

October 23, 2019

To: Teri Silva and Jacqueline Guzman

cc: Marguerite Leoni and John Nagel

Fr: Douglas Johnson, President

Re: Sunnyvale Datasets

Unfortunately there was an issue in the dataset used to generate numbers and percentages reported in the NDC-generated spreadsheets for the draft districting maps that were published on the City's districting webpage. The issue has been identified and corrected, and the new dataset tested for other potential issues. This is National Demographics' report on the discovery and analysis of the issue, its impact on the demographic reports for the City, its correction, and testing to identify whether there are other issues in the data set. No other defects have been identified.

On Thursday, October 17<sup>th</sup>, I received an email from Nicole Wong of Asian Americans Advancing Justice – Asian Law Caucus. Her email very politely “wanted to bring to [my] attention” some odd differences in the demographic data for Sunnyvale, in particular how the Citizen Voting Age Population counts for District 2 of Map 120C changed between the online mapping tool numbers and the numbers reported in the NDC-generated spreadsheets for the map that were published on the City's districting webpage.

My initial impression was that the differences were due to the change from the online mapping tool's use of 2012-2016 American Community Survey (ACS) data and NDC's use of the more-recent 2013-2017 ACS data when processing the maps (the more recent data were not available at the time the online tool was built). But as we dug into the question in more detail, we discovered some Census Blocks were missing part of their dataset.

Our districting datasets consist of data from five sources:

1. The 2010 decennial Census;
2. California Statewide Database voter data from one gubernatorial general election;
3. California Statewide Database voter data from one presidential general election;
4. The most recent five-year American Community Survey data;
5. The Census Bureau's “Special Tabulation of Citizen Voting Age Population Data” from the most recent five-year American Community Survey.

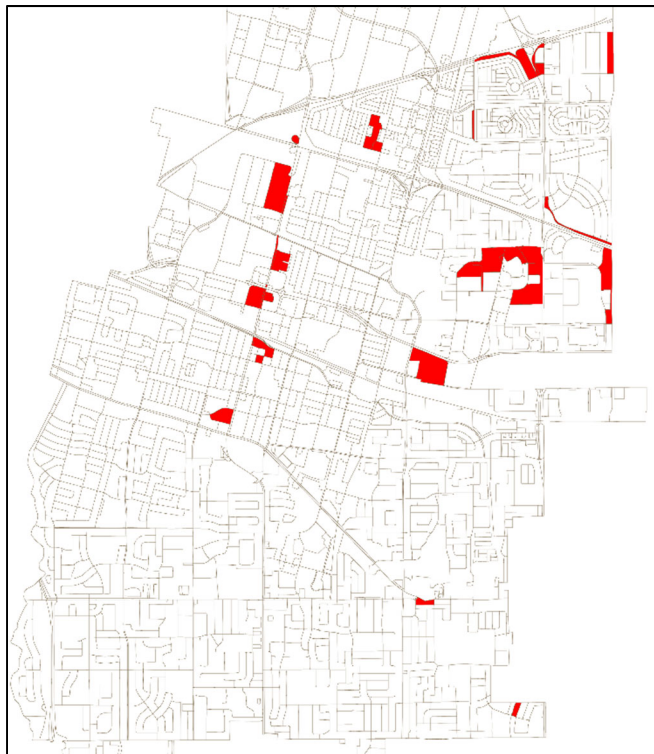
Census Blocks (essentially the equivalent of city blocks) are the small building blocks used to build districting plans. But only the decennial Census data are provided at the Census Block level of geography. The California Statewide Database data are provided at the “consolidated election precinct” level of geography. The five-year ACS data is provided at the “Tract” level of geography. And the Special Tabulation is provided at the “Block Group” level of geography. NDC takes each of these databases and “disaggregates” the data down to the Census Block level of geography. Each disaggregated dataset has a Census Block identifier for each Census Block containing data from that dataset. The last step in the process is that NDC joins together all of those different datasets into one big Census Block dataset, using those Census Block numbers as the common link among the datasets.

Unfortunately, when we at NDC updated the dataset in August to move from the 2014/2016 elections to 2016/2018, and to move from the 2012-2016 to 2013-2017 ACS and Special Tabulation data, the field

used to join the data was missing some of the Census Block identifiers. The Census Blocks missing identifiers did not receive the updated ACS, 2018 election, and Special Tabulation data, leaving blanks in those parts of the dataset for those Census Blocks.

As you will see in the map below, there is not a systematic pattern to which area of the City was impacted – the impacted Census Blocks (shown in red) are everywhere from Mountain View/Alviso Rd. in the north to the southeast corner of the City. But more of them are in the north-central portion of the City than in other areas of the city. In most of the proposed districts in the four preferred maps, the defect makes either no difference in the CVAP percentages reported, or makes a difference of only a percentage point or two (well within the existing margin of error for percentages calculated from disaggregated ACS, Statewide Database and Special Tabulation data). Given the concentration of Census Blocks lacking data in the northern and central part of the City, correcting the dataset will result in a greater change in the Citizen Voting Age Population percentages for districts in those areas. For example, in Map 120C, the change in the District 2 percentage of Asian/Pacific Islander CVAP is about 7%. Correcting the data brings the numbers and percentages in line with the numbers and percentages a residents saw when he/she created the map in the online mapping tool (although they are not exactly the same, because the data is more current), as the percentages and counts residents saw as they drew the maps were correct.

**Map of Census Blocks where Data Errors Occurred**



Some notes on the scope of this issue:

The data used by residents to draw their maps in the online mapping tool are correct and accurate.

The vital “total population” counts are not involved in the problem, so those remain exact and correct for each district in each map.

The data does not affect the placement of district boundaries because the total population counts are correct.

The error only appeared in NDC’s demographic spreadsheets created after each map was submitted – the data residents saw as they drew their maps using the online tool or the Participation Kit was correct. In each map the district most affected is the district covering the triangle bordered by Mountain View/ Alviso Rd (State Route 237) on the north, Highway 101 on the south, and the City’s eastern border: In the

current focus maps, these are District 1 in map 107A; District 2 in 120A and 120C; and District 4 in 121A.

Updated demographic profiles for each of the four maps are attached, labeled “rebuild.” For comparison, the original NDC demographic profiles (labeled only with the map number) are also included. And updated PDF packets (with unchanged maps but updated demographic sheet second pages) are ready for posting on the city website.

## National Demographics Corporation

To ensure the updated dataset is accurate, we have performed the following quality control tests:

1. Compared the sum of all the Census Blocks to the Census Bureau's reported totals for the City as a whole. The results were accurate within 1%, which is an expected range given the imprecision inherent to disaggregation of data down to the Census Block level of geography.
2. Compared the rebuilt Census Block database to the Maptitude Online Census Block database (since we know the latter is correct). All Census Blocks that had ACS and Special Tabulation data in the Maptitude Online database also have ACS and Special Tabulation data in the rebuilt Census Block database, which is what we hoped to see.
3. Compared the CVAP reported from the 2012-2016 ACS for each Census Block, to the CVAP reported from the 2013-2017 ACS for each Census Block. Small increases were prevalent (as expected), and where we found decreases we confirmed those decreases are in the source Census Special Tabulation Census Block Group level data.
4. Confirmed that in all Census Blocks the 2016 and 2018 counts of registered voters are higher than the respective counts of voter turnout.

All of these data checks confirmed the accuracy of the new database.

Again, my apologies both personally and on behalf of NDC for the mistake. I am happy to provide any additional information or answer any questions that you may have about these datasets and what happened with them at the October 29 Council Hearing, or whenever you wish to discuss these issues.