

## **How is Your Health? Using SAS® Macros, ODS Graphics, and GIS Mapping to Monitor Neighborhood and Small-Area Health Outcomes**

Roshni Shah, Santa Clara County Public Health Department

### **ABSTRACT**

With the constant need to inform researchers about neighborhood health data, the Santa Clara County Health Department created socio-demographic and health profiles for 109 neighborhoods in the county. Data was pulled from many public and county data sets, compiled, analyzed, and automated using SAS. With over 60 indicators and 109 profiles, an efficient set of macros was used to automate the calculation of percentages, rates, and mean statistics for all of the indicators. Macros were also used to automate individual census tracts into pre-decided neighborhoods to avoid data entry errors. Simple SQL procedures were used to calculate and format percentages within the macros, and output was pushed out using Output Delivery System (ODS) Graphics. This output was exported to Microsoft Excel, which was used to create a sortable database for end users to compare cities and/or neighborhoods. Finally, the automated SAS output was used to map the demographic data using geographic information system (GIS) software at three geographies: city, neighborhood, and census tract. This presentation describes the use of simple macros and SAS procedures to reduce resources and time spent on checking data for quality assurance purposes. It also highlights the simple use of ODS Graphics to export data to an Excel file, which was used to mail merge the data into 109 unique profiles. The presentation is aimed at introductory SAS users at local and state health departments who might be interested in finding an efficient way to run and present health statistics given limited staff and resources.

### **INTRODUCTION**

The environment where individuals work, live, and play affects their health and well-being, and hence public health agencies are designing various geographically targeted interventions which require data for small geographic areas. As the demand for meaningful data on small area geographies has increased, local health departments and other public health agencies need efficient and simple ways to provide small area data that are easy to access and can be updated on a regular basis.

This paper discusses the data automation process for a neighborhood profile tool created to access sub-county data and inform interventions, support program planning and address resources disparities in Santa Clara County, California. Santa Clara County is a diverse community with 15 cities and nearly 2 million residents, and though the county has one of the highest median household incomes in California, many disparities exist in different parts of the county.

The Santa Clara County Neighborhood Profiles provide data for 109 neighborhoods countywide. Most neighborhoods are in San Jose, the largest city in the county, with nearly 1 million residents; all other cities have fewer than 10 neighborhoods each. Various criteria were established to define neighborhoods within city boundaries including maintain existing boundaries and names, clustering census tracts to use U.S. Census data, and having a target size of 10,000 to 15,000 residents in each neighborhood. The larger population resulting from combining census tracts better enabled meeting reporting thresholds for various data sources and facilitated reporting on a smaller population than a zip code. A number of new neighborhoods were defined by using spatial features like major roads, and selected generic names that included only the city name and a directional indicator, e.g., "Cupertino—Eastside". Areas were designated as "small areas/neighborhoods" in the tool to reflect that some were not based on pre-existing neighborhoods and others contain multiple neighborhoods. This paper will show the process used to create the neighborhood profiles and how these profiles can be used by other organizations seeking to improve access to small area data.

### **ORGANIZING THE DATA**

The first step in preparing the neighborhood specific profiles is to select indicators of interest and download publicly available data. In this paper, all publicly available data come from the United States

Census Bureau and the American Community Survey. Data can be downloaded from the American Fact Finder query system, located at [www.factfinder.census.gov](http://www.factfinder.census.gov). Download the indicators of interest for the geographies that you are interested in. For detailed information on how to download publicly available data from the American Fact Finder query system, see Appendix A.

Once all of the data of interest are imported into SAS as SAS datasets using the import wizard in Base SAS, the first step is to check the datasets to make sure that the datasets have been read in properly. At this time, numerator and denominator variables are also created for some indicators, such as foreign born status, where data may not have been formatted in Excel in order to automate the calculation of percentages later in the syntax.

```
libname profile 'xyz';

data edu;
  set profile.edu;
run;

data foreignborn;
  set profile.foreignborn;
  numerator = estimate_total_foreign_born;
  denominator = estimate_total;
run;
```

## AUTOMATING THE CALCULATIONS: RUNNING THE MACROS

Once the datasets have been checked and the observations match, the next step is to add a variable called “neighborhood” to each dataset and write a macro to convert the 372 census tracts in Santa Clara County into 109 pre-defined neighborhoods. The Santa Clara County Public Health Department worked with its Planning departments to divide the county into specific neighborhoods. Each neighborhood was given a specific name, and neighborhoods were typically made up of an average of 4 to 5 census tracts. The macro below creates a variable named “dat” and assign it to a series of variables from the data we imported into SAS.

```
%let dat=
  edu
  foreignborn;
```

Next, we will automate assigning each census tract in the datasets to its appropriate neighborhood. There are 372 census tracts in Santa Clara County. Since all of the publicly available data were downloaded at the census tract level, an efficient method of renaming the census tracts for all of the indicators was needed. In the macro below, a new variable called “neighborhood” has been created and takes every census tract assigned to a specific neighborhood and gives the neighborhood its pre-existing name. At the end of the macro, a PROC CONTENTS is run for each variable included in the previous macro. This allows the programmer to check that there are 372 observations in each dataset and that the neighborhood variable has been properly added to each dataset as well.

```
%macro var (num);
%let k=1;
%let dat1=%scan (&dat, &k);
%do %while("&dat1" NE "");

  data &dat1;
  set &dat1;
  length neighborhood $60;
  if ct in ('1' '2' '3') then neighborhood=ABC;
  if ct in ('4' '5' '6') then neighborhood=DEF;

  proc contents data=&dat1 noprint;
  run;
```

```

proc contents data=&dat1;
run;

%let k = %eval(&k + 1);
%let dat1=%scan(&dat,&k);
%end;
%mend var;
%var;

```

Now that all of the census tracts in the publicly available datasets have been assigned to a neighborhood name in the county, the next step is to check to make sure that the correct number of neighborhoods have been created and that the variable called “neighborhood” runs properly using basic PROC PRINT and PROC FREQ syntax.

```

proc print; var neighborhood; run;
proc freq; tables neighborhood; run;

```

Once the neighborhood variable runs successfully, the next step is to automate the calculation of basic statistics which will be used in the neighborhood profiles. For this paper, we have chosen to demonstrate how percentages can be automated using macros. We first create a new variable called “dat2”, which defines the selected variables for which percentages will need to be automated and calculated. Note that these are the same variables for which the “numerator” and “denominator” variable was created at the very beginning of the syntax. The “numerator” and “denominator” variables will later be used in the PROC SQL macro code to automate the calculation of the percentages.

```

%let dat2=
    foreignborn;

```

Now that the “dat2” variable is created, we will work on reading in each dataset and calculating the appropriate statistic. Note that the profiles would like to show the percentage statistic at a neighborhood level, so we must select and group by the neighborhood variable in our PROC SQL code. The PROC SQL code inside of the macro takes each dataset, provides a title for the output using the variable name, calculates the percentage statistic, formats the statistic as a percentage with a percentage sign and groups it by neighborhood to provide us with a percentage of each of the 109 neighborhoods in Santa Clara County.

```

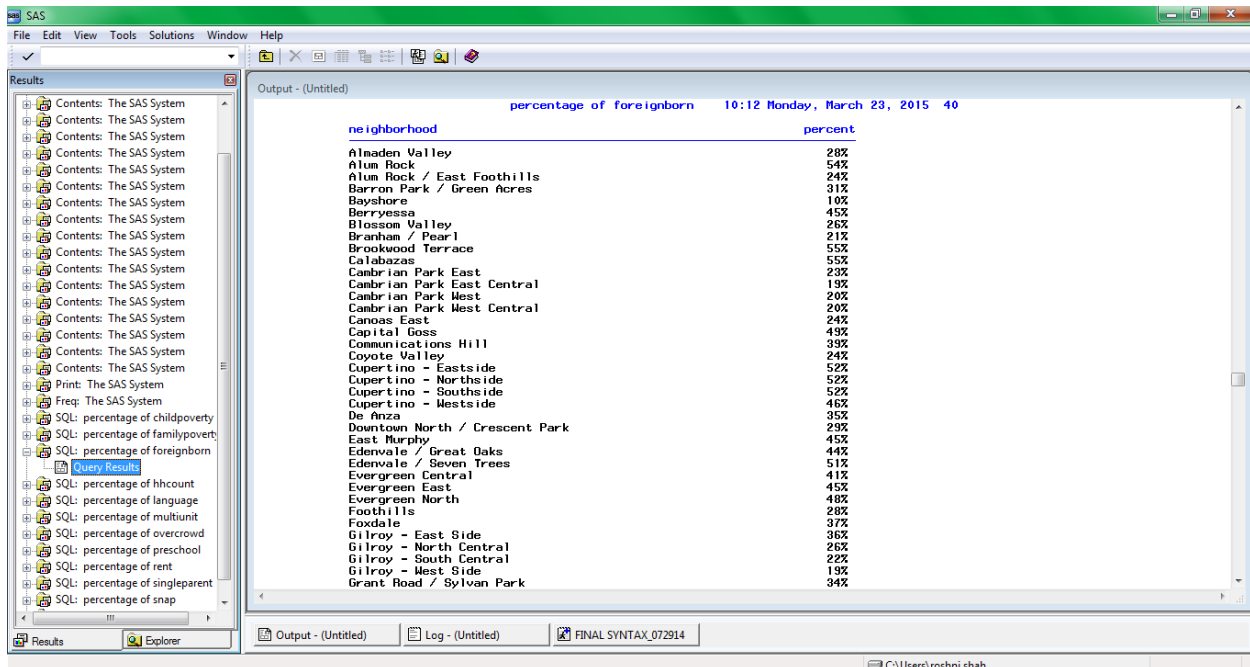
%macro percent (num);
%let k=1;
%let dat3=%scan (&dat2, &k);
%do %while("&dat3" NE "");

    title "percentage of &dat3";
    proc sql;
    select neighborhood, (sum(numerator)/sum(denominator)) as percent
    format percent6.
    from &dat3
    group by neighborhood;
    quit;

%let k = %eval(&k + 1);
%let dat3=%scan(&dat2,&k);
%end;
%mend percent;
%percent;

```

When this macro is run, the PROC SQL output will be displayed in the output window. The output below is an example of the data for foreign born status.



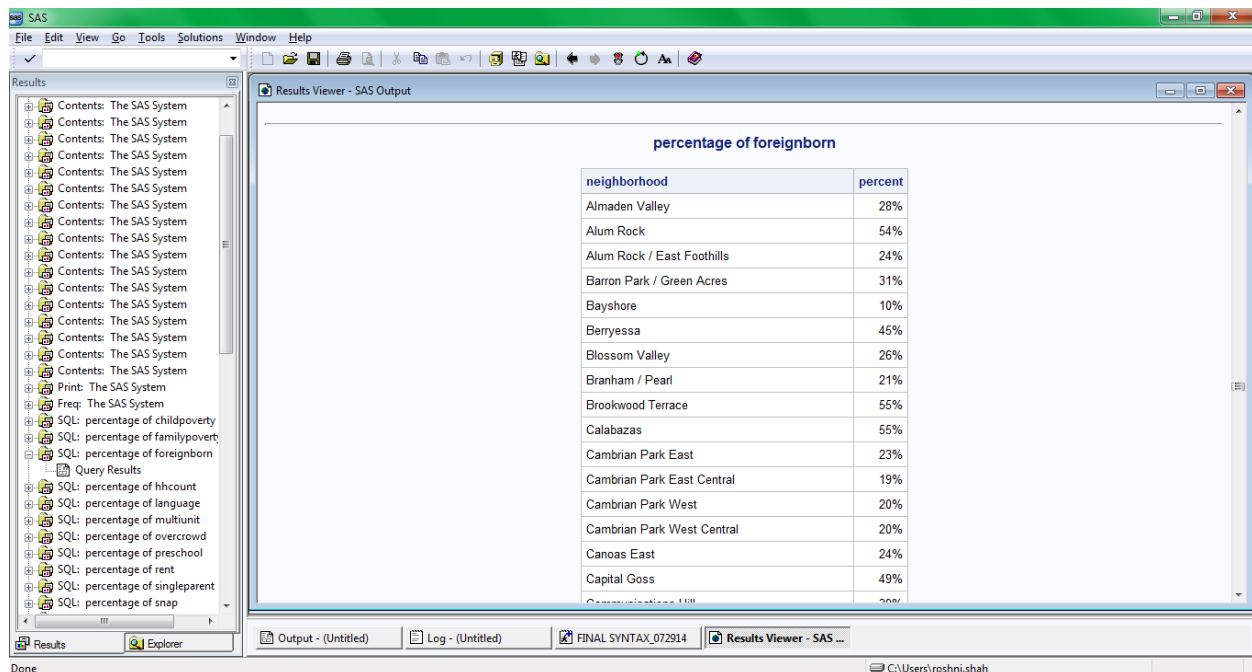
neighborhood	percent
Almaden Valley	28%
Alum Rock	54%
Alum Rock / East Foothills	24%
Barron Park / Green Acres	31%
Bayshore	10%
Berryessa	45%
Blossom Valley	26%
Branham / Pearl	21%
Brookwood Terrace	55%
Calabazas	55%
Cambrion Park East	23%
Cambrion Park East Central	19%
Cambrion Park West	20%
Cambrion Park West Central	20%
Canoas East	24%
Capital Goss	49%
Communications Hill	39%
Coyote Valley	24%
Cupertino - Eastside	52%
Cupertino - Northside	52%
Cupertino - Southside	52%
Cupertino - Westside	46%
De Anza	35%
Downtown North / Crescent Park	25%
East Murphy	45%
Edenvale / Great Oaks	44%
Edenvale / Seven Trees	51%
Evergreen Central	41%
Evergreen East	45%
Evergreen North	48%
Foothills	28%
Foxdale	37%
Gilroy - East Side	36%
Gilroy - North Central	26%
Gilroy - South Central	22%
Gilroy - West Side	19%
Grant Road / Sylvan Park	34%

**Display 1. Output of Foreign Born Status by Neighborhood using PROC SQL**

Notice how the default output destination does not convert to Excel easily; therefore we use the ODS Graphics feature to find a more structured way to output the data. In Base SAS, the option to output results using ODS Graphics can be selected by following the steps below:

1. Click "Tools" at the top of the ribbon
2. Click on "Options"
3. Click on "Preferences"
4. When the "Preferences" window pops up, click on the "Results" tab
5. Check "Use ODS Graphics" under "Results options" and "Create HTML" if you wish to create an HTML link in your work or any other specific folder

Once the ODS Graphics option is checked and selected, rerun the final macro which created the neighborhood specific percentages. The resulting output will look like cleaner and can then be exported into Microsoft Excel in order to compile the data by neighborhood.



**Display 2. Output of Foreign Born Status by Neighborhood using PROC SQL with ODS Graphics**

Once exported into Microsoft Excel, the data can be compiled into a single sheet in a workbook. It is crucial that data be labeled with clear column headers that are identical to the template of indicators originally created in a Microsoft Word file. See Appendix B for a copy of the template of indicators that was used to merge the data into for each neighborhood in Santa Clara County. Once all of the data have been compiled into a single sheet in an Excel workbook, the data can be mail-merged into Microsoft Word documents using Microsoft's simple guide (see Appendix C). Each file can then be converted into a PDF file and distributed for planning, programmatic and policy use.

Visit <http://www.sccgov.org/sites/sccphd/en-us/Partners/Data/Pages/City-Profiles.aspx> to view the detailed city and neighborhood profiles produced for all 15 cities and 109 neighborhoods in Santa Clara County.

## Evergreen Central profile 2014

### Demographic Snapshot

	Evergreen Central	SCC
Population Size	16,561	1,781,642
<b>Race/Ethnicity</b>		
African American	3%	2%
Asian/Pacific Islander	50%	32%
Latino	30%	27%
White	14%	35%
Foreign-born	41%	37%
Speaks a language other than English at home	64%	51%
Single parent households	11%	7%
Households with children	50%	38%
Average household size	3.795	2.90
<b>Age Groups</b>		
0-5 years	8%	8%
6-11 years	9%	8%
12-17 years	9%	8%
18-24 years	9%	9%
25-34 years	13%	15%
35-44 years	15%	16%
45-54 years	15%	15%
55-64 years	12%	10%
Ages 65 and older	10%	11%

Sources: U.S. Census Bureau, 2010 Census and 2007-2011 American Community Survey 5-Year Estimates



Display 3. Sample Screenshot of Final Neighborhood Profile Output

## GOING BEYOND THE NUMBERS: MAPPING THE DATA

Visual statistics tend to be more compelling and interesting to individuals from various disciplines. With the growing interest in online mapping, we integrated our small geography data with the power of mapping tools, such as ArcGIS online, in order to create an interactive thematic atlas tool to help programs, stakeholders and various other healthcare organizations plan future interventions and campaigns.

Since data were downloaded at the census tract level and analyzed at both the neighborhood and city level, indicators could be mapped at three different geographies to identify various disparities. ArcGIS online offers a collaborative cloud-based platform to create and share maps and data. The data were analyzed and prepared for webmap creation in Excel. Map shapefiles were created in ArcGIS software for desktop and then tiled into ArcGIS online, where the layers were used to create an online atlas for maps. The atlas included a set of demographic indicators, with pop-ups and the ability to share maps, for three levels of geographies: 15 cities, 109 neighborhoods and 372 census tracts in Santa Clara County. To view the Santa Clara County Thematic Atlas, visit

<http://www.sccgov.org/arcgis/SantaClaraCountyThematicAtlas>.

## CONCLUSION

There is a great need for small area statistics in order to plan appropriate local healthcare interventions. SAS can be a great tool for researchers to efficiently automate the display and calculation of statistics therefore avoiding manual data entry and repeated quality assurance checks. This method can be a good way for health jurisdictions with limited resources to quickly produce reports and profiles and continue their public health outreach and efforts to improve the health status of their jurisdictions.

## ACKNOWLEDGMENTS

I would like to thank the team of epidemiologists at the Santa Clara County Public Health Department and the GIS analysts at the Santa Clara County Information Systems Department for their support, hard work and technical skills in the entire process of creating the Santa Clara County Neighborhood Profiles. I would also like to thank my mentor, R. Scott Leslie, for his encouragement, motivation and expertise throughout the SAS Global Forum application, abstract approval, and paper writing and presentation creation process. Finally, I would like to thank the SAS Global Forum 2015 Team for selecting my abstract and acknowledging my hard work by rewarding me with the 2015 Junior Professional Award.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Roshni Shah  
Santa Clara County Public Health Department  
408-792-5021  
Roshni.Shah@phd.sccgov.org  
[www.sccphd.org/statistics2](http://www.sccphd.org/statistics2)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.



## APPENDIX A

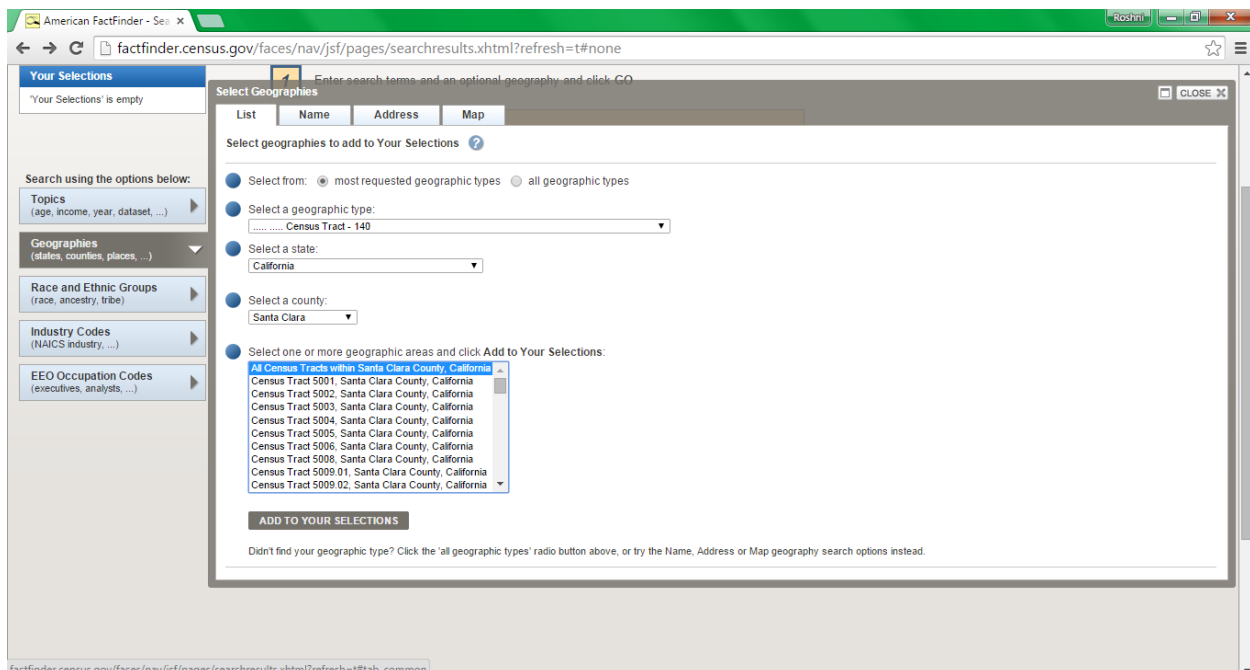
### Downloading publicly available data from the American Fact Finder Query System

Educational attainment for Santa Clara County is used as an example to show how Census and American Community Survey data can be easily downloaded for various geographies of interest.

Using the sidebar, select “Geographies”.



Select “Census Tract” as your geographic type, followed by “California” for State”, “Santa Clara” for county and “All Census Tract within Santa Clara County, California” for your geographic areas. Click “Add to your selections”.





Using the options on the left, select “Topics” and click on “Dataset”. For the purposes of this project, the 2011 ACS 5-year estimates were used. It is recommended that the most recently available data be used for your planning and project purposes.

The screenshot shows the American FactFinder website interface. The top navigation bar includes links for MAIN, COMMUNITY FACTS, GUIDED SEARCH, ADVANCED SEARCH, and DOWNLOAD CENTER. The left sidebar contains search filters for Topics, Geographies, Race and Ethnic Groups, Industry Codes, and EEO Occupation Codes. The main content area displays search results for '2011 ACS 5-year estimates' with a table listing various datasets and their corresponding years.

Dataset	About
2014 Population Estimates	?
2014 Population Estimates	?
2013 ACS 1-year estimates	?
2013 ACS 3-year estimates	?
2013 ACS 5-year estimates	?
2013 ACS 1-year estimates	?
2013 ACS 3-year estimates	?
2013 ACS 5-year estimates	?
2013 ACS 1-year estimates	?
2013 ACS 3-year estimates	?
2013 ACS 5-year estimates	?
2013 ACS 1-year estimates	?
2013 ACS 3-year estimates	?

Search for your indicator of interest using the search tool and download the table. Educational attainment by census tract is table number B14002. Download this table as a CSV file so that it can easily be read into SAS, selecting the option to download the data and annotations in a single file. Once downloaded, four files will be available to you: two CSV files and two text files. Note that the metadata and annotated data are separated into two files, which have to be combined in Excel prior to importing in SAS as a CSV. Once the file is compiled, all additional calculations to collapse categories at the census tract level were done in Excel. Fields were compiled into four major categories: less than high school, high school graduate, some college or associates degree, and college graduate or higher. Percentages were calculated for each of the four categories at the census tract level, and the final CSV file was imported into Base SAS using the import wizard. This process was repeated for all of the variables of interest where Census and/or American Community Survey data for the indicators of interest were available.

SAS - [VIEWTABLE: Profile.Edu]

File Edit View Tools Data Solutions Window Help

Explorer

Contents of 'Profile'

Age\_grps Avg\_hhsize Childpoverty

Commute Ct Edu

Familypov... Foreignborn Hhcount

Hhsize Income Language

Multunit Overcrowd Preschool

Race Rent Singleparent

Snap Tot\_housi... Unemploy...

	ct	Geography	Estimate: Total:	Margin of Error, Total:	less than high school	high school graduate	some college or associates degree	college graduate or higher	pct_less than high school	pct_high school graduate
1	06085500100	Census Tract 5001, Santa Clara County, California	4066	439	1341	1084	736	905	0.3298081653	*****
2	06085500200	Census Tract 5002, Santa Clara County, California	4181	465	537	762	1120	1762	0.1284381727	*****
3	06085500300	Census Tract 5003, Santa Clara County, California	2199	209	289	306	526	1078	0.1314233743	0.139154161
4	06085500400	Census Tract 5004, Santa Clara County, California	1725	182	366	321	375	663	0.212173913	*****
5	06085500500	Census Tract 5005, Santa Clara County, California	3826	292	395	683	816	1932	0.1032409827	*****
6	06085500600	Census Tract 5006, Santa Clara County, California	3495	360	364	719	711	1701	0.104148784	*****
7	06085500800	Census Tract 5008, Santa Clara County, California	1828	207	340	346	461	681	0.1859956236	*****
8	06085500901	Census Tract 5009 01, Santa Clara County, California	2144	222	253	384	450	1057	0.1180037313	*****
9	06085500902	Census Tract 5009 02, Santa Clara County, California	870	132	199	135	207	329	0.2287356322	*****
10	06085501000	Census Tract 5010, Santa Clara County, California	3214	475	995	545	600	1074	0.3095830741	*****
11	06085501101	Census Tract 5011 01, Santa Clara County, California	2737	417	469	628	360	1280	0.1713554987	*****
12	06085501102	Census Tract 5011 02, Santa Clara County, California	2934	381	966	635	758	575	0.3292433538	*****
13	06085501200	Census Tract 5012, Santa Clara County, California	2456	254	636	463	890	467	0.2589576547	*****
14	06085501300	Census Tract 5013, Santa Clara County, California	2321	325	318	338	428	1237	0.1370099095	0.145626885
15	06085501401	Census Tract 5014 01, Santa Clara County, California	1756	345	828	423	326	179	0.4715261959	*****
16	06085501402	Census Tract 5014 02, Santa Clara County, California	1658	225	767	305	273	313	0.4626055489	*****
17	06085501501	Census Tract 5015 01, Santa Clara County, California	2593	300	845	743	474	531	0.3258773621	*****

Results Explorer

Output - (Untitled) Log - (Untitled) FINAL SYNTAX\_072914 VIEWTABLE: Profile.Edu

## APPENDIX B

### Sample Template of Santa Clara County Neighborhood Profile Indicators

PROFILE 2014

#### DEMOGRAPHIC SNAPSHOT

	CITY	SCC
Population Size		1,781,642
<i>Race/Ethnicity</i>		
African American		2%
Asian/Pacific Islander		32%
Latino		27%
White		35%
Foreign-born		37%
Speaks a language other than English at home		51%
Single parent households		7%
Households with children		38%
Average household size		2.90
<i>Age Groups</i>		
0-5 years		8%
6-11 years		8%
12-17 years		8%
18-24 years		9%
25-34 years		15%
35-44 years		16%
45-54 years		15%
55-64 years		10%
Ages 65 and older		11%

**Sources:** U.S. Census Bureau, 2010 Census and 2007-2011 American Community Survey 5-Year Estimates

#### ECONOMIC AND EDUCATIONAL OPPORTUNITIES

##### Income and job opportunities

	NAME	SCC
Median household income		\$89,064
Unemployed (ages ≥ 16 years)		6%
<i>Poverty</i>		
Families below 185% FPL		16%
Children (ages 0-17) below 185% FPL		24%

**Sources:** U.S. Census Bureau, 2010 Census and 2007-2011 American Community Survey 5-Year Estimates

## APPENDIX C

### **Mail-Merging Excel Data into Microsoft Word 2010**

In order to merge the data compiled in Microsoft into the final neighborhood profile temple (see Appendix A for portion of the template), a mail-merging process was used. The first step was to use Dynamic Data Exchange (DDE) to connect to our Excel file of data. In order to do this and complete the mail merge process, Microsoft has provided a step-by-tutorial for various versions of Microsoft Word. The detailed tutorial used for our process can be found here: <https://support.microsoft.com/en-us/kb/320473>.