Dynamic Census Data Reports: Methodology

About this Methodology

Our dynamic Census data reports allow users to obtain a wide variety of socioeconomic and demographic indicators for existing places, custom user-defined polygons, or areas within a specific distance from a point. Census data is made available at predefined geographic boundaries such as blocks, block groups, tracts, cities, counties and zip codes. When users want to generate a report for a predefined area to which the data has already been summarized, it is a simple task to return the data as originally provided by the Census Bureau (see Attribute Queries). However, allowing users to generate a report for an area to which the data has not already been summarized is a more complex task, especially at quick speeds and for huge geographies that can span across the state (see Geometry Queries). This methodology explains how this is accomplished, and why it can result in discrepancies between the numbers provided directly by the Census Bureau.

2010 Census SF1 and Census ACS 5-Year

U.S. Census data is loaded for all counties in the State of Texas. For Summary File 1 (SF1), eleven files were loaded that provide data needed to populate Census Summary Reports. For the 2010 American Community Survey (ACS) 5-Year data, 51 tables were loaded that populate the ACS data reports.

Attribute Queries

An attribute query returns data for a predefined boundary to which the data was originally summarized. Both 2010 Census SF1 and 2010 ACS are summarized by a variety of predefined geographies. In our reports, the Census predefined geographies are used and data is assigned to blocks, block groups, tracts, cities, counties, and zip codes (zip codes are not included in ACS). When standard Census geographies are selected the values in the data are reported. For example, when a user queries data for Austin, Texas the report will include the same summarized totals provided by the Census Bureau for that city.

Geometry Queries

All reports for areas to which the Census data has **not** already been summarized are powered by geometry queries that use spatial database tools to dynamically calculate proportional totals based on geographic intersections of polygons. In order to facilitate custom geography reports at the fastest possible speeds, data has been proportionately assigned to over 68 million small grid polygons across Texas, approximately 500 ft x 500 ft each. These grids were assigned numerical attributes based on the proportion of census geography land area intersecting each grid. For SF1 data, the primary geography is census blocks. For ACS, the primary geography is block groups or tracts. For example, if fifty percent of a block group with a total population of 500 overlaps a grid cell, this grid cell is assigned a total population value of 250. When a user runs a report for a custom shape they have drawn on the map, for a specified distance around a point they have clicked, or for a named place that was not predefined by the Census Bureau, the system selects all of the grids **whose center points (centroids) are contained inside the query area** and generates a sum total for each indicator. **A variable margin of error is introduced in this step because we include any grid cells that may not be entirely contained by the query area but whose centroids are included, and because we likewise exclude any grid cells who are partially contained by the query area but whose centroids are not.** This margin of error is typically less than 1%. To continue on the previous example, if a user selects an area containing the centroid of the grid with a total population of 250 along with the centroids of 3 more grids that have total population values of 100 each, the resulting report total would be 550. If we attempted to generate these totals without first

summarizing to this grid-based system, these reports could take several minutes to calculate and would require much more computing and server power.

The variety of summarized geographies (blocks, block groups, tracts) provided by the Census can be used in different situations. For larger areas, we can take advantage of the data summarized at block group or tract level, again improving the performance and the representation of the original census data. In other words, when larger areas are selected, data provided by the Census for larger areas are used.

In addition to allowing a user to draw a polygon on the map, the system also has access to a database of special geographies. These include school districts, legislative districts, and more. These can also be intersected with the grid and provides a large number of summary options.

Explanation for Differences in Reporting Totals:

ACS versus SF1 General Reports

ACS provides different numbers than SF1 because they are gathered using different methods and for different purposes.

- SF1 is a total count of the population on a specific date.
- ACS is a sample survey meant to provide the characteristics of the population over a time period.

Over time, ACS and SF1 numbers will begin to differ more and more, as ACS reflects updates, whereas SF1 remains a 2010 sum.

1. SF1 is a total count based on the smallest possible census tabulation geography (block) at one year, 2010.
2. ACS is an estimate across a larger census area (block group) based on a sampling of data, and is updated every year going forward.

For example, 735,512 may be the estimate for a county based on the samples taken for the ACS, but the SF1 value of 778,803 may be the actual number based on the counts made in 2010. Next year, SF1 will still say 778,803 (the 2010 value), but ACS may change to say 752,000 or something similar based on the sampling technique used for a year later update.

In addition, each of the ACS figures reported by the Census Bureau is accompanied by a margin of error. For the Connector, the reported estimates are used in the calculations and treated as likely values, but in reality the true value lies somewhere in the reported margin of error range.

Helpful Links

http://www.census.gov/acs/www/about_the_survey/american_community_survey_and_2010_census/

http://www.census.gov/acs/www/guidance_for_data_users/estimates/

Difference in Summary Levels
(Specifically zip codes)

The Census Bureau provides ZIP code tabulation for SF1, but does not do so for ACS.

Within the system, the ZIP code summary showing in the ACS report is an estimate of the ACS data that intersects the chosen ZIP code boundary (75070). A comparison of total population between the SF1 ZIP code summary and an ACS ZIP code summary will most likely show differences.

Because of the differences, ACS and Decennial Census figures are difficult to compare. Unlike the 2010 Census, the purpose of ACS is not to count population but to estimate characteristics of the population. The numbers in SF1 and ACS will not be exactly the same, but each can be used to describe an area.

ACS data Summarized at different levels:

Some of the data in the general ACS reports are summaries of smaller geographies than others. Some of the data is available from the Census Bureau at block group level. Some data are available only at larger geographies like tracts. The report might contain data from both block groups and tracts. In this example, a query for the same polygon reveals:

- Total Population by Sex and Age as 735,512
- Total Population in Households by Relationship as 810,038

  Naturally, these represent two different population numbers, summarized by different criteria. The reason they are different is due to the level at with the data are available.

- Based on the data provided by **Block Groups** in Collin County, the population is estimated at 735,512

- Based on the data provided by **Tracts** in Collin County, the population is estimated at 810,038

The data for the block groups are available at smaller, specific areas, and tend to look more precise. The data for the tracts are summarized at larger areas, so the estimated proportion in the user polygon query is smaller.

Differences in Standard Geography:

One final source of difference is related to the digital geographic files themselves. A polygon drawn on the map or stored in the database might not exactly match the digital files the Census Bureau used when performing its tabulations. This is a common geographic issue. Boundary files from different sources were used in the preprocessing steps and are also being used in the dynamic actions. Generally, the differences resulting from inconsistencies in the digital geographic files are small.

Disclaimer

NCTCOG is not responsible for the accuracy of data reported in the Texas Connector or any of its products. The U.S. Census data used in this tool comes from the Decennial Census and the continuously-administered American Community Survey (ACS), which surveys a subset of the population. ACS responses are collected over a period of time from a randomly-selected sample of households. Along with

techniques used to derive characteristics for the whole population from the responses of the sample population, the Census Bureau employs a variety of disclosure avoidance techniques to protect individual responses. Due to the nature of the data collection process, there is a level of uncertainty associated with the data. For each ACS data item, the Census Bureau releases an associated “margin of error” to assist data users in drawing conclusions about the data. Unlike the decennial census, which attempts to provide a complete *count* of the population, the ACS was designed to provide estimates of the *characteristics* of the population. Because of this, the Census Bureau recommends that users compare derived measures such as percents, means, medians, and rates rather than estimates of the population totals. Neither the North Central Texas Council of Governments nor OneStar Foundation makes any warranty, express or implied, including warranties of merchantability and fitness for a particular purpose. Responsibility for the use of these data lies solely with the user.