ACS Small Area Estimates for Doing Neighborhood Analysis

Joe Francis
Director Emeritus

Cornell University
Program on Applied Demographics

Acknowledgements

Funding for this research was provided by Cornell University's Emeritus Faculty funds.

Appreciation to **Laura Close** for the opportunity to present this work.

Special thanks to **Warren Brown** and **Jan Vink** for help, insights and advice.

The author is solely responsible for any mistakes, mischaracterizations and errors of omission.

Introduction

- Information about local demographic processes (births, deaths and migration) as well as characteristics (income, poverty, education, employment status, household occupancy, etc.) are always in demand.
- Local government agencies, community based research organizations, academic researchers and private foundations constantly seek reliable and accurate local demographic information for some project or grant application.
- Indeed the demand for local demographic information has increased over the past couple of decades.

Limitations

- However, researchers and practitioners alike face three problems in providing reliable estimates at the local level (especially census tract and block group levels).
- One problem is the **effective sample size** at subcounty (tracts, block groups) levels for many of the demographic and housing characteristics of interest.
- A second problem is the **absence of accurate measures of error** of the estimates, particularly for non-standard geographies (neighborhoods, CDPs. Tribal areas).
- A third factor, receiving renewed attention, is application of disclosure avoidance systems employed to protect confidentiality.
- The first two of these limitations will be addressed in sequence, beginning with the sample size issue.

Effective ACS Sample Size

- The 2000 decennial Census, collected socio-economic and housing information on a large sample (1 out of 6 households or 16.7%) through an instrument known as the “Long Form.”
- The sample size behind the Long Form, was approximately 47 million households in the 2000 Census.
- By contrast, for a **given year**, the ACS has a **target sample size** of 3.5 million of the currently estimated households of the United States with a goal of capturing 3.5*5=17.5 million households across a five year period, but lately the ACS yields an **effective sample size** only about 2.2 million.
- As Table 1 shows, in **2010** the ACS was based on finalized interviews with a little more than 1.9 million households or about 4% the size of the sample size yielded from the 2000 Census.

Effective ACS Sample Size

Table 1

Initial Addresses and Sample Selected and Final Interviews

	Housing Units		Group Quarters People		
Year	Initial Addresses Selected	Final Interviews	Initial Sample Selected	Final Actual Interviews	Final Synthetic Interviews #
2016	3,527,047	2,229,872	206,415	160,572	131,915
2015	3,540,307	2,305,707	206,630	161,865	134,224
2014	3,540,532	2,322,722	207,403	165,116	129,913
2013 *	3,551,227	2,208,513	207,410	163,663	135,758
2012	3,539,552	2,375,715	208,551	154,182	137,086
2011	3,272,520	2,128,104	204,553	148,486	150,052
2010	2,899,676	1,917,799	197,045	144,948	N/A
2009	2,897,256	1,917,748	198,808	146,716	N/A
2008	2,894,711	1,931,955	186,862	145,974	N/A
2007	2,886,453	1,937,659	187,012	142,468	N/A
2006	2,885,384	1,968,362	189,641	145,311	N/A
2005	2,922,656	1,924,527	N/A	N/A	N/A
2004	838,293	568,966	N/A	N/A	N/A
2003	828,590	572,447	N/A	N/A	N/A
2002	742,409	512,768	N/A	N/A	N/A
2001	858,058	601,875	N/A	N/A	N/A
2000	890,698	587,519	N/A	N/A	N/A

Effective ACS Sample Size

- Moreover, the effective sample size in the 2016 Vintage 1-year ACS is not much larger, with only approximately 2.3 million units (households and group quarters combined).
- Of course, for the ACS 2016 5-year data the effective sample sizes are cumulated for the years 2012 thru 2016, yielding a **period** sample size of 12,111,425, or 8.9% of all housing units estimated to exist in 2016 for the US, about half the size of the old "long form.
- The **decrease** in effective sample size from the Decennial "Long Form" to the ACS has brought major consequences for achieving reliable estimates at the local area levels of geography—e.g. at the tract and block group level.
- And, please note, that effective sample size reported in Table 1 is for the entire US, not states or counties or lower levels of census geography.

Effective ACS Sample Size

Analysts and practioners, working with local data, quickly encounter the consequences of effective sample size of the ACS in their work--large MOEs for variables of interest.

As an example, consider Table 2 showing estimates of housing units and associated margins of error by **occupancy status** from the 2012-2016 ACS for Syracuse NY Tract 19

Table 2

Subject	Census Tract 19, Onondaga County, New York			
	Estimate	Margin of Error	Percent	Percent Margin of Error
HOUSING OCCUPANCY				
Total housing units	2,072	+/-73	2,072	(X)
Occupied housing units	1,968	+/-98	95.0%	+/-3.6
Vacant housing units	104	+/-74	5.0%	+/-3.6

Effective ACS Sample Size

Looking at Table 2, you can quickly see that the ACS 2012-16 estimated **vacant housing units** for tract 19 can be anywhere between 30 and 178 or a spread of 148 units.

Moreover, **occupied housing** units are estimated, with 90% confidence, to be anywhere between 1870 and 2066 or a spread of almost 200 units.

These are quite dramatic differences in such a small geographic area (tract), don't you agree? And, I think its only going to get worse.

To see why the MOEs are so large, consider the ACS 2012-16 unweighted sample size for that same Census tract--illustrated in Tables 3 (next side) for housing units.

Effective ACS Sample Size

Table 3

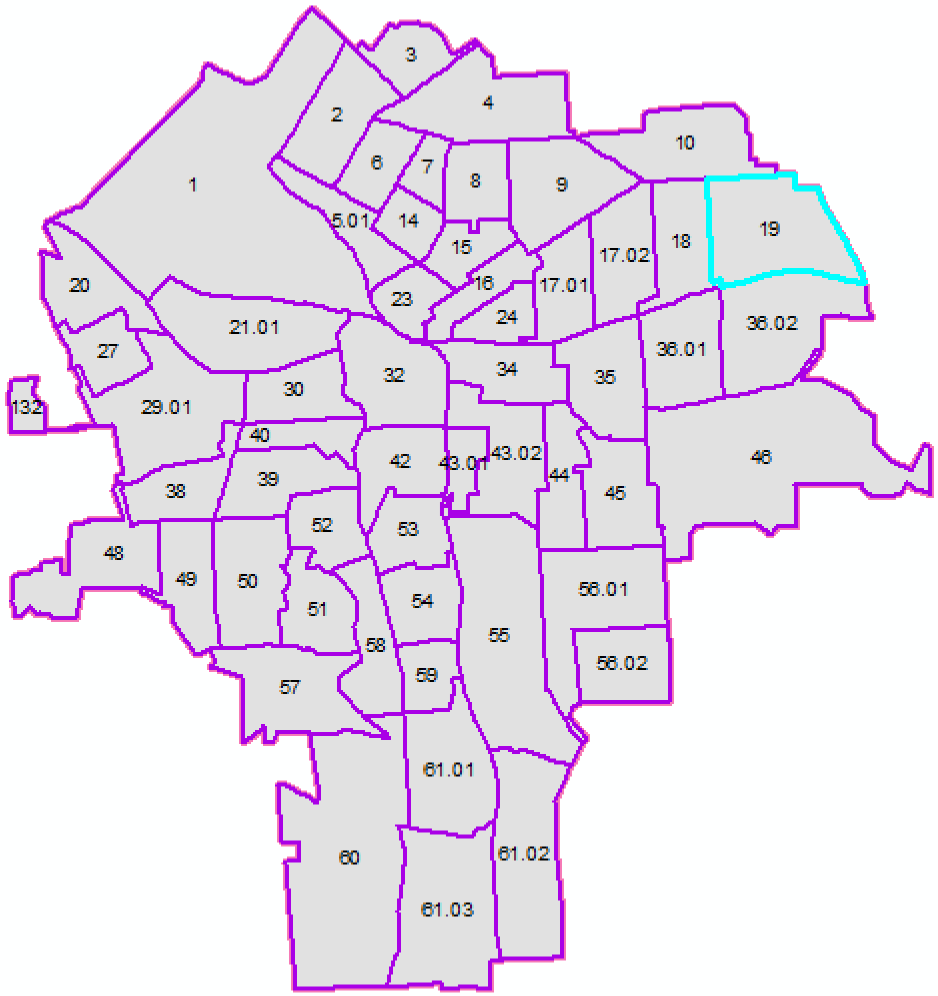
	Census Tract 19, Onondaga County, New York
	Estimate
Total	173

So, the ACS 2012-1 estimate of 2072 total HUs in Tract 19, given above in Table 2 is based on 173 actual sample points. Pretty small! For block groups, the unweighted sample situation is even worse, as shown below.

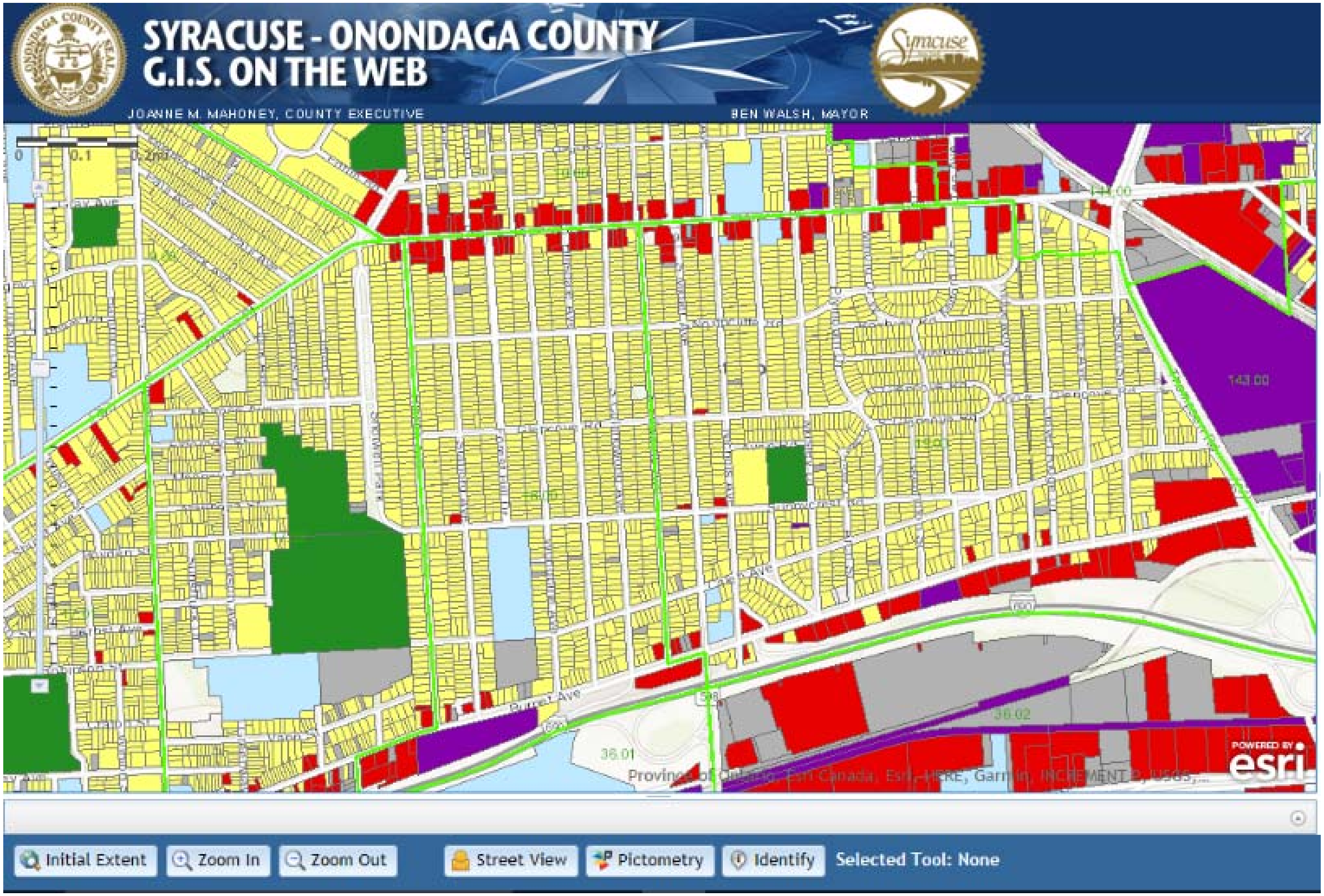
Table 4

Block Group 1, Census Tract 19, Onondaga County, New York	Block Group 2, Census Tract 19, Onondaga County, New York	Block Group 3, Census Tract 19, Onondaga County, New York	Block Group 4, Census Tract 19, Onondaga County, New York	Block Group 5, Census Tract 19, Onondaga County, New York
Estimate	Estimate	Estimate	Estimate	Estimate
31	31	37	38	36

Syracuse Tract 19



Actual Housing Units



Suggested Solutions

Typical suggestions given to counter the vicissitudes of small sample sizes (yielding large MOEs) are:

1. collapse all or selected categories provided in Census ACS summary tables for a given variable, or
2. combine adjacent (or similarly structured non-adjacent) geographies, or
3. do both.

I would like to shift now the focus to the second of the limitations identified above when working with local data—accurate estimates as well as improved measures of reliability of the estimates (MOEs).

The MOE Problem

Since the publication of the ACS Handbooks for Data Users in 2008, users were presented with well worked out procedures for producing aggregated counts, percentages, ratios etc. for combined geographies or collapsed categories.

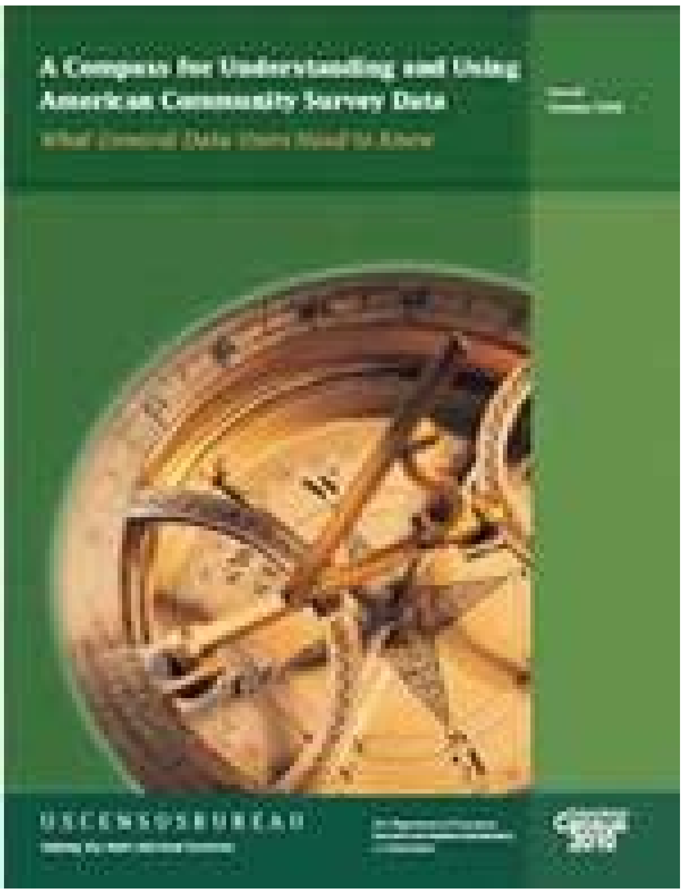
However, the Handbooks' advice was **not** well formulated for measures of the sampling reliability of these aggregate estimates.

The chief reason was the absence of information of an important component for computing the variances of any newly aggregated MOE estimates derived from published summary ACS tables.

To wit, when aggregating counts like total population or total housing units, the Handbooks suggested aggregating up the counts of selected block groups to achieve an accurate estimate of those units for the combined geographies (e.g. tract).

Census Bureau References

Compass Series



ACS Methods Page

Accuracy of the Data

A basic explanation of the sample design, estimation methodology, and accuracy of the data

- 2010-2012 & 2008-2012 Multiyear Accuracy (US) [PDF 319KB]
- 2010-2012 & 2008-2012 Multiyear Accuracy (Puerto Rico) [PDF 361KB]
- 2012 ACS 1-year Accuracy of the Data (US) [PDF 625KB]
- 2012 PRCS 1-year Accuracy of the Data (Puerto Rico) [PDF 604KB]

The MOE Problem

The Handbooks suggested, specifically, that the appropriate way to compute an estimated MOE for that aggregated count was to find the variances of the block groups being combined for the variable of interest, say total population, add them together, and take the square root. See formula 1 below.

$$SE\left(\hat{X}_1 \pm \hat{X}_2\right) = \sqrt{\left[SE\left(\hat{X}_1\right)\right]^2 + \left[SE\left(\hat{X}_2\right)\right]^2} \qquad (1)$$

Unfortunately, in the absence of information about the **covariance**, the above approximation formula often yielded wildly incorrect standard errors and MOEs for the aggregated estimate.

Bias Due to Missing Term

Moreover, the above approximation formula can produce bias in calculation of such Standard Errors, and thereby, in the MOEs. See formula 2 below.

$$SE(\hat{X}_1 \pm \hat{X}_2) = \sqrt{[SE(\hat{X}_1)]^2 + [SE(\hat{X}_2)]^2 \pm 2cov(\hat{X}_1, \hat{X}_2)} \quad (2)$$

The direction of bias on the standard error may be positive or negative depending on the sign of the covariance.

Bias Due to Missing Term

To illustrate the issue of possible bias, consider aggregating **total population** from the five block groups in the Tract 19 we have been examining. The ACS 2012-16 estimates and MOEs for each block group are shown in Table 5.

Table 5

Block Group	Tot Pop Estimate	MOE	Variance
1	689	218	17689
2	571	139	7225
3	945	230	19600
4	766	149	8281
5	866	251	23104
Total	3837	453.19	75899

Bias Due to Missing Term

For comparison to the computations in Table 5, we note that the Census Bureau's published estimates for tract 19 had a total population of 3837 and a MOE for that estimate of 266 in the ACS 2012-16.

Table 6

	Census Tract 19, Onondaga County, New York	
	Estimate	Margin of Error
Total	3,837	+/-266

Comparing the estimate for the aggregated 5 block groups in Tract 19 with the published summary table values for that tract, we see that the estimates agree (previous slide).

Not so for the MOEs (453 vs 266).

Bias Due to Missing Term

As mentioned above, the reason for the huge difference between these two estimates of MOEs in Tables 5 & 6 is the absence of information for the covariances.

That is, the covariances (the -2cov(Xi,Xj)) couldn't be subtracted from the sum of the variances as shown in formula 2 above because of lack of such information.

Fortunately, the Census Bureau recognized the problem and subsequently did three things to correct this problem.

First, they produced what are called **replicate variance tables** for the 2010-14 ACS, the 2011-15 ACS and the 2012-16 ACS, the later of which I will show in a moment.

Variance Replicate Tables

Secondly, the Census Bureau produced documentation about the variance replicate tables including a document showing how to use the tables.
https://www.census.gov/programs-surveys/acs/data/variance-tables.html

Variance Replicate Tables

Variance replicate estimate tables include estimates, margins of error, and 80 variance replicates for selected American Community Survey 5-year detailed tables.

View documentation and table shells on the Variance Replicate Tables Documentation page.

Users should be aware that issues may arise when opening large files in Excel due to the file exceeding the row limit (1,048,576 in current versions), causing Excel to truncate the data. Not all files will have this issue. Data users may need to use other programs to examine the variance replicate estimates in some of these large files.

2016 | 2015 | 2014

2012-2016 ACS 5-year Estimates

5-year Variance Replicate Tables

New ACS User Handbook

Thirdly, the Census Bureau published a new version of their **Understanding and Using American Community Survey Data: What All Data Users Need to Know** in July 2018 which includes at least a mention of the variance replicate tables on page 50.

Additional Background Information and Tools

Sample Size and Data Quality
<www.census.gov/acs/www/methodology/sample-size-and-data-quality/>
This Web page describes the steps the Census Bureau takes to ensure that ACS da
also includes several measures of ACS data quality for the nation and states.

Statistical Testing Tool
<www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>
The Statistical Testing Tool is a spreadsheet that tests whether ACS estimates are s
another. Simply copy or download ACS estimates and their MOEs into the tool to g
tests.

Variance Replicate Tables
<www.census.gov/programs-surveys/acs/data/variance-tables.html>
Variance replicate estimate tables include estimates, MOEs, and 80 variance replica
Detailed Tables. The tables are intended for advanced users who are aggregating A
across geographic areas. Unlike available approximation formulas, this method resu
incorporating the covariance.

Variance Replicate Table B01003

Without going into detail about how to use the tables, let me show a replicate variance table for total population, Table B01003.

The major point, without getting into the weeds, is that employing replicate variance tables in your research does produce the correct MOE values for combined geographies.

Table 7

NAME	TITLE	estimate	moe	Var_Rep1	Var_Rep2	Var_Rep3	Var_Rep4	Var_Rep5
	TOTAL POPULATION							
	Universe: Total population							
Block Group 1, Census Tract 1, Albany	Total	846	212	968	756	842	801	832
Block Group 2, Census Tract 1, Albany	Total	1088	229	1139	1173	1035	1143	1142
Block Group 1, Census Tract 2, Albany	Total	1854	382	1759	1852	1886	1840	1686
Block Group 2, Census Tract 2, Albany	Total	694	349	654	815	392	738	641
Block Group 3, Census Tract 2, Albany	Total	1211	399	1273	990	1268	1192	1267
Block Group 4, Census Tract 2, Albany	Total	672	253	710	701	610	759	627

Using Variance Replicates

Table 8, below, shows the correct results for the aggregated total population estimate and associated MOE for Tract 19 in Onondaga computed using the 2012-16 variance replicate table B01003.

Table 8

Block Group	Total Pop	MOE BG	VarRep 1	VarRep 2	…	VarRep 80	MOE Tract
1	689	218	768	725		708	
2	571	139	643	674		483	
3	945	230	876	825		918	
4	766	149	751	777		833	
5	866	251	936	862		774	
Sum	3837		3974	3863		3716	
Diff			137	26		-121	
$Diff^2$			18769	676		14641	266.1

Using Variance Replicates

The value of having an accurate means of producing estimates and MOEs for combined block group geographies encourages one to use these techniques for more interesting local area analyses.

One example of is combining clusters of block groups into neighborhoods or other "nonstandard" geographies. Another example could be their use in PSAP work or redistricting.

I would like to show some work producing accurate estimates of both the total population and MOEs for Syracuse's envisioned 8 TNT neighborhoods.

The next slide shows a spatial delineation of those envisioned neighborhoods.

Urban Neighborhoods

City of Syracuse and "Tomorrow's Neighborhoods Today" (TNT)

Using Variance Replicates

Using the ACS 2012-16 variance replicate table for total population, table B01003, both population estimates and MOEs were produced for these 8 neighborhoods by combining BGs.

In turn, this allowed us to determine significance of population change in these neighborhoods since Census 2010.

Neighborhood	Census 2010 Pop	ACS 2012-16 Pop	Change	% Change	Lower Confidence Limit	Upper Confidence Limit	Sig
Lakefront	579	674	95	16.4%	589	759	Yes
Northside	43881	44116	235	0.5%	43059.2	45172.8	No
Downtown	4257	4754	497	11.7%	4357.8	5150.2	Yes
Eastwood	6576	6792	216	3.3%	6385.6	7198.4	No
Eastside	36671	36610	-61	-0.2%	25468.2	47751.8	No
Westside	17319	16549	-770	-4.4%	15878.8	17219.2	Yes
Southside	25331	24394	-937	-3.7%	23586.8	25201.2	Yes
Valley	10554	10461	-93	-0.9%	9975.9	10946.1	No

Block Groups with Significant Loss

Another analysis that can be undertaken using this approach is to examine variation of changes. For example, the Census Bureau's Quick Facts, indicatesthe City of Syracuse has lost 1.2% population between April 1,2010 and July 1, 2017.

As the previous table shows, there is considerable variation in population change across different areas of the city.

The Lakefront and Downtown neighborhoods showed considerable percentage gains (16.4% and 11.7%).

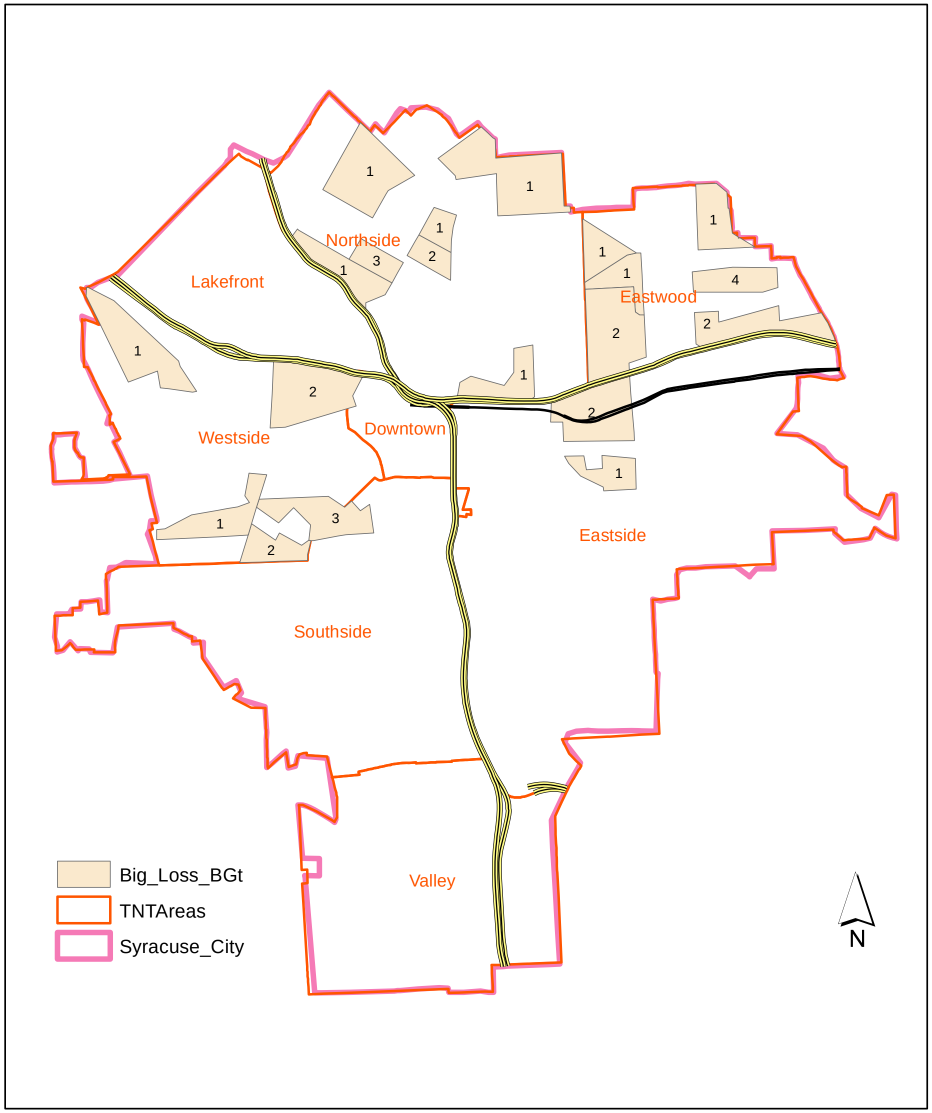
In contrast, the Westside and Southside neighborhoods showed significant percentage losses (-4.4% and -3.7%).

Also, if you are doing a detailed analysis of local areas, you could consider which block groups have the highest loss. The next two slides demonstrate that interest.

Block Groups with Significant Loss

Geography	DEC_10_TotPop	ACS_2012-16_TotPop	ACS2012-216 minus DEC10	TNT	%change
Block Group 1, Census Tract 2, Onondaga	1912	1448	-464	Northside	-0.24
Block Group 1, Census Tract 4, Onondaga	1798	1288	-510	Northside	-0.28
Block Group 1, Census Tract 5.01, Onondaga	1260	1113	-147	Northside	-0.12
Block Group 3, Census Tract 6, Onondaga	1171	1019	-152	Northside	-0.13
Block Group 1, Census Tract 7, Onondaga	960	836	-124	Northside	-0.13
Block Group 2, Census Tract 7, Onondaga	893	769	-124	Northside	-0.14
Block Group 1, Census Tract 9, Onondaga	987	815	-172	Northside	-0.17
Block Group 1, Census Tract 10, Onondaga	756	615	-141	Northside	-0.19
Block Group 2, Census Tract 17.02, Onondaga	1761	1532	-229	Northside	-0.13
Block Group 2, Census Tract 19, Onondaga	689	571	-118	Eastwood	-0.17
Block Group 4, Census Tract 19, Onondaga	901	766	-135	Eastwood	-0.15
Block Group 1, Census Tract 20, Onondaga	1045	753	-292	Westside	-0.28
Block Group 2, Census Tract 21.01, Onondaga	987	836	-151	Westside	-0.15
Block Group 1, Census Tract 24, Onondaga	1296	1167	-129	Northside	-0.10
Block Group 1, Census Tract 27, Onondaga	690	620	-70	Westside	-0.10
Block Group 1, Census Tract 35, Onondaga	1011	878	-133	Eastside	-0.13
Block Group 2, Census Tract 35, Onondaga	844	728	-116	Eastside	-0.14
Block Group 1, Census Tract 38, Onondaga	1217	940	-277	Westside	-0.23
Block Group 2, Census Tract 39, Onondaga	913	631	-282	Westside	-0.31
Block Group 3, Census Tract 39, Onondaga	1909	1483	-426	Westside	-0.22

Syracuse Block Groups with Significant Population Loss from 2010 to 2012-16 Period



Future Work

1. Allocate population in block groups belonging to two or more neighborhoods.
2. Employ the housing unit methodology for estimating population in local areas like neighborhoods.
3. Explore use of variance replicate weighting with variables--occupied housing, population per household and population in group quarters.

Questions