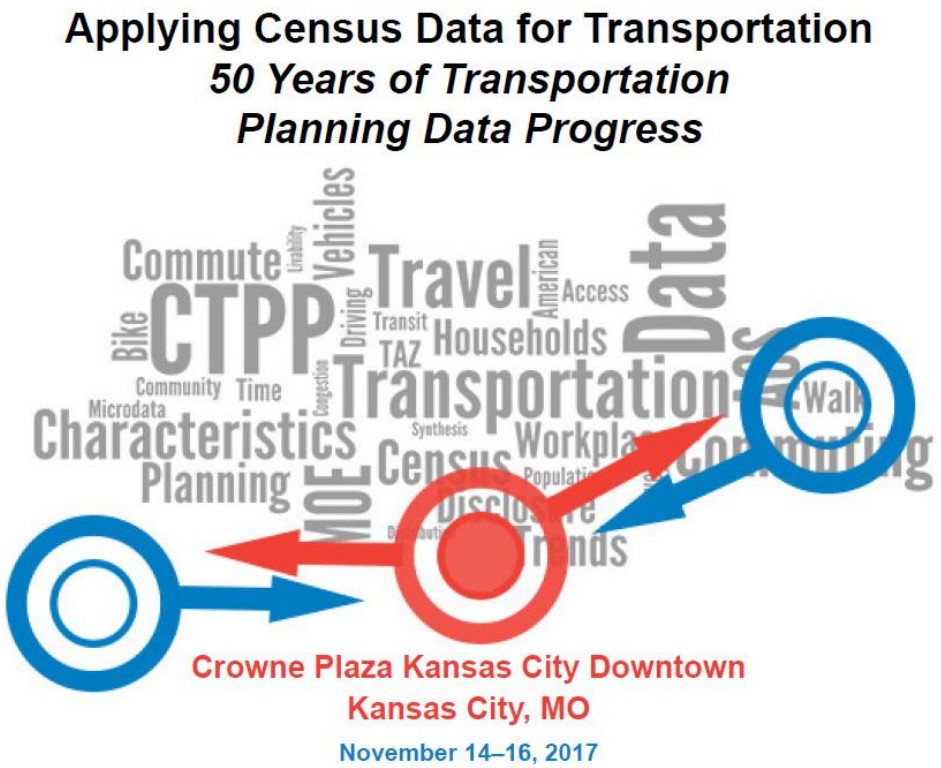
Understanding the Role and Relevance of the Census in a Changing Transportation Data Landscape

Gregory D. Erhardt
University of Kentucky
greg.erhardt@uky.edu

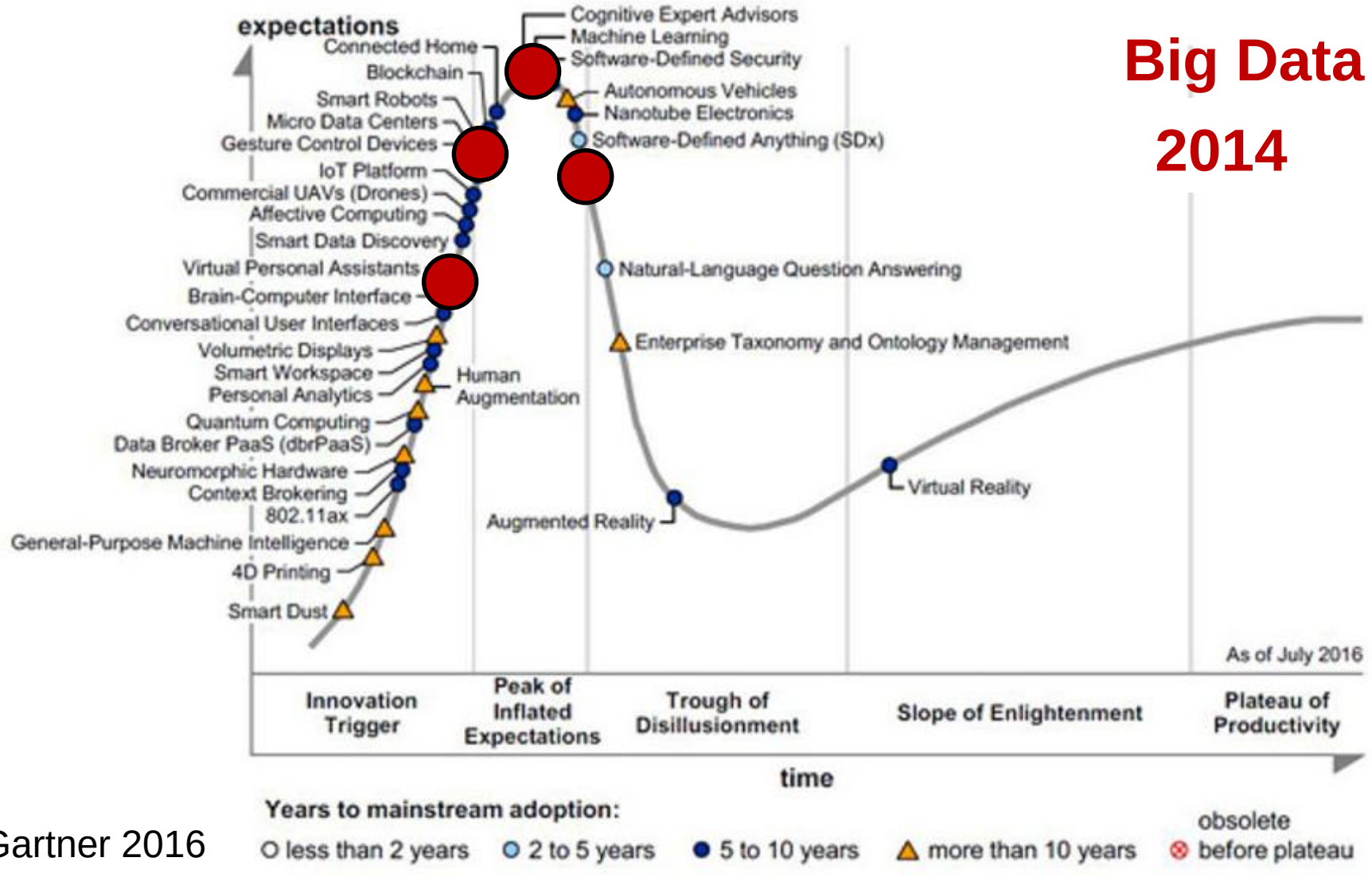
Adam Dennett
University College London (UCL)
adennett@ucl.ac.uk



Big Data


CONSULTANTS SAY THREE QUINTILLION BYTES OF DATA ARE CREATED EVERY DAY.
IT COMES FROM EVERYWHERE. IT KNOWS ALL.
ACCORDING TO THE BOOK OF WIKIPEDIA, ITS NAME IS "BIG DATA."
BIG DATA
BIG DATA LIVES IN THE CLOUD. IT KNOWS WHAT WE DO.
IN THE PAST, OUR COMPANY DID MANY EVIL THINGS.
BUT IF WE ACCEPT BIG DATA IN OUR SERVERS, WE WILL BE SAVED FROM BANKRUPTCY.
LET US PAY.
$
IS IT TOO LATE TO SIDE WITH EVIL?
SHHHH! IT HEARS YOU.
DilbertCartoonist@gmail.com
www.dilbert.com

Big Data's place in the hype cycle



Big Data 2014

Gartner 2016

Still on the upswing in transportation…

Articles by year for "Census" and "Big Data" search terms in TRID database

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

What are Big Data anyway?

Image from: wikipedia.com

Image from: appliedcomm.com

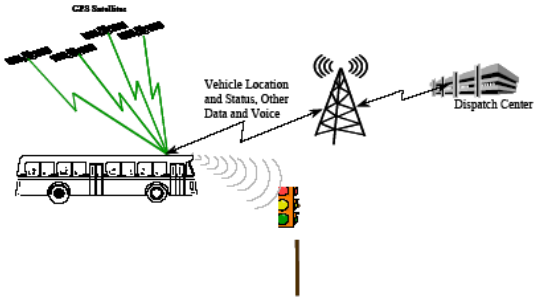


Image from: commons.wikimedia.org

Any data that cannot fit in an Excel spreadsheet (Batty 2013)

Image from: commons.wikimedia.org

Image from: sfpark.org

A bit of history…

Models originally based on large OD surveys:

- CATS Home Interview Survey (1956):
  - Personal interviews of 50,000 households (1 in 30)
  - $675,000 (~$6 million in 2017 dollars)
- KY Statewide Model (1975 by Voorhees)
  - $5-6 million (~$25 million in 2017 dollars)
- OH Roadside Surveys into the 2000s
- Toronto through present day

The gravity model allowed smaller surveys

Sources: Black 1990, Weiner 2016, Miller 2017, correspondence with Rob Bostrom (KY) and Greg Giaimo/Rebekah Anderson (OH)

A forecast:

Big Data will continue to be simultaneously **disappointing** and **useful**.

The question then becomes:

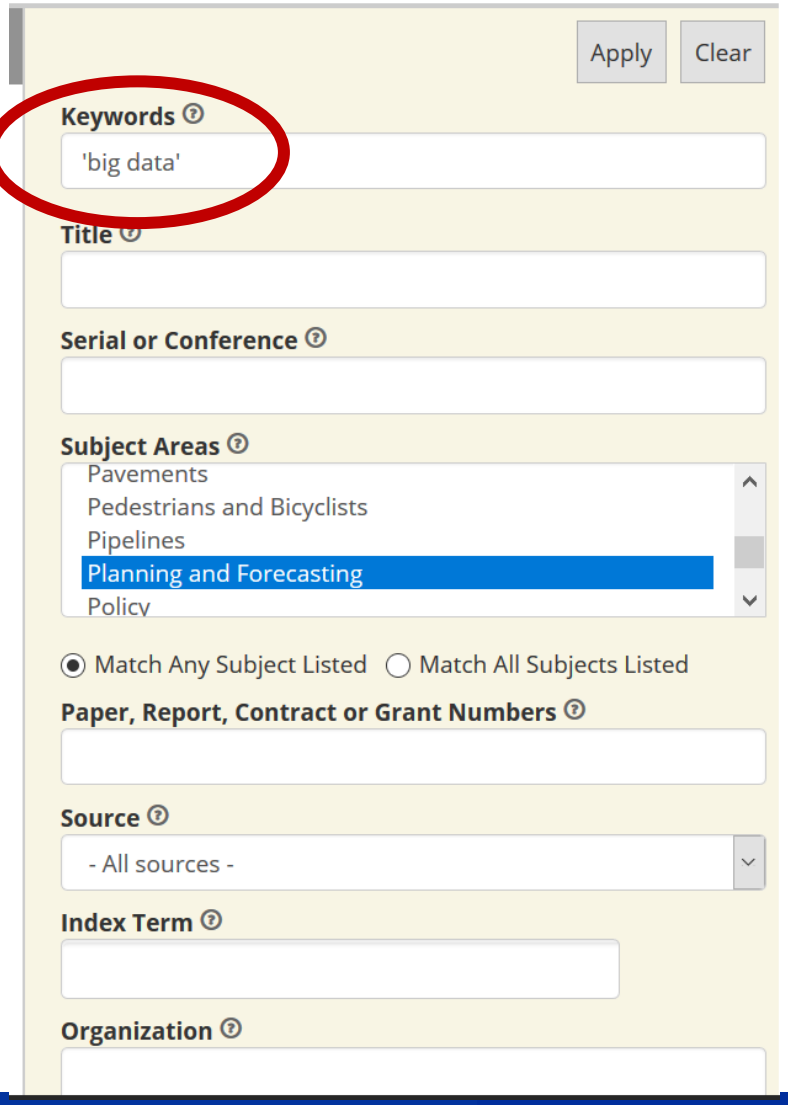
How should we use them, and how should we adapt our existing data?

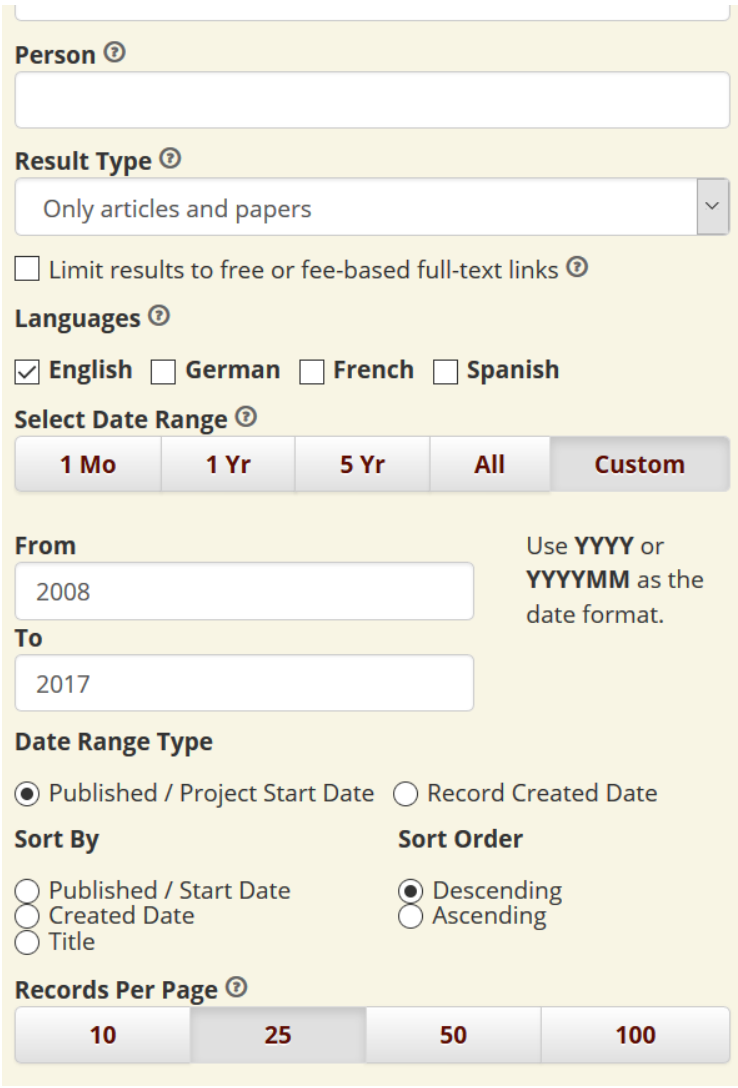
Key Questions

In the context of transportation planning:

1. **How do the uses of Big Data and Census data relate?**
2. Are other nations adapting their Census given the advent of Big Data?
3. What new policy questions are should planners anticipate over the next decade?
4. What are the options for keeping the Census relevant?

Method: TRID Search & Keyword Analysis





Overlapping key words

Selected overlapping articles

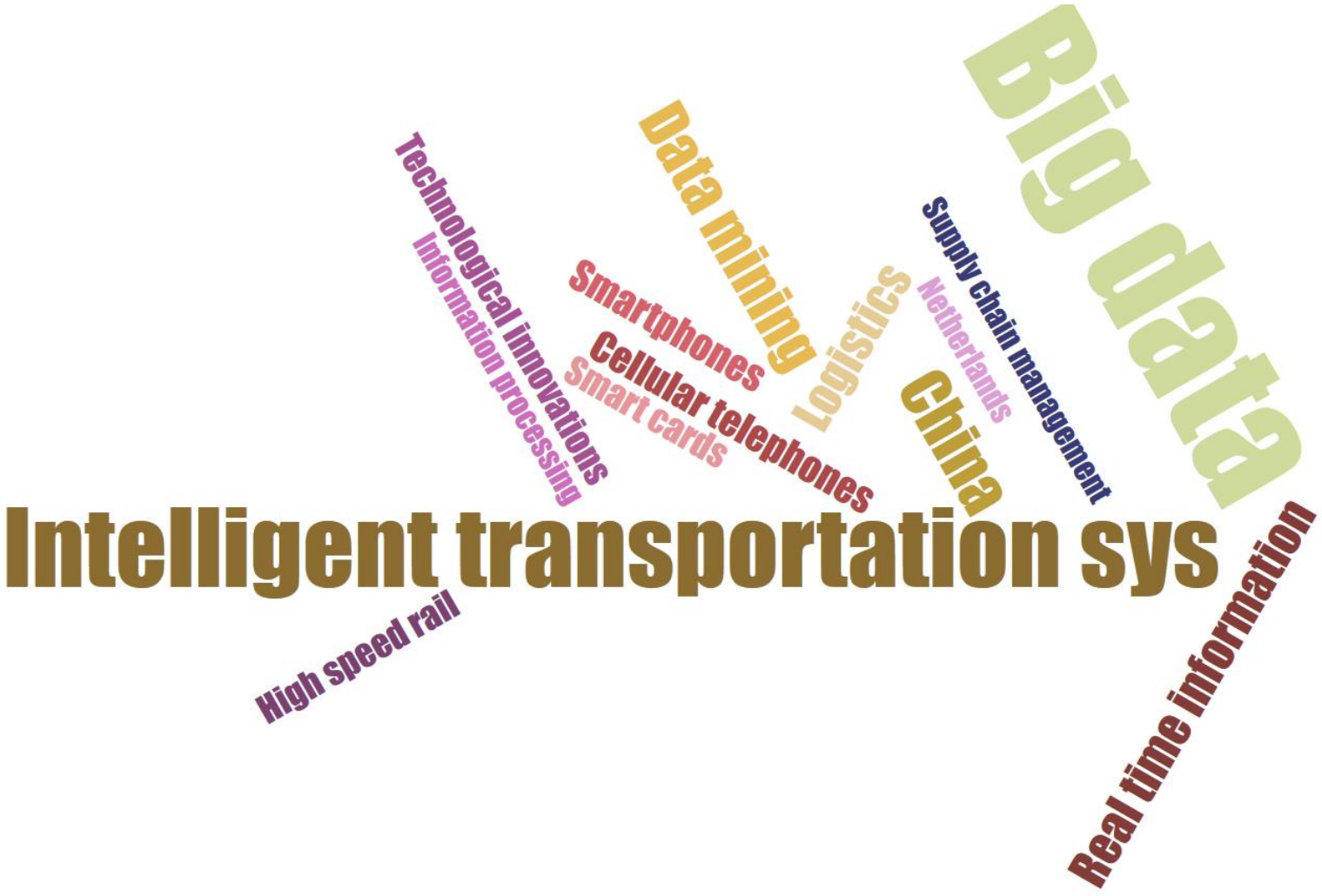
Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Travel demand & Census	Yasmin, Morency, and Roorda 2017	Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model	Activity based models, Activity choices, Montreal (Canada), Origin and destination, Travel demand, Validation	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.

Census dominant key words

Selected Census-dominant articles

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.

Big Data dominant key words



Selected Big Data articles

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.

Themes & observations

Census Data

- Demographics, socioeconomics, commutes
- Commute flow & mode
- Uniform & publicly available
- Richer
- Privacy constraints

Big Data

- Operational, traffic & logistics
- OD matrices
- Location-dependent, license restricted
- Larger sample
- Privacy constraints

Often combined → Census serves as a basis for expansion & for adding demographic & socioeconomic attributes

Key Questions

In the context of transportation planning:

1. How do the uses of Big Data and Census data relate?
2. **Are other nations adapting their Census given the advent of Big Data?**
3. What new policy questions are should planners anticipate over the next decade?
4. What are the options for keeping the Census relevant?

Census 'could use mobile phone data instead of questions'

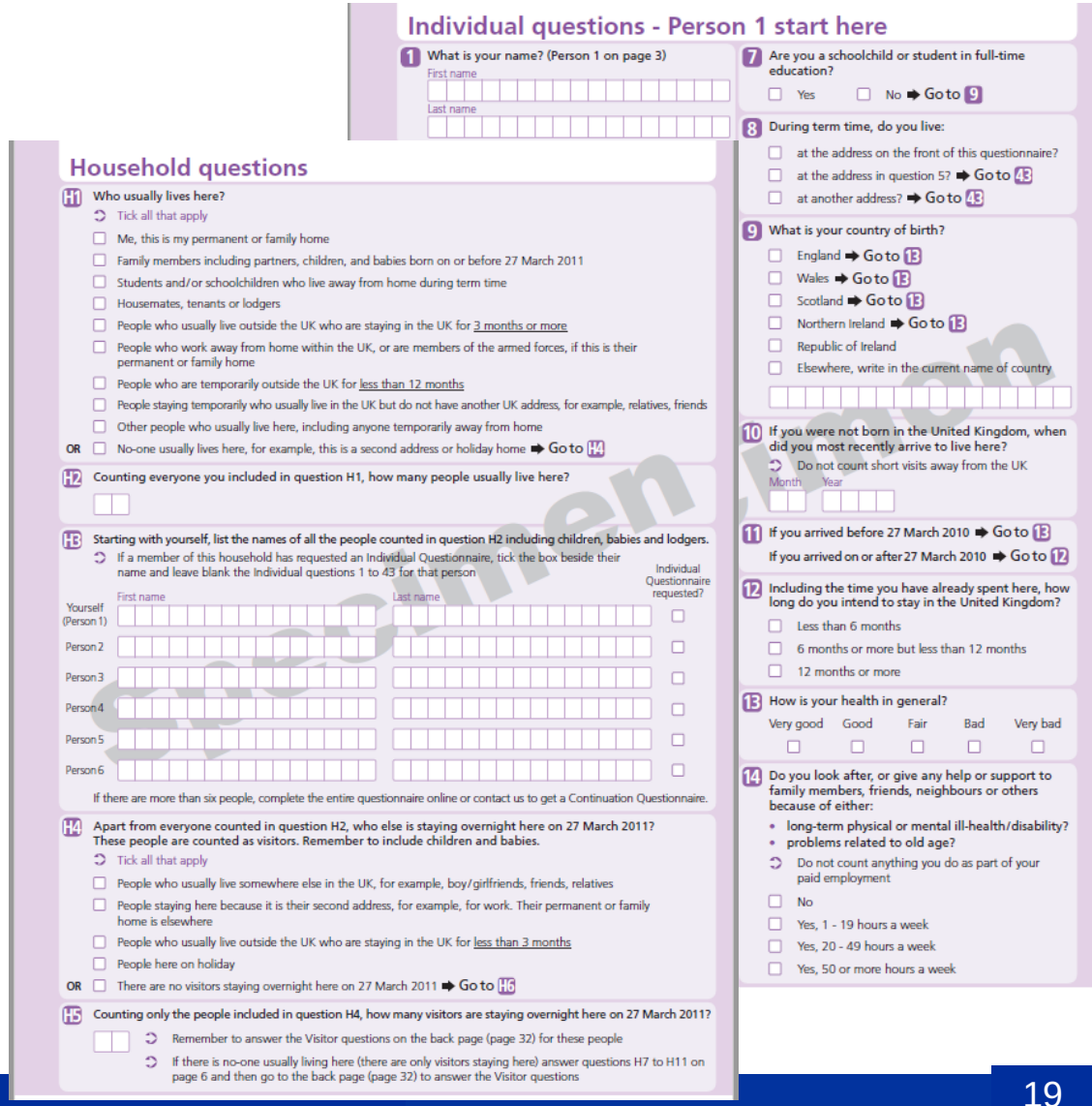
7 November 2017 | UK

Mobile phone data could be used in place of census questions in the future, a report from the Office for National Statistics (ONS) suggests.

http://www.bbc.com/news/uk-41898318

UK Census

- No constitution → no constitutional requirement
- Decennial since 1801
- Mail-back "long form" with 94% response rate
- £480 million in 2011 (US Census $13 billion in 2010)



Individual questions - Person 1 start here

1 What is your name? (Person 1 on page 3)
First name
Last name

7 Are you a schoolchild or student in full-time education?
☐ Yes ☐ No ➡ Go to 9

8 During term time, do you live:
☐ at the address on the front of this questionnaire?
☐ at the address in question 5? ➡ Go to 43
☐ at another address? ➡ Go to 43

9 What is your country of birth?
☐ England ➡ Go to 13
☐ Wales ➡ Go to 13
☐ Scotland ➡ Go to 13
☐ Northern Ireland ➡ Go to 13
☐ Republic of Ireland
☐ Elsewhere, write in the current name of country

10 If you were not born in the United Kingdom, when did you most recently arrive to live here?
Do not count short visits away from the UK
Month Year

11 If you arrived before 27 March 2010 ➡ Go to 13
If you arrived on or after 27 March 2010 ➡ Go to 12

12 Including the time you have already spent here, how long do you intend to stay in the United Kingdom?
☐ Less than 6 months
☐ 6 months or more but less than 12 months
☐ 12 months or more

13 How is your health in general?
Very good ☐ Good ☐ Fair ☐ Bad ☐ Very bad ☐

14 Do you look after, or give any help or support to family members, friends, neighbours or others because of either:
- long-term physical or mental ill-health/disability?
- problems related to old age?

Do not count anything you do as part of your paid employment
☐ No
☐ Yes, 1 - 19 hours a week
☐ Yes, 20 - 49 hours a week
☐ Yes, 50 or more hours a week

Household questions

H1 Who usually lives here?
Tick all that apply
☐ Me, this is my permanent or family home
☐ Family members including partners, children, and babies born on or before 27 March 2011
☐ Students and/or schoolchildren who live away from home during term time
☐ Housemates, tenants or lodgers
☐ People who usually live outside the UK who are staying in the UK for 3 months or more
☐ People who work away from home within the UK, or are members of the armed forces, if this is their permanent or family home
☐ People who are temporarily outside the UK for less than 12 months
☐ People staying temporarily who usually live in the UK but do not have another UK address, for example, relatives, friends
☐ Other people who usually live here, including anyone temporarily away from home
OR ☐ No-one usually lives here, for example, this is a second address or holiday home ➡ Go to H4

H2 Counting everyone you included in question H1, how many people usually live here?

H3 Starting with yourself, list the names of all the people counted in question H2 including children, babies and lodgers.
If a member of this household has requested an Individual Questionnaire, tick the box beside their name and leave blank the Individual questions 1 to 43 for that person

	First name	Last name	Individual Questionnaire requested?
Yourself (Person 1)			☐
Person 2			☐
Person 3			☐
Person 4			☐
Person 5			☐
Person 6			☐

If there are more than six people, complete the entire questionnaire online or contact us to get a Continuation Questionnaire.

H4 Apart from everyone counted in question H2, who else is staying overnight here on 27 March 2011? These people are counted as visitors. Remember to include children and babies.
Tick all that apply
☐ People who usually live somewhere else in the UK, for example, boy/girlfriends, friends, relatives
☐ People staying here because it is their second address, for example, for work. Their permanent or family home is elsewhere
☐ People who usually live outside the UK who are staying in the UK for less than 3 months
☐ People here on holiday
OR ☐ There are no visitors staying overnight here on 27 March 2011 ➡ Go to H6

H5 Counting only the people included in question H4, how many visitors are staying overnight here on 27 March 2011?
Remember to answer the Visitor questions on the back page (page 32) for these people
If there is no-one usually living here (there are only visitors staying here) answer questions H7 to H11 on page 6 and then go to the back page (page 32) to answer the Visitor questions

Office for National Statistics

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

March 2014

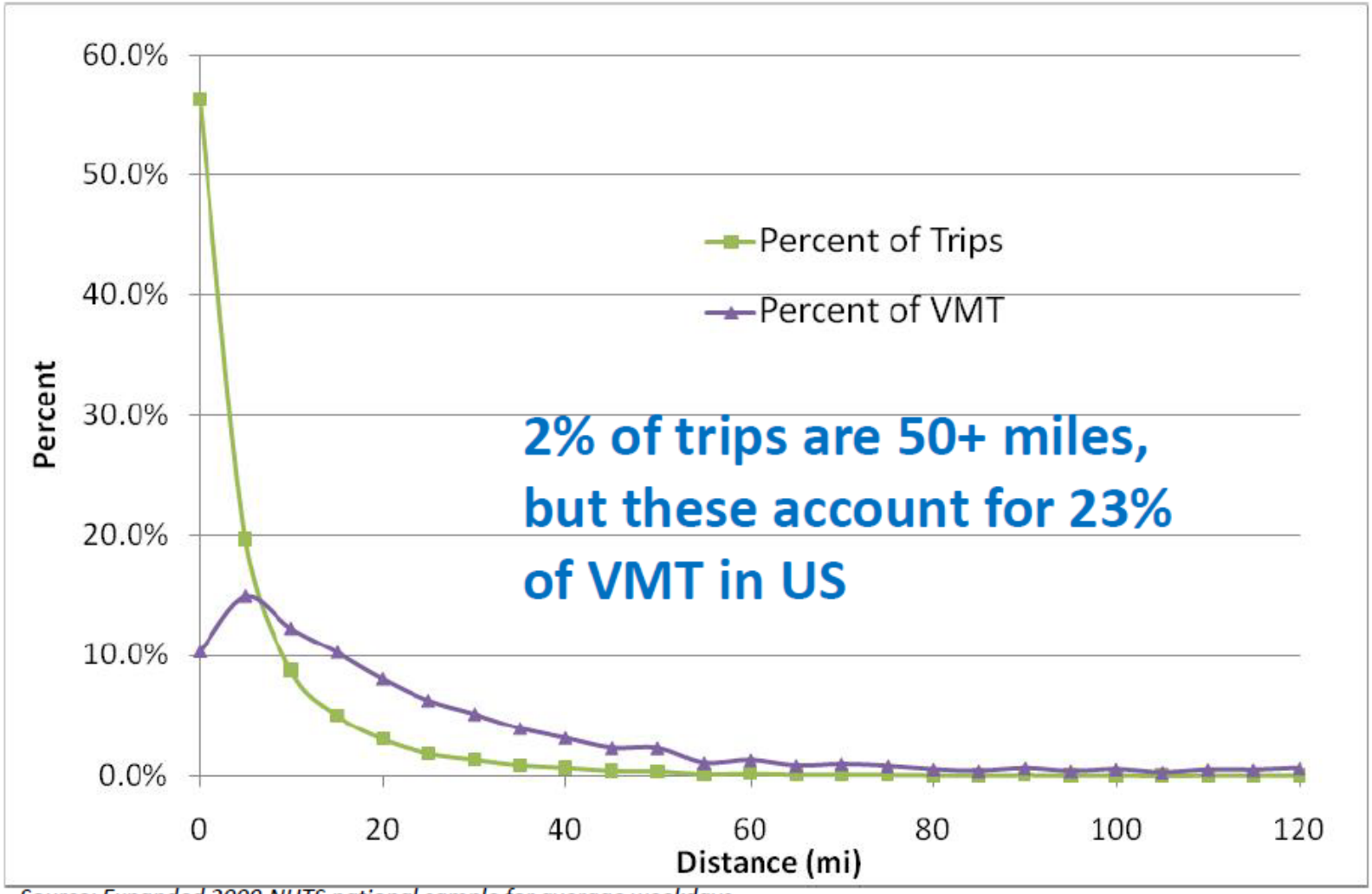
- Considers the practicality of using **administrative data** (such as tax & benefit records) to produce the statistics currently generated by the Census
- Countries in Northern Europe take this approach, but they have a **population register**
- Recommends **online Census in 2021** + increased use of administrative data

Key Questions

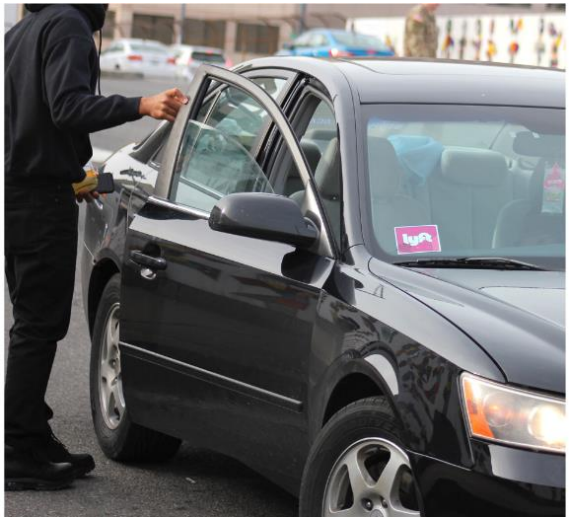
In the context of transportation planning:

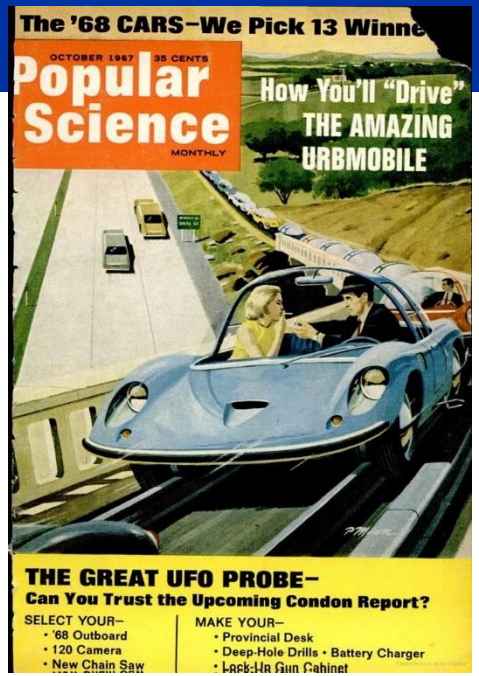
1. How do the uses of Big Data and Census data relate?
2. Are other nations adapting their Census given the advent of Big Data?
3. **What new policy questions are should planners anticipate over the next decade?**
4. What are the options for keeping the Census relevant?

A question for today: Long-distance travel

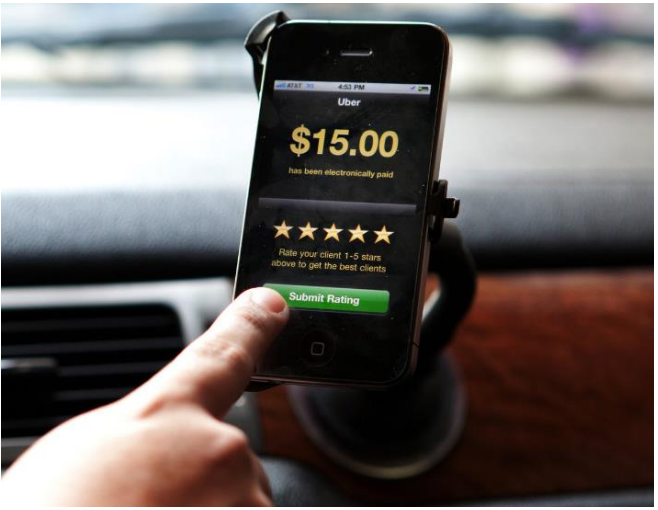


Source: Expanded 2009 NHTS national sample for average weekdays

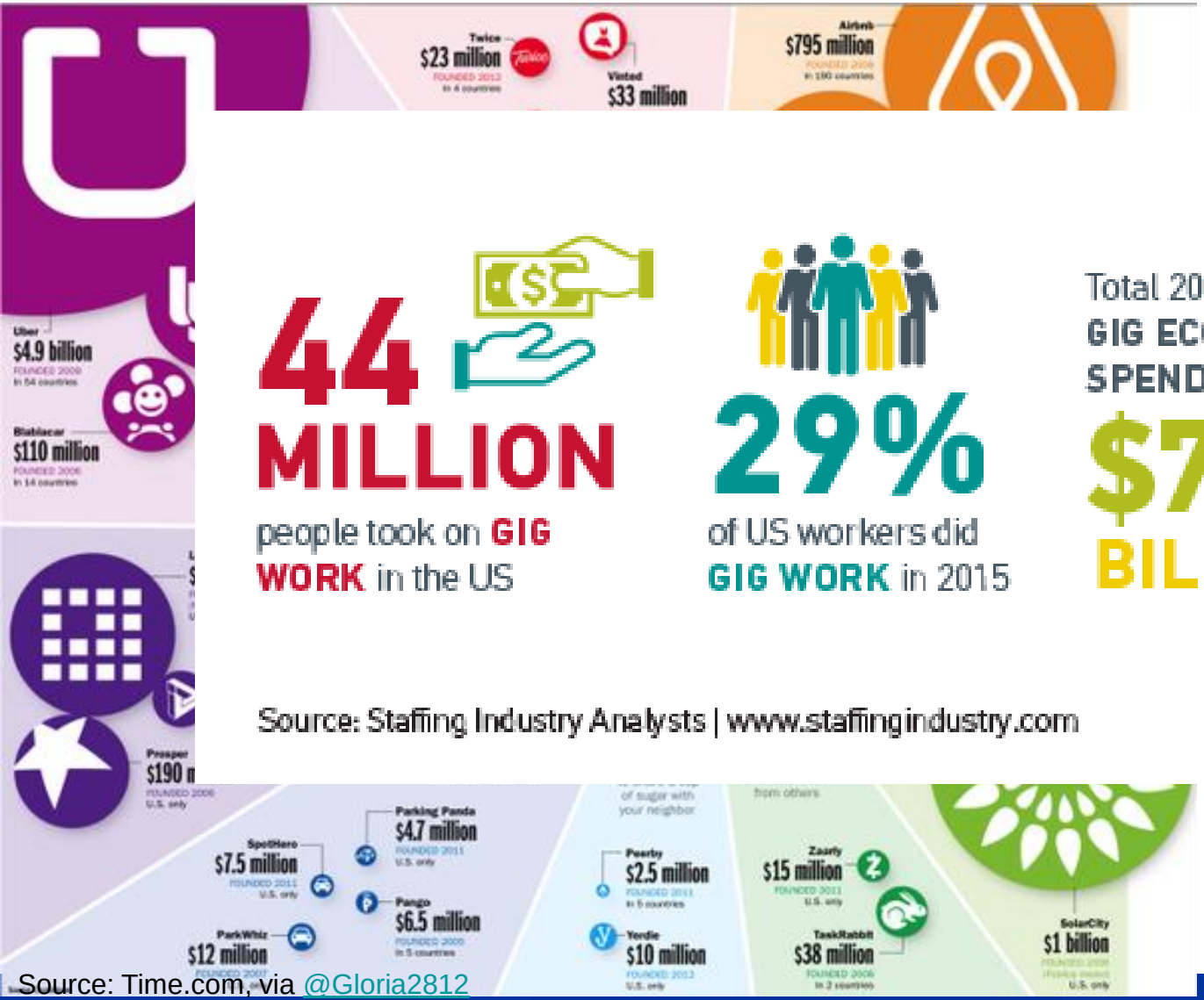




The future of mobility



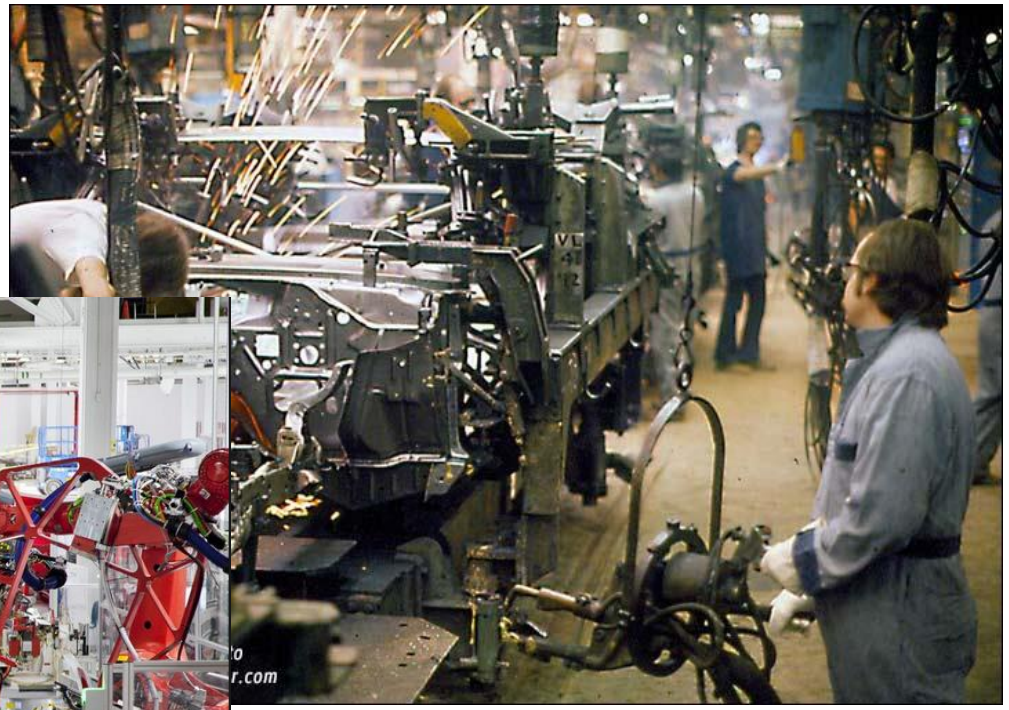
The future of work



Source: Time.com, via @Gloria2812

The future of work

Chrysler 1980s

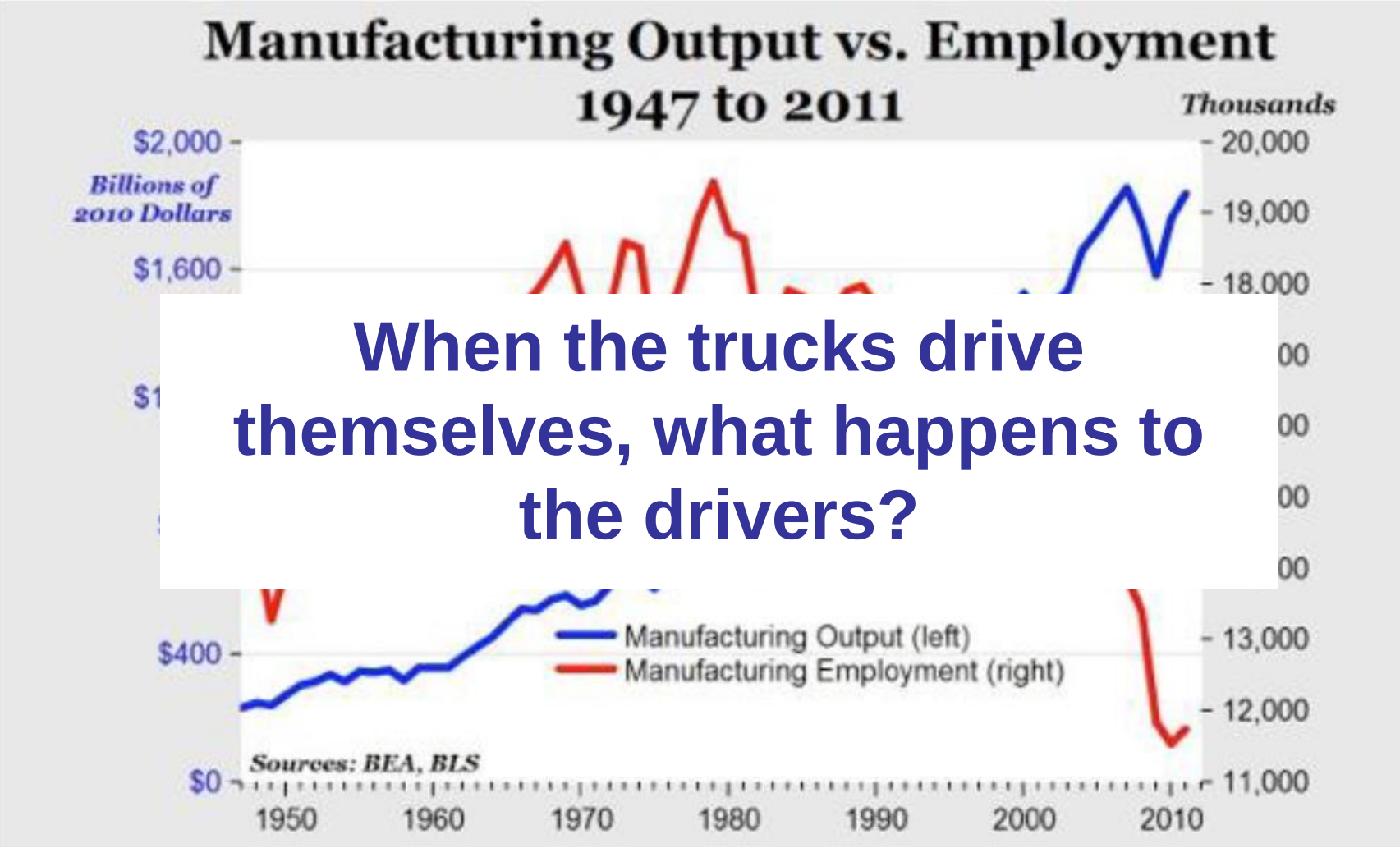


https://www.allpar.com/corporate/factories/canada/minivans-1980s-1990s.html

Tesla 2018

https://www.theverge.com/2017/11/9/16627528/tesla-perbix-acquisition-automated-factory-machine-builds-machines

The future of work



When the trucks drive themselves, what happens to the drivers?

Vardi 2017

Key Questions

In the context of transportation planning:

1. How do the uses of Big Data and Census data relate?
2. Are other nations adapting their Census given the advent of Big Data?
3. What new policy questions are should planners anticipate over the next decade?
4. **What are the options for keeping the Census relevant?**

Strategy 1: Give up and go home

Assume that Big Data will solve everything. Do the minimum required, and leave it at that.

Bad idea!

Strategy 2: Keep calm and carry on

It's not really broken, so don't try to fix it. Consider a few tweaks:

- Add journey-to-school
- Ensure consistency with external data were possible

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

Consider a major purchase of mobility data.

Employ back-end (disaggregate) integration.

Census 'could use mobile phone data instead of questions'

7 November 2017 | UK

Share

Mobile phone data could be used in place of census questions in the future, a report from the Office for National Statistics (ONS) suggests.

Strategy 4: Administrative integration

Major push to augment the Census and ACS with other administrative data.

Similar to the LEHD.

Strategy 5: Capture the future

Modify data collection to capture emerging modes of travel, and emerging modes of work.

Strategy 6: Go long (distance)

Integrate a retrospective long-distance survey with the ACS.

Strategy 7: Go long(itudinal)

Adapt the ACS into a panel survey to provide more value-added.

Similar to the German and Dutch mobility panels.

What do you recommend?

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)

Question 1

What has been your experience in integrating Census Data with other data sources? How do these complement/supplement what Greg and Adam have found?

Question 2

Strategies. Anything missing?

Audience ranking – importance vs. most likely

Vote: Which strategy is most IMPORTANT?

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)

Vote: Which strategy is most LIKELY?

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)

Question 3

What role do these strategies and recommendations from our session today play in CTPP Board and Census Bureau decisions?

Question 4

What are the opportunities for data fusion/integration?

Example: Integrating toll transponder data or transit smartcard data with Census.

Question 5

What is one thing that the CTPP program/Census could do to better help you with your transportation planning efforts?

- Tabulations
- Tools
- Products
- Data
- Other (specify)

What do you recommend?

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)

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

Census dominant keywords

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

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