao, Zhu, and Zhong (2015) provide an extensive review of Big Data applications in planning practice in China. It is recommended reading for anyone who wants a good overview of the range of applications of Big Data to planning. They note that Chinese language papers are more likely to focus on plan making and plan management than English language papers. It is interesting to consider why that may be—it could be a different research focus, that China lacks the same availability of other data sets, or that there are institutional differences in the planning structure that make Big Data more relevant.

Papers with the logistics keyword generally focused on supply chains, freight transportation or railroads. Coyle, Ruamsook, and Symon (2016), for example, considers the issue of delivering adequate supplies to stores prior to a coming storm.

Papers with the real time information keyword are generally about traffic flow, speed predictions or methodological developments. Fusco, Colombaroni, and Isaenko (2016) use GPS floating car data for short-term traffic predictions.

2.5 Common Themes and Observations

Several themes and observations emerge from the above review:

- There is substantial overlap between the use of Census data and the use of Big Data. The greatest overlap occurs in areas related to transportation modeling and public transit. Often, Census data and Big Data are used in combination, with the Census serving as a basis for expansion, or providing demographic and socioeconomic information.
- Census data remain the dominant source of demographic and socioeconomic information, as well as a widely available and widely used source of commute data.
- Big Data dominant topics tend to focus on shorter-term operational, traffic, and logistics issues.
- Due to their large sample sizes, Big Data also excel as the basis for generating OD matrices.
- Big Data tend to be much less rich than survey data or Census data in terms of information content per observation. They generally lack information on demographics, household composition, trip purpose, mode, etc.
- The methods for inferring mode, purpose and other attributes from GPS or mobile phone traces remains weak, and the errors can offset the value of the additional observations.
- The quality of Big Data and the biases inherent in those data are often unknown and difficult to assess. This is especially true when commercial data are purchased, since the methods used in processing those data are often proprietary. This makes it especially important to have some external data source that they can be expanded to or validated against.
- Longitudinal data can overcome important limitations of cross-sectional data sets and open up new applications.
- The availability of Big Data remains sporadic, and even as they become more widely available, there is a risk that "data monopolies" will result in high prices (Erhardt, Batty, and

Arcaute 2018). In contrast, the Census remains a widely available public resource, and the consistency across cities is important to allowing larger-scale analyses.

3. Beyond 2011: The Future of the UK Census

In the United Kingdom, the decennial Census (which is actually comprised of three separate censuses with some country-specific questions asked in England & Wales, Scotland and Northern Ireland and separate statistical authorities governing the collection and dissemination of the data) has captured information on the residential and workplace addresses of respondents since 1921 (Office for National Statistics 2012). From this locational information, estimates of the journey-to-work have been derived and are available to access in digital form as origin-destination matrices dating back to 1981 (UK Data Service 2017).

Information on the journey to work is derived from the home address of census respondent and then, historically, a question relating to their place of usual work. In 2011, journey-to-work statistics were joined by 'journey-to-learn' statistics relating to students and their location of educational establishment.

One of the major advantages of the Census travel-to-work data over any other measurement of commuting (apart from it being free to use and open) is its coverage. It is a legal requirement to complete a Census return in the UK and in 2011 a national 94% response rate was achieved (Office for National Statistics 2017c), meaning that even before estimation and imputation, nearly all geographic and demographic dimensions of the population were covered. This is clearly a significant benefit to anyone using the data for travel-to-work analysis, as volumes and close to the full range of origins and destinations are well represented. Taking advantage of this feature of the data, for a number of decades now, Travel to Work Areas (TTWAs) have been defined using these flow data for the purpose of local labor market analysis and statistical reporting leading to policy decisions made by the Department for Work and Pensions in relation to out of work benefits.

Clearly, however, Census travel to work data is not without its issues. Aside from the well-established issues such as errors in recording peripatetic working / other irregular travel to work patterns and timeliness; origin-destination data contain no routing information, detail on modal shifts and reveal little about other important travel activities not associated with work (such as shopping, 'school runs' and leisure). All of this means that researchers are starting to explore the potential of other datasets in conjunction with Census data to enhance our understanding of travel patterns.

Work is underway to determine whether detailed route and mode data captured continuously from a mobile application, can be used to validate modelled detailed journal estimates using Census origin-destination data (Innovate UK 2017). The smartphone application, TravelAi (http://www.travelai.info/), provides recommended routing across travel modes, but also monitors the location of the user to provide that data to transportation agencies. The Office for National Statistics in England and Wales are also actively looking at the potential of other mobile telephone related data for mobility- transportation research (Office for National Statistics 2017b). They propose to evaluate the comparability of flows derived from mobile telephone data and those estimated from the Census, however this is no indication of whether any headway has been made with this as yet.

The last Census in 2011 cost the UK government around £480 million to run (Office for National Statistics 2017a), which despite being a very low cost per capita over the 10-year lifespan of the data, contributed the opening of a conversation on whether the Census is still value for money or even necessary in a world where alternative population data exist amongst the myriad of administrative, commercial and survey datasets now in existence. There is no constitutional requirement for a Census to take place in the UK and the 'Beyond 2011' programme explored the potential for replacing all of part of the Census using these data sources, as well as other options such as short-form and rolling censuses.

After an extensive research and consultation period, the National Statistician recommended that the 2021 Census would be a 'full' census, however the data collection methods would be entirely online (Office for National Statistics 2014). This approach eliminates the need to post paper forms out to households – the feature of previous censuses which had created the most cost. The National Statistician also recommended that the 2021 Census feature an increased use of administrative data and surveys to enhance statistics from the Census and improve statistics between censuses. The report recommended against an approach that eliminated the Census and instead used only administrative data to construct population statistics. While other countries successfully use such an approach, those countries have a population register, which the UK does not. A population register is a centralized data system for recording, and keeping current, vital statistics for all residents of a country (United Nations Statistical Office 2014). Such registers are common in northern Europe, with the vital statistics recorded typically including births, deaths, marriages, name changes and other changes of interest. Assuming it is accurate, a population register would make the Census function of counting people unnecessary because the register contains that count, although address and other attributes may or may not be recorded. The administrative data approach was viewed as a risky endeavor without a population register. The government accepted the recommendation, but expressed interest in moving towards an administrative approach in the future (Maude 2014).

4. The Policy Questions of Tomorrow

When planning on 10 year Census timeframes, it is valuable to consider not just competing and complementary data sources, but also how the relevant policy questions may change over those timeframes. This section discusses policy areas that should be on Census planners' minds. It does not suggest that these issues are resolved, or will definitively come to be—just that they are questions worth grappling with.

4.1 The Future of Mobility

The past several years have seen both the rise of new shared mobility modes, and massive investment in developing the technology of self-driving cars.

Over the past decade, advances in payment and smartphone technology have enabled new uses for old transportation modes. The literature review above has already identified car sharing as a mode of interest (Tyndall 2017), but bike sharing systems have proliferated as well (Shaheen et al. 2012). The option to share vehicles has the potential to re-shape decisions about owning a vehicle and the demand for parking (Martin, Shaheen, and Lidicker 2010).

Transportation Network Companies (TNCs), such as Uber and Lyft, also represent a re-invention of an old mode. TNCs allow a user to book and pay for a ride with a smartphone app, with the ride delivered by an independent driver in their personal vehicle. At current rates, the cost to the user is generally much lower than a taxi, and some drivers prefer the convenience of the app and payment system. They did not exist a decade ago (Uber was founded in 2009), but they are no longer a niche mode, at least in major cities. In San Francisco, for example, TNCs make over 170,000 vehicle trips within the city, which is approximately 12 times the number of taxi trips, and 15% of all intra-San Francisco vehicle trips (San Francisco Count Transportation Authority 2017).

TRB Special Report 319 identifies, but does not resolve, many of the policy questions related to shared mobility and technology-enabled transportation services (Transportation Research Board 2016b). Among these are questions of regulation, safety and security, the impact on congestion and transit ridership, equity of access and the effects on the labor market.

In the future, drivers themselves may become unnecessary. Both technology companies and traditional auto makers are investing billions of dollars in developing self-driving cars or autonomous vehicles. The prospects and timeframe for broad adoption of the technology remain uncertain (Litman 2014; Bansal and Kockelman 2017; Rohr et al. 2016), but the implications for the transportation system and transportation policy are profound (Fagnant and Kockelman 2015; Anderson et al. 2014). The effects depend in part on how they are used. Will households replace their personal vehicles with self-driving cars? Will they be used as fleet vehicles by TNCs? Perhaps they will first become common for freight transportation, as opposed to personal travel? These are important questions that transportation planners must grapple with, and as the technology emerges, it is important to have the data to understand these trends.

4.2 The Future of Work

The future of mobility highlights issues related to the future of work that extend beyond transportation.

Arguably, TNCs biggest innovations have happened not in transportation, per se, but in the labor market. Special Report 319 (Transportation Research Board 2016b) considers these employment and labor issues. Drivers are not treated as employees, but as independent contractors who own and maintain their own vehicles, pay for their own health insurance and manage their own payroll/self-employment taxes. This represents an important shift from a traditional employer-employee relationship, with looser ties between the two. There are implications not only on the levels of net compensation, but also brings potential for less regularity of working hours, lower stability of employment, a higher share of part-time works, an increased ability to engage in multiple jobs, and a decreased stability of employment. It is easy to see how these trends may extend beyond transportation to a wide range of jobs, and it is sometimes referred to as the "gig economy". From a transportation perspective, such a situation is very different than commuting to regular shift work.

Self-driving cars and trucks may have an even bigger impact on labor markets. According to the Bureau of Labor Statistics, the US has 1.8 million heavy truck drivers, 1.3 million delivery truck drivers, 665,000 bus drivers, and 233,000 taxi and chauffer drivers. As self-driving vehicles emerge, it is logical to expect that these workers will be displaced, that the cost to consumers of delivering goods is reduced, and that the firms that own the vehicles see their profit margins increase. These trends are likely to increase income and wealth inequality in the US. As drivers are pushed out of regular employment, they may also engage in the gig economy, accentuating the trends discussed above.

It easy to dismiss such concerns as speculation, and future employment is indeed difficult to forecast, but Vardi (2017) argues that the future is already here. He notes the combination of high manufacturing output with low manufacturing employment and stagnant wages over the past several decades. While it is difficult to pinpoint the exact reasons for such trends, increasing automation is likely a contributing factor.

While Vardi uses an example of the shift from horse-powered transportation to automobiles a century ago, a better analogy may be the rise of containerization 40 years ago. Containerization dramatically reduced the labor involved in shipping, greatly reducing its cost. Beyond the direct labor market implications, this contributed to the rise of global trade, a major shift in the nature of our nations ports, and the re-purposing of waterfront areas and entire neighborhoods in many cities.

The shifting nature of work and shifting mobility options may also contribute to regional disparities in several dimensions. While TNCs and bike sharing systems are popular in large cities, they are most effective when combined with a certain level of density. It is easy to envision fleets of autonomous cars shuttling people around Pittsburg (as Uber is doing today) or San Francisco, but their market may be more limited in the smaller cities in Kentucky where auto ownership is higher and the distances are greater. Changes in labor markets and employment are likely to be geographically uneven, and there is evidence that people are less likely to move to follow jobs than in the past (Cooke 2013; Molloy, Smith, and Wozniak 2017).

4.3 Long-Distance Travel: A Policy Question of Today

Rather than a policy question of tomorrow, accommodating the demand for long-distance travel is a commonly overlooked policy question of today. In the United States, personal vehicle trips longer than 50 miles account for 2% of total trips, but 23% of vehicle miles traveled (VMT). While a precise estimate of resources is not available, it is clear that long-distance travel commands far less than 23% of the effort involved in transportation planning, data and forecasting. In spite of the fact that there appears to be a renewed call for spending billions of dollars on inter-city high-speed rail every few years, and the huge portion of our roadway system dedicated to intercity travel, the data and resources available for long-distance planning are woefully inadequate, as illustrated by the reliance of recent long-distance models on either the 2001-2002 National Household Travel Survey (NHTS), or the 1995 American Travel Survey (ATS) (Moeckel, Fussell, and Donnelly 2015; Outwater et al. 2015).

The TRB Executive Committee recognized this deficiency and commissioned Special Report 320 on Interregional Travel: A New Perspective for Policy Making (Transportation Research Board 2016a). Two of the reports key findings are especially noteworthy here:

> "Because of outdated travel behavior survey data, long-distance travel is not nearly as well understood as local travel."
>
> "To encourage the development of urban transportation systems that are integrated and function well across a metropolitan region, the federal government has long required state

> and local authorities to coordinate their urban highway and transit investments. The goal of this coordination, which is often challenging to implement, is to guide transportation investments from a multimodal and multijurisdictional perspective that is informed by sound data and objective analysis. Because interregional travel corridors often span multiple states, many lack the coordinated planning and funding structures needed to ensure that investments in transportation capacity are made from a corridor-level perspective."

In other words, there is both a lack of reliable data on the topic, and a challenge in overcoming the institutional and jurisdictional coordination problems associated with investing in new data.

5. Options for Keeping the Census Relevant

In this final section, we consider several options for keeping the Census data relevant to transportation planning. These are grouped into two general tracks: the competition track considers strategies where the Census data is directly competing against other data sources, while the complementary track considers strategies associated with identifying a unique niche for the Census to fill. The first three strategies constitute the competition track, while strategies four through seven are on the complementary track.

Strategy 1: Give up and go home

Strategy 1 is based on the premise that emerging Big Data sources are becoming so good and so cheap that they are making the Census obsolete. This represents a vision of the future (or the present) where technology is so omnipresent that our every movement is recorded in a database, where it is linked to every credit card purchase we have ever made, every social media comment we have ever posted, and a facial-recognition database of every photo we have ever been in. This is a vision of total knowledge, where it is unnecessary to ask about travel behavior because we already know the answers.

In such a future, the Census may very well become obsolete. Even in a world that only partially approximates this vision, it may seem a reasonable strategy to decide that the Census is irrelevant and therefore the prudent course of action is to give up and go home. There are two problems with this strategy.

First, it is clear from the literature review above that, regardless of grand visions for where the world may be heading, we are not nearly at the point where Big Data can be considered "all knowing." The Big Data studies identified above are limited in scope to specific applications and specific geographies. They often have limitations and biases that arise from the way the data are collected, such as the tendency of transit smart card data to underrepresent minority and low-income travelers (Erhardt 2016b). Those biases and limitations can be difficult to detect and evaluate, especially when the methods are not fully transparent (Huntsinger 2017), and those data limitations can easily offset the value of a larger sample size (Vij and Shankari 2015). For these reasons, it is common for Big Data to be used in combination with Census data or other actively collected data, as illustrated by many of the studies cited above.

Second, to the extent that such a vision of the future is viable, it is much closer to reality in the private sector than it is for transportation planners shaping public infrastructure, services and policy.

Technology companies are in a position to invest heavily in acquiring data resources and the computing infrastructure necessary to support them, and to hire talented engineers and computer scientists. They also operate with a different set of political and legal constraints than the public sector—what may be viewed as inappropriate government intrusion in Washington might be perfectly acceptable in Silicon Valley. Will the role of transport planners be that of a customer purchasing these data? Or will it be to work with the companies providing transportation (via self-driving vehicles) to develop optimization strategies to regulate traffic volumes along routes so that congestion is avoided and efficiency maximized?

While these are reasonable and appropriate roles for transportation planners to play, there are risks in limiting the planning role in this way. One such risk is the danger of a "data monopolies" (Erhardt, Batty, and Arcaute 2018). A data monopoly can occur when a single company has exclusive rights to all the data of a certain type or on a certain topic. In such cases, that company can exert control over the price, at the expense of those purchasing the data.

A second issue is that private sector interests may or may not align with the interest of serving the public good. Shuldiner and Shuldiner (2013) consider how the public interest can be best served when the transportation data of greatest value is collected by private entities, and how the current situation differs from the historic development of transportation planning models based on public data. If the data show a picture of the real world that is inconsistent with a company's public image or corporate strategy, what incentive do they have to share those data? The Freedom of Information Act would not apply, so a company would be within its rights to filter the data that it releases.

The current experience with TNCs illustrates the types of issues that can arise. Uber has been in conflict with multiple cities over regulatory issues, most recently resulting in a Transport for London's (TfL's) decision ending its ability to operate in London (Rao and Isaac 2017). From its own operations, Uber has extensive data about travel in London that may be useful to planners at TfL, but it is not realistic to expect Uber to provide those data to planners at TfL while it appeals TfL's decision in court, nor is it realistic to expect planners at TfL to trust those data should they be made available.

Going beyond appropriate restrictions to protect privacy, which all good data stewards have an obligation to uphold, do we really want to put ourselves in a position where private interests can control and filter the data that shapes our understanding of the world? It is precisely to avoid this situation that there will always be a role for data as a public resource, and the authors urge the Census Bureau to continue its historic role providing this resource.

Strategy 2: Keep calm and carry on

The second strategy considered is for the Census to continue its transportation data program in its current state, a strategy we label "Keep calm and carry on." The rationale for this strategy is that the review of research studies show that the Census is clearly continuing to play a role in transportation planning. In particular, it has an important role in providing context with respect to household characteristics, socioeconomic and demographic characteristics, and it is often use in combination with Big Data as a basis for expanding or supplementing those data. The fact that it is universally available as a public data resource ensures that a wide variety of actors can each conduct independent analyses using these data, contributing to a diversity of ideas and viewpoints, and a rich environment for innovation.

This strategy may be combined with some minor adjustments to the existing approach. For example, the UK Census includes questions on the journey-to-school in addition to the journey-to-work, and the American Census could benefit from the same. This would provide planners with more complete travel information, particularly in locations where colleges or universities are major attractors, such as Arizona State University which contributes a substantial portion of ridership to Phoenix's light rail line (Federal Transit Administration 2013).

It may also be beneficial to add questions designed to provide consistency with external data sources. For example, the Census asks about usual place of work and usual mode to work, whereas most travel surveys record the destination and mode of work commutes for a designated travel day. This makes it difficult to compare the data between the two, and can be particularly important when reflecting the variability of travel, particularly for something like a bicycle commute which can be affected by the weather (Nosal et al. 2015).

If such questions are added, they should be supplemental to, not in place of, the existing journey-to-work questions. Consistency with past Census and ACS data is important to ensure that trends can be monitored cleanly.

Strategy 3: If you can't beat 'em, buy 'em

The third strategy considered on the competition track is labeled, "If you can't beat 'em, buy 'em." The goal here is to use the relative advantages of similar data sources to get a more complete picture of the journey to work.

Currently, the main advantage of mobile phone data is that the large sample size provides a strong basis for creating trip tables at a reasonable level of geographic detail. In contrast, the ACS journey to work data becomes noisy for more detailed geographies simply because there are a limited number of observations. The ACS data, however, provide more information than mobile phone trip tables, such as the usual mode to work, and characteristics of the workers.

This strategy would involve purchasing mobile phone data for a region, specifically focusing on work commutes, which are expected to be the most reliable purpose that can be extracted due to their regularity. These data would be compared in detail to the Census journey to work data, and the expansion factors would be adjusted for each to create a unified, best-estimate trip table. It is expected that this approach would be most effective if the adjustments could be made on disaggregate data, and then released as aggregate trip tables to protect privacy. Such an approach would require an appropriate licensing arrangement with the mobile phone data vendor, and if Census restricted data were to be used, it would need to be conducted in an established secure data center.

There are a few possible paths toward making this happen. One is for the Census bureau to do the analysis and expansion on their end, and then release it as part of the JTW data products. Alternatively, an arrangement can be made where the data vendors better incorporate the Census data into their own products. A third option would be to do the analysis as a post-processing step, starting from both sources. This third option could be done as a pilot test for a single region.

A more sophisticated approach would be to manage the integration as part of the data collection process, rather than after-the-fact. For example, when the ACS surveys a household, the questionnaire could ask for permission to access the mobile phone records for individuals in that household. Those data would be combined with the survey to improve the data quality. Alternatively, the mobile phone records could serve as a sampling frame for the survey, allowing for integration in that direction. The privacy elements of such an approach would need to be carefully managed.

Strategy 4: Administrative integration

The "Administrative integration" strategy draws from the future envisioned for the UK Census to integrate appropriate administrative data sets for the purpose of improving the Census. Already, the Census is doing this through its Longitudinal Employment-Household Dynamics (LEHD) data product that integrates unemployment insurance records, tax records, and other data to create spatially detailed estimates of employment, workers, and commute flows.

The types of data integrated could be expanded in several directions. Feeney et al. (2015) provide a useful overview of the types of administrative data that researchers have used in the past, and could potentially be integrated with Census transportation data. Some promising options include:

- Birth and death records for monitoring population changes.
- School enrollment data, both for primary schools and for colleges and universities. Most institutions can be expected to have address lists of their students which would be useful for developing journey to school matrices.
- School districting data, which may be valuable for restricting the journey to school matrices based on district boundaries.
- Incarceration records, representing a portion of the group quarters population that does not travel.
- Social security administration data, which can be used both as a means of merging data across multiple sources and as a means of linking age, income and retirement status.
- State vehicle registration data, as a means of linking auto ownership information.
- Utility records, particularly power usage data which could potentially be used to identify when a unit is occupied either seasonally or by time-of-day.
- Parcel data from county assessors' offices, which are already public, as a means of integrating land use.
- Transaction data from toll transponders, transit farecards, and similar transportation transactions.
- Credit report information, which is widely available, and could be used to infer information about income, housing tenure and vehicle ownership.
- Credit card transaction data, which may provide information on the location and type of purchases.

For some of these data, the value contributed may be outweighed by privacy concerns or by the trouble of compiling the data. It is, however, worth being deliberate in assessing that trade-off.

For the administrative integration strategy, the role of the Census Bureau (or another agency that took on the task), would be that of a data aggregator. As envisioned, it would:

1. Gather disaggregate data from multiple jurisdictions.
2. Code the data to be as consistent as possible across jurisdictions, and merge them into a unified data set.
3. Link those data across types. For example, vehicle registration data could be linked to utility usage data, and parcel records to improve the estimates of car ownership currently included in the ACS.
4. Clean and check the unified data.
5. Aggregate them in such a way as to protect the privacy of individual records.

The data aggregator is able to add value both by working with disaggregate records, but keeping those records hidden behind a firewall, and by ensuring consistency across regions allowing for larger scale analyses.

Strategy 5: Capture the future

The "Capture the future" strategy is aimed at adjustments to the Census journey to work data collection to better reflect current, and possible future trends in mobility and work.

The key change needed for capturing the future of mobility is simply to expand the list of modes included in the journey-to-work questionnaire. Already, the change in travel modes is prominent, at least in certain cities. For example, between 2005 and 2015, the ACS show that in San Francisco, the share of work commutes by taxi, bike and other modes more than doubled, from 3.4% to 6.9% (Erhardt 2016a). This represents a combination of what has been called a "bicycle renaissance" and (Pucher, Buehler, and Seinen 2011) and the emergence of TNCs, which as of 2016 composed 15% of intra-San Francisco vehicle trips (San Francisco Count Transportation Authority 2017). While still small shares relative to other modes, these are important trends in their own right. It would be valuable to split the other category to explicitly consider TNCs, or at least to clarify that they are included in the taxi category. Moving to the future, it would be valuable to break out autonomous modes, both in the commute mode choice questions and in vehicle ownership. This would be most effective ahead of the trend, such that the annual ACS data be used to monitor trends in those modes.

For the future of work, the key issue is how to account for informal and irregular work. Options here include the option to collect more than one workplace, with a usual mode associated with each, as well as further clarifying the definition of work. It may be that respondents have different understandings of whether a "gig" should be reported as work, leading to ambiguity in the responses. However it is counted, there is value in consistency.

5.6 Strategy 6: Go long (distance)

The "Go long (distance)" strategy deviates from the focus on work commutes and considers an important, but neglected, travel market—long distance travel. Because it spans state and municipal boundaries, it is important that long distance travel data be collected at a national level. Extending the Census transportation data offerings could be a natural way to accomplish this.

Such a survey approach would likely be a retrospective question asking respondents to list long-distance trips made by members of their household in the last month. While the definition offered for long-distance trips can vary—often 50+ miles, sometimes 100+ miles—defining the question based on overnight trips would define a clean breakpoint for respondents in terms of identifying and remembering those trips. The information collected could be very simple and would include:

- Destination, recorded with city/county, state and ZIP code,
- Mode of travel,
- Departure and return dates, and
- Purpose: business versus leisure.

Such a data set, collected across the country for a reasonably large sample size, would be a tremendous resource for this important component of travel demand.

Strategy 7: Go long(itudinal)

The general lack of longitudinal data has been recognized as a limitation of transportation research for nearly 30 years (Kitamura 1990). This is a problem because cross-sectional correlations among different variables can make it difficult to detect the effects of certain policy interventions or other changes. For example, a time-of-day model might wish to consider the effect of congestion on changes in the temporal distribution of trips. A model estimated from cross-sectional data would likely find that congestion is higher in the peak period, and people prefer to travel in the peak period, so more congestion would lead to a higher likelihood of traveling in the peak period. Of course, the directionality of this assessment is wrong, but the model estimation cannot distinguish that. Conversely, if longitudinal data were available where the same households were observed in subsequent years, the data and resulting models would correctly show that an increase in congestion between those two years would make the travelers in that household less likely to travel in the peak period.

There are a range of other examples that can be used to illustrate this effect, but the issue broadly is that our interest as transportation planners extends beyond describing the state of the system as it is today. Our interest in transportation data is also in understanding the factors that cause the system to change, and applying that understanding to predict how the system will change in response to our interventions. For this purpose, cross-sectional data that does not observe change is inherently limited.

As discussed above, Big Data do offer some advantage in this area. Because they tend to be continuously collected, they provide an opportunity to measure change, which can be leveraged to measure the impacts of transportation projects (Erhardt 2016a).

The American Community Survey could evolve into a panel survey, where a portion of the households are re-surveyed in subsequent years. The German Mobility Panel has taken this approach since 1994 (Weiss et al. 2017). In Germany, this approach has enabled a range of applications and analyses that otherwise would be difficult or impossible, such as assessing the individual-level stability in commute patterns (Hilgert et al. 2016) and studying the effect of life changes on travel behavior (Scheiner, Chatterjee, and Heinen 2016). Together these provide a means of understanding the levers that can be most effectively used to induce changes in travel behavior.

6. Recommendations and Next Steps

This paper has found that in spite of the emergence of a variety of Big Data sources, the Census remains relevant to transportation planning. The paper considered the types of applications where one or another is more commonly applied, and found a large area of overlap where the two are used together as complementary data sources, even in studies that are labeled as "Big Data" studies. In spite of this relevance, the Census faces challenges in maximizing its relevance and value for transportation applications going forward, and these challenges are not unique to the American context. They include a natural desire for cost effectiveness, and the evolving nature of mobility and work. There are also opportunities, such as the dearth of long-distance and longitudinal data where the Census is in a position where it could step up to provide important resources.

Seven strategies are considered for keeping the Census data relevant to transportation planning. Three consider the Census' role in direct competition with Big Data, and four consider the ways in which it could be more complementary.

- Strategy 1: Give up and go home
- Strategy 2: Keep calm and carry on
- Strategy 3: If you can't beat 'em, buy 'em
- Strategy 4: Administrative integration
- Strategy 5: Capture the future
- Strategy 6: Go long (distance)
- Strategy 7: Go long(itudinal)

Among the strategies considered, this paper provides a clear and strong recommendation on only the first. **We strongly recommend against the "Give up and go home" strategy, and we urge the Census Bureau, transportation planning organizations, and universities to continue their historic role of providing data as a public resource.**

The remaining strategies are intended to provide a menu of options, which are not necessarily mutually exclusive. They will serve as a starting point for discussion at the Transportation Research Board Conference on Applying Census Data for Transportation in Kansas City, Missouri in November 2017. The authors hope that that discussion will continue in the broader community as we renew our effort to keep the Census relevant and valuable for transportation planning purposes.

Acknowledgments

Thank you to Penelope Weinberger, Erik Sabina, Ed Weiner and other reviewers for their contributions and feedback. Thank you to the conference organizers and American Association of State Highway and Transportation Officials (AASHTO) for sponsoring this research.

References

Allos, Anan, Andrew Merrall, Robert Smithies, Rebecca Fishburn, John Bates, and Roger Himlin. 2014. "New Data Sources and Data Fusion." In European Transport Conference 2014, Association for European Transport (AET).

Anderson, John Michael, Nidhi Kalra, Karlyn D. Stanley, Paul Sorensen, Constantine Samaras, and Oluwatobi A. Oluwatola. 2014. *Autonomous Vehicle Technology: A Guide for Policymakers*.

Bansal, Prateek, and Kara M. Kockelman. 2017. "Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies." *Transportation Research Part A: Policy and Practice* 95 (January): 49–63. doi:10.1016/j.tra.2016.10.013.

Bolbol, Adel, Tao Cheng, Ioannis Tsapakis, and James Haworth. 2012. "Inferring Hybrid Transportation Modes from Sparse GPS Data Using a Moving Window SVM Classification." *Computers, Environment and Urban Systems*, Special Issue: Advances in Geocomputation, 36 (6): 526–37. doi:10.1016/j.compenvurbsys.2012.06.001.

Chen, Cynthia, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang. 2016. "The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis." *Transportation Research Part C: Emerging Technologies* 68 (July): 285–99. doi:10.1016/j.trc.2016.04.005.

Clark, Ben, Kiron Chatterjee, Steve Melia, Gundi Knies, and Heather Laurie. 2014. "Life Events and Travel Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 2413 (September): 54–64. doi:10.3141/2413-06.

Çolak, Serdar, Lauren P. Alexander, Bernardo G. Alvim, Shomik R. Mehndiratta, and Marta C. González. 2015. "Analyzing Cell Phone Location Data for Urban Travel." *Transportation Research Record: Journal of the Transportation Research Board* 2526 (January): 126–35. doi:10.3141/2526-14.

Çolak, Serdar, Lauren P. Alexander, Bernardo Guatimosim Alvim, Shomik R. Mehndiretta, and Marta C. Gonzalez. 2015. "Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities." In Transportation Research Board Annual Meeting. Washington, D.C.

Cooke, Thomas J. 2013. "Internal Migration in Decline." *The Professional Geographer*, September. doi/abs/10.1080/00330124.2012.724343.

Coyle, John J., Kusumal Ruamsook, and Eric J. Symon. 2016. "Weatherproofing Supply Chains: Enable Intelligent Preparedness with Data Analytics." *Transportation Journal* 55 (2): 190–207. doi:10.5325/transportationj.55.2.0190.

El Esawey, Mohamed. 2016. "Toward a Better Estimation of Annual Average Daily Bicycle Traffic." *Transportation Research Record: Journal of the Transportation Research Board* 2593 (January): 28–36. doi:10.3141/2593-04.

Erhardt, Gregory D. 2016a. "Fusion of Large Continuously Collected Data Sources: Understanding Travel Demand Trends and Measuring Transport Project Impacts." PhD Thesis, University College London (UCL).

Erhardt, Gregory D. 2016b. "How Smart Is Your Smart Card?: Evaluating Transit Smart Card Data with Privacy Restrictions and Limited Penetration Rates." *Transportation Research Record: Journal of the Transportation Research Board* 2544 (January): 81–89. doi:10.3141/2544-10.

Erhardt, Gregory D., Michael Batty, and Elsa Arcaute. 2018. "Recommendations for Big Data Programs at Transportation Agencies." In *Big Data for Regional Science*, edited by Laurie A. Schintler and Zhenhua Chen. New York: Routledge.

Fagnant, Daniel J., and Kara Kockelman. 2015. "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." *Transportation Research Part A: Policy and Practice* 77 (July): 167–81. doi:10.1016/j.tra.2015.04.003.

Federal Transit Administration. 2013. "Before-and-After Studies of New Starts Projects: Report to Congress."
Feeney, Laura, Jason Bauman, Julia Chabrier, Geeti Mehra, Kyle Murphy, and Michelle Woodford. 2015. "Catalog of Administrative Data Sets." J-PAL North America. https://www.povertyactionlab.org/sites/default/files/documents/AdminDataCatalog.pdf.
Fusco, Gaetano, Chiara Colombaroni, and Natalia Isaenko. 2016. "Short-Term Speed Predictions Exploiting Big Data on Large Urban Road Networks." *Transportation Research Part C: Emerging Technologies* 73 (December): 183–201. doi:10.1016/j.trc.2016.10.019.
Hao, Jinwei, Jin Zhu, and Rui Zhong. 2015. "The Rise of Big Data on Urban Studies and Planning Practices in China: Review and Open Research Issues." *Journal of Urban Management*, Big/Open Data for Urban Management, 4 (2): 92–124. doi:10.1016/j.jum.2015.11.002.
Hilgert, Tim, Christine Weiss, Martin Kagerbauer, Bastian Chlond, and Peter Vortisch. 2016. "Stability and Flexibility in Commuting Behavior: Analyses of Mode Choice Patterns in Germany." In Transportation Research Board 95th Annual Meeting, Washington, D.C..
Huntsinger, Leta F. 2017. "The Lure of Big Data: Evaluating the Efficacy of Mobile Phone Data for Travel Model Validation." In Transportation Research Board 96th Annual Meeting, Washington, D.C.
Innovate UK. 2017. "Local Authority Solutions for Integrated Transport - Innovateuk." Accessed August 21. https://connect.innovateuk.org/web/local-authority-solutions-for-integrated-transport.
Jacques, Cynthia, and Ahmed El-Geneidy. 2014. "Does Travel Behavior Matter in Defining Urban Form? A Quantitative Analysis Characterizing Distinct Areas within a Region." *Journal of Transport and Land Use* 7 (1): 1–14.
Johnson, Timothy P., and Tom W. Smith. 2017. "Big Data and Survey Research: Supplement or Substitute?" In *Seeing Cities Through Big Data*, edited by Piyushimita (Vonu) Thakuriah, Nebiyou Tilahun, and Moira Zellner, 113–25. Springer Geography. Springer International Publishing. doi:10.1007/978-3-319-40902-3_7.
Kitamura, Ryuichi. 1990. "Panel Analysis in Transportation Planning: An Overview." *Transportation Research Part A: General* 24 (6): 401–15. doi:10.1016/0191-2607(90)90032-2.
Litman, Todd. 2014. "Autonomous Vehicle Implementation Predictions." *Victoria Transport Policy Institute* 28. http://leempo.com/wp-content/uploads/2017/03/M09.pdf.
Liu, Xiaoli, Matthew C. Roberts, and Ramteen Sioshansi. 2017. "Spatial Effects on Hybrid Electric Vehicle Adoption." *Transportation Research Part D: Transport and Environment* 52, Part A (May): 85–97. doi:10.1016/j.trd.2017.02.014.
Martin, Elliot, Susan Shaheen, and Jeffrey Lidicker. 2010. "Impact of Carsharing on Household Vehicle Holdings: Results from North American Shared-Use Vehicle Survey." *Transportation Research Record: Journal of the Transportation Research Board* 2143 (December): 150–58. doi:10.3141/2143-19.
Maude, Francis. 2014. "Government's Response to the National Statistician's Recommendation," July 18. http://webarchive.nationalarchives.gov.uk/20160204145156/https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/letterfromrthonfrancismaudemptosirandrewdilnot18071_tcm97-43946.pdf.
Moeckel, Rolf, Rhett Fussell, and Rick Donnelly. 2015. "Mode Choice Modeling for Long-Distance Travel." *Transportation Letters* 7 (1): 35–46. doi:10.1179/1942787514Y.0000000031.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2017. "Job Changing and the Decline in Long-Distance Migration in the United States." *Demography* 54 (2): 631–53. doi:10.1007/s13524-017-0551-9.

Nosal, Thomas, Luis F. Miranda-Moreno, Zlatko Krstulic, and Thomas Götschi. 2015. "Accounting for Weather Conditions When Comparing Multiple Years of Bicycle Demand Data." In Transportation Research Board 94th Annual Meeting, Washington, D.C.

Office for National Statistics. 2012. "Census 1911-2001." Text. *Office for National Statistics*. May 18. http://webarchive.nationalarchives.gov.uk/20160111030756/http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/about-censuses/census-history/200-years-of-the-census/census-1911-2001/index.html.

Office for National Statistics (ONS). 2017a. "Frequently Asked Questions - Office for National Statistics." Accessed August 21. https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/frequentlyaskedquestions.

Office for National Statistics (ONS). 2017b. "ONS Methodology Working Paper Series No. 8: Statistical Uses for Mobile Phone Data: Literature Review." Accessed August 21. https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno8statisticalusesformobilephonedataliteraturereview.

Office for National Statistics (ONS). 2017c. "Response Rates - Office for National Statistics." Accessed August 21. https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/qualityofthecensusdata/responserates.

Office for National Statistics (ONS). 2014. "The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority." United Kingdom. www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/national-statisticians-recommendation.pdf.

Oort, N. v, and O. Cats. 2015. "Improving Public Transport Decision Making, Planning and Operations by Using Big Data: Cases from Sweden and the Netherlands." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 19–24. doi:10.1109/ITSC.2015.11.

Outwater, Maren, Mark Bradley, Nazneen Ferdous, Chandra Bhat, Ram Pendyala, Stephane Hess, Andrew Daly, and Jeff LaMondia. 2015. "Tour-Based National Model System to Forecast Long-Distance Passenger Travel in the United States." In Transportation Research Board 94th Annual Meeting, Washington, D.C.

Owen, Andrew, and David M. Levinson. 2017. "Developing a Comprehensive U.S. Transit Accessibility Database." In *Seeing Cities Through Big Data*, edited by Piyushimita (Vonu) Thakuriah, Nebiyou Tilahun, and Moira Zellner, 279–90. Springer Geography. Springer International Publishing. doi:10.1007/978-3-319-40902-3_16.

Pucher, John, Ralph Buehler, and Mark Seinen. 2011. "Bicycling Renaissance in North America? An Update and Re-Appraisal of Cycling Trends and Policies." *Transportation Research Part A: Policy and Practice* 45 (6): 451–75. doi:10.1016/j.tra.2011.03.001.

Rao, Prashant S., and Mike Isaac. 2017. "Uber Loses License to Operate in London." *The New York Times*, September 22, sec. Business Day. https://www.nytimes.com/2017/09/22/business/uber-london.html.

Rohr, Charlene, Liisa Ecola, Johanna Zmud, Fay Dunkerley, James Black, and Eleanor Baker. 2016. "Travel in Britain in 2035: Future Scenarios and Their Implications for Technology

Innovation." RAND Corporation. http://www.rand.org/pubs/research_reports/RR1377.html.
San Francisco Count Transportation Authority. 2017. "TNCs Today: A Profile of San Francisco Transportation Network Company Activity." http://www.sfcta.org/sites/default/files/content/Planning/TNCs/TNCs_Today_061317.pdf.
Scheiner, Joachim, Kiron Chatterjee, and Eva Heinen. 2016. "Key Events and Multimodality: A Life Course Approach." *Transportation Research Part A: Policy and Practice* 91 (September).
Shaheen, Susan A., Elliot W. Martin, Adam P. Cohen, and Richard S. Finson. 2012. "Public Bikesharing in North America: Early Operator and User Understanding." Mineta Transportation Institute, San José State University http://transweb.sjsu.edu/PDFs/research/1029-public-bikesharing-understanding-early-operators-users.pdf.
Shuldiner, Alec T., and Paul W. Shuldiner. 2013. "The Measure of All Things: Reflections on Changing Conceptions of the Individual in Travel Demand Modeling." *Transportation* 40 (6): 1117–31. doi:10.1007/s11116-013-9490-5.
Smith, T. W. 2013. "Survey-Research Paradigms Old and New." *International Journal of Public Opinion Research* 25 (2): 218–29. doi:10.1093/ijpor/eds040.
Transportation Research Board. 2016a. "Interregional Travel: A New Perspective for Policy Making." Special Report 320. Washington, D.C.: National Academies Press. doi:10.17226/21887.
Transportation Research Board. 2016b. "Between Public and Private Mobility: Examining the Rise of Technology-Enabled Transportation Services." Special Report 319. Washington, D.C.: National Academies Press. doi:10.17226/21875.
Tyndall, Justin. 2017. "Where No Cars Go: Free-Floating Carshare and Inequality of Access." *International Journal of Sustainable Transportation* 11 (6): 433–42. doi:10.1080/15568318.2016.1266425.
UK Data Service. 2017. "WICID: About the Data Sets." Accessed August 21. http://wicid.ukdataservice.ac.uk/cider/about/data_int.php?type=2.
United Nations Statistical Office. 2014. *Principles and Recommendations for a Vital Statistics System*. New York: United Nations.
Vardi, Moshe Y. 2017. "Humans, Machines and the Future of Work." presented at the College of Engineering Distinguished Lecture, University of Kentucky, Lexington, KY, April 13.
Vij, Akshay, and K. Shankari. 2015. "When Is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis." *Transportation Research Part C: Emerging Technologies* 56: 446–62. doi:10.1016/j.trc.2015.04.025.
Walker, Joan. 2017. "NSF Award #1648930 - Workshop: Advancing the Science of Transportation Demand Modeling." National Science Foundation, Washington, D.C. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1648930.
Wang, Ted, Alex Lu, and Alla Reddy. 2013. "Maintaining Key Services While Retaining Core Values: NYC Transit's Environmental Justice Strategies." *Journal of Public Transportation* 16 (1). https://trid.trb.org/View/1248230?ajax=1.
Wang, Xize. 2017. "Peak Car in the Car Capital? Double-Cohort Analysis for Commute Mode Choice in Los Angeles County, California, Using Census and ACS Microdata." In . https://trid.trb.org/View/1439710?ajax=1.
Weiss, Christine, Bastian Chlond, Clotilde Minster, Christian Jödden, and Peter Vortisch. 2017. "Assessing Effects of Mixed-Mode Design in a Longitudinal Household Travel Survey." In Transportation Research Board 96th Annual Meeting, Washington, D.C.

Xiao, S., X. C. Liu, and Y. Wang. 2015. "Data-Driven Geospatial-Enabled Transportation Platform for Freeway Performance Analysis." *IEEE Intelligent Transportation Systems Magazine* 7 (2): 10–21. doi:10.1109/MITS.2014.2388367.

Yasmin, Farhana, Catherine Morency, and Matthew J. Roorda. 2017. "Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model." *Transportmetrica A: Transport Science* 13 (3): 222–49. doi:10.1080/23249935.2016.1249437.

Zhang, Qi, Hongfei Zhan, and Junhe Yu. 2017. "Car Sales Analysis Based on the Application of Big Data." *Procedia Computer Science*, Advances in Information and Communication Technology: Proceedings of 7th International Congress of Information and Communication Technology (ICICT2017), 107 (January): 436–41. doi:10.1016/j.procs.2017.03.137.

Appendix

Supplemental Table 1: All keywords with 6 or more appearances in search results

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Traffic counts	147	0	147	High	Low
2	Travel demand	84	21	105	High	High
3	Origin and destination	74	19	93	High	High
4	Data collection	46	39	85	High	High
5	Travel behavior	62	19	81	High	High
6	Public transit	57	19	76	High	High
7	Travel surveys	55	10	65	High	High
8	Mode choice	50	9	59	High	High
9	Case studies	34	22	56	High	High
10	Urban areas	44	11	55	High	High
11	Commuting	52	1	53	High	Low
12	Demographics	49	2	51	High	Low
13	Transportation planning	34	17	51	High	High
14	Socioeconomic factors	47	2	49	High	Low
15	Spatial analysis	41	5	46	High	Low
16	Travel time	29	16	45	High	High
17	Data analysis	15	30	45	High	High
18	Traffic data	24	20	44	High	High
19	Big data	2	42	44	Low	High
20	Geographic information systems	36	7	43	High	High
21	Mobility	25	18	43	High	High
22	Planning	36	6	42	High	High
23	Travel patterns	26	16	42	High	High
24	Land use	39	1	40	High	Low
25	Accessibility	36	4	40	High	Low
26	Traffic flow	28	12	40	High	High
27	Households	33	3	36	High	Low
28	Work trips	33	1	34	High	Low
29	Mathematical models	30	3	33	High	Low
30	Traffic volume	27	6	33	High	High
31	Traffic models	20	13	33	High	High
32	Bicycling	27	4	31	High	Low
33	Forecasting	22	8	30	High	High
34	Traffic congestion	16	14	30	High	High
35	Census	29	0	29	High	Low
36	Traffic estimation	25	4	29	High	Low
37	Neighborhoods	27	1	28	High	Low
38	Algorithms	13	15	28	High	High
39	Intelligent transportation systems	2	26	28	Low	High
40	Automobile ownership	24	3	27	High	Low
41	Commuters	23	4	27	High	Low
42	United States	22	4	26	High	Low
43	Traffic forecasting	18	8	26	High	High
44	Walking	23	2	25	High	Low
45	City planning	20	5	25	High	Low
46	Modal split	20	3	23	High	Low
47	Surveys	19	4	23	High	Low
48	Choice models	15	8	23	High	High

49	Freight transportation	14	9	23	High	High
50	Global Positioning System	10	13	23	High	High
51	Trip generation	20	2	22	High	Low
52	Canada	20	2	22	High	Low
53	Microsimulation	18	4	22	High	Low
54	Vehicle sharing	15	7	22	High	High
55	Simulation	13	9	22	High	High
56	Nonmotorized transportation	21	0	21	High	Low
57	Land use planning	19	2	21	High	Low
58	Ridership	12	9	21	High	High
59	Optimization	10	10	20	High	High
60	Decision making	8	12	20	High	High
61	Annual average daily traffic	18	1	19	High	Low
62	Metropolitan areas	17	2	19	High	Low
63	Activity choices	15	4	19	High	Low
64	Sustainable development	12	7	19	High	High
65	Data mining	5	14	19	Low	High
66	Equity (Justice)	17	1	18	High	Low
67	Location	17	1	18	High	Low
68	Regression analysis	17	1	18	High	Low
69	Estimation theory	16	2	18	High	Low
70	Residential location	16	2	18	High	Low
71	Trip matrices	14	4	18	High	Low
72	Demand	13	5	18	High	Low
73	Bicycles	15	2	17	High	Low
74	Methodology	12	5	17	High	Low
75	Infrastructure	11	6	17	High	High
76	China	2	15	17	Low	High
77	Statistical analysis	13	3	16	High	Low
78	Traffic simulation	9	7	16	High	High
79	Route choice	8	8	16	High	High
80	Bicycle facilities	14	1	15	High	Low
81	Social factors	13	2	15	High	Low
82	Data quality	13	2	15	High	Low
83	Pedestrians	12	3	15	High	Low
84	United Kingdom	11	4	15	High	Low
85	Behavior	11	4	15	High	Low
86	Cluster analysis	11	4	15	High	Low
87	Traffic assignment	11	4	15	High	Low
88	Trip length	11	4	15	High	Low
89	Networks	10	5	15	High	Low
90	Logistics	4	11	15	Low	High
91	Housing	14	0	14	High	Low
92	Conferences	13	1	14	High	Low
93	Automobile travel	12	2	14	High	Low
94	Built environment	12	2	14	High	Low
95	New York (New York)	8	6	14	High	High
96	Urban transportation	8	6	14	High	High
97	Sustainable transportation	8	6	14	High	High
98	Real time information	3	11	14	Low	High
99	Low income groups	13	0	13	High	Low
100	Vehicle miles of travel	12	1	13	High	Low
101	Multinomial logits	10	3	13	High	Low
102	Policy	9	4	13	High	Low

103	Cellular telephones	5	8	13	Low	High
104	Traffic count	12	0	12	High	Low
105	Rural areas	11	1	12	High	Low
106	California	11	1	12	High	Low
107	Transportation	11	1	12	High	Low
108	Transportation modes	10	2	12	High	Low
109	Traffic surveillance	9	3	12	High	Low
110	Validation	9	3	12	High	Low
111	Cyclists	9	3	12	High	Low
112	Logistic regression analysis	9	3	12	High	Low
113	Level of service	9	3	12	High	Low
114	Disaggregate analysis	10	1	11	High	Low
115	Aged	10	1	11	High	Low
116	Transit oriented development	10	1	11	High	Low
117	Conference	10	1	11	High	Low
118	Population	9	2	11	High	Low
119	Travel	8	3	11	High	Low
120	Network analysis (Planning)	7	4	11	High	Low
121	Performance measurement	6	5	11	High	Low
122	Multimodal transportation	6	5	11	High	Low
123	Logits	6	5	11	High	Low
124	Information processing	5	6	11	Low	High
125	Smartphones	3	8	11	Low	High
126	Montreal (Canada)	10	0	10	High	Low
127	Commodity flow	10	0	10	High	Low
128	Bicycle commuting	10	0	10	High	Low
129	Population density	10	0	10	High	Low
130	Employment	9	1	10	High	Low
131	Australia	9	1	10	High	Low
132	Gender	9	1	10	High	Low
133	Policy analysis	8	2	10	High	Low
134	Modal shift	7	3	10	High	Low
135	Data fusion	6	4	10	High	Low
136	Trend (Statistics)	6	4	10	High	Low
137	Trip purpose	6	4	10	High	Low
138	Railroad transportation	5	5	10	Low	Low
139	Neural networks	5	5	10	Low	Low
140	Environmental impacts	5	5	10	Low	Low
141	Smart cards	3	7	10	Low	High
142	Jobs	9	0	9	High	Low
143	Minneapolis (Minnesota)	9	0	9	High	Low
144	Activity based modeling	9	0	9	High	Low
145	Monte Carlo method	9	0	9	High	Low
146	Land use models	8	1	9	High	Low
147	Accuracy	8	1	9	High	Low
148	Pedestrian safety	7	2	9	High	Low
149	Toronto (Canada)	7	2	9	High	Low
150	Pollutants	6	3	9	High	Low
151	Traffic distribution	6	3	9	High	Low
152	Calibration	6	3	9	High	Low
153	Costs	6	3	9	High	Low
154	Regional planning	5	4	9	Low	Low
155	Bus transit	5	4	9	Low	Low
156	Mathematical prediction	5	4	9	Low	Low

157	Stochastic processes	5	4	9	Low	Low
158	Strategic planning	5	4	9	Low	Low
159	Rail transit	4	5	9	Low	Low
160	Data files	4	5	9	Low	Low
161	Transit operating agencies	4	5	9	Low	Low
162	High speed rail	2	7	9	Low	High
163	Technological innovations	2	7	9	Low	High
164	Transportation disadvantaged persons	8	0	8	High	Low
165	Dublin (Ireland)	8	0	8	High	Low
166	Population forecasting	8	0	8	High	Low
167	Chicago (Illinois)	8	0	8	High	Low
168	Least squares method	8	0	8	High	Low
169	Immigrants	8	0	8	High	Low
170	Road networks	7	1	8	High	Low
171	Errors	7	1	8	High	Low
172	Central business districts	7	1	8	High	Low
173	Bicycle travel	7	1	8	High	Low
174	Freight traffic	6	2	8	High	Low
175	Traffic safety	6	2	8	High	Low
176	Highway traffic control	6	2	8	High	Low
177	Residential areas	6	2	8	High	Low
178	Arterial highways	6	2	8	High	Low
179	Systems analysis	6	2	8	High	Low
180	Beijing (China)	4	4	8	Low	Low
181	Mobile telephones	4	4	8	Low	Low
182	Sensors	4	4	8	Low	Low
183	Policy making	4	4	8	Low	Low
184	Netherlands	2	6	8	Low	High
185	Multivariate analysis	7	0	7	High	Low
186	Walkability	7	0	7	High	Low
187	Economic factors	7	0	7	High	Low
188	Synthetic populations	7	0	7	High	Low
189	Data acquisition	7	0	7	High	Low
190	Urban area	7	0	7	High	Low
191	Bayes' theorem	6	1	7	High	Low
192	Probits	6	1	7	High	Low
193	England	6	1	7	High	Low
194	Transport planning	6	1	7	High	Low
195	Energy consumption	6	1	7	High	Low
196	Databases	5	2	7	Low	Low
197	Pedestrian-vehicle crashes	5	2	7	Low	Low
198	Planning and design	5	2	7	Low	Low
199	Electric vehicles	4	3	7	Low	Low
200	Evacuation	4	3	7	Low	Low
201	France	4	3	7	Low	Low
202	Stated preferences	4	3	7	Low	Low
203	London (England)	3	4	7	Low	Low
204	Quality of service	3	4	7	Low	Low
205	Cities	3	4	7	Low	Low
206	Alternatives analysis	3	4	7	Low	Low
207	Hybrid vehicles	6	0	6	High	Low
208	Seasons	6	0	6	High	Low
209	School trips	6	0	6	High	Low
210	Ireland	6	0	6	High	Low

211	Freight Analysis Framework	6	0	6	High	Low
212	Carpools	6	0	6	High	Low
213	Peak hour traffic	6	0	6	High	Low
214	Multi-agent systems	6	0	6	High	Low
215	Traffic counting	6	0	6	High	Low
216	Links (Networks)	6	0	6	High	Low
217	Population synthesis	6	0	6	High	Low
218	Texas	6	0	6	High	Low
219	Loop detectors	6	0	6	High	Low
220	Hamilton (Canada)	6	0	6	High	Low
221	Peak periods	6	0	6	High	Low
222	Urban transportation policy	6	0	6	High	Low
223	Agent based models	6	0	6	High	Low
224	Geography	6	0	6	High	Low
225	Urban development	5	1	6	Low	Low
226	Light rail transit	5	1	6	Low	Low
227	Traffic analysis zones	5	1	6	Low	Low
228	Days	5	1	6	Low	Low
229	Markov chains	5	1	6	Low	Low
230	Signalized intersections	5	1	6	Low	Low
231	Railroad commuter service	5	1	6	Low	Low
232	Intersections	5	1	6	Low	Low
233	Crashes	5	1	6	Low	Low
234	Estimating	5	1	6	Low	Low
235	Revealed preferences	4	2	6	Low	Low
236	Suburbs	4	2	6	Low	Low
237	Paratransit services	4	2	6	Low	Low
238	Income	4	2	6	Low	Low
239	Parking	3	3	6	Low	Low
240	Greenhouse gases	3	3	6	Low	Low
241	Data banks	3	3	6	Low	Low
242	Rapid transit	3	3	6	Low	Low
243	India	3	3	6	Low	Low
244	Routes	3	3	6	Low	Low
245	Developing countries	3	3	6	Low	Low
246	Medium sized cities	3	3	6	Low	Low
247	Bluetooth technology	2	4	6	Low	Low
248	Floating car data	2	4	6	Low	Low
249	Urban highways	2	4	6	Low	Low
250	Routing	1	5	6	Low	Low
251	Traveler information and communication systems	1	5	6	Low	Low
252	Special events	1	5	6	Low	Low
253	Supply chain management	0	6	6	Low	High