# Can They Count on You?

**What does a depository librarian need to know about statistical data and data sets to provide basic service to users?**

- **Beth Williams, Library Director, Stanford Law School**
- **Rich Gause, Government Information Librarian, University of Central Florida**

Beth Williams
Library Director
Stanford Law School

Rich Gause
Government Information Librarian
University of Central Florida

# GODORT Competencies

**U.S. Federal Government Information Competencies for Beginning Government Information & General Reference Librarians**

GODORT Education Committee, approved at ALA Annual 2011

http://wikis.ala.org/godort/index.php/Education

(Projects, Competencies, Pt. 1)

# GODORT Competencies

Essential (learn as quickly as possible)

- Knows how to locate reliable sources of statistical information

- Knows where to seek training

- Knows how to reach live people in the U.S. Government who can help with information

# **GODORT Competencies**

Recommended (learn within two to four years)

- Understands issues and terminology related to statistical information

- Acquires a deep and broad knowledge of the resources and tools pertaining to Government information

Focus of this session:

- Are there levels of competency?

- What are some troubleshooting issues?

- Where can depository librarians learn more about specific resources and tools?

**Questions for virtual participants:**

- Are there terms regarding data and statistical analysis that are confusing to users?

- What are some typical issues you deal with regarding data questions at your library?

# Possible Levels of Competency

- **Awareness** – recognizes issues & terminology; knows some places to begin searching

- **Understanding** – can explain issues & terminology with examples; can match questions with potential sources

- **Data retrieval skills** – can guide users in retrieving results from specific databases & sites

- **Data analysis skills** – can guide users regarding how to prepare data for analysis, which types of analysis to apply, how to use specific analysis programs (e.g., DataFerrett, SPSS, SAS, Stata)

# Competency levels will vary

**Data retrieval skills** – focus on topical areas relevant to your specific users

- Population demographics
- Economics & business & labor
- Health
- Education
- Criminal justice
- Agriculture
- Other scientific data sets

# Competency levels will vary

**Data analysis skills**
can guide users regarding how to prepare data for analysis, which types of analysis to apply, how to use specific analysis programs (e.g., DataFerrett, SPSS, SAS, Stata)

- These may be more complex skills than most depository librarians will want to tackle. BUT…

- Do you know where to send the user for assistance? Is there someone else in your library, e.g., data librarian, business librarian, social sciences librarian? At your institution? At another depository? At a Government agency?

# Competency levels will vary

You might decide to not progress beyond basic awareness for some aspects.

For example, you might be aware that *data cleaning* is something to be done to retrieved data before it can be analyzed by a program. You might be aware that data cleaning can be a time-consuming process. However, you might never pursue developing an understanding of the specifics.

# Troubleshooting Issues

- Does requested data exist?
- Data tables/reports vs data sets
- Comparing data over time
- Currency of data
- Choice of geography

# Why is the user not finding data?

- Looking in the wrong place
- Searching with the wrong terms
- Data was not collected for that time period or that population or that geographic area
- Data is not freely accessible
- Data does not exist

# **Data tables/reports vs data sets**

- Ready-made statistical tables, charts & graphs summarize and interpret data. The raw data has been cross-tabulated by characteristics.

- Data sets are the raw data in machine-readable files that users can analyze with software programs.

# Existing data tables vs PUMS

- Public Use Microdata Sample (PUMS) files are made available so that data users can create custom tables that are not available through pre-tabulated or summary data products.

- Social scientists often use the PUMS for regression analysis and modeling applications.

- Statistical software is needed to use the PUMS files.

# Comparing data over time

Decennial Census

- Limited census tracts 1950-1960

- Census tracts/BNAs for entire U.S., 1990+

- Long form (detailed topics) ceased with 2000

American Community Survey

- 5-year data begins with 2009

- 3-year data began in 2007 & ceased in 2013

# Which years had which questions?

*Measuring America: The Decennial Censuses from 1790 to 2000*
https://www.census.gov/history/pdf/measuringamerica.pdf
pages 119-124, tables showing which population questions were asked

**Population Items on Principal Census Questionnaires: 1900 to 2000**—Con.

| Economic characteristics | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Industry 5 years ago | - | - | - | - | - | - | - | - | - | - | ˢX |
| Occupation 5 years ago | - | - | - | - | - | - | - | ˢX | - | - | - |
| Class of worker 5 years ago | - | - | - | - | - | - | - | ˢX | - | - | - |
| Value of real estate | - | - | - | - | - | - | - | - | ˢ ⁹X | ˢ ⁹X | ˢ ⁹X |
| Value of personal property | - | - | - | - | - | - | - | - | - | - | - |
| Income | - | - | - | - | X | ˢX | ˢX | ˢX | ˢX | ˢX | ˢX |

# Changes in census topics

- Hispanic
  - 1960 – Spanish surname in five states: Arizona, California, Colorado, New Mexico, & Texas
  - 1970 – Origin or descent: Mexican, Puerto Rican, Cuban, Central or South American, Other Spanish, None of These
  - 1980 – Hispanic category on short form
- Poverty
  - 1960 – not defined
  - 1970 – % families in poverty
  - 1980 – % living in poverty

# Changes in census topics

- Same-sex married couples
  - 1990 – if reported as a same-sex married couple, the sex of one individual was edited to report as opposite-sex married couple
  - 2000 – if reported as same-sex married couple, the relationship of the couple was edited to reflect a same-sex unmarried partnership
  - 2010 – estimates of same-sex married couples were published in special reports/tables

# Currency of data

Where is the data for 2016-2017?

- American Community Survey release schedule
  - 2016 1-year data – Sep 14, 2017
  - 2016 1-year supplemental & PUMS – Oct 19, 2017
  - 2012-2016 5-year data – Dec 7, 2017
  - 2012-2016 5-year PUMS – Jan 18, 2018

- Economic Census release schedule
  - 2012 reports – March 2014 through June 2016
  - 2017 reports – 2019+

- American Housing Survey, biennial in odd years
  - 2015 – Miami-Ft. Lauderdale
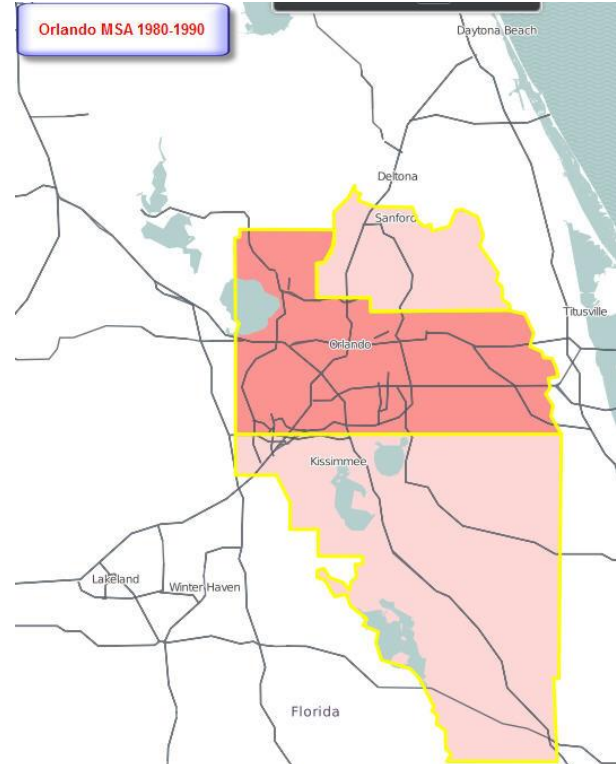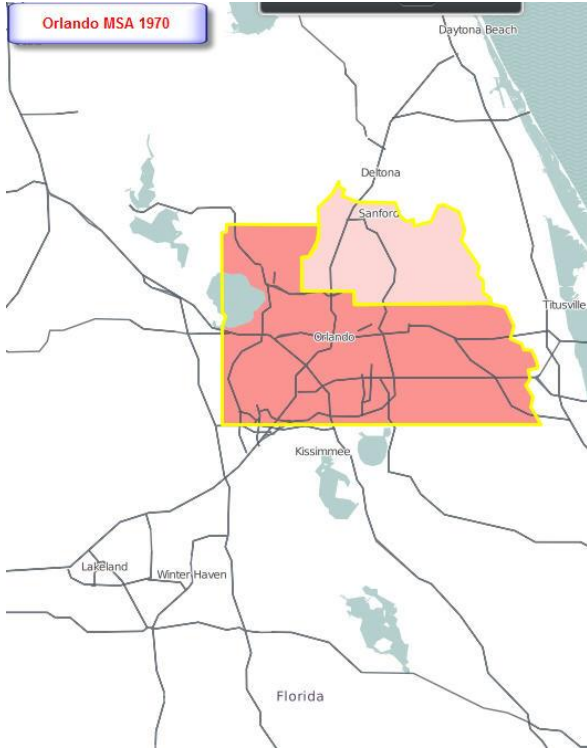  - 2013 – Miami-Ft. Lauderdale, Orlando, Tampa-St. Petersburg

# Choice of Geography

- County vs Metropolitan Area
- Census Tract vs ZIP Code

# Growth of Orlando MSA 1970-2000
## - aggregate the counties instead of using MSA
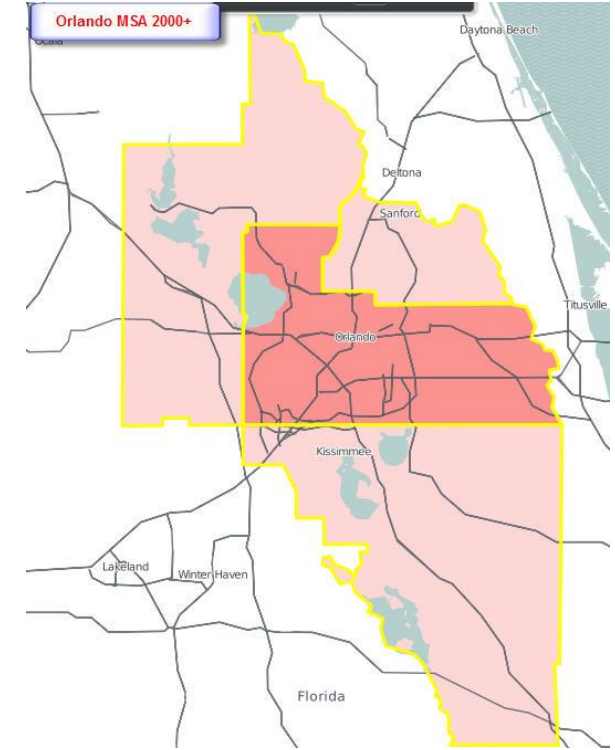
# Reasons to use Census Tracts instead of ZIP Codes in analysis

- 50% more census tracts than ZIP codes

- More complete coverage of U.S.; ZIP codes don't cover geographic areas without mail delivery

- Better aggregation of geographic boundaries for comparisons with municipal & county data. Census tracts don't cross county lines; 15% of Orange County, FL ZIP codes cross into other counties

- In 2010 there were only 10 states with no ZIP codes crossing into other states

- Understandable boundary changes over time

# Training about Data

- GPO Webinars & Webcasts / FDLP Academy
  http://login.icohere.com/public/topics.cfm?cseq=1172

- Help! I'm an Accidental Government Information Librarian Webinars
  http://www.nclaonline.org/government-resources/help-im-accidental-government-information-librarian-webinars

**FDLP Academy – a selection of previous depository library community webinars related to data:**

- Librarian's Guide to Trade Data
- Using OnTheMap for Community Analysis
- Mapping Census Data: Unleash your Sociodemographic Knowledge through the Power of GIS
- Introducing the U.S. Bureau of Labor Statistics' Databases, Tables, & Calculators Search Tool
- Using the BEA Interactive Data Tool to get Local Data
- CDC American Indian & Alaska Native Health Data Online
- Selected Advanced Features of USA Trade Online & DataWeb
- Hitting Pay Dirt - Text Mining Government Sources
- Census Bureau Statistics in Schools
- StatsAmerica - A Portal to Apps & Data Rich Tools for Economic & Community Development

**FDLP Academy – a selection of previous Federal Agency webinars related to data:**

- Accessing Public School System Finances Data

- Measuring America Series
  - Accessing County Business Patterns
  - Accessing Fertility Statistics
  - Accessing International Data

- Federal Government Open Data Update

- Introduction to the U.S. Census Bureau's DataFerrett

- Using American Community Survey (ACS) Estimates & Margins

- Introduction to Veteran Statistics: Decennial Census & American Community Survey

**Help! I'm an Accidental Government Information Librarian webinars – a selection of previous webinars related to data:**

- NASA Astrophysics Data System: Linking Publications to Data
- Introduction to Research Data Management for Librarians
- International Government Survey Data: How to Find & Use It
- The APIs of Data.gov
- Statistics Canada data for researchers beyond the Canadian borders
- Making Peace with United Nations data: Learn to navigate UN data
- Accessing Datasets for the Data Curious

# Volunteer to present FDLP virtual training
# or
# Request a specific topic be presented via the FDLP Academy

https://www.fdlp.gov/component/rsform/form/32-request-for-fdlp-training-and-use-of-gpo-s-virtual-training-room

# **Links to Additional Resources**

- Data Sources from the Federal Government
  http://guides.ucf.edu/govdata