



**Title:** "Data Visualizations of the ACS, Redistricting file, Shapefiles, Bayesian Improved Surname Geocoding (BISG) proxy method and Animated Venn Diagrams using the Census Surname Lists".

**Date/Time:** February 27, 2020 10:00 to 11:30

**Speakers:** Martha Dusenberry Pohl

**Chair:** Wendy L. Martinez, Bureau of Labor Statistics

**Sponsor:** WSS Defense and National Security & the FCSM Geospatial Interest Group

**Abstract:** Data visualization examples including animation will be provided using public Census data at the ZCTA (Zip Code Tabulation area), County and Block levels. Census data sources include Shapefiles, Redistricting file, American Community Survey (ACS), and the Decennial Census.

The methodology to compute a proxy for missing race and ethnicity released in the summer of 2014 uses the most frequently occurring surname in the United States of SMITH and the state of California as an example of this computation. This methodology, called the "Bayesian Improved Surname Geocoding (BISG) proxy method", uses Bayes' Theorem, the Census Surname List by Race Ethnicity and the racial/ethnic composition of the population of the geographic area. This presentation will look at the Surname of SMITH (as well as other surnames) at the County and Zip Code (ZCTA) levels nationwide to examine these changes by geographic location and concentration. Data Scientists continually look to enhance the capability of analyzing data. This can be done by dynamic mapping over time to visualize changes. Geographic visualizations will include animations of the racial/ethnic composition of the population by county over time using the ACS 5-year Estimates. An example using the Census Redistricting file using Shapefiles at the Block level will be conducted.

We will take a preliminary look at 4 of the Census Surname lists ("Names Files", listed below) using animated Venn diagrams.

- #1. 2010 Census Surname List ("2010 List") consisting of 162,253 observations of those surnames appearing on the 2010 Census at least 100 times;
- #2. Census 2000 Surnames List ("2000 List") consisting of 151,671 entries of those surnames appearing 100 times or More in the Census 2000;
- #3. 1990 Census Heavily Hispanic List contains 639 surnames "Heavily Hispanic List";
- #5. 1980 Census "Passel-Word Spanish surname list with 12,497 entries.

All calculations and visualizations were conducted using SAS 9.4. The SAS code for these procedures is included.

\* Disclaimer: Any opinions expressed in this paper are those of the author and do not constitute policy or opinion of the U.S. Department of Justice or any of its subcomponents.

**Location:** Bureau of Labor Statistics Janet Norwood Conference Center, Rooms 7 & 8

To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to e-mail your name, affiliation, and seminar name to [wss\\_seminar@bls.gov](mailto:wss_seminar@bls.gov) (underscore after 'wss') by noon at least two days in advance of the seminar. Please bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station. Parking in the area of BLS is available at Union Station. For parking information see <http://www.unionstationdc.com/parking>. No validation is available from BLS for reduced parking rates.

**IMPORTANT:** BLS is now following enhanced security protocols and screenings for access to the facility. All personnel need to remove all items from their pockets/person (to include belts) and electronic devices from their person or clothing and place the items in the designated bin to be screened through the x-ray machine. Please allow extra time to get through security.

**Remote Access:**

Join Zoom Meeting

<https://zoom.us/j/870812001?pwd=NkI3aWc2eDBWODE1QWITbnRUQWJndz09>

Meeting ID: 870 812 001

Password: 298637

Find your local number: <https://zoom.us/u/arOBjDFU7>

Audio available via computer or phone

Note: Particular computer configurations might not be compatible with Zoom.