

Local Employment Dynamics: Synthetic Data for OnTheMap

Jeremy S. Wu, Ph. D.

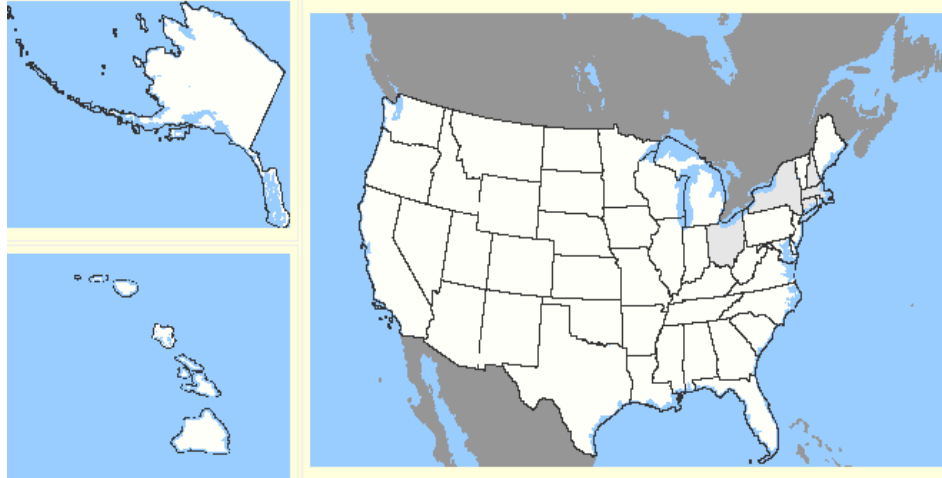
COPAFS
Washington, DC
December 4, 2009

Outline

- OnTheMap – the online tool
 - History
 - Features
- OnTheMap – the underlying data
 - Beyond censuses and surveys
 - Integrated data
 - Synthetic data
 - Future role in statistics

History of OnTheMap Releases

Major Versions	Release Date	Participating States	Data Coverage
1.0	02/03/2006	14	2002-2003
1.1.2	12/05/2006	17	2002-2003
2.0	04/15/2007	3	2002-2004
2.3	09/30/2007	42	2002-2004
3.0	09/04/2008	45	2002-2006
3.2	12/15/2008	46	2002-2006
4.0	12/14/2009	47	2002-2008
5.0	12/2010	47+	2002-2009



OnTheMap

LED's online dynamic mapping and reporting tool

- ✓ Where do workers live?
- ✓ Where do residents work?
- ✓ Reports on age, earnings, and industries
- ✓ Cross-state flows
- ✓ 46 states online
- ✓ 2002-2006 annual data
- ✓ User-selected areas
- ✓ Base unit is Census Block
- ✓ Innovative disclosure protection
- ✓ Free and available 24/7

OnTheMap: Construction/Manufacturing Workers in the Las Vegas Area

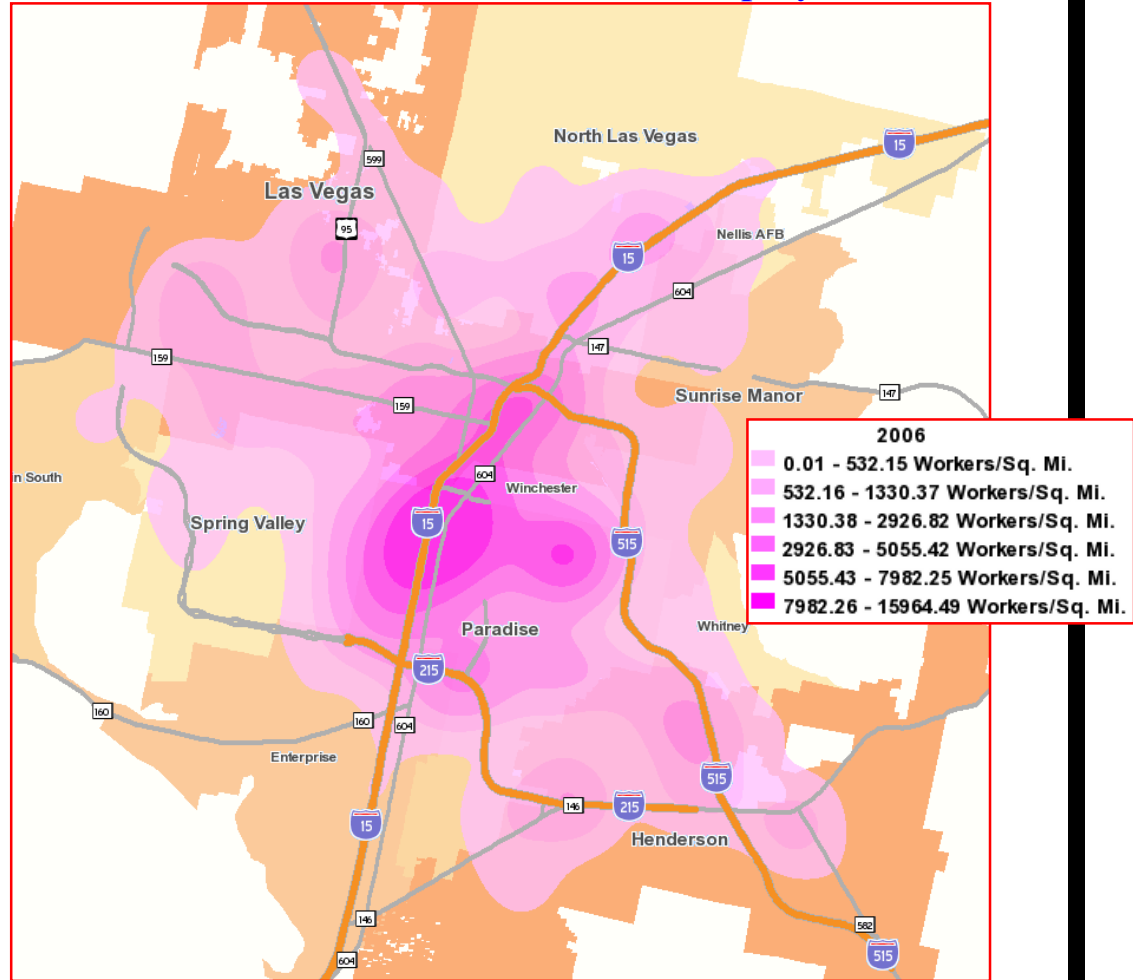
Profile of All Workers

2006		
	Count	Share
Total Primary Jobs	782,084	100.0%
Jobs by Worker Age		
2006		
	Count	Share
Age 30 or younger	213,119	27.3%
Age 31 to 54	437,018	55.9%
Age 55 or older	131,944	16.9%
Jobs by Earnings Paid		
2006		
	Count	Share
\$1,200 per month or less	152,615	19.5%
\$1,201 to \$3,400 per month	362,910	46.4%
More than \$3,400 per month	266,555	34.1%

Profile of Construction/Manufacturing Workers

2006		
	Count	Share
Total Primary Jobs	118,113	100.0%
Jobs by Worker Age		
2006		
	Count	Share
Age 30 or younger	34,862	29.5%
Age 31 to 54	67,725	57.3%
Age 55 or older	15,526	13.1%
Jobs by Earnings Paid		
2006		
	Count	Share
\$1,200 per month or less	15,127	12.8%
\$1,201 to \$3,400 per month	47,925	40.6%
More than \$3,400 per month	55,061	46.6%

Where All Workers are Employed

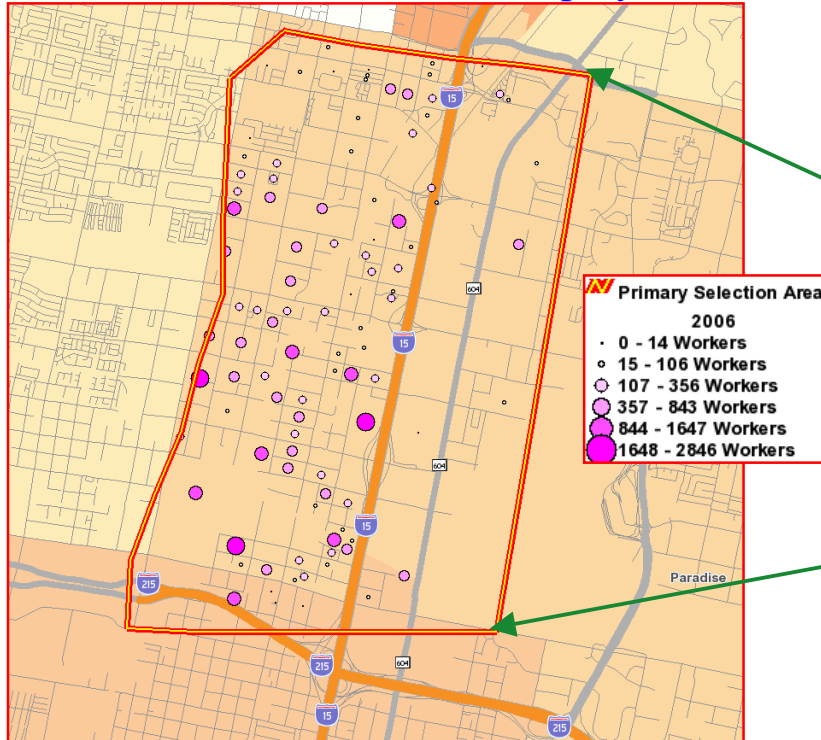


Construction/Manufacturing in the Las Vegas Area (2006-2002)

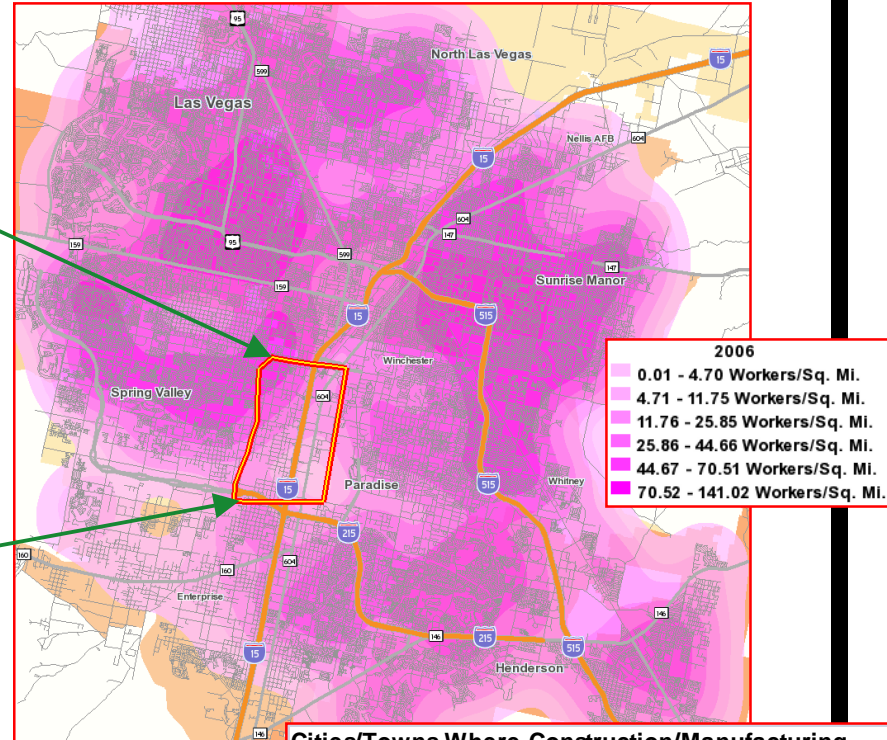
	2006		2005		2004		2003		2002	
	Count	Share	Count	Share	Count	Share	Count	Share	Count	Share
Total Primary Jobs	782,084	100.0%	743,477	100.0%	686,994	100.0%	650,254	100.0%	627,358	100.0%
Total Primary Jobs in Construction/Manufacturing	118,113	15.1%	107,175	14.4%	92,784	13.5%	81,422	12.4%	77,475	12.4%

OnTheMap: Construction/Manufacturing Workers Employed near "The Strip" in Las Vegas

Where Workers are Employed



Where Workers Live



Profile of Construction/Manufacturing Workers in NW Paradise, NV

	2006		2005		2004		2003		2002	
	Count	Share	Count	Share	Count	Share	Count	Share	Count	Share
Total Primary Jobs	32,316	100.0%	29,274	100.0%	24,776	100.0%	23,621	100.0%	22,161	100.0%
Jobs by Worker Age										
	2006		2005		2004		2003		2002	
	Count	Share	Count	Share	Count	Share	Count	Share	Count	Share
Age 30 or younger	9,450	29.2%	7,986	27.3%	6,548	26.4%	5,835	24.7%	5,275	23.8%
Age 31 to 54	18,807	58.2%	17,521	59.9%	15,189	61.3%	14,911	63.1%	14,233	64.2%
Age 55 or older	4,059	12.6%	3,767	12.9%	3,039	12.3%	2,875	12.2%	2,653	12.0%
Jobs by Earnings Paid										
	2006		2005		2004		2003		2002	
	Count	Share	Count	Share	Count	Share	Count	Share	Count	Share
\$1,200 per month or less	3,960	12.3%	3,915	13.4%	3,253	13.1%	3,135	13.3%	3,047	13.7%
\$1,201 to \$3,400 per month	12,408	38.4%	11,166	38.1%	10,023	40.5%	9,916	42.0%	9,283	41.9%
More than \$3,400 per month	15,948	49.4%	14,193	48.5%	11,500	46.4%	10,570	44.7%	9,831	44.4%

Cities/Towns Where Construction/Manufacturing Workers (employed in NW Paradise, NV) Live

	2006	
	Count	Share
Las Vegas, Nevada	9,710	30.0%
Sunrise Manor, Nevada	3,930	12.2%
Henderson, Nevada	3,852	11.9%
Paradise, Nevada	3,408	10.5%
North Las Vegas, Nevada	3,200	9.9%
Spring Valley, Nevada	2,950	9.1%
Enterprise, Nevada	733	2.3%
Whitney, Nevada	453	1.4%
Pahrump, Nevada	399	1.2%
Winchester, Nevada	319	1.0%
All Other Locations	3,362	10.4%

Census – A long History

History of the Census From Ancient Romans to GPS Technology

© [Elizabeth Linehan](#)

May 27, 2009

Today's US Census is a far cry from its humble beginnings some 2500 years ago. And it's still evolving.



The first recorded census took place in the 5th century b.c. in Rome. Under the rule of Servius

http://americanaffairs.suite101.com/article.cfm/history_of_the_census

Chinese Census during Western Han Dynasty (206 BC – 220 AD)

Population ~ 56-57 million

京兆尹，故秦内史，高帝元年属塞国，二年更为渭南郡，九年罢，复为内史。武帝建元六年分为右内史，太初元年更为京兆君。元始二年，户十九万五千七百二，口六十八万二千四百六十八。县十二：长安，高帝五年置。惠帝元年初城，六年成。户八万八百，口二十四万六千二百。王莽曰常安。新丰，骊山在南，故骊戎国。秦曰骊泉。高祖七年置。船司空，莽曰船利。蓝田，山出美玉，有虎侯山祠，秦孝公置也。华阴，故阴晋，秦惠文王五年更名宁秦，高帝八年更名华阴。太华山在南，有祠，豫州山。集灵宫，武帝起。莽曰平坛也。郑，周宣王弟郑桓公邑。有铁官。湖，有周天子祠二所。故曰胡，武帝建元元年更名湖。下BD6A，南陵，文帝七年置。沂水出蓝田谷，北至霸陵入霸。有铁官。北入海。有铁官。更名以章霸功。视子孙。奉明，宣帝置也。霸陵，故芷阳，文帝更名。莽曰水章也。杜陵。故杜伯国，宣帝更名。有周右将军杜主祠四所。莽曰饶安也。

http://www.guoxue.com/shibu/24shi/hansu/hsu_038.htm

左冯翊，故秦内史，高帝元年属塞国，二年更名河上郡，九年罢，复为内史。武帝建元六年分为左内史，太初元年更名左冯翊。户二十三万五千一百一，口九十一万七千八百二十二。县二十四：高陵，左辅都尉治。莽曰千春。栌阳，秦献公自雍徙。莽曰师亭。翟道，莽曰涣。池阳，惠帝四年置。E742DE65山在北。夏阳，故少梁，秦惠文王十一年更名。《禹贡》梁山在西北，龙门山在北。有铁官。莽曰冀亭。衙，莽曰达昌。粟邑，莽曰粟城。谷口，九B077山在西。有天

Sampling – The Introduction

International Statistical Institute, 55th Session 2005
International Statistical Institute, 55th Session 2005: Colm O'Muircheartaigh

Balancing Statistical Theory, Sampling Concepts, and Practicality in the Teaching of Survey Sampling

Colm O'Muircheartaigh
University of Chicago, NORC and Harris Graduate School of Public Policy,
1155 East 60th Street, Chicago IL 60637, USA

colm@norc.uchicago.edu

抽样调查理论与实践 (1995), 中国统计出版社。
ISBN 7-5037-1670-3

抽样调查一百周年

胡善庆 著
张尧庭 于祥 译

1995 年将是统计史上一个重大转折的一百周年纪念。依照 You Poh Seng (1951)、以及卡尔顿 (Graham Kalton) (1988) 的论述, 这是因为 1895 年, 在瑞士伯尔尼召开的国际统计学会的报告会上, 挪威统计局局长 A. N. Kiaer (凯尔) 在阐明他的“代表性调查”的思想时, 在历史上首次引入了抽样的概念, 并引起了国际统计学界的关注。

2. History of Survey Sampling

The origins of survey sampling are in official government statistics in the late 19th and early 20th centuries; the provenance is non-mathematical, and the motivation is practical and operational. [A more extensive version of the material in this section can be found in O'Muircheartaigh and Soon (1981).]

One of the finest discussions of the philosophy of sampling took place at the International Statistical Institute (ISI) meetings in Berne in 1895. This was the first recorded occasion on which the statistics establishment (then primarily the collectors of data for government purposes) was confronted with a coherent plea for the use of samples in data collection. A N **Kiaer**, the director of the Norwegian Bureau of Statistics, presented a report on his experience with sampling surveys and advocated further investigation of the field. Kiaer's definition of a representative investigation was that it was a partial inquiry in which observation was made on a large number of units distributed

Sampling – The Reception

Michael Cowles (1936, 2001). *Statistics in Psychology: an historical perspective*. Lawrence Erlbaum Associates, Inc. New Jersey.
http://books.google.com/books?id=8pRuRm6WAp8C&dq=a.+n.+kiaer&source=gbs_navlinks_s

pling, pressing for complete enumeration. It took almost 30 years for the utility and benefits of the methods to be appreciated. Seng (1951) and Kruskal and Mosteller (1979) give accounts of this most interesting period in statistical history. The latter authors give a translation and paraphrase of the remarks of Georg von Mayer, Professor at the University of Munich, on Kiaer's work on the *representative method*, which was presented at a meeting of the Institute in Berne in 1895:

I regard as most dangerous the point of view found in his work. I understand that representative samples can have some value, but it is a value restricted to terrain already illuminated by full coverage. One cannot replace by calculation the real observation of facts. A sample provides statistics for the units actually observed, but not true statistics for the entire terrain.

It is especially dangerous to propose representative sampling in the midst of an assembly of statisticians. Perhaps for legislative or administrative goals sampling may have uses – but one must never forget that it cannot replace a complete survey. It is necessary to add that there is among us these days a current in the minds of mathematicians that would, in many ways, have us calculate rather than observe. We must remain firm and say: no calculations when observations can be made. (von Mayer, quoted by Kruskal & Mosteller, 1979, pp. 174–175)

Sampling – Acceptance in the U.S.

Developing Sampling Techniques

The Census Bureau first used statistical sampling methods in the 1937 test survey United States during the Great Depression, but it also served as a "check" on a lar

The Census Bureau implemented statistical sampling in a decennial census for the increasing cost or respondent burden. Enumerators asked a random sample of the demographic data for the entire United States.

Sampling became a fixture of the decennial censuses, with a certain pe century. In fact, because the American Community Survey is now the in asked of a sample of the population.

The success of statistical sampling in the decennial census contributed expanded and renamed the Current Population Survey in 1947. The Cer figures.

http://www.census.gov/history/www/innovations/data_collection/developing_sampling_techniques.html



[View larger image](#)

The Census Bureau used statistical sampling in the 1937 Survey of Unemployment.

Sampling: Evolution in U.S. Census

- 1940 5% sample on occupation, ...
- 1950 20% sample on income, schools, ...
1 in 6 sample of 20% on marriage, ...
- 1960 25% sample ...
- 2000 “Short” form – for all
“Long” form – 1 in 6 sample
- 2005 American Community Survey
~1 in 8 sample over 5 years

Two Lessons

A 5% random sample
is “better” than
a 5% non-random sample
in measurable terms

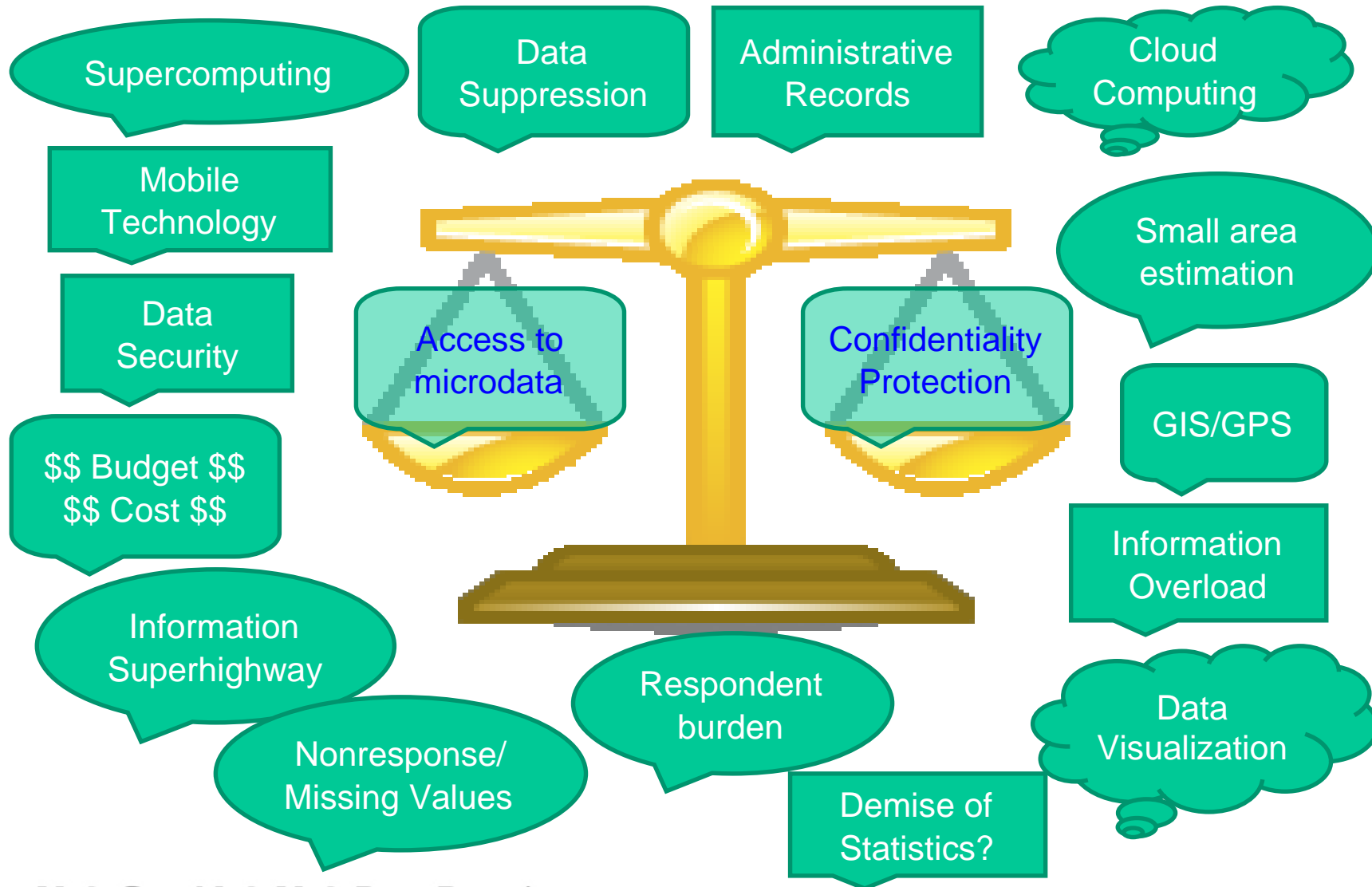
Study of statistics was saved;
Mathematical Statistics was born

Study of Statistics

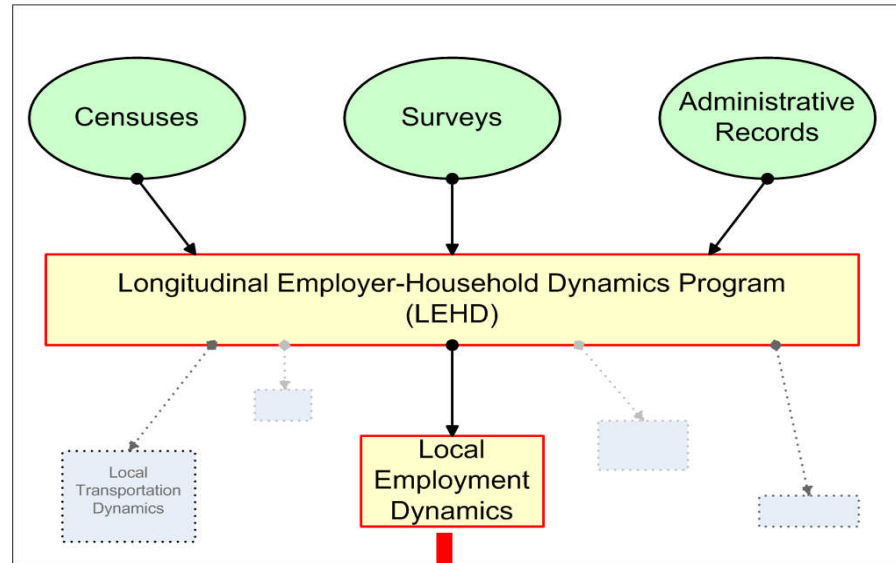
We all live happily ever after!

Or do we?

The Information Re/Evolution



The LED Approach

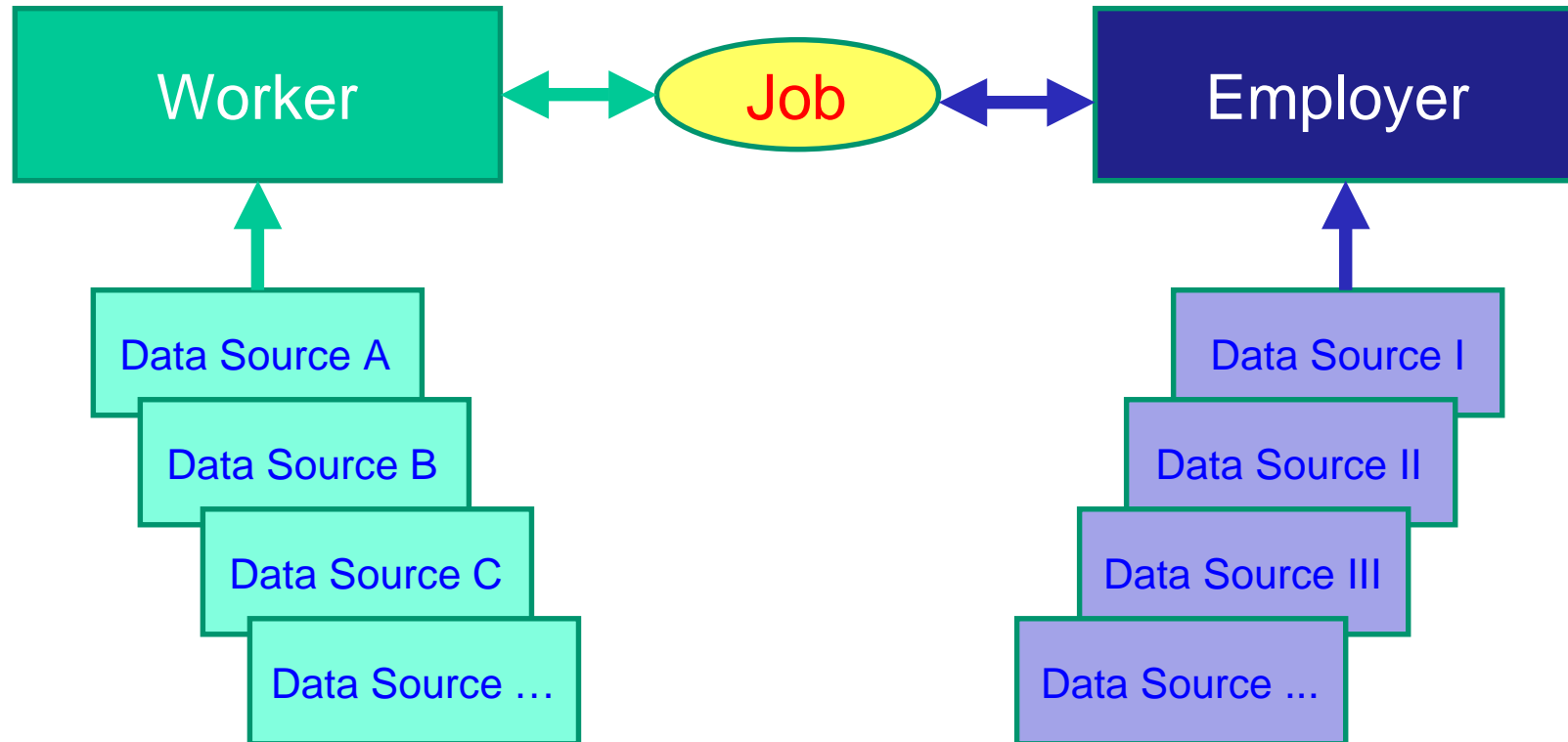


- Started in 1998
- Use existing data
- Create new data and products
- Make valid detailed data available and protect confidentiality

Longitudinal National
Frame of Jobs

New data and products

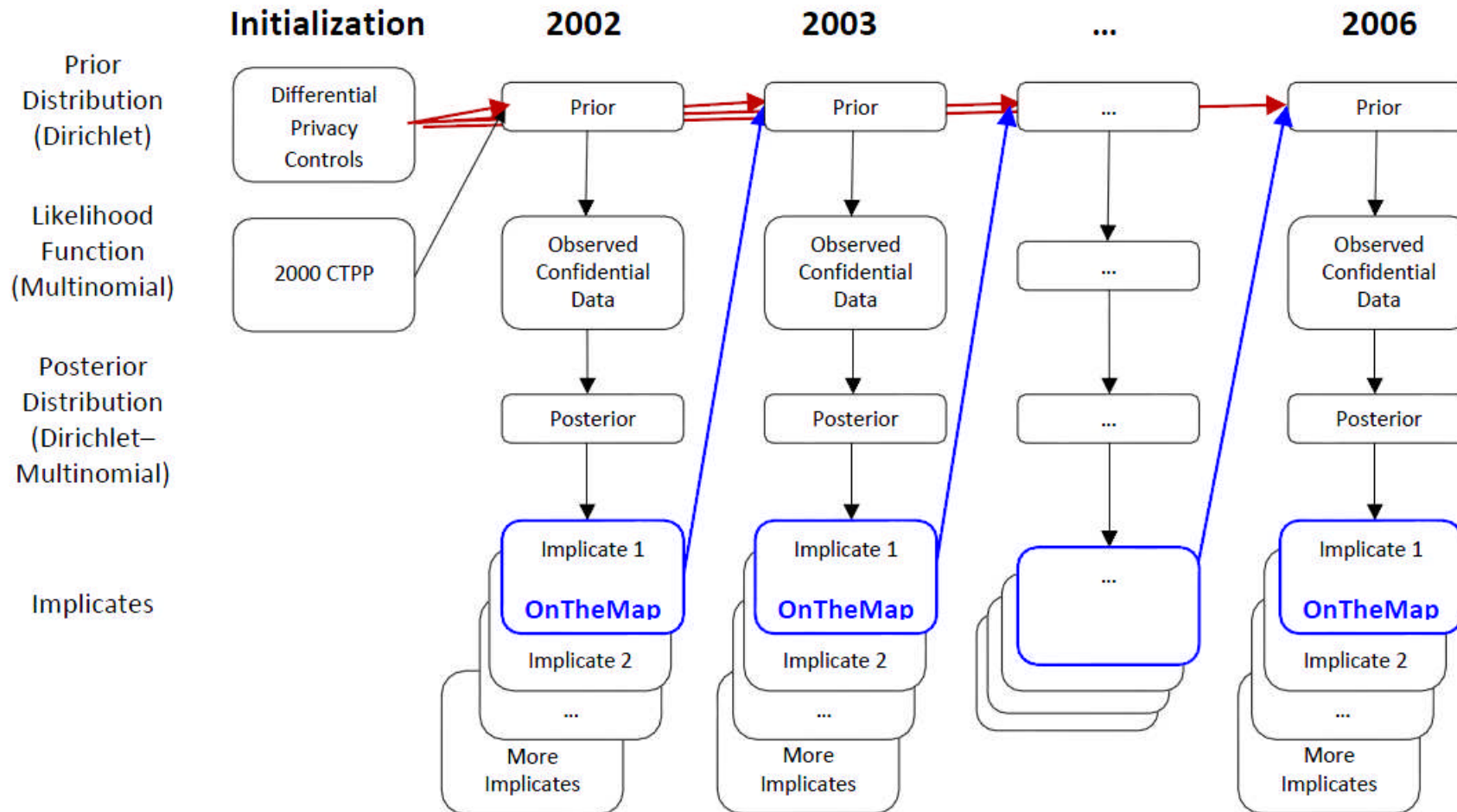
LED: Longitudinal Integrated Data



LED: Key Methodologies

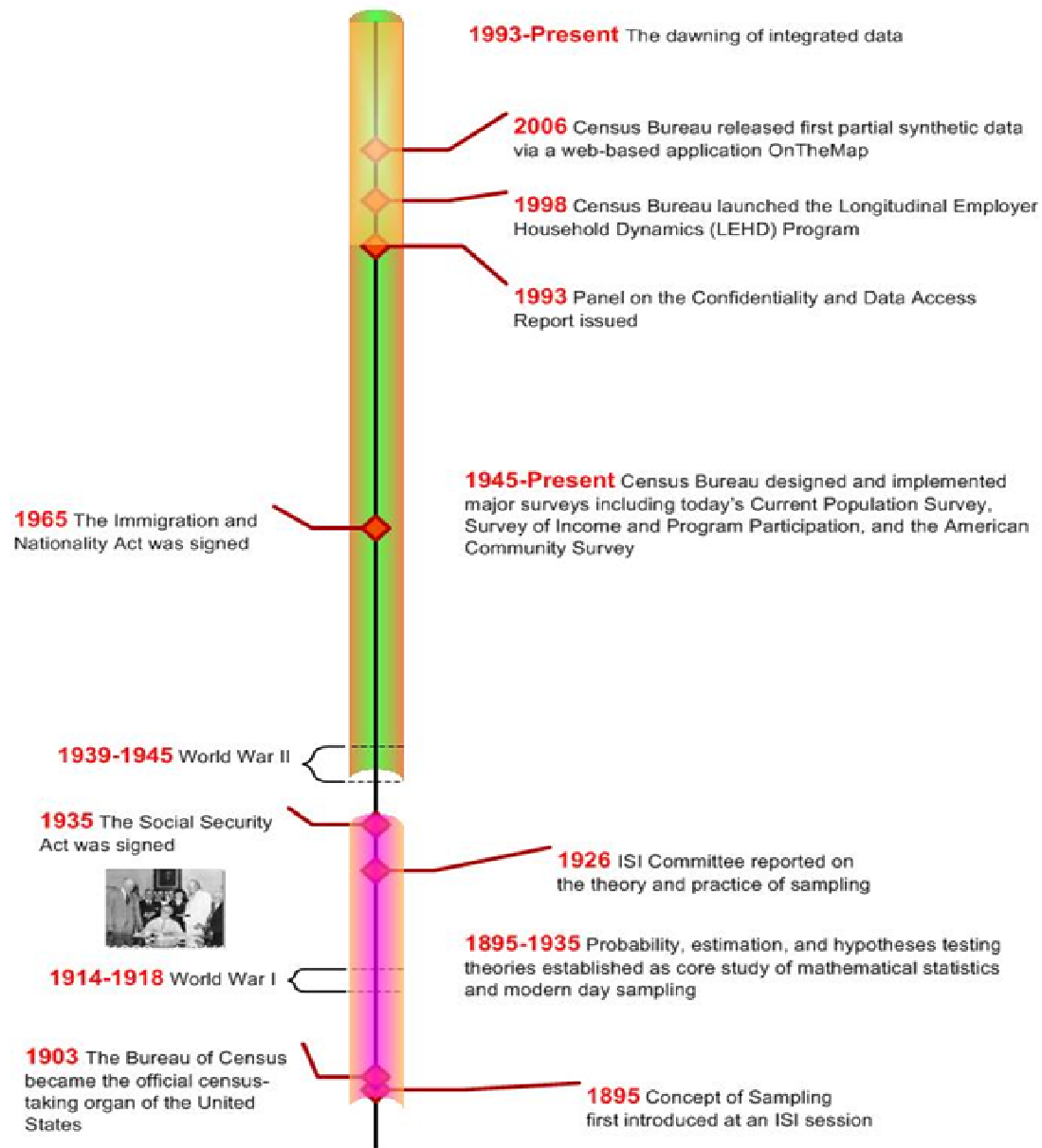
- System Design
- Record Linkage
- Noise Infusion
- Imputation
- Synthetic Data Modeling
- Measure of Goodness and Quality

OnTheMap: Partial Synthetic Data Product



Comparison of ACS and LEHD Data

- NCHRP 08-36/Task 81 Report
- “Enhancing the American Community Survey Data as a Source for Home-to-Work Flows,” Cambridge Systematics, Inc, 10/01/2009
 - [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(81\)_FR.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(81)_FR.pdf)
 - 2004 OnTheMap data (Version 2)
 - 2006 ACS data



The Unknowns

- Is an 80% non-random sample “better” than a 5% random sample in measurable terms? 90%? 95%? 99%?
- Is this the dawning of another new field of statistical knowledge?

Access to LED via <http://www.census.gov>

The screenshot displays the U.S. Census Bureau website interface. At the top left, the logo reads "U.S. Census Bureau". On the right side of the top navigation bar, there are links for "FAQs", "Subjects A to Z", "Help", and a search box with a "GO" button.

The left sidebar contains several navigation categories: "Families and Living Arrangements", "New on the Site", "Data Tools", "American FactFinder", "Jobs@Census", "Catalog", "Publications", "Are You in a Survey?", "About the Bureau", "Regional Offices", "Doing Business with Us", and "Related Sites".

The main content area is organized into several sections:

- United States Census 2010**: Includes links for "2010 Census", "News", "American Community Survey", and "Census 2000".
- People & Households**: Includes links for "Estimates", "Projections", "Housing", "Income | State Median Income", "Poverty", "Health Insurance", "International", "Genealogy", and "More".
- Business & Industry**: Includes links for "Economic Census", "Get Help with Your Form", "Economic Indicators", "NAICS", "Survey of Business Owners", "Government", "E-Stats", "Foreign Trade", "Export Codes", "Local Employment Dynamics", and "More". The "Local Employment Dynamics" link is circled in red.
- Geography**: Includes links for "Maps", "TIGER", "Gazetteer", and "More".
- Newsroom**: Includes links for "Releases", "Facts For Features", "Minority Links", "Broadcast & Photo Services", "Embargo/News Release Subscription", and "More".
- Special Topics**: Includes links for "Phishing & Email Scams", "Census Bureau Data and Emergency Preparedness", "Events Calendar", "Training", "For Teachers & Students", "Statistical Abstract", "FedStats", "USA.gov", and "Recovery Act at the Census Bureau".

On the right side, the "Data Finders" section is prominent, showing "Population Clocks" with the following statistics:

- U.S. **307,616,761**
- World **6,788,321,738**
- 13:28 GMT (EST+5) Oct 04, 2009

Below the population clock is a "Population Finder" section with input fields for "city/ town, county, or zip" and "or state", and a "Select a state" dropdown menu with a "GO" button.

Further down is a "Find An Area Profile with QuickFacts" section with "Select a state to begin" and "Select a state" dropdown menus.

The "Latest Economic Indicators" section lists:

- Manufacturers' Shipments, Inventories, and Orders
- Construction Spending

At the bottom of the "Data Finders" section, there is an "Economic Indicators" section with two "Select an indicator" dropdown menus.

The footer of the website includes the text "USCENSUSBUREAU Helping You Make Informed Decisions" on the left, and "Accessibility | Information Quality | FOIA | Data Protection & Privacy Policy | U.S. Dept of Commerce" on the right.

Contact Us

Comments/Suggestions

CES.local.employment.dynamics@census.gov

Jeremy.S.Wu@census.gov

Local Employment Dynamics Website

<http://lehd.did.census.gov>

Join the Listservs

Lehd-general@lists.census.gov

Lehd-onthemap@lists.census.gov