3. CASE STUDIES USING ACS DATA

Case Study #1: Minnesota State Demographic Center Analysis of Earnings in Urban and Rural Areas

Skill Level: Intermediate/Advanced
Subject: Earnings, Rural-Urban Geographic Areas
Type of Analysis: Making comparisons across geographic areas
Creating custom geographic areas from census tracts
Calculating margins of error for derived estimates
Tools Used: Data.census.gov, spreadsheet, U.S. Census Bureau's Statistical Testing Tool
Author: Susan Brower, State Demographer of Minnesota

Susan is the State Demographer of Minnesota. She wants to study how earnings differ across geographic regions of the state. She plans to use a rural-urban typology that corresponds to the characteristics of individual census tracts.

Susan will use Rural-Urban Commuting Area (RUCA) classification codes developed by the U.S. Department of Agriculture's (USDA) Economic Research Service (ERS) to examine economic characteristics of Minnesota residents living in a range of settings—from remote, rural areas to dense, urban cities. RUCA codes classify census tracts using measures of population density, urbanization, and commuting patterns. Susan will aggregate characteristics of residents across the state, based on the RUCA code of the census tract in which they live. (More information about RUCA codes can be found on the ERS Web page on Rural-Urban Commuting Area Codes.)[16]

[16] U.S. Department of Agriculture, Economic Research Service, "Rural-Urban Commuting Area Codes," <www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>.

Census tracts are roughly equivalent to neighborhoods. They contain 2,500 to 8,000 people per tract. Since detailed American Community Survey (ACS) 1-year estimates are only available for geographic areas with at least 65,000 residents, Susan will use ACS 5-year estimates, which she will download from data.census.gov.

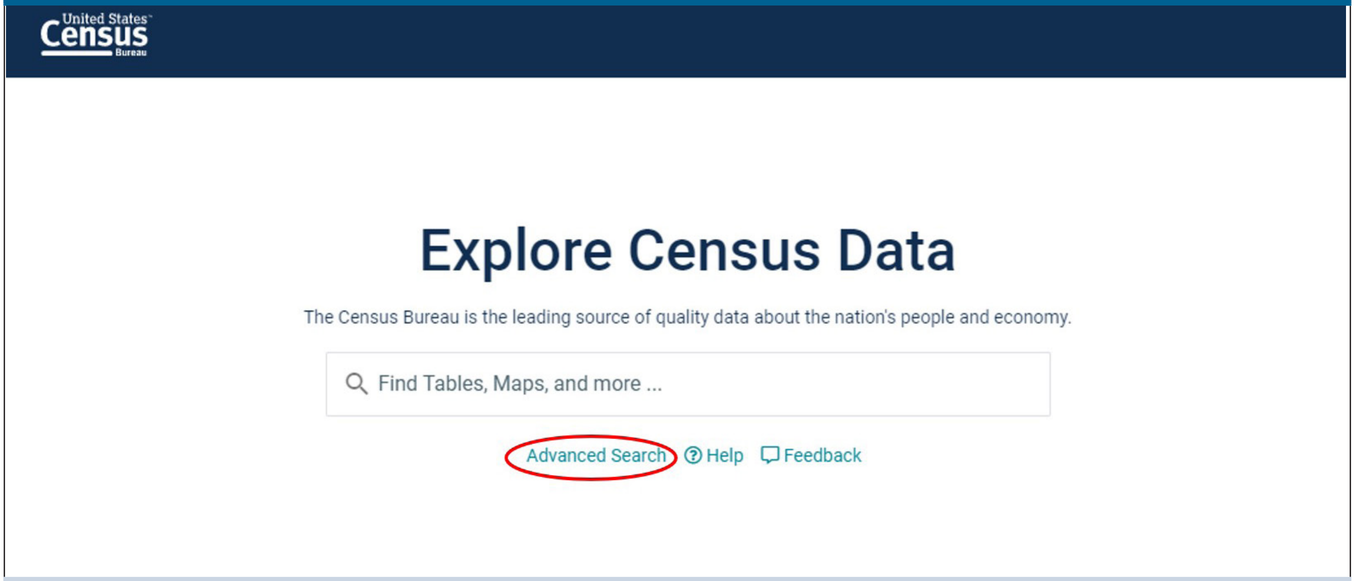
There are roughly 1,300 census tracts in Minnesota. Susan will aggregate these tracts into four RUCA-based areas—Rural, Small Town, Large Town, and Urban. She will also estimate how much uncertainty is associated with the new estimates she has created.

The U.S. Census Bureau provides a number of formulas that can be used to estimate uncertainty, or margins of error (MOEs), for estimates that are produced from calculations based on published data tables. Calculating the estimates of uncertainty will allow her to make judgments about whether observed differences in earnings are real or whether they are within the expected variations that result from survey sampling.

Susan starts her analysis by going to the data.census.gov Web site at <https://data.census.gov>.

- She clicks on "Advanced Search" under the search bar (see Figure 3.1).

Figure 3.1. **Selecting Advanced Search in Data.census.gov**



Source: U.S. Census Bureau, data.census.gov <https://data.census.gov>.

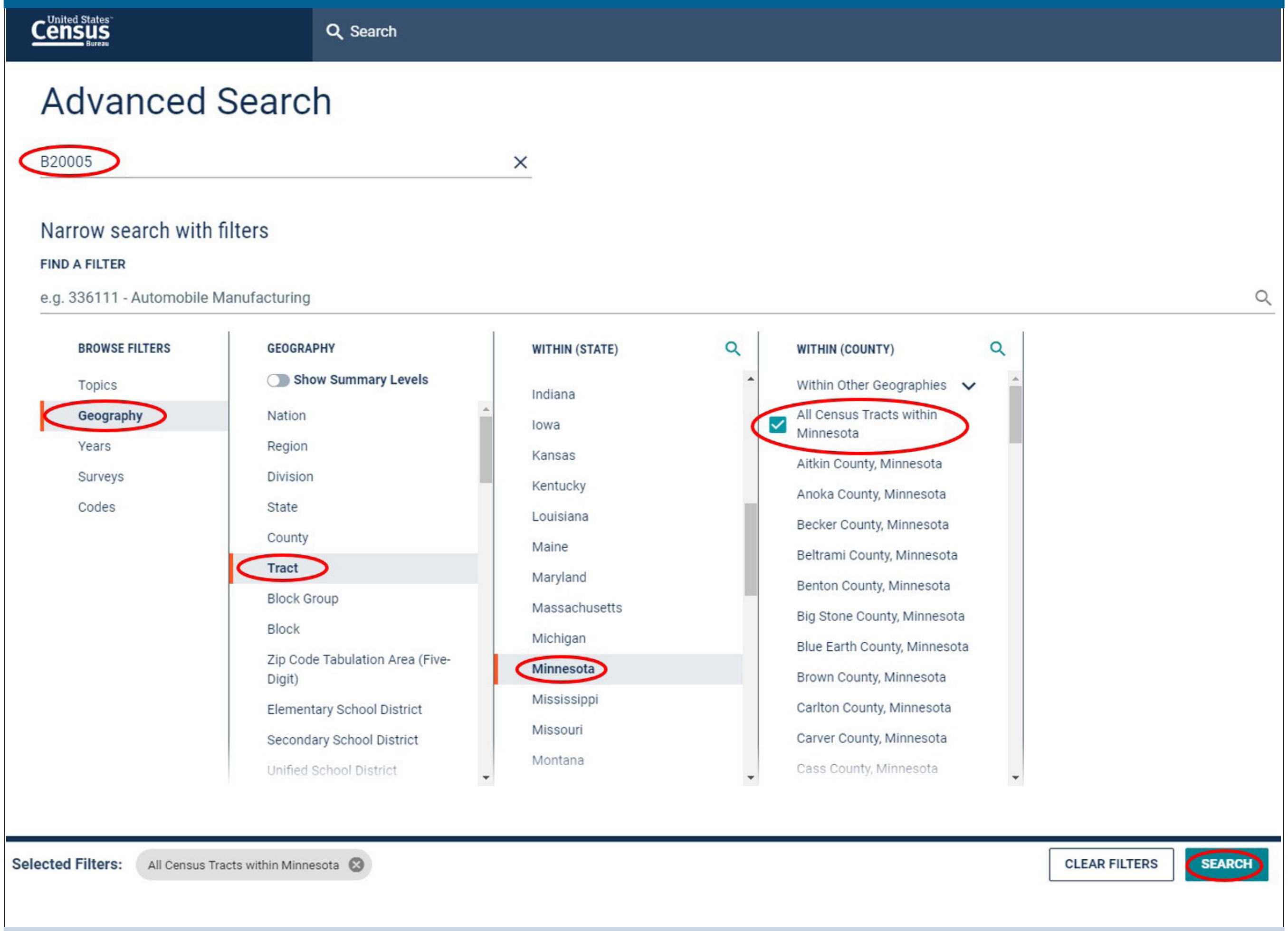
- Since Susan already knows her desired table ID, she types "B20005" in the first text box directly under the Advanced Search heading. B20005 is the table ID for "Sex by Work Experience in the Past 12 Months by Earnings in the Past 12 Months (in 2018 Inflation-Adjusted Dollars) for the Population 16 Years and Over."

- She selects "Geography" to view the geography filters.

- Next, she selects "Tract," and scrolls to select "Minnesota" from the "Within (State)" filter.

- Susan then checks the box for "All Census Tracts within Minnesota" from the "Within (County)" filter and clicks "Search" in the lower right corner (see Figure 3.2).

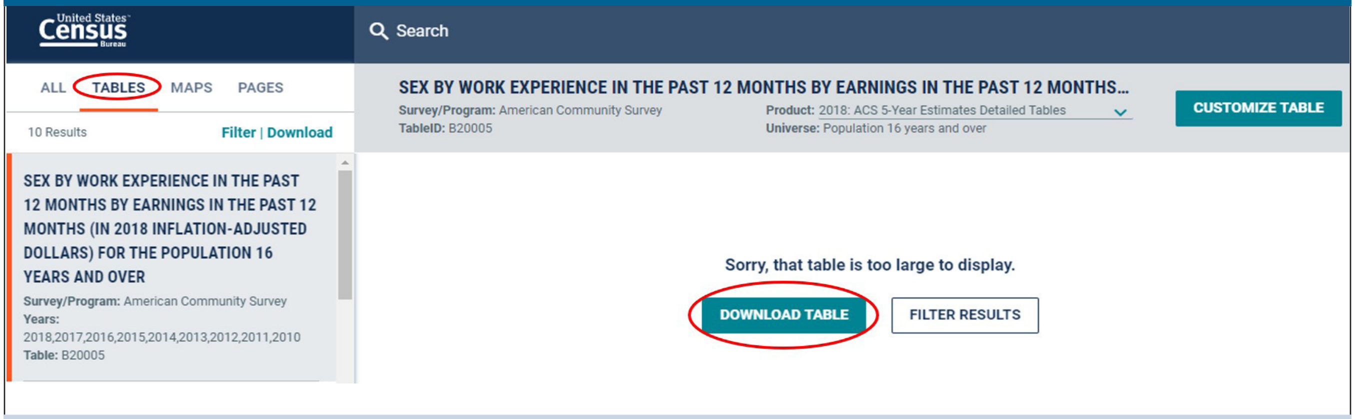
Figure 3.2. **Selecting a Table and Geographies in Data.census.gov**



Source: U.S. Census Bureau, data.census.gov, <https://data.census.gov>.

- On the next page, Susan clicks "Tables" in the upper left corner.
- Then, she selects "Download Table" under the message that the "table is too large to display" (see Figure 3.3).

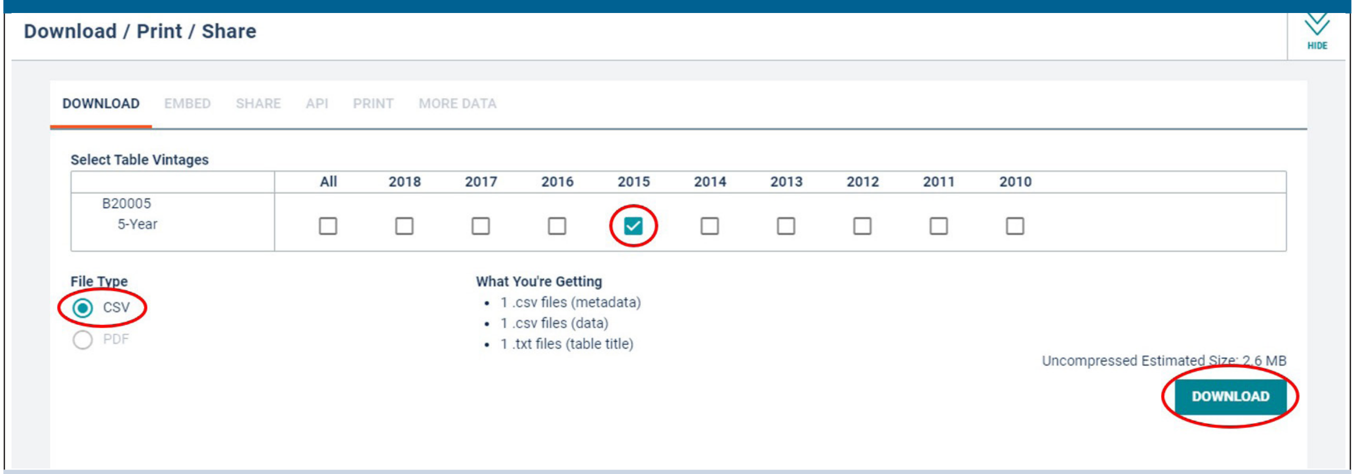
Figure 3.3. **Downloading a Table From Data.census.gov**



Source: U.S. Census Bureau, data.census.gov, <https://data.census.gov>.

- Next, she uses the checkboxes to select the 2015 ACS 5-year data.
- She chooses the File Type "CSV."
- Then, she clicks "Download" in the lower right corner (see Figure 3.4).

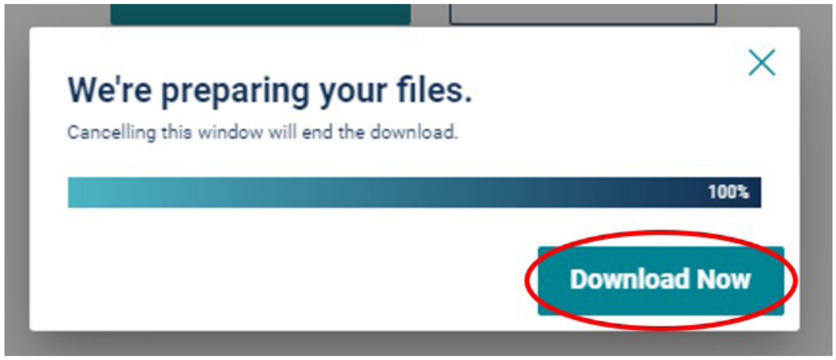
Figure 3.4. **Changing the Survey Year and Choosing a File Type in Data.census.gov**



Source: U.S. Census Bureau, data.census.gov, <https://data.census.gov>.

- Susan selects “Download Now” after the file is prepared (see Figure 3.5).

Figure 3.5. **Downloading the Compressed (ZIP) File**



Source: U.S. Census Bureau, data.census.gov, <https://data.census.gov>.

- From the compressed folder, Susan opens the file with “data_with_overlays” shown in the file name. Documentation relating to the data table is also included in her zipped file.

- Now that Susan has her data file, she analyzes how earnings vary across the rural-urban areas of her state. The USDA publishes 10 primary RUCA codes that delineate census tracts.[17] Susan adds these codes to the ACS data file that she has sorted by geographic identifier (GEO_ID).

- Using Excel, Susan “copies” two columns of Minnesota census tract data from the USDA RUCA file—the “Primary RUCA Code 2010” and the “State-County-Tract FIPS Code.” (FIPS refers to Federal Information Processing Standards.) She then pastes these two columns of data into the ACS data file.

- To verify that the census tract data from the two files are properly matched in the new file, Susan creates a column with the final 11 numbers of the ACS file’s “id” column.[18] The last 11 numbers in the “id” column are the tract’s FIPS code. Susan then subtracts this new column (FIPS code) from her ACS file from the “State-County-Tract FIPS Codes” column from her RUCA file. If the rows match, the resulting difference will be zero.

- Susan analyzes earnings for a collapsed version of the RUCA codes. She creates a new column of data with four string values: “Urban” for RUCA codes 1–3, “Large Town” for codes 4–6, “Small Town” for codes 7–9, and “Rural” for code 10.

[17] U.S. Department of Agriculture, Economic Research Service, “Documentation: 2010 Rural-Urban Commuting Area (RUCA) Codes,” <www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/>.

[18] The Census Bureau has published detailed instructions for matching these GEOIDs: <https://ask.census.gov/prweb/PRServletCustom/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pyActivity=pyMobileSnapStart&ArticleID=KCP-5651>.

- The resulting data file now looks like this, with the highlighted cells added from the USDA RUCA file and Susan's subsequent recoding and match verification (see Figure 3.6).

Figure 3.6. **Adding Rural-Urban Codes to Downloaded Data**

GEO_ID	NAME					B20005_001E	B20005_001M	B20005_002E	B20005_002M
id	Geographic Area Name	State-County-Tract FIPS Code (lookup by address at http://www.ffiec.gov/Geocode/)	Verify match	Primary RUCA Code 2010	RUCA SDC	Estimate!!Total	Margin of Error!!Total	Estimate!!Total!!Male	Margin of Error!!Total!!Male
1400000US27001770100	Census Tract 7701, Aitkin County, Minnesota	27001770100	0	10	Rural	1970	102	1017	60
1400000US27001770200	Census Tract 7702, Aitkin County, Minnesota	27001770200	0	10	Rural	1756	124	915	81
1400000US27001770300	Census Tract 7703, Aitkin County, Minnesota	27001770300	0	10	Rural	2748	157	1331	105
1400000US27001770400	Census Tract 7704, Aitkin County, Minnesota	27001770400	0	10	Rural	2621	140	1346	81
1400000US27001790501	Census Tract 7905.01, Aitkin County, Minnesota	27001790501	0	10	Rural	1608	95	820	63
1400000US27001790502	Census Tract 7905.02, Aitkin County, Minnesota	27001790502	0	10	Rural	2752	114	1401	75
1400000US27003050107	Census Tract 501.07, Anoka County, Minnesota	27003050107	0	2	Urban	2139	165	1079	118
1400000US27003050108	Census Tract 501.08, Anoka County, Minnesota	27003050108	0	2	Urban	3739	176	1888	141
1400000US27003050109	Census Tract 501.09, Anoka County, Minnesota	27003050109	0	2	Urban	4151	194	2223	152
1400000US27003050110	Census Tract 501.10, Anoka County, Minnesota	27003050110	0	2	Urban	2136	135	1074	88
1400000US27003050111	Census Tract 501.11, Anoka County, Minnesota	27003050111	0	2	Urban	3052	169	1653	115
1400000US27003050114	Census Tract 501.14, Anoka County, Minnesota	27003050114	0	2	Urban	2209	176	1150	105

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates; and USDA RUCA codes.

- Susan uses "PivotTables" in Excel to aggregate the earnings distribution across census tracts. The PivotTables sums the number of males working full-time, year-round by rural, small town, large town, and urban census tracts within each earnings distribution category (see Figure 3.7).

Figure 3.7. **Cross-Tabulating Persons Per Income Level and Rural-Urban Category**

Values	Large Town	Rural	Small Town	Urban
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings	114903	82448	71267	819602
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$1 to $2,499 or loss	377	565	345	1503
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$2,500 to $4,999	255	289	186	997
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$5,000 to $7,499	750	687	436	4336
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$7,500 to $9,999	703	542	404	3296
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$10,000 to $12,499	1921	1688	1307	9195
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$12,500 to $14,999	1682	1002	732	7078
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$15,000 to $17,499	2718	1750	1383	12054
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$17,500 to $19,999	1976	1426	1284	11497
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$20,000 to $22,499	4456	3403	2539	21296
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$22,500 to $24,999	2641	2457	2040	16528
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$25,000 to $29,999	8349	6228	5543	41505
Sum of Estimate!!Total!!Male!!Worked full-time, year-round in the past 12 months!!With earnings!!$30,000 to $34,999	10005	7936	6868	55597

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates; and USDA RUCA codes.

- Next, Susan estimates the median earnings of men who work full-time, year-round and live in rural areas. The Census Bureau provides guidance on how to interpolate a median from a weighted distribution in its Accuracy of the PUMS documentation.[19] Susan creates an Excel spreadsheet to estimate a median using the method described in the Census Bureau's documentation. The documentation also describes how to calculate standard errors and confidence intervals for her estimates.[20]

[19] U.S. Census Bureau, American Community Survey (ACS), PUMS Technical Documentation, Accuracy of the PUMS, <www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>.

[20] The method described in this case study to approximate a median estimate will not match medians published in data.census.gov, as the published medians are calculated using different and more detailed distributions than are available to users. Also, the approximated MOE of the median using this method may underestimate or overestimate the true MOE, due to the limitations of using the PUMS design factor methodology.

- She repeats these calculations for men's and women's earnings in each of her four geographic areas (see Figure 3.8).

Figure 3.8. **Calculating a Median From a Weighted Distribution**

Male, With Earnings, Worked Full-Time, Year Round, Rural, 2011-2015	Rural	Cumulative frequency	Cumulative Percent		
Sum of Estimates	82448				
$1 to $2,499 or loss	565	565	0.7%	SE(50 percent)	0.599
$2,500 to $4,999	289	854	1.0%	p_lower	49.401
$5,000 to $7,499	687	1541	1.9%	p_upper	50.599
$7,500 to $9,999	542	2083	2.5%	p_median	50
$10,000 to $12,499	1688	3771	4.6%	A1	40,000
$12,500 to $14,999	1002	4773	5.8%	A2	45,000
$15,000 to $17,499	1750	6523	7.9%	C1	42.40
$17,500 to $19,999	1426	7949	9.6%	C2	52.30
$20,000 to $22,499	3403	11352	13.8%	lower bound percent	0.707
$22,500 to $24,999	2457	13809	16.7%	width of interval	5,000
$25,000 to $29,999	6228	20037	24.3%	lower bound value	43,536
$30,000 to $34,999	7936	27973	33.9%	upper bound percent	0.828
$35,000 to $39,999	6976	34949	42.4%	width of interval	5,000
$40,000 to $44,999	8196	43145	52.3%	upper bound value	44,141
$45,000 to $49,999	5815	48960	59.4%	SE	302
$50,000 to $54,999	7188	56148	68.1%		
$55,000 to $64,999	7952	64100	77.7%		
$65,000 to $74,999	4835	68935	83.6%	p_50	0.768
$75,000 to $99,999	6376	75311	91.3%	width of the interval	5,000
$100,000 or more	7137	82448	100.0%	median	43,838

Note: This example uses unrounded values in the calculations, but the values displayed are rounded to two or three decimal places as appropriate.

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates; and USDA RUCA codes.

- When Susan compiles the calculated medians and their standard errors into a single table, she can see that median earnings for men in urban areas ($55,064) appear to be higher than the median earnings for men in rural, small town, and large town regions of the state. Similarly, urban women's median earnings ($45,053) are considerably higher than those for women living outside of urban areas. To calculate MOEs for the approximate median earnings, Susan multiplies 1.645 by the standard error of each median. This creates an MOE at the 90 percent confidence level (see Figure 3.9).

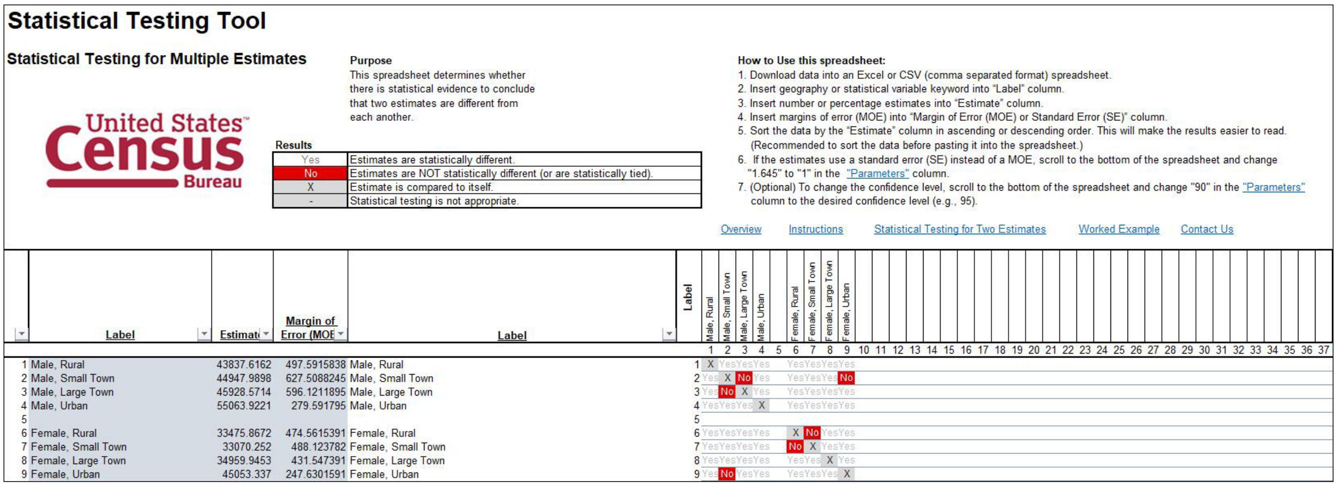
Figure 3.9. **Median Earnings for Men and Women in Minnesota by Rural-Urban Location**

Median Earnings, Men, Full-Time, Year-Round Workers Minnesota: 2011–2015			
	Median earnings	Standard error	Margin of error (90%)
Male, rural	$43,838	$302	$498
Male, small town	$44,948	$381	$628
Male, large town.	$45,929	$362	$596
Male, urban	$55,064	$170	$280
Median Earnings, Women, Full-Time, Year-Round Workers			
	Median	Standard	Margin of
Female, rural.	$33,476	$288	$475
Female, small town . . .	$33,070	$297	$488
Female, large town . . .	$34,960	$262	$432
Female, urban.	$45,053	$151	$248

Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates; and USDA RUCA codes.

- Susan then tests whether the calculated differences in median earnings across geographic areas are statistically significant. She pastes the estimated medians and MOEs into the Census Bureau's Statistical Testing Tool and learns that, as expected, urban men's median earnings are significantly different from their counterparts in rural areas, small towns, and large towns.[21] She also confirms that urban women's median earnings are statistically different from those of women in other areas of the state (see Figure 3.10).

Figure 3.10. **Statistical Testing Tool for Multiple Estimates (90 Percent Confidence Level)**



Statistical Testing Tool

Statistical Testing for Multiple Estimates

Purpose
This spreadsheet determines whether there is statistical evidence to conclude that two estimates are different from each another.

Results

Yes	Estimates are statistically different.
No	Estimates are NOT statistically different (or are statistically tied).
X	Estimate is compared to itself.
-	Statistical testing is not appropriate.

How to Use this spreadsheet:
1. Download data into an Excel or CSV (comma separated format) spreadsheet.
2. Insert geography or statistical variable keyword into "Label" column.
3. Insert number or percentage estimates into "Estimate" column.
4. Insert margins of error (MOE) into "Margin of Error (MOE) or Standard Error (SE)" column.
5. Sort the data by the "Estimate" column in ascending or descending order. This will make the results easier to read. (Recommended to sort the data before pasting it into the spreadsheet.)
6. If the estimates use a standard error (SE) instead of a MOE, scroll to the bottom of the spreadsheet and change "1.645" to "1" in the "Parameters" column.
7. (Optional) To change the confidence level, scroll to the bottom of the spreadsheet and change "90" in the "Parameters" column to the desired confidence level (e.g., 95).

Overview Instructions Statistical Testing for Two Estimates Worked Example Contact Us

	Label	Estimate	Margin of Error (MOE	Label	Label	Male, Rural	Male, Small Town	Male, Large Town	Male, Urban		Female, Rural	Female, Small Town	Female, Large Town	Female, Urban																											
						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1	Male, Rural	43837.6162	497.5915838	Male, Rural	1	X	Yes	Yes	Yes		Yes	Yes	Yes	Yes																											
2	Male, Small Town	44947.9898	627.5088245	Male, Small Town	2	Yes	X	No	Yes		Yes	Yes	Yes	No																											
3	Male, Large Town	45928.5714	596.1211895	Male, Large Town	3	Yes	No	X	Yes		Yes	Yes	Yes	Yes																											
4	Male, Urban	55063.9221	279.591795	Male, Urban	4	Yes	Yes	Yes	X		Yes	Yes	Yes	Yes																											
5					5																																				
6	Female, Rural	33475.8672	474.5615391	Female, Rural	6	Yes	Yes	Yes	Yes		X	No	Yes	Yes																											
7	Female, Small Town	33070.252	488.123782	Female, Small Town	7	Yes	Yes	Yes	Yes		No	X	Yes	Yes																											
8	Female, Large Town	34959.9453	431.547391	Female, Large Town	8	Yes	Yes	Yes	Yes		Yes	Yes	X	Yes																											
9	Female, Urban	45053.337	247.6301591	Female, Urban	9	Yes	No	Yes	Yes		Yes	Yes	Yes	X																											

Source: U.S. Census Bureau, Statistical Testing Tool, <www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>.

Susan uses this analysis to help her convey differences in earnings among residents of rural, small town, large town, and urban areas in reports that her office produces for state policymakers. While she will not always report the numeric results of statistical tests, knowing which differences are significant helps her know which differences she can highlight in her narrative. Conversely, knowing which differences are not statistically significant helps her know which differences she should downplay in her reporting.

An example of a report that was informed by this type of analysis is Greater Minnesota: Refined & Revisited.[22] (This report was produced using 2010–2014 ACS 5-year estimates, so the medians are somewhat different, but the results are consistent with what is described here.) This report has been used by policymakers working on rural health care initiatives, on Equal Employment Opportunity activities, and by legislators working to create policies that align with current economic conditions in different areas of the state.

[21] U.S. Census Bureau, Statistical Testing Tool, <www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>.

[22] Minnesota State Demographic Center, Greater Minnesota: Refined & Revisited, <https://mn.gov/admin/demography/reports-resources/greater-mn-refined-and-revisited.jsp>.

Case Study #2: New York City, Department of City Planning, Uncertainty in Mapping ACS Data

Skill Level: Intermediate/Advanced
Subject: Uncertainty in Mapping American Community Survey (ACS) Data
Type of Analysis: Assessment of statistical reliability of ACS maps
Tool Used: Map Reliability Calculator
Authors: Joel A. Alvarez, Senior Analyst, NYC City Planning, Population Division; and Joseph J. Salvo, Director, NYC City Planning, Population Division

In the summer of 2017, New York City established the New York Works plan—a series of 25 initiatives to promote the creation of 100,000 new jobs with good wages over the next decade.[23] In support of the plan, the Department of City Planning (DCP) produced a series of maps informing the public about employment patterns in New York City. In this case study, we walk ACS data users through the process we used to assess the reliability of map classification schemes when producing maps for general consumption.

ACS data provide city planners with unique insights into the socioeconomic characteristics of local populations, including information about employment. Mapping the data is one way to examine differences in employment across geographic areas. However, ACS estimates are subject to sampling variability, so reality on the ground may differ from survey results.[24] Given the uncertainty associated with ACS estimates, data users should exercise caution when producing maps to avoid misrepresenting the characteristic(s) being displayed. The following case study provides guidance in this regard, demonstrating how we produced statistically reliable maps of employment and unemployment using an online Map Reliability Calculator.[25]

Mapping Employment

In support of a mayoral jobs creation initiative, DCP was asked to create a series of maps showing the latest information on employment and unemployment. One possible approach was to examine administrative data from unemployment insurance filings. However, this data set excludes many self-employed workers and those working "off-the-books," so we turned to the ACS as a more comprehensive source of data on local employment patterns.

First, we examined overall employment in New York City. Our preference was to produce a map using small geographic units, making census tracts ideal. However, in New York City, census tracts typically consist of only six to eight city blocks and have populations of about 3,000 to 4,000. Consequently, ACS 5-year estimates for census tracts are based on small sample sizes—typically 250 to 300 people surveyed in each tract. To ensure that our map was reliable and would

Box 3.1. Establishing a Minimum Reliability Threshold for Maps

Subjects covered in the ACS often display meaningful spatial patterning at very fine levels of geography. ACS data users may be tempted to present these data in maps using the smallest available geographic units. However, the reliability of ACS estimates typically decreases as units of analysis get smaller, because of diminishing sample sizes. When mapping ACS data, users must decide whether to use small geographic areas and see all the fine detail, but risk false conclusions due to data uncertainty; or to use large, statistically reliable geographic areas, but risk overlooking the most salient spatial distributions.

This dilemma can be resolved by establishing a minimum reliability threshold. Once map quality is assured by passing the threshold, ACS data users can pursue mapping at the smallest geographic area for which reliable data are available. New York City's Department of City Planning (DCP) has adopted a threshold of a 10 percent error rate, under which a map is considered suitable for general use. A 10 percent error rate means that any given geographic area would have a 1 in 10 chance of being erroneously classed, placing it at odds with reality on the ground. This threshold was adopted because it matches the Census Bureau's standard of 90 percent confidence intervals. Additionally, the DCP standard is to ensure that no individual map category has an error rate of 20 percent or more, so that map users can trust the reliability of each respective map class. While this is a lower standard than that used for the overall map, it helps ensure that even categories with relatively few values—and therefore little influence on the overall reliability—can still be trusted by end users.

[23] City of New York, New York Works, <https://newyorkworks.cityofnewyork.us/introduction/>.
[24] Sampling variability is the difference between an estimate based on a sample and the corresponding value that would be obtained if the estimate were based on the entire population.
[25] Statistical reliability refers to the ability of a measurement tool to consistently produce the same results. When used in reference to the ACS, the measurement tool is the survey itself.

not mislead people into making false conclusions, we tested the preliminary map using an online Map Reliability Calculator developed by DCP (see Box 3.1 on Establishing a Minimum Reliability Threshold for Maps).[26]

To conduct this analysis of map reliability, we first went to the U.S. Census Bureau's data.census.gov located at <https://data.census.gov> and downloaded data on the employed population aged 16 and older in the civilian labor force, at the census tract level (from Table B23025).[27] The data were then imported into a Geographic Information System (GIS) to produce a map with seven categories using a natural breaks classification scheme.[28] We then tested the results using the Map Reliability Calculator.

The reliability calculator has three required inputs:

- The estimates and associated margins of error (MOEs).
- The number of classes or map categories.
- The lower limit for each class.

After inserting this information into the tool, we examined the results and found that our proposed map was not reliable (see Figure 3.11). When the reliability calculator marks a set of map categories as "not reliable," it means that 10 percent or more of the geographic areas are potentially misclassified (that is, included in the wrong category). In our example, shown in Figure 3.11, the overall reliability of the map was 14.2 percent. This means that of New York City's 2,167 census tracts, more than 300 may have been incorrectly classified. Further, the second- and fourth-highest map classes in our proposed map had reliability scores of more than 20 percent. As with the overall map, reliability scores for individual map classes tell users the percentage of geographic areas that are likely to be misclassified based on the published MOEs. These excessive scores for individual map categories also marked our proposed map as too unreliable for general use.

[26] New York City Department of City Planning, Map Reliability Calculator, <https://www1.nyc.gov/site/planning/data-maps/nyc-population/geographic-reference.page>.

[27] U.S. Census Bureau, data.census.gov, <https://data.census.gov/>.

[28] A Geographic Information System, or GIS, is an application used for mapping, managing, and analyzing spatial data. Various map classification schemes can be employed when creating categories for quantitative data. We used a natural breaks scheme for our employment analysis. This scheme maximizes the variance between classes, while minimizing variance within classes.

Figure 3.11. **Results From Map Reliability Calculator (Seven Class Breaks)**

MAP RELIABILITY CALCULATOR

(Fill in boxes to get map reliability* for classification schemes)

STEP 1
Insert estimates & Margins of Error (MOEs)
(Insert up to 2,500 lines)

	Estimates	MOEs
1	0.00	
2	1105.00	222.00
3	2667.00	358.00
4	0.00	
5	4028.00	998.00
6	6463.00	737.00
7	4132.00	619.00
8	1120.00	162.00
9	695.00	110.00
10	1882.00	560.00
11	1709.00	246.00
12	3419.00	400.00
13	1734.00	231.00

STEP 2
Select number of classes
(Type in number from 2 to 7)

Classes: 7

STEP 3
Select class breaks
(Type in lower limit for each class)

	User Defined Class Breaks	Count	Reliability *
top class	5,981.00	20	16.6
lowest (2 classes)	4,832.00	42	29.5
lowest (3 classes)	3,259.00	147	16.1
lowest (4 classes)	2,390.00	280	21.4
lowest (5 classes)	1,634.00	550	17.6
lowest (6 classes)	789.00	854	10.1
lowest (7 classes)	0.00	274	8.9
Total		2,167	14.2

Not Reliable

*The reliability score is the probability that any given estimate is erroneously classed.

Note: Estimates with blank margins of error (MOEs) are treated as having MOEs of zero.
Source: New York City Department of City Planning, Map Reliability Calculator, <https://www1.nyc.gov/site/planning/data-maps/nyc-population/geographic-reference.page>.

One method of improving map reliability is to reduce the number of map classes. Based on this logic, we decreased the number of categories in the proposed map to six, but the map was still not reliable. It wasn't until the map was reduced to four categories that it qualified as reliable. Further, to make the categories more presentable, we rounded the class breaks and checked to confirm that the map was still reliable (see Figure 3.12).

Figure 3.12. **Results From Map Reliability Calculator (Four Class Breaks)**

MAP RELIABILITY CALCULATOR

(Fill in boxes to get map reliability* for classification schemes)

STEP 1
Insert estimates & Margins of Error (MOEs)
(Insert up to 2,500 lines)

	Estimates	MOEs
1	0.00	
2	1105.00	222.00
3	2667.00	358.00
4	0.00	
5	4028.00	998.00
6	6463.00	737.00
7	4132.00	619.00
8	1120.00	162.00
9	695.00	110.00
10	1882.00	560.00
11	1709.00	246.00
12	3419.00	400.00
13	1734.00	231.00
14	1470.00	244.00
15	3663.00	390.00
16	5231.00	428.00
17	3813.00	583.00
18	4951.00	778.00
19	1373.00	256.00

STEP 2
Select number of classes
(Type in number from 2 to 7)

Classes
4

STEP 3
Select class breaks
(Type in lower limit for each class)

	User Defined Class Breaks	Count	Reliability *
top class	5,000.00	52	11.9
lowest (2 classes)	2,500.00	384	11.0
lowest (3 classes)	1,000.00	1,241	6.4
lowest (4 classes)	0.00	490	7.3
lowest (5 classes)			
lowest (6 classes)			
lowest (7 classes)			
Total		2,167	7.6

Reliable

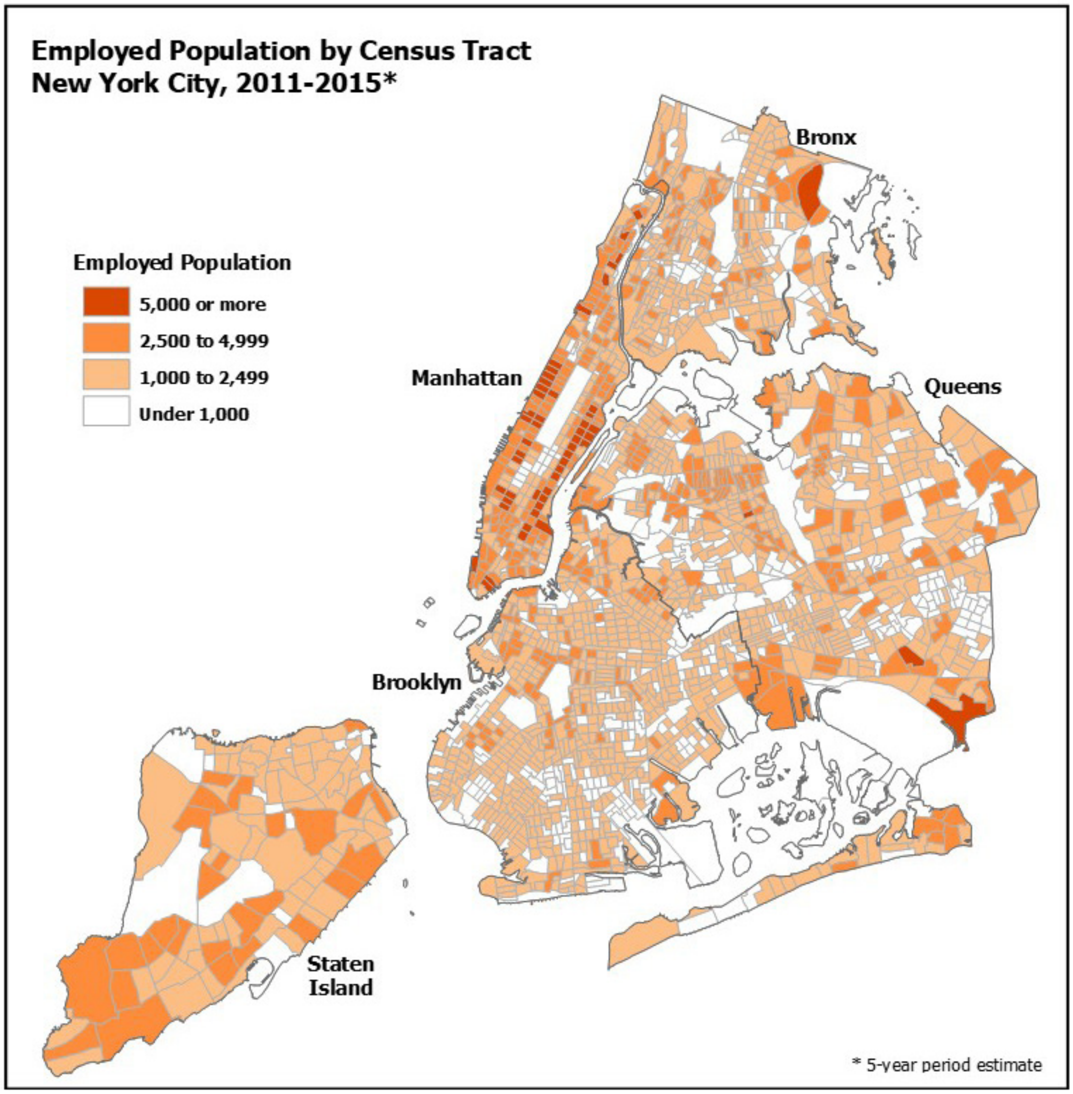
*The reliability score is the probability that any given estimate is erroneously classed.
A map classification is considered to be reliable if the total reliability score is less 10% and all individual classes are less than 20%.

Suggested notation for maps that pass this reliability test:
For each individual category, less than 20% of geographies are likely to be misclassified due to sampling error;
fewer than 10% of all geographies are likely to be misclassified across all categories.
Source: New York City Department of City Planning, Population Division

Note: Estimates with blank margins of error (MOEs) are treated as having MOEs of zero.
Source: New York City Department of City Planning, Map Reliability Calculator, <https://www1.nyc.gov/site/planning/data-maps/nyc-population/geographic-reference.page>.

With this evaluation, we were confident that our map provided a relatively reliable depiction of reality on the ground and went ahead with its use supporting the mayoral initiative (see Figure 3.13).

Figure 3.13. **Map of Employed Population**



Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates.

Mapping Unemployment

Generally, the relative size of ACS MOEs increases in relation to associated ACS estimates as count estimates get smaller. It follows that smaller estimates are often less reliable, in a relative sense.[29] Consequently, maps built using smaller ACS estimates are typically less reliable than those built using large estimates. We confronted this issue when we attempted to map unemployment estimates for New York City, since the unemployed population is usually much smaller than the employed population. (The unemployed population is only about one-tenth the size of the employed population in New York City.) Because of the relatively large MOEs, we could only produce a reliable map of census tracts if we sorted them into two categories—one for tracts with 250 or more unemployed persons and one for tracts with fewer than 250 unemployed. While such a map would be informative, we wanted

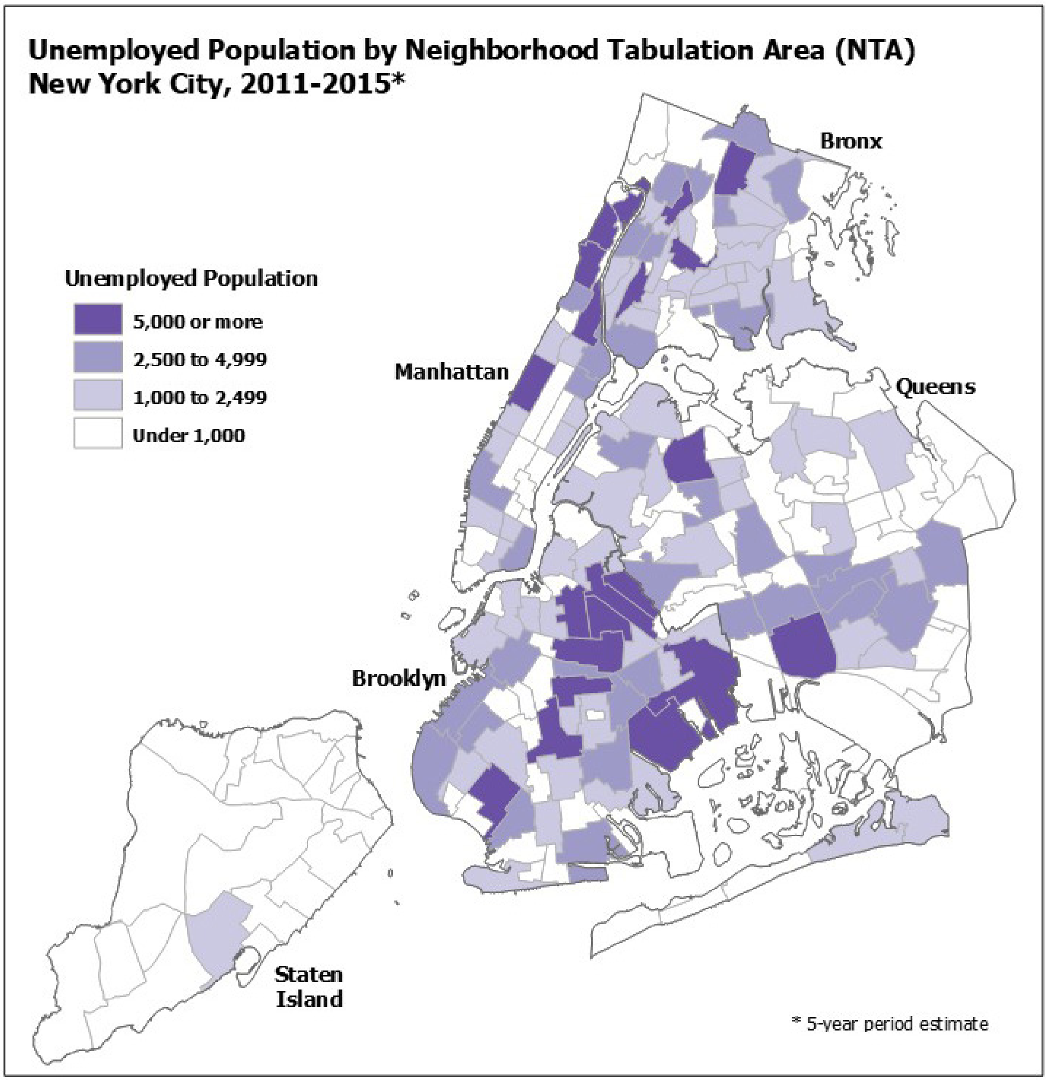
[29] Because estimates and associated MOEs vary greatly in size, it helps to examine the size of MOEs in relation to estimates to better understand the relative reliability of ACS estimates. ACS analysts often use Coefficients of Variation (CVs) as a measure of relative reliability—making it possible to compare the reliability of ACS estimates across different years, periods (1-year vs. 5-year periods), geographic areas, and variables. For more information on CVs, see the section on "Understanding Error and Determining Statistical Significance" in *Understanding and Using American Community Survey Data: What All Data Users Need to Know,* <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

to give the public a greater understanding of the differences in unemployment across our city. For this reason, a higher-order geographic area, Neighborhood Tabulation Areas (NTAs), was evaluated for mapping suitability.

NTAs were created by DCP using aggregates of census tracts that approximate New York City neighborhoods and fit perfectly within Public Use Microdata Area (PUMA) boundaries. This geographic area has gained widespread acceptance and use in New York City because of its relative statistical reliability and because New Yorkers tend to think in terms of neighborhoods. However, since the Census Bureau does not publish data at the NTA level, we needed to calculate new estimates and MOEs aggregating from published, tract-level, unemployment data.[30] Using NTAs, a reliable map of unemployment was produced with four categories—as with employment, breakpoints were rounded to make the map more presentable (see Figure 3.14).

[30] For more information on calculating MOEs for aggregated count estimates, see the section on "Calculating Measures of Error for Derived Estimates" in *Understanding and Using American Community Survey Data: What All Data Users Need to Know,* <www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>.

Figure 3.14. **Map of Unemployed Population**



Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates.

Mapping Change in Employment

The 2011–2015 ACS data release provided us with our first opportunity to compare two nonoverlapping 5-year period estimates (2006–2010 and 2011–2015) based on common population controls derived from the 2010 Census and, for the most part, common geographic boundaries. Therefore, we wanted to map change in the employed population as well. To conduct an evaluation of map reliability, it was necessary to first calculate the tract-level changes in employment and calculate the MOEs associated with those changes.[31] These calculations were quite simple, because we could use the same formula we used when calculating the MOEs for aggregate areas: the square root of the sum of the squared MOEs.[32] Again, it was our preference to create a tract-level map, so we first calculated employment change and associated MOEs for census tracts. Once calculated, estimates and MOEs were inserted into the Map Reliability Calculator.

Employment had increased substantially across the city (up nearly 200,000 or 5 percent), so we were surprised to find that a reliable tract map could not be produced, no matter how few categories were used. As with the map of unemployment, we turned to NTAs, a higher-order geographic area, to see if change could be reliably mapped. Change in employment, however, could not pass reliability thresholds using a natural breaks classification scheme. Therefore, PUMAs, the next higher order statistical geography, were considered. PUMA employment estimates and MOEs from 2006–2010 had to be calculated using census tract aggregations (as with NTAs), because PUMA boundaries changed in 2012, and 2011–2015 estimates were based on the 2012 boundaries.

Unfortunately, as with census tracts and NTAs, the PUMA geographic level proved to be unreliable for a natural breaks classification scheme.

With no reliable results, we re-examined our calculator analysis for all three geographic areas. Map classification schemes that were close to being reliable were manipulated to test whether they could pass reliability thresholds with a set of alternate breakpoints. We found that we could produce a reliable NTA map by slightly adjusting the breakpoint between the first and second categories of a two-class, natural-breaks map (see Figure 3.15).

[31] The Census Bureau endorses the use of statistical testing to gauge the reliability of change over time. This testing tells users that the directionality of change has a 9 in 10 chance of being correct. However, to gauge the reliability of the magnitude of change, it is important that ACS data users go beyond this basic test and consider the MOE associated with the estimate of change.

[32] For detailed guidance on "Comparing Estimates for Nonoverlapping Periods" see page 4 in the Census Bureau's "Instructions for Applying Statistical Testing to the 2011–2015 ACS 5-Year Data," available at <www.census.gov/programs-surveys/acs/technical-documentation/code-lists.2015.html>.

Figure 3.15. **Results From Map Reliability Calculator (Two Class Breaks)**

MAP RELIABILITY CALCULATOR
(Fill in boxes to get map reliability* for classification schemes)

STEP 1
Insert estimates & Margins of Error (MOEs)
(Insert up to 2,500 lines)

STEP 2
Select number of classes
(Type in number from 2 to 7)

STEP 3
Select class breaks
(Type in lower limit for each class)

	Estimates	*MOEs*
1	-453.00	975.22
2	997.00	1486.07
3	1168.00	1081.80
4	-596.00	1048.04
5	457.00	823.59
6	2571.00	1111.90
7	1183.00	1123.19
8	3491.00	1058.02
9	5673.00	1837.22
10	2897.00	1581.40
11	1709.00	1259.39
12	1501.00	1568.27
13	2187.00	1563.64
14	758.00	1246.68
15	92.00	1748.86

Classes: 2

	User Defined *Class Breaks*	*Count*	*Reliability **
top class	2,000.00	42	18.7
lowest (2 classes)	-3,402.00	146	6.6
lowest (3 classes)			
lowest (4 classes)			
lowest (5 classes)			
lowest (6 classes)			
lowest (7 classes)			
Total		188	9.3

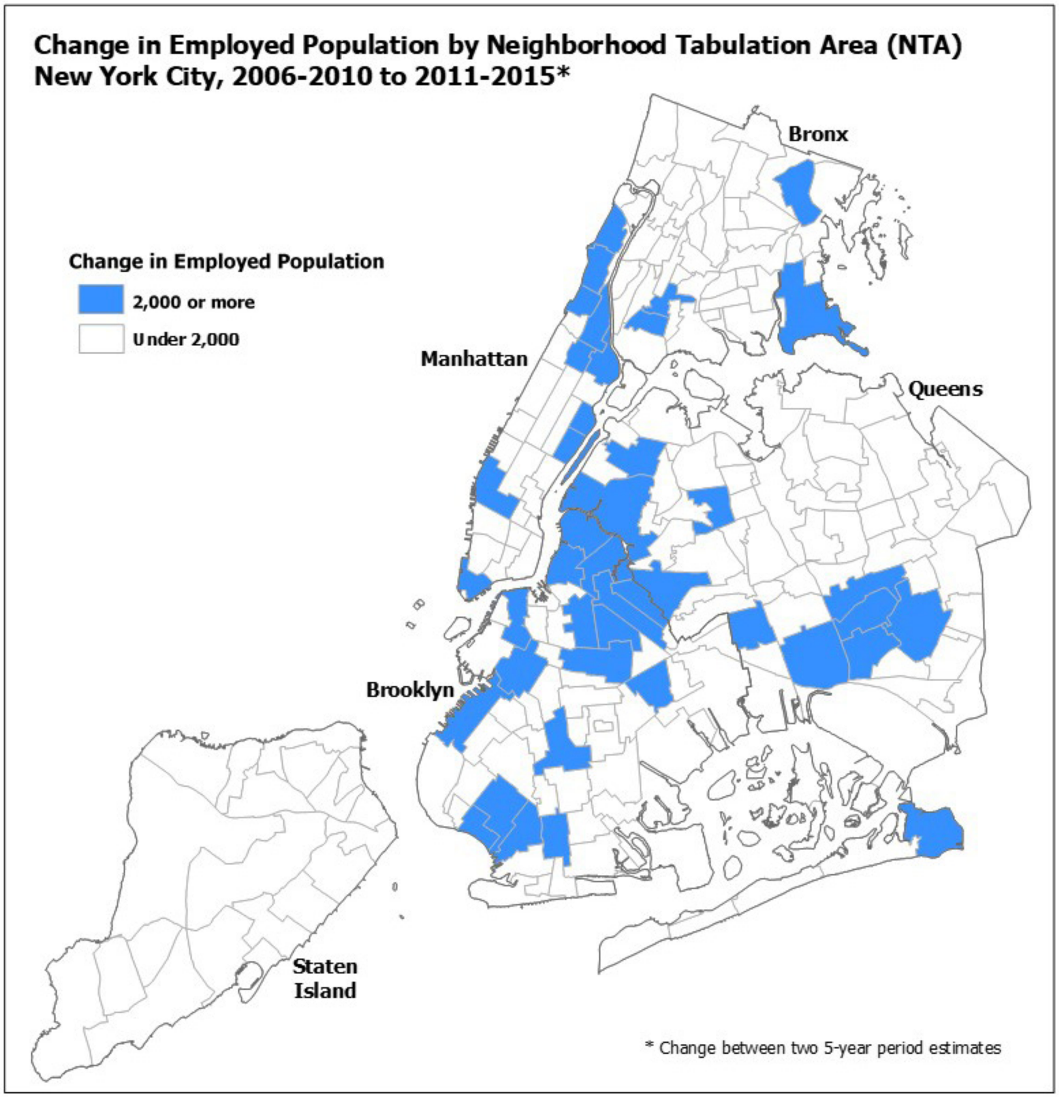
Reliable

*The reliability score is the probability that any given estimate is erroneously classed.
A map classification is considered to be reliable if the total reliability score is less 10% and all individual classes are less than 20%.

Source: New York City Department of City Planning, Map Reliability Calculator, <https://www1.nyc.gov/site/planning/data-maps/nyc-population/geographic-reference.page>.

Because our lowest map category encompassed both positive and negative change in employment, we chose to only emphasize the top category, where change was equal to, or exceeded, an employment increase of 2,000 (see Figure 3.16).

Figure 3.16. **Map of Change in Employed Population**



Source: Author's analysis of data from the U.S. Census Bureau, American Community Survey, 5-Year Estimates.

Conclusion

In producing this series of maps depicting dimensions of employment in New York City, we learned quite a bit about producing reliable maps for general use. In creating a tract-level map of employment, we learned that map reliability can typically be improved by reducing the number of map categories. Additionally, through the production of the unemployment map, we found that map reliability can usually be improved by using higher-order geographic areas, because the reliability of underlying estimates is improved. Finally, while generating a map showing change in employment, we discovered that category breakpoints can be adjusted to make a map statistically reliable.

This was an important lesson, because it is ultimately up to each end user to decide which breaks work best for their purposes.

While we decided to use a mix of different geographic types in our maps, others might opt for uniformity in their publication summary level. In fact, data users have several different options in mapping ACS data. For example:

- Choosing different classification schemes, such as equal interval or quantile schemes.
- Selecting fewer map categories to reduce the risk of misclassification.
- Normalizing data using percentages (as opposed to using counts).
- Loosening map reliability standards to gain insight into a very generalized spatial distribution—acknowledging that such a map is more prone to error.[33]

Regardless of your approach, it is essential that ACS data mappers pursue their cartographic endeavors with a full understanding that uncertainty is inherent in all survey data, including ACS data, and will impact the quality of maps. It is ultimately up to each end user to decide which standards are most appropriate for their applications.

[33] The NYC Department of City Planning's Map Reliability Calculator provides reliability scores so that users can select alternative thresholds if they choose.

Case Study #3: King County Housing Assessment

Skill Level: Intermediate/Advanced
Subject: Evaluating Housing Program Participation
Type of Analysis: American Community Survey (ACS) microdata analysis
Tools Used: ACS Public Use Microdata Sample File and data.census.gov
Authors: John Wilson, Assessor, King County, WA, Department of Assessments; Chandler Felt, Demographer, King County, WA; and Susan Kinne, Epidemiologist at Public Health-Seattle and King County

John Wilson:

When I became King County (WA) assessor in 2016, housing affordability was headed towards a crisis level—especially for low-income seniors, disabled veterans, and other disabled individuals. King County has 2.1 million residents, and real estate values had been rising at a double-digit pace annually.

I was curious how many people were enrolled in a state-authorized property tax exemption program. It turned out to be only about 15,000 countywide. That number seemed low to me, so I contacted Chandler Felt, King County's demographer.

I asked Chandler, knowing how familiar he was with U.S. Census Bureau data, if he knew of any way to determine how many people in King County might be eligible for the program. Chandler suggested the latest American Community Survey (ACS) data.

Chandler Felt:

As demographer for the county, I turned to the Census Bureau's ACS via data.census.gov.[34] I looked through the available tables in data.census.gov using the 2014 ACS 1-year data set and the 2010–2014 ACS 5-year data set, but soon realized that the data.census.gov tables would not provide the entire list of eligibility criteria for the exemption. The ACS Public Use Microdata Sample (PUMS) data set would be required to slice our population precisely enough to answer the question, and I do not have experience using the PUMS data.[35] I forwarded John's request to my colleague Susan Kinne, Epidemiologist at Public Health-Seattle and King County, who is a skilled PUMS user.

King County's senior tax exemption is based on three eligibility criteria, all from household data:

- Household tenure = owner (as opposed to renter).
- Age of householder is 62 or older.
- Household income is less than $40,000.

Using the regular data.census.gov tables, I could only report and analyze these criteria two at a time—and not very precisely at that. Income by age is available for householders aged 65 and over, and the cross tabulation of owners by age was likewise for 65-year-olds. Generating a series of data.census.gov tables, I developed a rough estimate that up to 34,000 households—4.2 percent of the over 800,000 households in the county—might be eligible as of 2014. Assessor John Wilson and I agreed that a more reliable estimate was needed, so we asked Susan Kinne to conduct a PUMS analysis, using the three eligibility criteria listed above.

For this analysis, Susan used data from the 2010–2014 ACS 5-year PUMS file because it was the most recent data available at the time. The 5-year PUMS files are multiyear combinations of the 1-year PUMS file with appropriate adjustments to the weights and inflation adjustment factors. She chose to use the 5-year file because it yields more reliable estimates than the 1-year file, and she was conducting an analysis for a relatively small geographic area and population subgroup (older homeowners living in King County).

[34] Data.census.gov was not available at the time this case study was written but is cited here because it is now the primary tool for accessing ACS data.

[35] The ACS PUMS files are a set of untabulated records with information about individual people or housing units. The Census Bureau produces the PUMS files so that data users can create custom tables that are not available through pretabulated (or summary) ACS data products.

Here are the steps she took to produce an estimate of the number of homeowners aged 62 and older in King County who may be eligible for a property tax exemption:

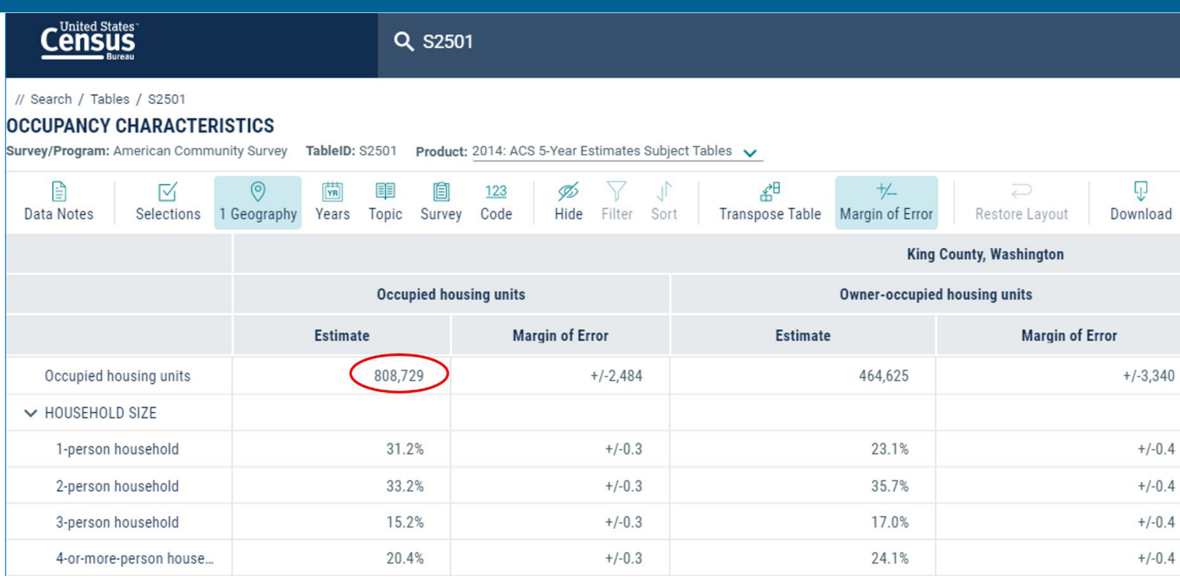
1. Using data.census.gov (Table S2501), Susan first found an estimate of the total number of occupied housing units in King County, WA, in 2010–2014 (808,729) (see Figure 3.17).
2. Using statistical software, she read in the data from the 2010–2014 ACS 5-year PUMS file.[36]
3. Next, she used the PUMS Data Dictionary to find the variables she needed to conduct her analysis.[37]
4. From her previous work with the PUMS data, she knew that King County was made up of 11 Public Use Microdata Areas, or PUMAs, ranging from PUMA 11606 through PUMA 11616. She selected these PUMAs using the PUMA10 variable in the data set.[38]
5. Next, she selected the PUMS variables and categories she needed to determine the percentage of occupied housing units in King County headed by homeowners aged 62 and older.
    a. AGEP (Age) >= 62
    b. RELP (Relationship) = 0 (Household reference person)
    c. TEN (Tenure) = 1 (Owned with a mortgage) or 2 (Owned without a mortgage)
6. A cross-tabulation of these variables showed that approximately 16.2 percent of occupied housing units were headed by homeowners aged 62 and older. Applying that estimate to the total number of occupied housing units from data.census.gov (808,729) yielded an estimate of about 131,000 occupied housing units headed by older homeowners.
7. As a final step, she used the HINCP (Household Income) variable to estimate that among the 131,000 housing units headed by older adults, approximately 40,000 (31 percent) had incomes below the $40,000 tax exemption threshold.

[36] U.S. Census Bureau, American Community Survey (ACS), PUMS Data, <www.census.gov/programs-surveys/acs/data/pums.html>.

[37] U.S. Census Bureau, American Community Survey (ACS), PUMS Technical Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>.

[38] PUMAs are special nonoverlapping areas that partition each state into contiguous geographic units. Each contains roughly 100,000 people.

Figure 3.17. **Table of Occupancy Characteristics in King County, Washington: 2010–2014**



// Search / Tables / S2501

OCCUPANCY CHARACTERISTICS

Survey/Program: American Community Survey TableID: S2501 Product: 2014: ACS 5-Year Estimates Subject Tables

	King County, Washington			
	Occupied housing units		Owner-occupied housing units	
	Estimate	Margin of Error	Estimate	Margin of Error
Occupied housing units	808,729	+/-2,484	464,625	+/-3,340
HOUSEHOLD SIZE				
1-person household	31.2%	+/-0.3	23.1%	+/-0.4
2-person household	33.2%	+/-0.3	35.7%	+/-0.4
3-person household	15.2%	+/-0.3	17.0%	+/-0.4
4-or-more-person house…	20.4%	+/-0.3	24.1%	+/-0.4

Source: U.S. Census Bureau, data.census.gov, <https://data.census.gov>.

Conclusion

The results suggested that there could be 25,000 low-income homeowners eligible to participate in the tax exemption program who were not enrolled (40,000 minus 15,000 currently enrolled).

We set into action an outreach plan to increase enrollment. By reaching into certain neighborhoods with large numbers of lower-income homeowners, we were able to increase the number of homeowners applying for the program.

After 18 months, the Department of Assessments has brought in nearly 7,500 new applications. That represents a nearly 50 percent increase in enrollment.