Final Report

Demographic Comparison: Exploring data from the U.S. Census Bureau and the NYC Department of Health and Mental Hygiene

Executive Summary

Harlem Children's Zone (HCZ) requested help in finding resources to answer the following question: "How many babies are born each year who reside in the Zone?" HCZ requested birth rate information for the past 5 years. The organization stated that their challenges included "finding data on population and demographic changes in the Zone, since it's not an area of land that maps perfectly into any tract, block group, school district, etc." HCZ stated: "often times we have to make educated guesses based on the information available."

The main project aim was to find better matching data in order to answer the question "How many babies are born each year who reside in the Zone?" as accurately as possible. HCZ stated that their ideal goal was to build a model to approximate the number of children born each year in the Zone for the past 5 years.

I used the U.S. Census Bureau as a starting point to see what data was publicly available to answer HCZ's project question. My search eventually led me to the NYC Department of Health and Mental Hygiene, Bureau of Vital Statistics. The Bureau of Vital Statistics publishes data based on birth records on an annual basis for different neighborhoods, technically known as community districts, within New York City. The agency also provides a Natality Module, which would address the need of a model to approximate number of births in the Zone.

Birth Data is only available at the community district level or the census tract level. The Zone does not fit into a community district perfectly, which is a challenge that HCZ has identified. Although data specific to the Zone is available utilizing the census tract data set (i.e. filtering for census tracts that fall into the Zone), that data set only has data for 5 year estimate chunks (as opposed to annual data). This means that there is significantly less data available to answer the question "How many babies are born each year who reside in the Zone?"

However, there is significant overlap between the Zone and Community District 10 of Manhattan. The project question now becomes, "Can data for the population in Community District 10 be used as a 'proxy' data set for the population within the Zone?" One of the advantages of using the Community District 10 Population as a proxy for the Zone is that there are several existing data resources available for this population. If Community District 10 is deemed a suitable proxy for the population within the Zone, HCZ would be able to use those resources to find out the number of births in the Zone each year.

This demographic analysis used data from the U.S. Census Bureau to compare the distributions of several population attributes (such as age and gender) between the sample populations found within the Zone and Community District 10.

In conclusion, the sample populations within the Zone and Community District 10 are similar enough to validate the use of Community District 10 data resources to the answer the question "How many babies are born each year who reside in the Zone?"

Introduction

Project Objectives:

- Determine whether or not the population in Community District 10 can be used as a 'proxy' for the population within the Zone?
- Share data resources.
- Recommend next steps for building a model to approximate the number of children born each year in the Zone.

HCZ provided a link (https://hcz.org/zone-map/) to a map of the area that the organization identifies as the Zone -- which encompasses the community that HCZ serves. The only publicly available data over birth rates specific to neighborhoods in New York City is from the NYC Department of Health and Mental Hygiene, Bureau of Vital Statistics. The project goal is to find birth rate data that is specific to the Zone, the area of interest. In terms of geographical units (e.g., State, City, Zip Codes, Blocks, etc.), the closest data available to the area of interest (The Zone) was for census tracts and NYC community districts. The following link leads to the web page with the birth rates data sets:

- http://www1.nyc.gov/site/doh/data/data-sets/vital-statistics-birth-micro-sas-datasets.page

The birth rate data available for census tracts could be filtered to include only census tracts that fall within the Zone in order to create a data set specific to area of interest. However, there is a major limitation to this data set. Unlike the data for the community districts, the data for census tracts are only available as 5-year aggregate counts. The community district data are available as 1-year aggregate counts. Furthermore, several of the available data sets for census tracts overlap with each other (e.g., 5-year estimates for 2004 - 2008 overlap with 5-year estimates for 2006 - 2010). This means that there is more robust birth rate data available for community districts than for census tracts. Another difference between the data available for tracts and community districts is that there are only one-way counts available for census tract data. This means that data for community districts has been verified by data from newborns as well as data from the mothers (two-way counts), while data for census tracts has only been verified by data from birth records (one-way counts).

Although the Zone does not map perfectly into any of the community districts, there is significant overlap between the Zone and Community District 10 of Manhattan (also identified as 'Neighborhood Central Harlem'). This overlap may be an advantage for the organization. If the population in Manhattan's Community District 10 is similar enough to the population in the Zone, HCZ can leverage the available data resources for Community District 10 -- Central Harlem. Due to its designation as a Public Use Microdata Area (PUMA), there are several existing data resources available for Community District 10 in addition to the birth rates data from the NYC Bureau of Vital Statistics.

Public Use Microdata Areas (PUMAs) are statistical geographic areas defined by the U.S. Census Bureau for the dissemination of Public Use Microdata Sample (PUMS) data. This data includes demographic data collected by the American Community Survey (which collects the same data as the U.S. Census, but more frequently). Starting in 2010, PUMAs were defined as areas that nest within states or equivalent entities, contain at least 100,000 people, are built on census tracts and counties, and should be geographically contiguous (https://www.census.gov/geo/reference/puma.html).

Besides the birth rates data, the NYC Bureau of Vital Statistics also provides an Epiquery Natality Module (https://a816-healthpsi.nyc.gov/epiquery/Birth/index.html). This module provides data for the total number of births in Community District 10 for the past 15 years (excluding 2016 and 2017). The module also provides data for the following: birth rate, fertility rate for women aged 15-44, whether the mother's ancestry is Hispanic, whether the mother is foreign born, percent of: first live births, low birth weight, preterm births, late or no

prenatal care, mothers not married, and several other factors. Results can be downloaded to a spreadsheet or another statistical program.

Moreover, there are existing resources for demographic data for the population in Community District 10. The following links provide data and visualizations for several types of data, including demographics, economics, family statuses, housing, and social factors:

- https://censusreporter.org/profiles/79500US3603803-nycmanhattan-community-district-10central-harlem-puma-ny/
- https://statisticalatlas.com/neighborhood/New-York/New-York/Harlem/Overview

The table below (*Table 1*) summarizes the advantages and disadvantages of using data available at the census tract level versus what is available at the community district level.

Geographic Level	Advantages	Disadvantages
Community District 10	1-year aggregate counts 15 datasets available - Other data available (including visualizations) - Model available - Two-way counts available	- Not specific to Zone - Imperfect overlap with the Zone - Need to compare with population within the Zone
Census Tract	- Specific to the Zone - Demographic data available	5-year aggregate counts - Overlap between datasets - Only one-way counts available - LESS data available

Table 1: Summary of the advantages and disadvantages of the available birth rate data.

Considering the advantages to using data available at the community district level, the project question now becomes: *Can data for the population in Community District 10 be used as a 'proxy' data set for the target population within the Zone?* In other words, is the Community District 10 population similar to the population in the Zone? In order to answer this question, a demographic analysis was conducted in which data from the U.S. Census Bureau's American Community Survey was used to compare the distributions of several population attributes (such as age and gender) between the sample populations found within the Zone and Community District 10.

Data

For this demographic comparison, data sets were obtained from the [American Fact Finder](#) web tool, provided by the U.S. Census Bureau. In terms of geographical units, the most granular demographic data publicly available for the areas of interest (i.e. The Zone and Community District 10) were 5-year estimates (aggregate counts for the time period) at the census block group level.

This demographic data was collected via the American Community Survey (ACS). The questions used in the ACS are very similar to those included on the long form used in Census Survey. The main difference between the ACS and Census Survey is that the ACS is conducted more frequently. The Census is a cross-sectional snapshot of the U.S. population, administered once every 10 years. ACS data is produced every year, but in exchange for this benefit, the sample size of the ACS is considerably smaller than that of the Census 2000 long-form sample. The main purpose of the Census is to provide counts of people for the purpose of congressional apportionment and legislative redistricting. While the main purpose of the ACS is to measure the changing social and economic characteristics of the U.S. population. ACS estimates are controlled to match the Census Bureau's annual population estimates, by age, sex, race, and Hispanic origin.

In 2010, the U.S. Census Bureau began publishing 5-year estimates for geographic areas down to the block group level. 5-year estimates are aggregates of counts for a 5-year period. Single-year estimates are only available for larger areas such as cities, towns, and census designated places (e.g., PUMAs). For geographic areas with smaller populations (i.e. zip codes, census tracts, census blocks), the ACS samples too few households to provide reliable single year estimates. For these communities, several years of data are pooled together to create reliable 3-year or 5-year estimates.

For this demographic comparison, the time period of 2011 – 2015 was chosen instead of the recently released 2012 – 2016 estimates. The rationale for this is that the available data sets for birth rates data (from the NYC Bureau of Vital Statistics) only includes data for up to year 2015. In an effort to ensure that the populations analyzed in this demographic comparison are the same populations for the birth rates data sets (which is the data that HCZ is requesting), demographic data for 2011 – 2015 will be analyzed. For the ACS General Handbook PDF (which is a guide for how to use ACS data) as well as a full description of the ACS data including data collection methods, difference between single year and multi-year data, and limitations of the survey data such as sampling bias, visit the following link:

- https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf

Defining "Areas of Interest" for the Demographic Comparison

Census block groups were used to filter ACS data sets using their GEO-IDs. Census block groups were the most granular geographical unit available that would allow us to refine data to be specific to the Zone and Community District 10. In order to identify which block groups were in each area of interest, I created a custom Google Map using shapefiles provided by the U.S. Census Bureau, a map from HCZ for the Zone, and Google Fusion Tables. Combining information from these sources, I created the following **Google Overlap Map**:

- [https://www.google.com/maps/d/u/0/edit?mid=15OzmkGJoOG5if4GdXYDAyL9sGqGcMGq4&ll=40.8197907
9676808%2C-73.9380606213989&z=14](https://www.google.com/maps/d/u/0/edit?mid=15OzmkGJoOG5if4GdXYDAyL9sGqGcMGq4&ll=40.81979079676808%2C-73.9380606213989&z=14)

The screenshots below display the various layers that can be selected and deselected in the custom Google Overlap Map. I identified which census tracts and block groups fell into each area of interest for this demographic comparison.

Census Block Groups:
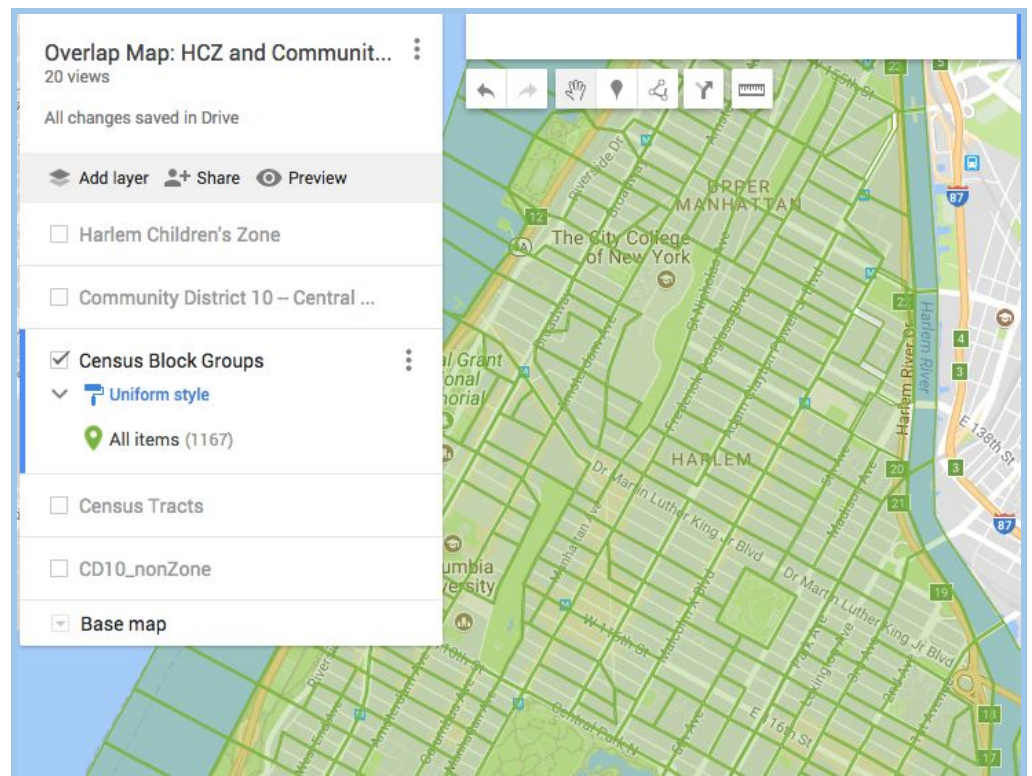
Census Tracts:
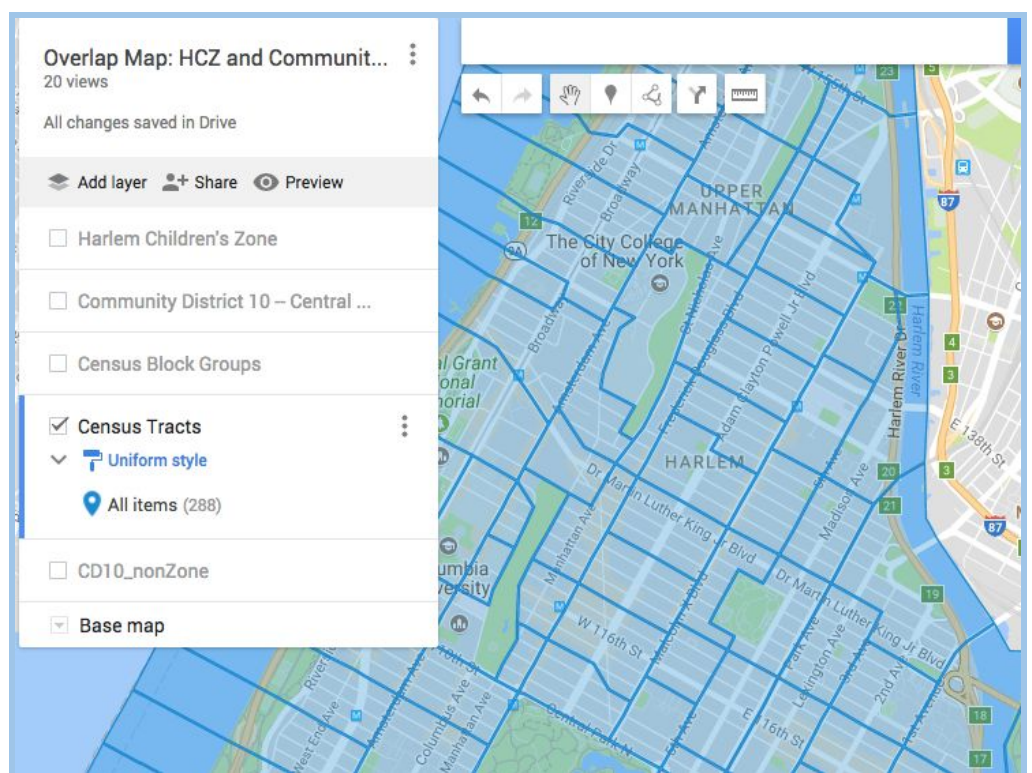
The Zone -- Defined by HCZ:

HCZ provided the following link to a map of the area of land identified as the Zone:

- https://hcz.org/zone-map/
- HCZ provided the following guidelines:
  - Please use the Zone map to identify which census blocks to focus in on.
  - Ignore the sites plotted on the map; focus on the shaded blocks.
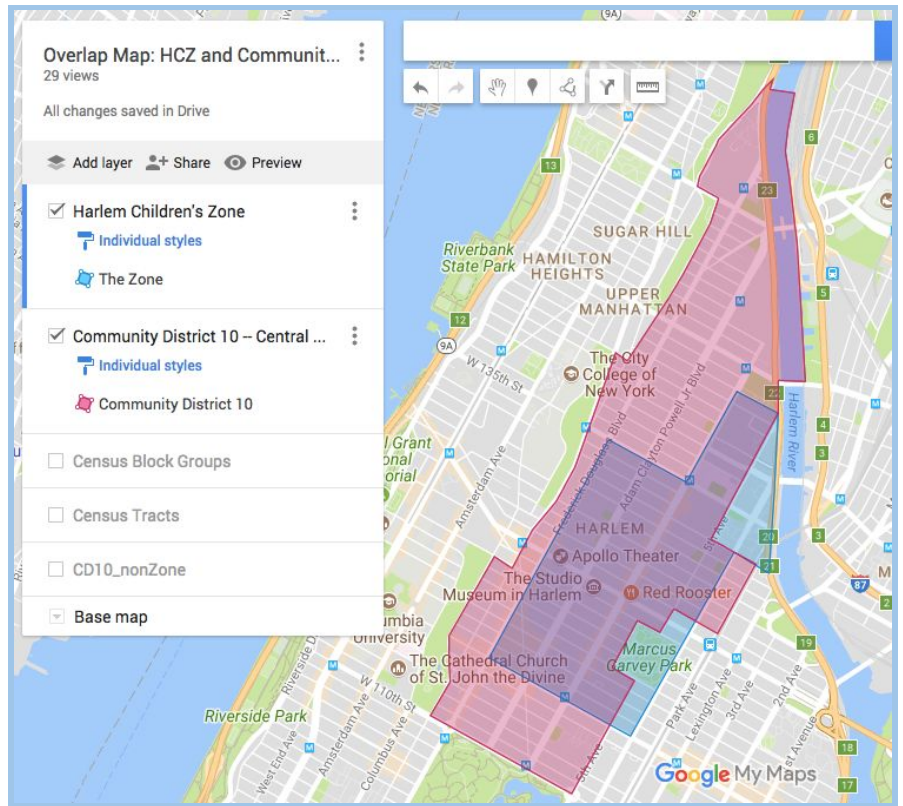 - The Zone does not include East Harlem.
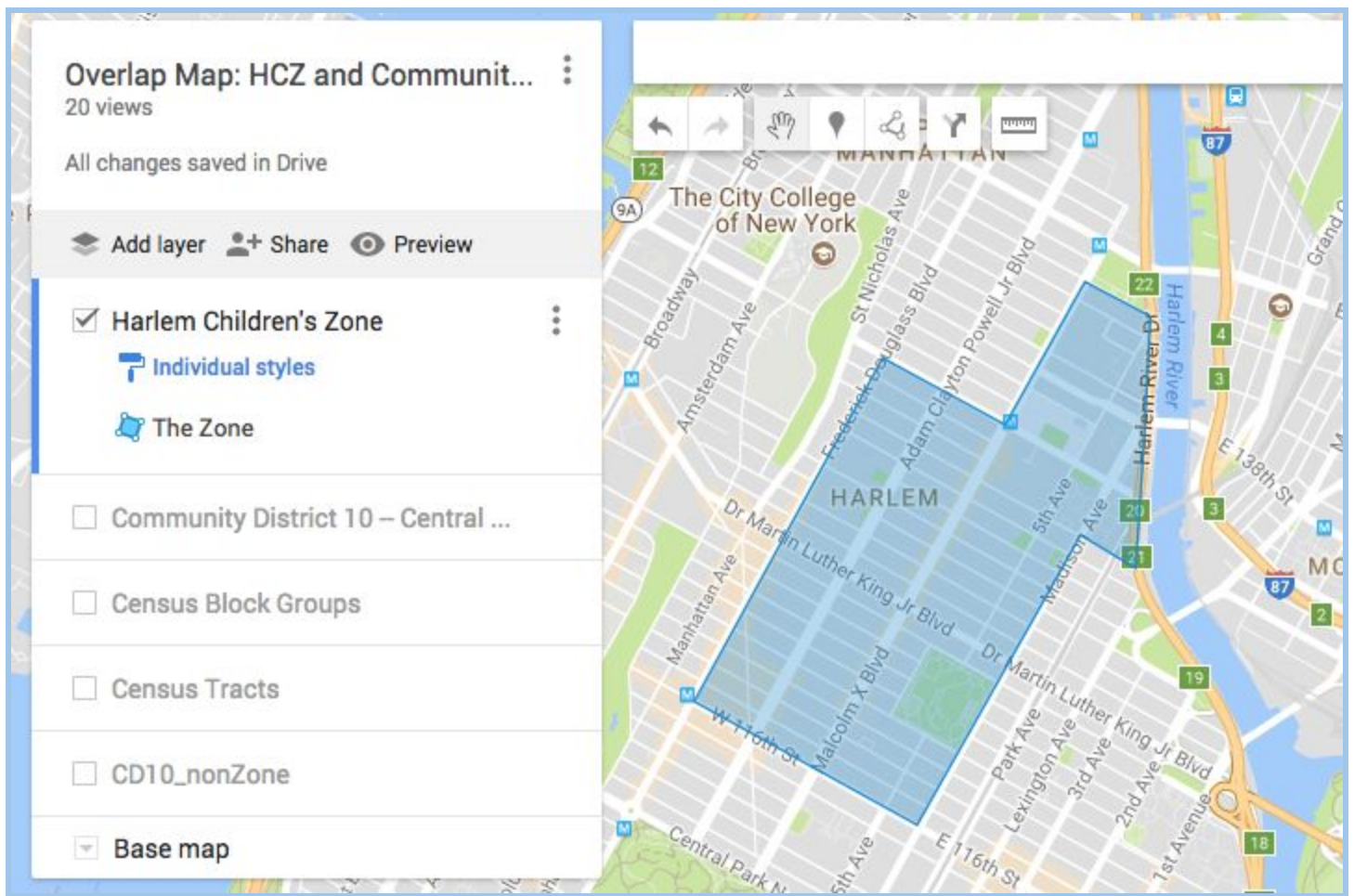
Community District 10 -- Central Harlem, PUMA, NY -- Defined by the U.S. Census Bureau:
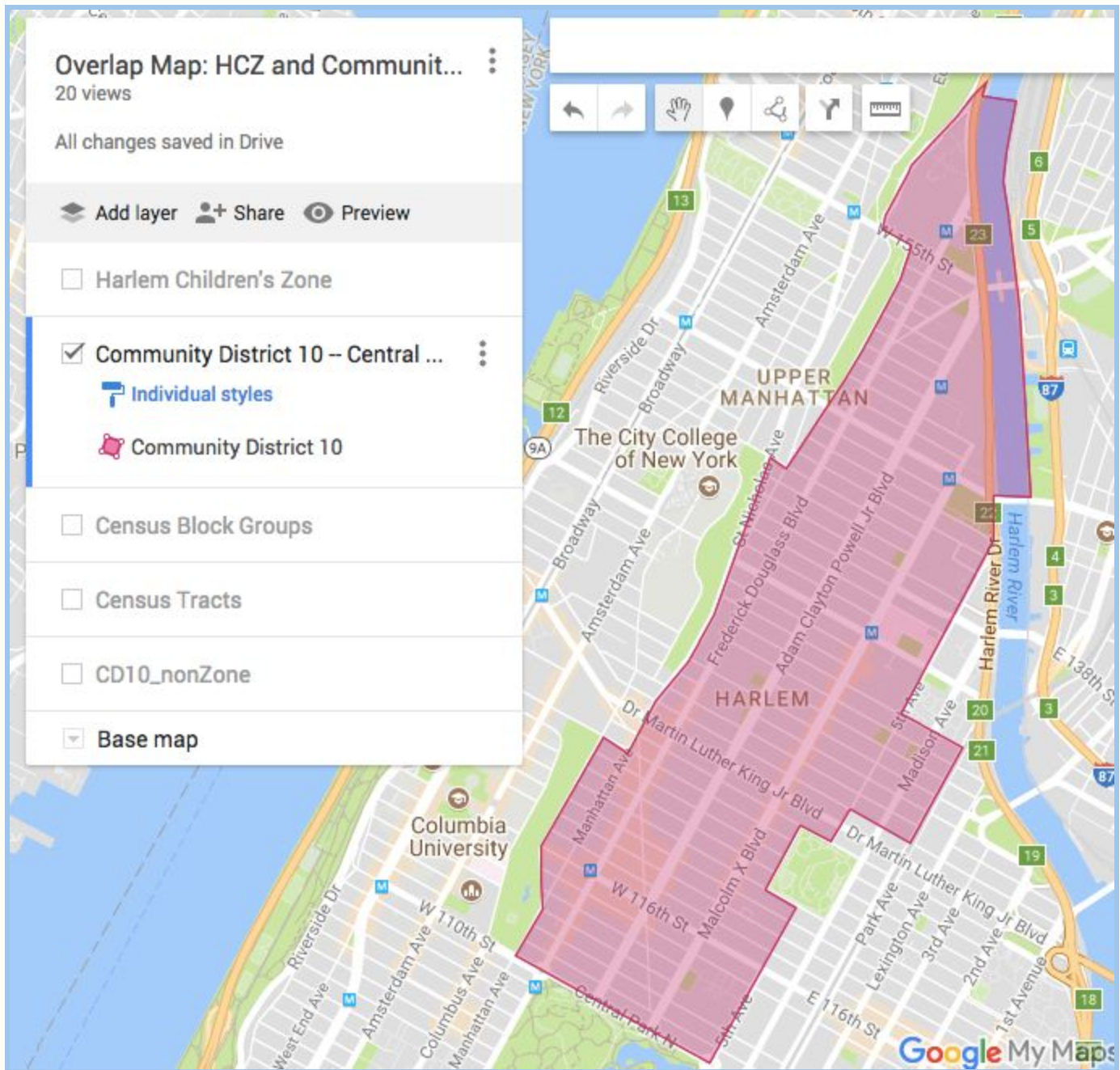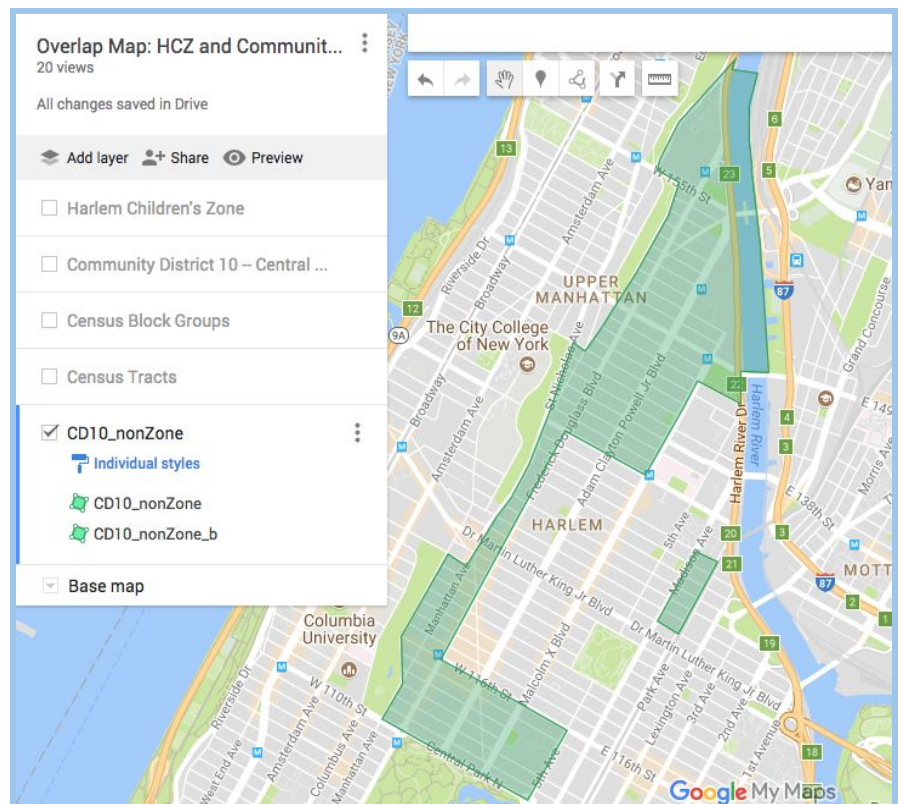
The screenshots below display the overlap and non-overlap areas between the Zone and Community District 10 (CD10).

Overlap between the Zone and CD10 (shown in purple, excluding the Harlem River portion):

Area that is in CD10, but not in the Zone:

Exclusion Criteria:

In an attempt to keep the different areas of interest as specific as possible, certain block groups were excluded. The block groups that were excluded fell into one of these categories:

- block group with portion in the Zone and portion in CD10_nonZone
- block group with portion in the Zone and portion in another Community District (that was not CD10--Central Harlem).

These are the excluded block groups (identified by their GEO-IDs):

- Part Zone/CD10 overlap, part CD10-only:
  ['360610190001','360610228003','360610228002','360610214001']

- Part Zone, part other community district:
  ['360610184002','360610184001','360610198001','360610242001']

SIDE NOTE: The Google Overlap Map can be further customized. Demographic data from the U.S. Census Bureau can be combined with the shapefiles data (map data) using GEO-IDs as the linking attributes. The combined data set can then be converted to a KML file using the Google Fusion Tables web tool. The KML file of the combined data set can then be uploaded into the custom Google Overlap Map.

Breakdown of Sample Data Set:

In the stacked bar chart below (*Figure 1*), we can see the breakdown of the proportions (in percentages) of the demographic data set provided by the U.S. Census Bureau (5-year estimates for 2011 - 2015) that fall into the following categories:
- 'Only-Zone' indicates the block groups that fall only in the Zone, and not in Community District 10.
- 'Overlap_ZCD10' indicates the block groups that fall in both the Zone and Community District 10.
- 'CD10_nonZone' indicates the block groups that fall only in Community District 10, and not in the Zone.
- 'Excluded' indicates the block groups that did not meet the criteria of being exclusively in the Zone or in CD10 non-zone areas.

The top bar displays the percentages out of the total population (denominator = sum of 5-yr population counts). The bottom bar displays the percentages out of the total number of rows (denominator = sum of number of rows in available dataset).
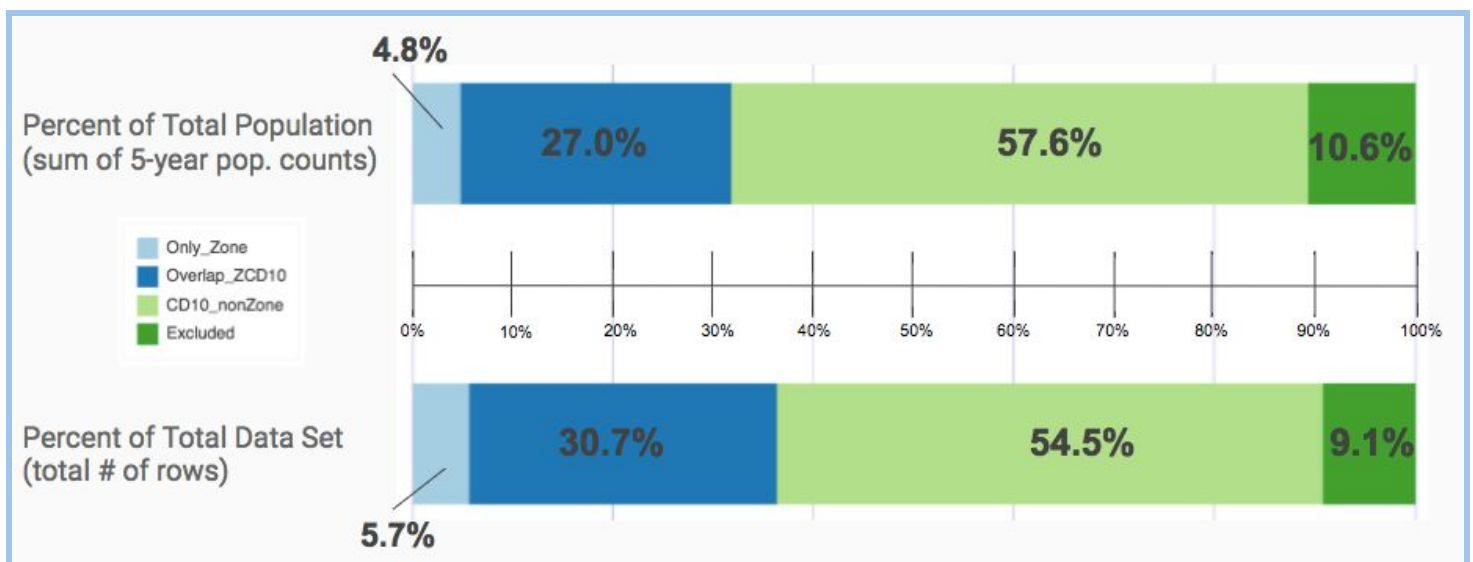

Figure 1: Stacked bar chart displaying the percent out of the total data set for each population of interest.

For this demographic comparison, the target population is the one within the Zone. The comparison populations are those within Community District 10 and Community District 10 Non-Zone. Considering the categories in *Figure 1*, the target and sample (or comparison) populations are defined as follows:

- The Zone = Only_Zone + Overlap_ZCD10
- CD10_all = CD10non_Zone + Overlap_ZCD10

In the stacked bar chart below (*Figure 2*), we can see the breakdown of the Community District 10 population (CD10_all). The denominator is equal to the sum of 5-yr population counts for CD10_all. Please Note: Only in Figure 2, the population for 'Only_Zone' (4.8%) was excluded because the purpose of this visual is to show how much of CD10_all is made up of the Zone versus Non-Zone population.
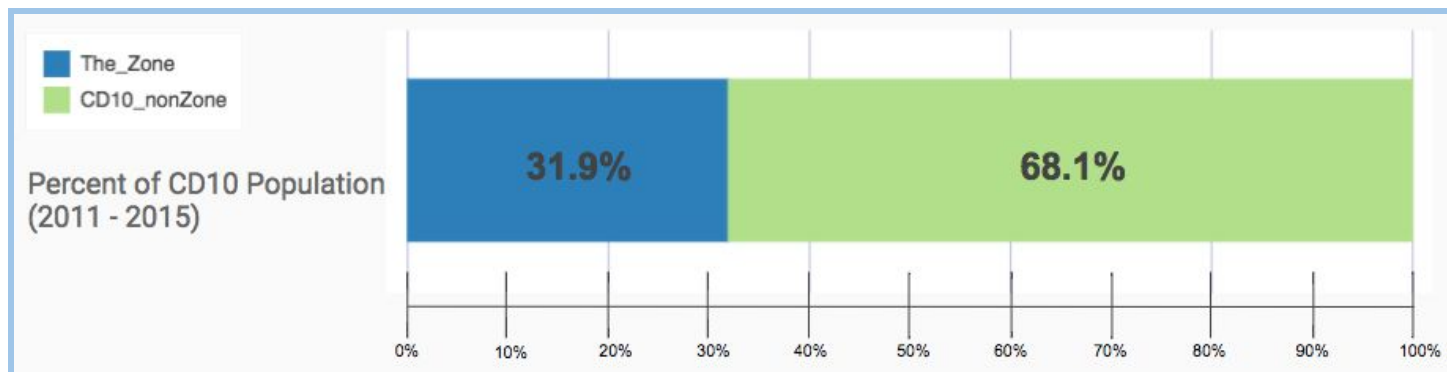

Figure 2: Stacked bar chart displaying the composition of CD10_all population, regarding Zone vs. Non-Zone populations.

For the following tables, population counts are the sum of the 5-yr estimates per area of interest.

Area of Interest	Total Count for Sample Population	Percent of Total Population	Percent of Data Set
Only_Zone	7084	4.815411	5.681818
Overlap_ZCD10	39732	27.008177	30.681818
CD10_nonZone	84717	57.587128	54.545455
Excluded	15578	10.589283	9.090909

Table 2: Sum of 5-yr population counts for area of interests and their percentage of the total population and percentage of the total data set that is available.

Area of Interest	Total Count for Population	Percent of CD10_all Population
The_Zone	46816	31.926331
CD10_nonZone	84717	68.073669
CD10_All	124449	100

Table 3: Sum of 5-yr population counts for target and sample populations and their percentage of CD10_all population.

Process

For this demographic comparison, we will be reviewing the similarities and differences in population distributions between the three populations of interest (i.e. The Zone, CD10 Non-Zone, and CD10 All) for several demographic attributes. The population distribution for all of NYC Manhattan will also be reviewed to gain a better perspective of the similarities and differences for some of the demographic attributes. In order to compare and analyze the population distributions, the following graphical tools were used when possible: a population pyramid, stacked bar charts (out of 100%), and Q-Q plots. In this analysis, we will be comparing population distributions for the following demographic attributes: Gender, Age Groups, Race, Hispanic or Latino Origin, Household Type, Family Type and Household Income.

Brief Overview of Q-Q Plots:

A Q-Q plot is a scatterplot created by plotting two sets of quantiles against one another. If both sets of quantiles came from the same distribution, we should see the points forming a line that's roughly straight. The main use for a q-q plot is to check the assumption that the data follows a theoretical distribution, such as a normal distribution (see *Figure 3*). The q-q plot is used to answer the following questions:
- Do two data sets come from populations with a common distribution?
- Do two data sets have common location and scale?
- Do two data sets have similar distributional shapes?
- Do two data sets have similar tail behavior?

Please Note: Q-Q Plots are just visual checks, so results are not entirely objective. More importantly, the sample size should be taken into account when judging how close the q-q plot is to the straight line. In *Figure 3*, observe that the q-q plot is almost identical to the line y = x when the sample size is equal to 1000. However, that is not the case when the sample size is only equal to 5. The data sets for this demographic comparison have extremely small sample sizes due to the scarcity of data points for our specific areas of interest. That being said, q-q plots do allow us to see 'at-a-glance' if our assumption of similar population distributions is plausible, and if not, how the assumption is violated and what data points contribute to the violation.
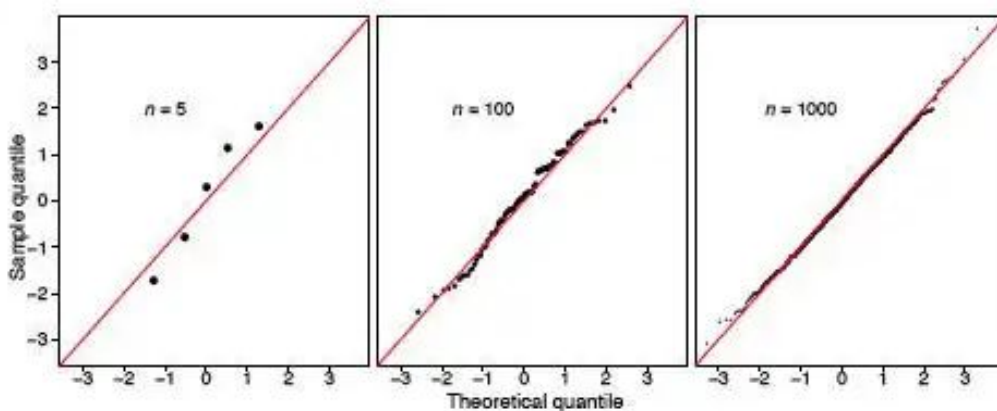

Figure 3: Q-Q plots of normal data for three different sample sizes.

Sources for information over Q-Q Plots:
- http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm
- http://data.library.virginia.edu/understanding-q-q-plots/
- http://onlinestatbook.com/2/advanced_graphs/q-q_plots.html
- https://www.youtube.com/watch?v=okjYjClSjOg

Findings

Table 4 summarizes the findings from the demographic comparison by categorizing the demographic attributes for the population in the Zone as either extremely similar, similar, or different than the population in Community District 10.

Overall, there were more similarities than differences between the target population in the Zone and the sample populations in Community District 10. The main differences in the demographic makeup of the populations was the distribution of the male population by age groups and the distribution of male-household families by age of children (see *Figures 5, 7, and 15*).

Table 4: Summary of Analysis Findings

Target Population	Distributions Extremely Similar to CD10	Distributions Similar to CD10	Distributions Different than CD10
The Zone	Gender Race	Female Age Groups Population By Household Income Population By Hispanic Ancestry Population By Household Type Population By Family Household Type Female-Household Families By Age of Children Married-Couple Families By Age of Children	Male Age Groups Male-Household Families By Age of Children

Recommendations

To recapitulate:

- The population in the Zone is more similar to Community District 10 than it is different.
- There are more robust data sets available for Community District 10 than for the Zone.
- There are several existing data resources for Community District 10 that HCZ can leverage for its own purposes.

Considering the findings listed above, I highly recommend that HCZ leverages the available data resources for Community District 10 to answer the question: "How many babies are born each year who reside in the Zone?"

Future project ideas or next steps the organization should consider:

- Improve the custom Google Overlap Map by including demographic data.
  - The Google Overlap Map can be further customized. Demographic data from the U.S. Census Bureau can be combined with the shapefiles data (map data) using GEO-IDs as the linking attributes. The combined data set can then be converted to a KML file using the Google Fusion Tables web tool. The KML file of the combined data set can then be uploaded into the custom Google Overlap Map.

- Use the EpiQuery Natality Module provided and maintained by the NYC Department of Health and Mental Hygiene, Bureau of Vital Statistics.
  - Step 1 -- Go to the following link: https://a816-healthpsi.nyc.gov/epiquery/Birth/index.html
  - Step 2 -- Choose year of interest and click 'SUBMIT'
  - Step 3 -- Choose 'Community District of Residence' on right-side menu and click 'SUBMIT'
  - Step 4 -- Download dataset and filter for Community District == Central Harlem (110)

- Explore available data resources for Community District 10:
  - https://statisticalatlas.com/neighborhood/New-York/New-York/Harlem/Overview
  - https://censusreporter.org/profiles/79500US3603803-nycmanhattan-community-district-10central-harlem-puma-ny/

Conclusion

In conclusion, the sample populations within the Zone and Community District 10 are similar enough to validate the use of Community District 10 data resources to the answer the question "How many babies are born each year who reside in the Zone?"

Recommended Next Steps:

- Use the EpiQuery Natality Module provided and maintained by the NYC Department of Health and Mental Hygiene, Bureau of Vital Statistics.
    - Step 1 -- Go to the following link: https://a816-healthpsi.nyc.gov/epiquery/Birth/index.html
    - Step 2 -- Choose year of interest and click 'SUBMIT'
    - Step 3 -- Choose 'Community District of Residence' on right-side menu and click 'SUBMIT'
    - Step 4 -- Download dataset and filter for Community District == Central Harlem (110)

- Explore available data resources for Community District 10:
    - https://statisticalatlas.com/neighborhood/New-York/New-York/Harlem/Overview
    - https://censusreporter.org/profiles/79500US3603803-nycmanhattan-community-district-10central-harlem-puma-ny/

If HCZ decides that the population in Community District 10 is not an appropriate 'proxy' for the target population in the Zone, then the following next steps are recommended for building a custom model to approximate the number of children born each year in the Zone for the past 5 years:

- Conduct a thorough exploratory data analysis (EDA) over the 5-year birth rates data available for census tracts.
    - Missing data will need to be imputed.

- Train the model using the 2006 - 2010 data, and test the model using 2010 - 2014 data.

- Reach out to the NYC Department of Health and Mental Hygiene, Bureau of Vital Statistics to see if more robust data sets are available upon request.
    - Contact: vsdata@health.nyc.gov

Findings -- Population by Gender

In *Figure 4*, we can see the breakdown (in percentages) of the sample populations by gender. The denominator used to obtain percent totals was the sum of population counts (5-yr estimates) for each area of interest.

- ● Observation: Extremely similar gender distribution among the three populations of interest.

Table 5: Population by Gender, Population Count Totals for 2011 - 2015

	The_Zone	CD10_nonZone	CD10_all
Male	21678	39280	57902
Female	25138	45437	66547

Table 6: Population by Gender, Population Percent Totals for 2011 - 2015

	The Zone	CD10 nonZone	CD10 all
Male	46.304682	46.366137	46.52669
Female	53.695318	53.633863	53.47331

Figure 4: Population by Gender, Stacked Bar Graph of Percent Totals for 2011 - 2015

Population By Gender, Percent Totals for 2011 - 2015
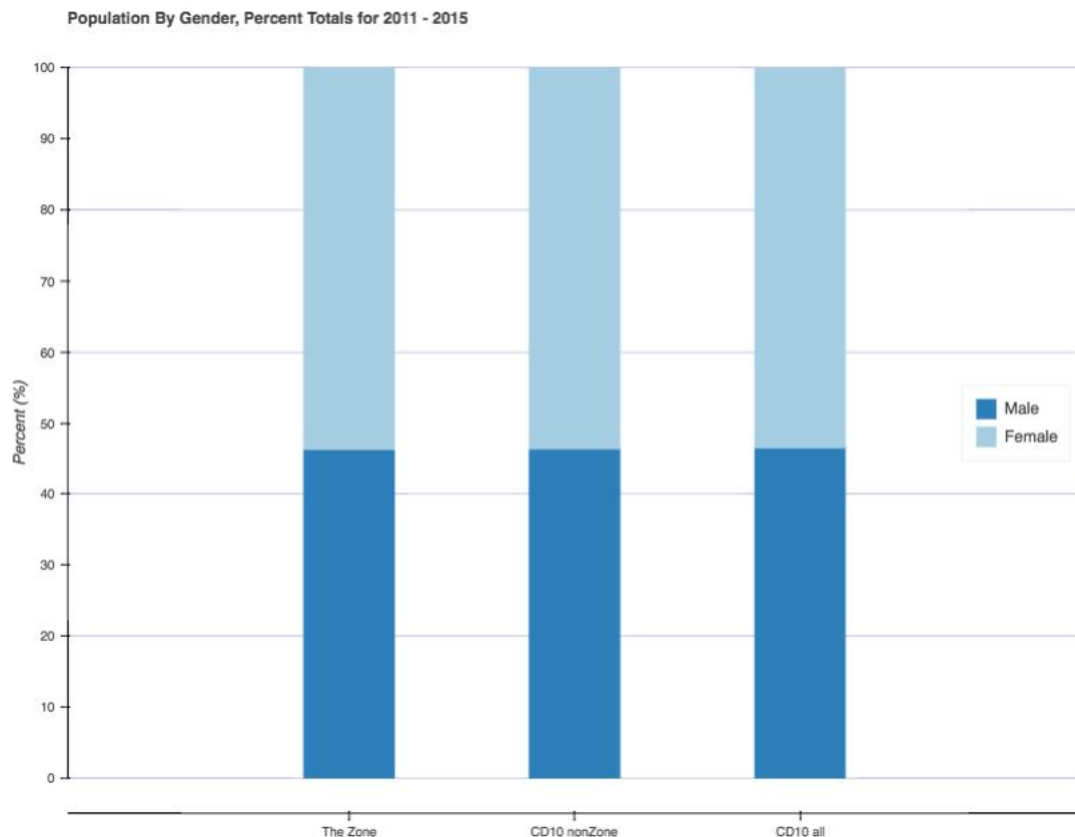
Findings -- Population by Gender AND Age Groups

In the population pyramid (*Figure 5*), we can see the breakdown (in counts) of the sample populations by gender and by age groups. The male population distribution by age group is displayed on the left side, and the female distribution on the right side. The darkest hues (blue and red) indicate data for the Zone. Medium hues (green and orange) indicate data for CD10 Non-Zone. The lightest hues (light green and light orange) indicate data for all of Community District 10. The goal is to compare the age-group population trend (2011 - 2015) for females and males between the Zone, CD10 Non-Zone, and CD10 all.

- Y-axis = Age Groups
  - Pre-Reproductive: 0 - 14 years
  - Reproductive: 15 - 44 years
  - Post-Reproductive: 45+ years
- X-axis = 5-year Aggregate Counts for 2011 - 2015. **Please Note:** All numbers are positive, regardless of negative sign or position on axis. Counts for male population were converted to negative numbers for the purpose of the population pyramid visual.

The structure of the pyramid indicates an overall stable (or stationary) birth rate for the population within the Zone, but a somewhat ageing population for the population within Community District 10. The narrower base for the CD10 population indicates a declining birthrate. However, the base seems to be broadening again, based on the two youngest age groups. CD10 yearly birth totals for 2010 - 2015 reflect this observation. The structure also indicates a higher dependency ratio. In other words, as the working population ages and retires, there will be fewer working age people (who are currently the youthful dependents/pre-reproductive population) to support them.

- Observation: Overall, the age-group distributions are similar among the three areas of interest. There are some age groups where the Zone has a noticeably different distribution than the other areas, especially for the male population. For example, look at age groups from 35 to 59 years for the male population and age groups from 5 to 24 years for the female population.

Figure 5: Population Pyramid for 2011 - 2015, Color Coded by Area of Interest
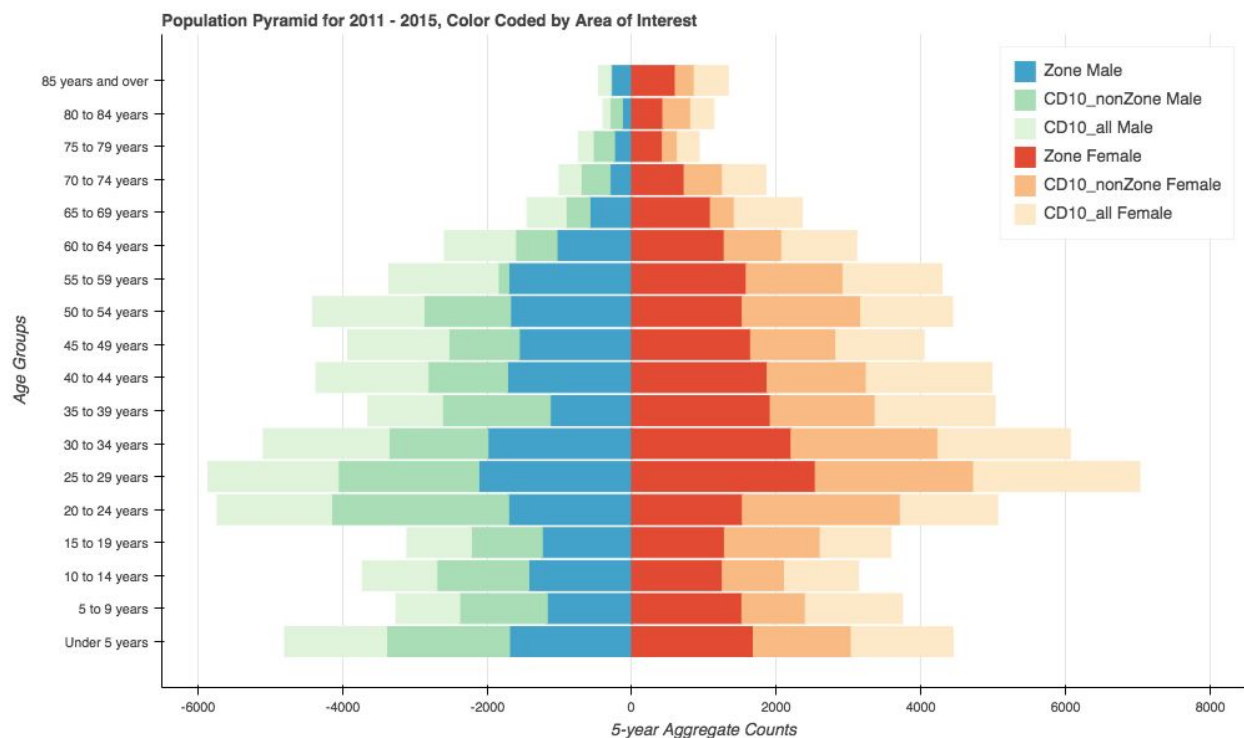
Please Note: All numbers are positive, regardless of negative sign or position on axis. Counts for male population were converted to negative numbers for the purpose of this visual.

Table 7: Female Population by Age Groups, Population Count Totals for 2011 - 2015

Age_Grps	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Under 5 years	1684	3037	4443	40419
5 to 9 years	1525	2402	3742	29992
10 to 14 years	1255	2115	3135	29511
15 to 19 years	1284	2608	3587	38205
20 to 24 years	1533	3716	5059	68280
25 to 29 years	2540	4729	7024	107630
30 to 34 years	2206	4235	6064	86770
35 to 39 years	1917	3368	5024	63333
40 to 44 years	1875	3245	4980	56463
45 to 49 years	1645	2825	4041	52068
50 to 54 years	1530	3167	4437	51047
55 to 59 years	1586	2923	4288	48439
60 to 64 years	1280	2076	3109	50951
65 to 69 years	1089	1419	2360	41095
70 to 74 years	727	1258	1854	31640
75 to 79 years	426	632	931	23097
80 to 84 years	432	816	1136	18419
85 years and over	604	866	1333	22714

Table 8: Female Population by Age Groups, Population Percent Totals for 2011 - 2015

Age_Grps	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Under 5 years	3.597061	3.584877	3.570137	2.480443
5 to 9 years	3.257433	2.835322	3.006854	1.840557
10 to 14 years	2.680707	2.496547	2.519104	1.811039
15 to 19 years	2.742652	3.078485	2.882305	2.344574
20 to 24 years	3.274522	4.386369	4.065119	4.190224
25 to 29 years	5.425496	5.582115	5.644079	6.605065
30 to 34 years	4.712064	4.998997	4.872679	5.324923
35 to 39 years	4.094754	3.975589	4.036995	3.886636
40 to 44 years	4.005041	3.8304	4.001639	3.465036
45 to 49 years	3.513756	3.334632	3.247113	3.195322
50 to 54 years	3.268113	3.738329	3.565316	3.132665
55 to 59 years	3.387731	3.450311	3.445588	2.972617
60 to 64 years	2.734108	2.450512	2.498212	3.126774
65 to 69 years	2.326128	1.674988	1.896359	2.521928
70 to 74 years	1.552888	1.484944	1.489767	1.941692
75 to 79 years	0.909945	0.746013	0.748098	1.417423
80 to 84 years	0.922761	0.963207	0.912824	1.130342
85 years and over	1.290157	1.022227	1.071122	1.393919

In *Figure 6*, q-q plots are displayed for five sample population comparisons (for female population by age group distribution): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for female age groups with Community District 10 as a whole. Comparing Figures 6-A and 6-B with 6-D and 6-E confirms that the Zone has a more similar distribution for female age groups with CD10 than with NYC Manhattan.

- Observation: Similar female age-group distribution between the Zone and Community District 10.

Figure 6: Q-Q Plot of Female Population by Age Groups for 2011 - 2015

Figure 6-A: Zone vs. CD10 Non-Zone
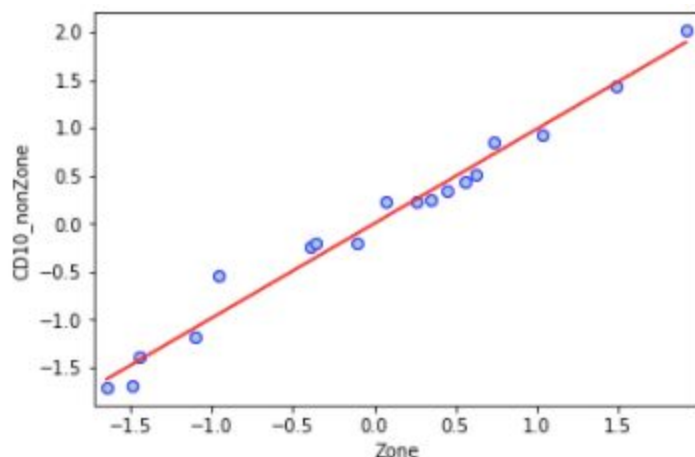
Figure 6-B: Zone vs. CD10 All
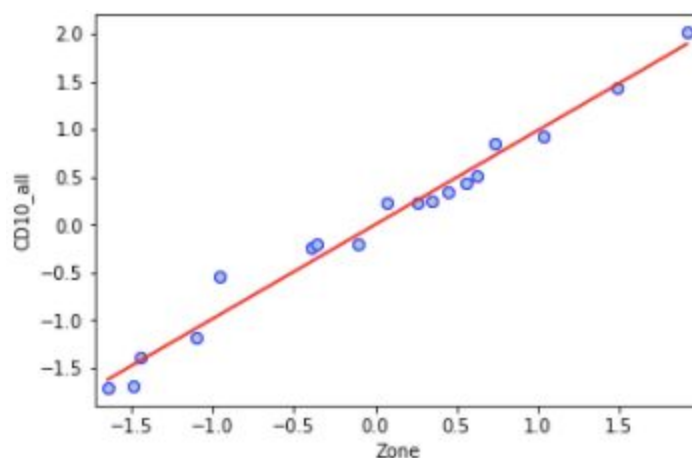
Figure 6-C: CD10_all vs. CD10 Non-Zone
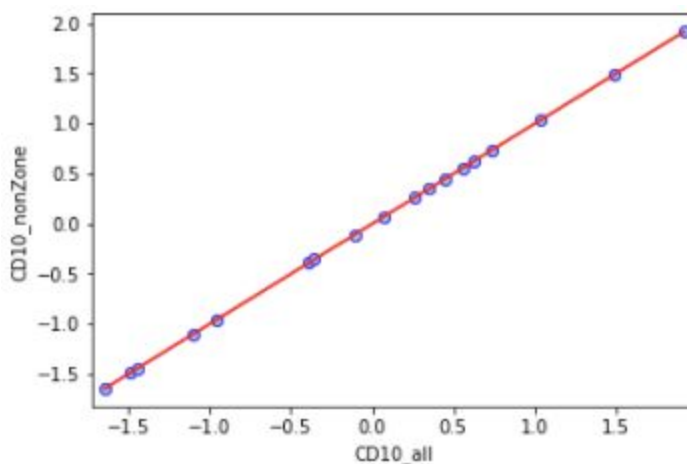
Figure 6-D: Zone vs. NY_Manhattan
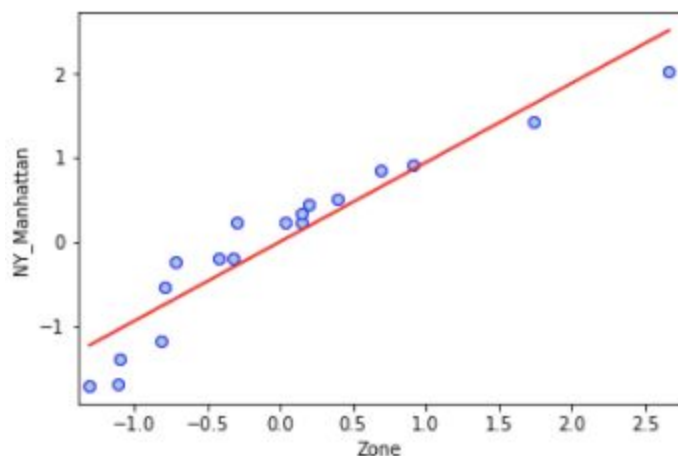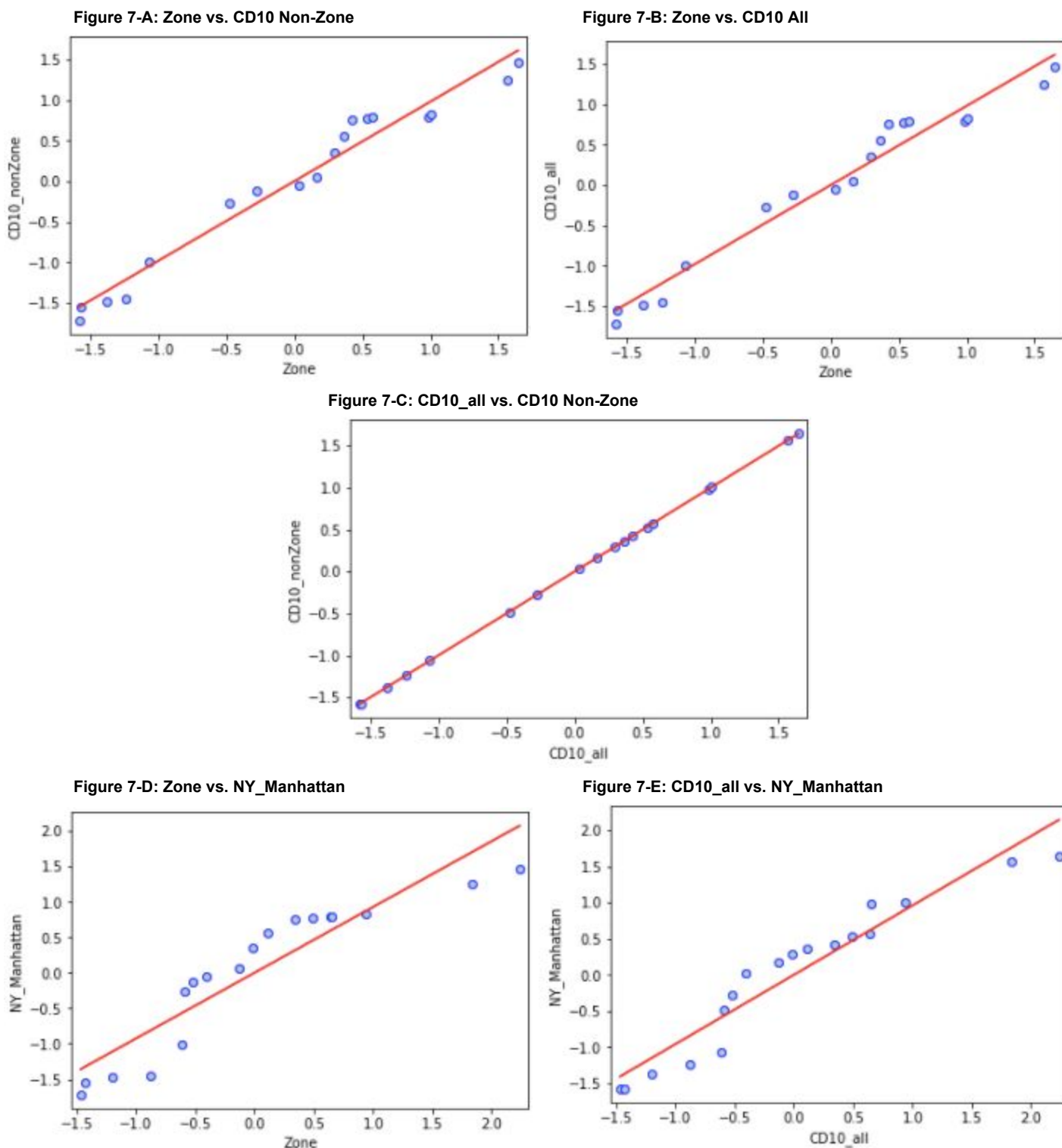
Figure 6-E: CD10_all vs. NY_Manhattan
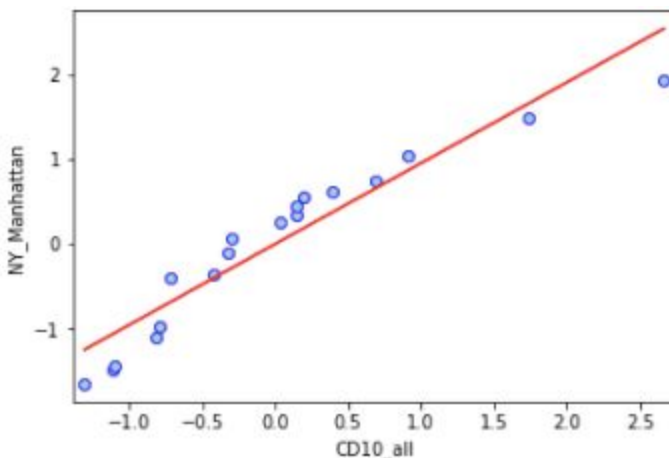
Table 9: Male Population by Age Groups, Population Count Totals for 2011 - 2015

Age_Grps	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Under 5 years	1691	3390	4799	42479
5 to 9 years	1170	2380	3258	31571
10 to 14 years	1425	2695	3725	29481
15 to 19 years	1239	2219	3106	34018
20 to 24 years	1704	4151	5732	56515
25 to 29 years	2116	4059	5858	91197
30 to 34 years	1989	3357	5093	82434
35 to 39 years	1127	2619	3647	63022
40 to 44 years	1718	2817	4364	56980
45 to 49 years	1556	2530	3925	53293
50 to 54 years	1677	2873	4412	50155
55 to 59 years	1701	1847	3358	45118
60 to 64 years	1035	1607	2589	39981
65 to 69 years	579	908	1446	30087
70 to 74 years	300	702	1002	23662
75 to 79 years	240	534	737	16759
80 to 84 years	131	300	395	11664
85 years and over	280	292	456	11018

Table 10: Male Population by Age Groups, Population Percent Totals for 2011 - 2015

Age_Grps	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Under 5 years	3.612013	4.001558	3.856198	2.606862
5 to 9 years	2.499146	2.809353	2.61794	1.937457
10 to 14 years	3.043831	3.18118	2.993194	1.809198
15 to 19 years	2.646531	2.619309	2.495801	2.087625
20 to 24 years	3.639781	4.899843	4.605903	3.468227
25 to 29 years	4.519822	4.791246	4.707149	5.596601
30 to 34 years	4.248548	3.962605	4.092439	5.058831
35 to 39 years	2.407297	3.091469	2.930518	3.86755
40 to 44 years	3.669686	3.325189	3.506657	3.496763
45 to 49 years	3.32365	2.986414	3.153902	3.270498
50 to 54 years	3.582109	3.391291	3.545227	3.077925
55 to 59 years	3.633373	2.1802	2.698294	2.768813
60 to 64 years	2.210783	1.896904	2.08037	2.453564
65 to 69 years	1.236757	1.071804	1.161922	1.846387
70 to 74 years	0.640807	0.828641	0.805149	1.452096
75 to 79 years	0.512645	0.630334	0.59221	1.028471
80 to 84 years	0.279819	0.35412	0.317399	0.715799
85 years and over	0.598086	0.344677	0.366415	0.676155

In *Figure 7*, q-q plots are displayed for five sample population comparisons (for male, age-group distribution): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for male age groups with Community District 10 as a whole. Comparing Figures 7-A and 7-B with 7-D and 7-E confirms that the Zone has a more similar distribution for male age groups with CD10 than with NYC Manhattan. However, Figures 7-A and 7-B seem to indicate that population distributions for male age groups are slightly different between the Zone and Community District 10.

● Observation: Different male age-group distribution between the Zone and Community District 10.

Figure 7: Q-Q Plot of Male Population by Age Groups for 2011 - 2015

Figure 7-A: Zone vs. CD10 Non-Zone

Figure 7-B: Zone vs. CD10 All

Figure 7-C: CD10_all vs. CD10 Non-Zone

Figure 7-D: Zone vs. NY_Manhattan

Figure 7-E: CD10_all vs. NY_Manhattan

Findings -- Population by Race

In *Figure 8*, we can see the breakdown (in percentages) of the sample populations by race. The denominator used to obtain percent totals was the sum of population counts (5-yr estimates) for each area of interest.

- Observation: Extremely similar race distribution among the three populations of interest.

**Table 11: Population by Race, Population Count Totals for 2011 - 2015**

Race	Zone	CD10_nonZone	CD10_all	NY_Manhattan
White	7923	15397	22754	919474
Black or African American	30016	52958	77963	244770
American Indian and Alaska Native	169	350	519	6155
Asian	1600	2246	3741	189892
Native Hawaiian and Other Pacific Islander	27	96	123	971
Some other race	4934	9555	13702	201754
Two or more races	2147	4115	5647	66491

Table 12: Population by Race, Population Percent Totals for 2011 - 2015

Race	Zone	CD10_nonZone	CD10_all	NY_Manhattan
White	16.923701	18.174628	18.283795	56.426514
Black or African American	64.114833	62.511656	62.646546	15.021108
American Indian and Alaska Native	0.360988	0.41314	0.417038	0.377722
Asian	3.417635	2.65118	3.006051	11.653341
Native Hawaiian and Other Pacific Islander	0.057673	0.113318	0.098836	0.059589
Some other race	10.539132	11.278728	11.010133	12.381291
Two or more races	4.586039	4.857349	4.537602	4.080437

Figure 8: Population by Race, Stacked Bar Graph of Percent Totals for 2011 - 2015
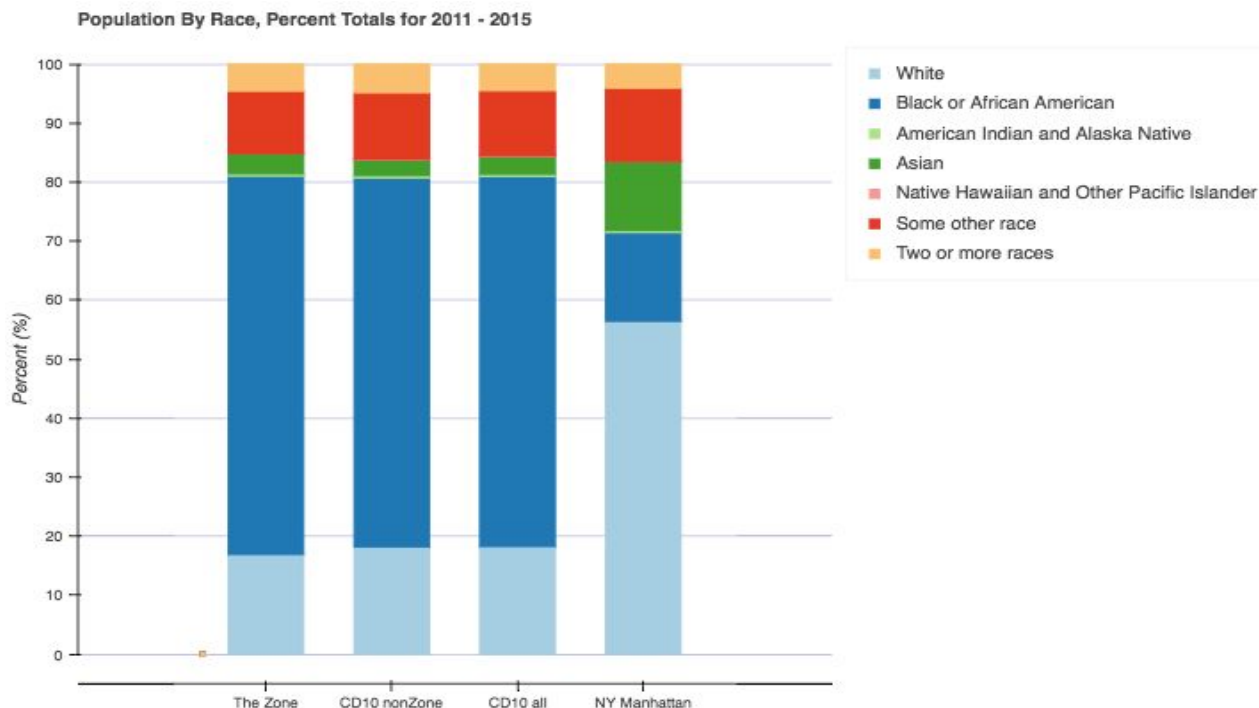
In *Figure 9*, q-q plots are displayed for five sample population comparisons (for race distribution): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for race with Community District 10 as a whole. Comparing Figures 9-A and 9-B with 9-D and 9-E confirms that the Zone has a similar distribution for race with CD10 than with NYC Manhattan. Comparing Figures 9-A and 9-B with Figure 9-C confirms that population distributions for race are extremely similar between the Zone and Community District 10.

- Observation: Extremely similar distribution of race between the Zone and Community District 10.

Figure 9: Q-Q Plot of Race for 2011 - 2015

Figure 9-A: Zone vs. CD10 Non-Zone

Figure 9-B: Zone vs. CD10 All

Figure 9-C: CD10_all vs. CD10 Non-Zone

Figure 9-D: Zone vs. NY_Manhattan

Figure 9-E: CD10_all vs. NY_Manhattan
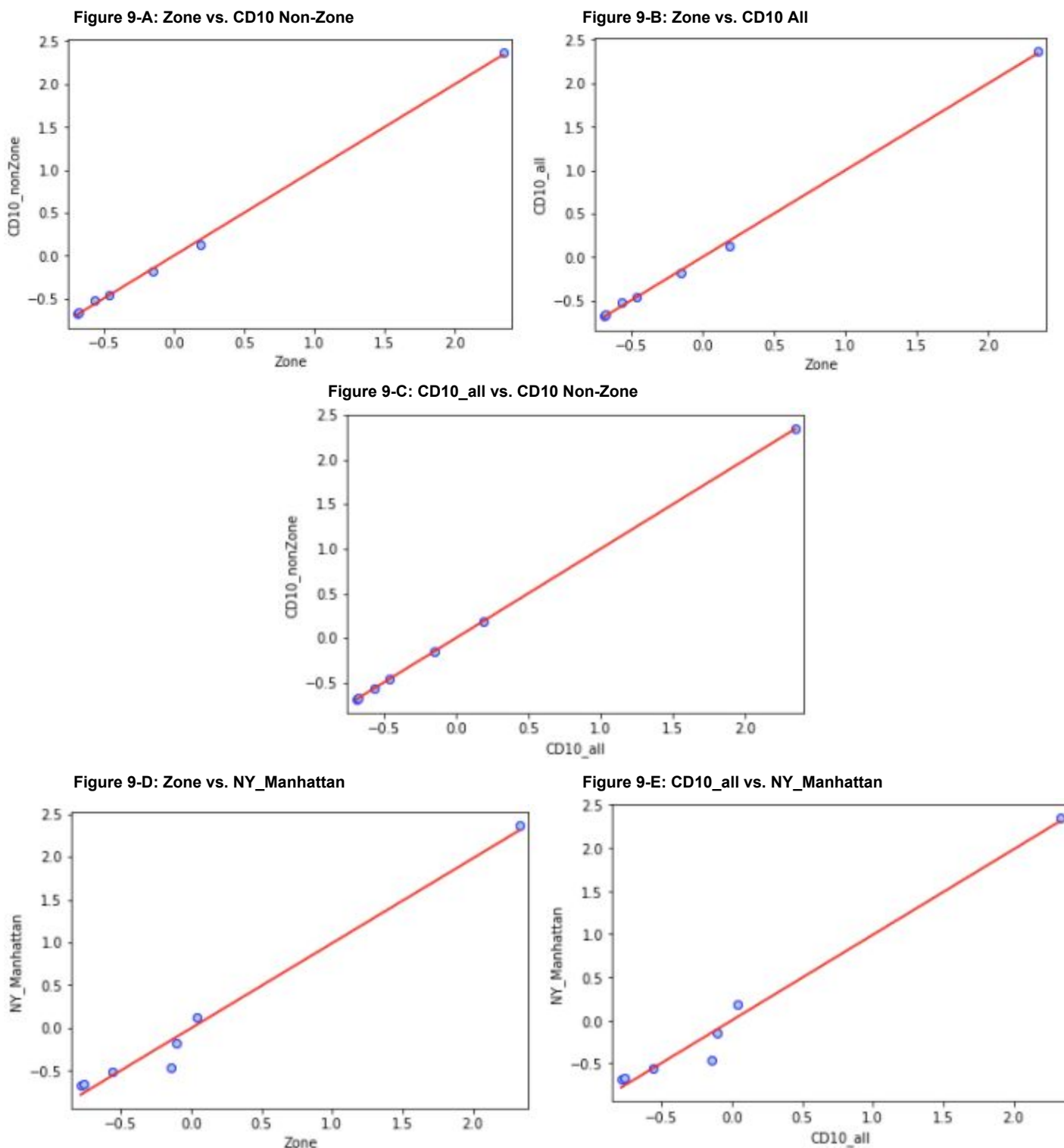
Findings -- Population by Hispanic or Latino Origin

In *Figure 10*, we can see the breakdown (in percentages) of the sample populations by hispanic or latino origin. The denominator used to obtain percent totals was the sum of population counts (5-yr estimates) for each area of interest.

- Observation: Similar hispanic or latino origin distribution among the three populations of interest.

Table 13: Population by Hispanic or Latino Origin, Population Count Totals for 2011 - 2015

	The_Zone	CD10_nonZone	CD10_all
Not Hispanic or Latino	36981	63378	95220
Hispanic or Latino	9835	21339	29229

Table 14: Population by Hispanic or Latino Origin, Population Percent Totals for 2011 - 2015

	The Zone	CD10 nonZone	CD10 all
Not Hispanic or Latino	78.992225	74.811431	76.51327
Hispanic or Latino	21.007775	25.188569	23.48673

Figure 10: Population by Hispanic or Latino Origin, Stacked Bar Graph of Percent Totals for 2011 - 2015
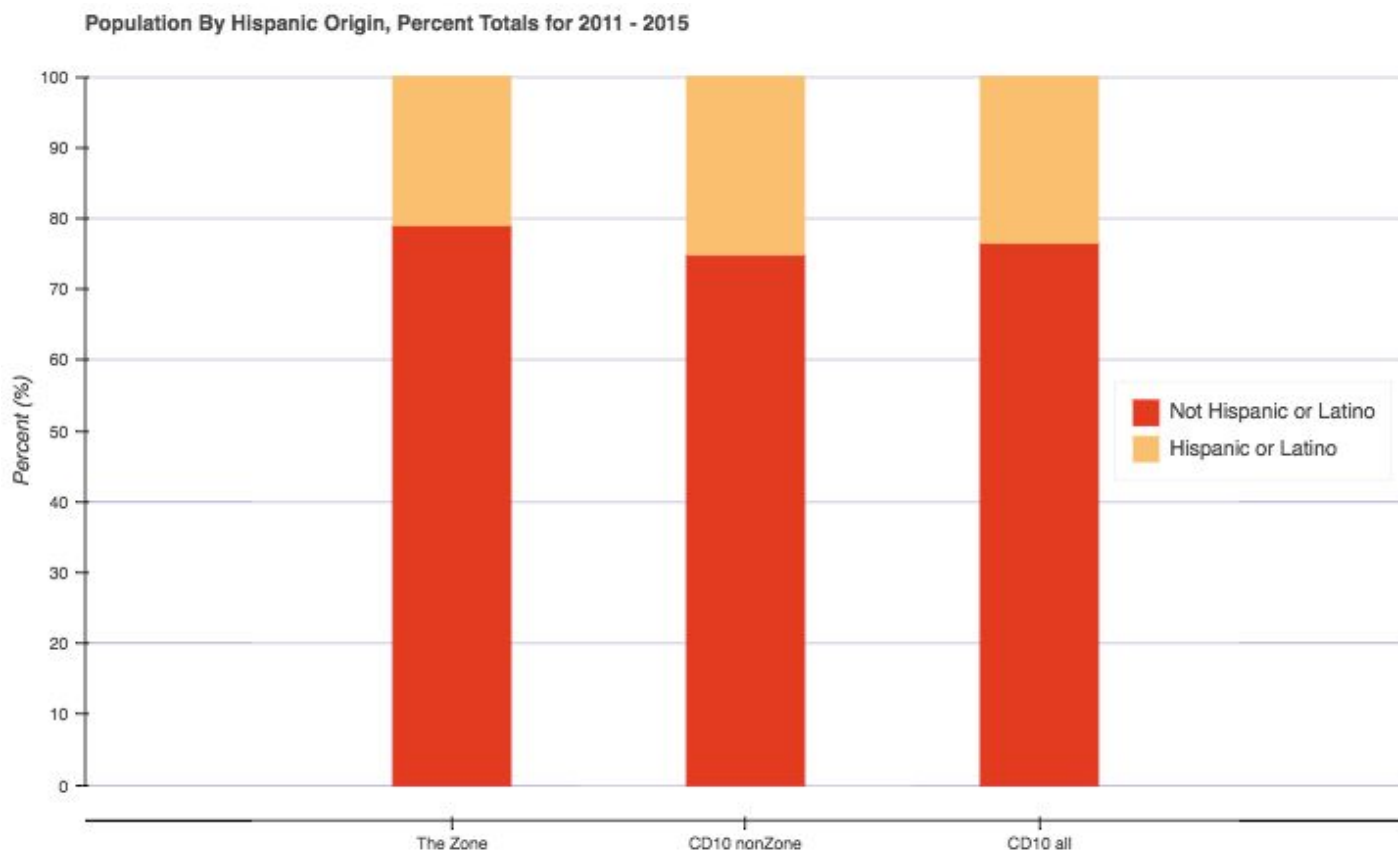
Findings -- Population by Household Type
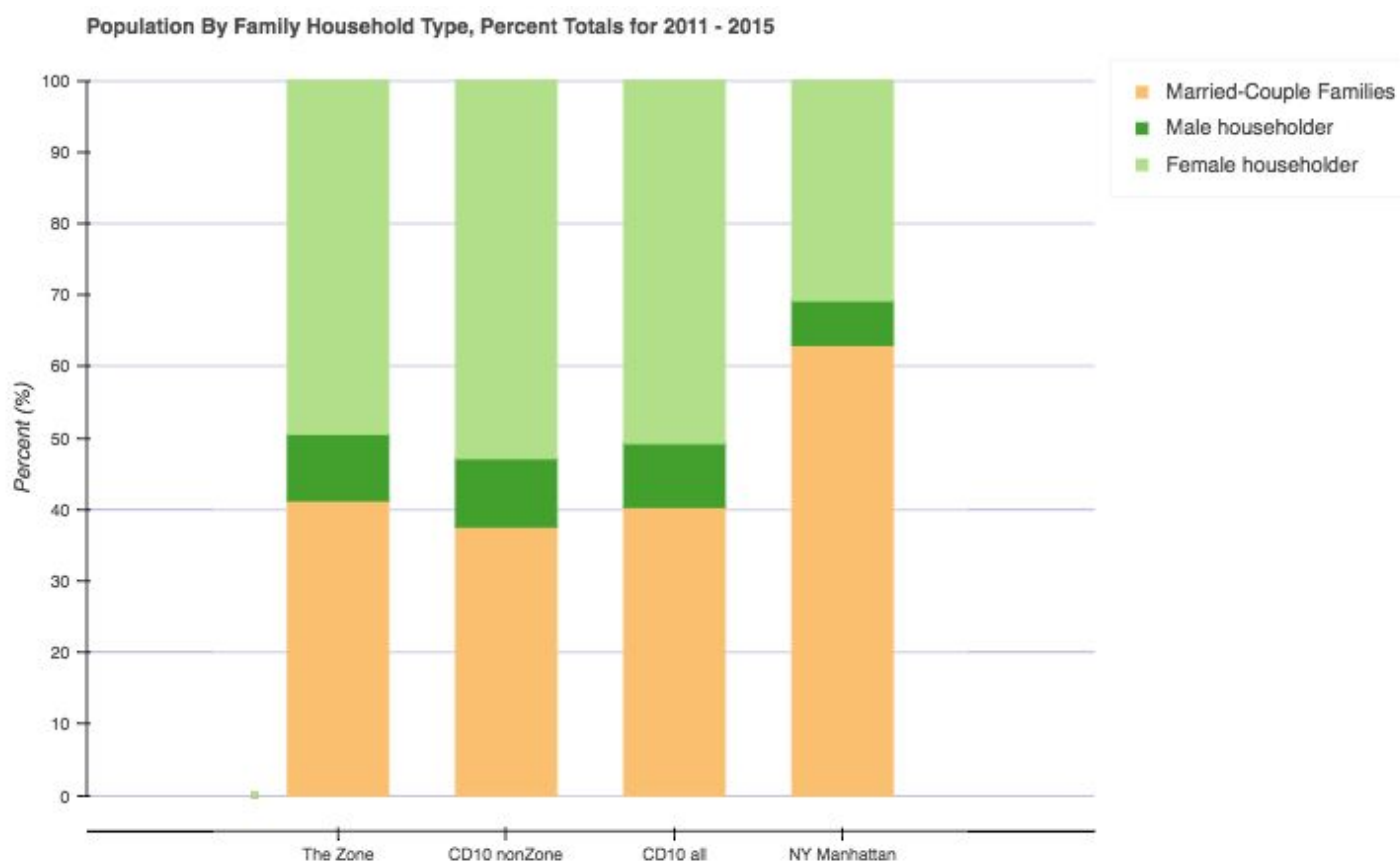
In *Figure 11*, we can see the breakdown (in percentages) of the sample populations by household type. The denominator used to obtain percent totals was the sum of household counts (5-yr estimates) for each area of interest.

- Observation: Slightly similar household type distribution among the three populations of interest.

Table 15: Population by Household Type, Household Count Totals for 2011 - 2015

Household_Type	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Family households	9477	16883	24684	310166
Nonfamily households	10719	15809	25137	440253

Table 16: Population by Household Type, Household Percent Totals for 2011 - 2015

Household_Type	Zone	CD10 nonZone	CD10 all	NY Manhattan
Family households	46.925134	51.642604	49.545372	41.332376
Nonfamily households	53.074866	48.357396	50.454628	58.667624

Figure 11: Population by Household Type, Stacked Bar Graph of Percent Totals for 2011 - 2015
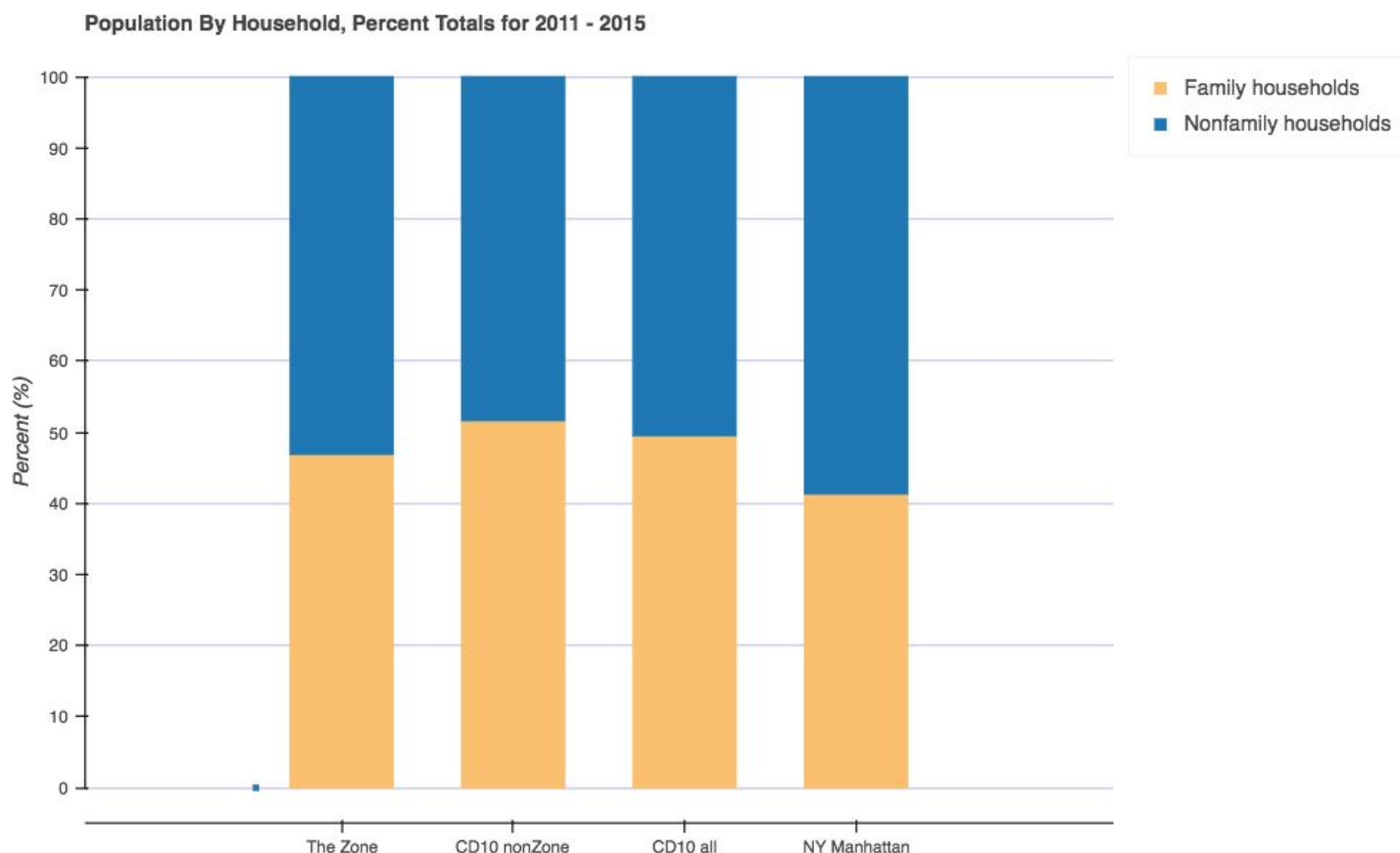
Findings -- Population by Family Household Type

In *Figure 12*, we can see the breakdown (in percentages) of the sample populations by family household type. The denominator used to obtain percent totals was the sum of family household counts (5-yr estimates) for each area of interest.

● Observation: Similar distribution of family household type among the three populations of interest.

Table 17: Population by Family Household Type, Household Percent Totals for 2011 - 2015

Family_Household	The Zone	CD10 nonZone	CD10 all	NY Manhattan
Married-Couple Families	41.24714	37.608442	40.354517	62.964386
Male householder, no wife present	9.279176	9.497605	8.903523	6.221253
Female householder, no husband present	49.473684	52.893953	50.74196	30.814361

Figure 12: Population by Family Household Type, Stacked Bar Graph of Percent Totals for 2011 - 2015

Findings -- Married-Couple (MC) Family Population by Children Age

Focusing on only the population of married-couple families, *Figure 13* visualizes the breakdown (in percentages) of the sample populations by age groups of the children in the family. In other words, out of the married-couple families in each area of interest, what percentage of them have children for the listed age groups. The denominator used to obtain percent totals was the sum of married-couple family household counts (5-yr estimates) for each area of interest.

- Observation: Similar married-couple family distribution among the three populations of interest.

Table 18: Married-Couple Family Population, Breakdown by Age of Children, Percent Totals for 2011 - 2015

Family_Household	The Zone	CD10 nonZone	CD10 all	NY Manhattan
(MC) Under 3 years	15.034674	21.914271	20.286129	25.777378
(MC) 3 and 4 years	20.638003	14.374247	16.765396	14.409894
(MC) 5 years	3.384189	5.663625	4.470772	5.942232
(MC) 6 to 11 years	32.676838	28.765708	31.228345	29.752337
(MC) 12 to 17 years	28.266297	29.282148	27.249357	24.118158

Figure 13: Married-Couple Family Population, Breakdown by Age of Children, Stacked Bar Graph of Percent Totals for 2011 - 2015
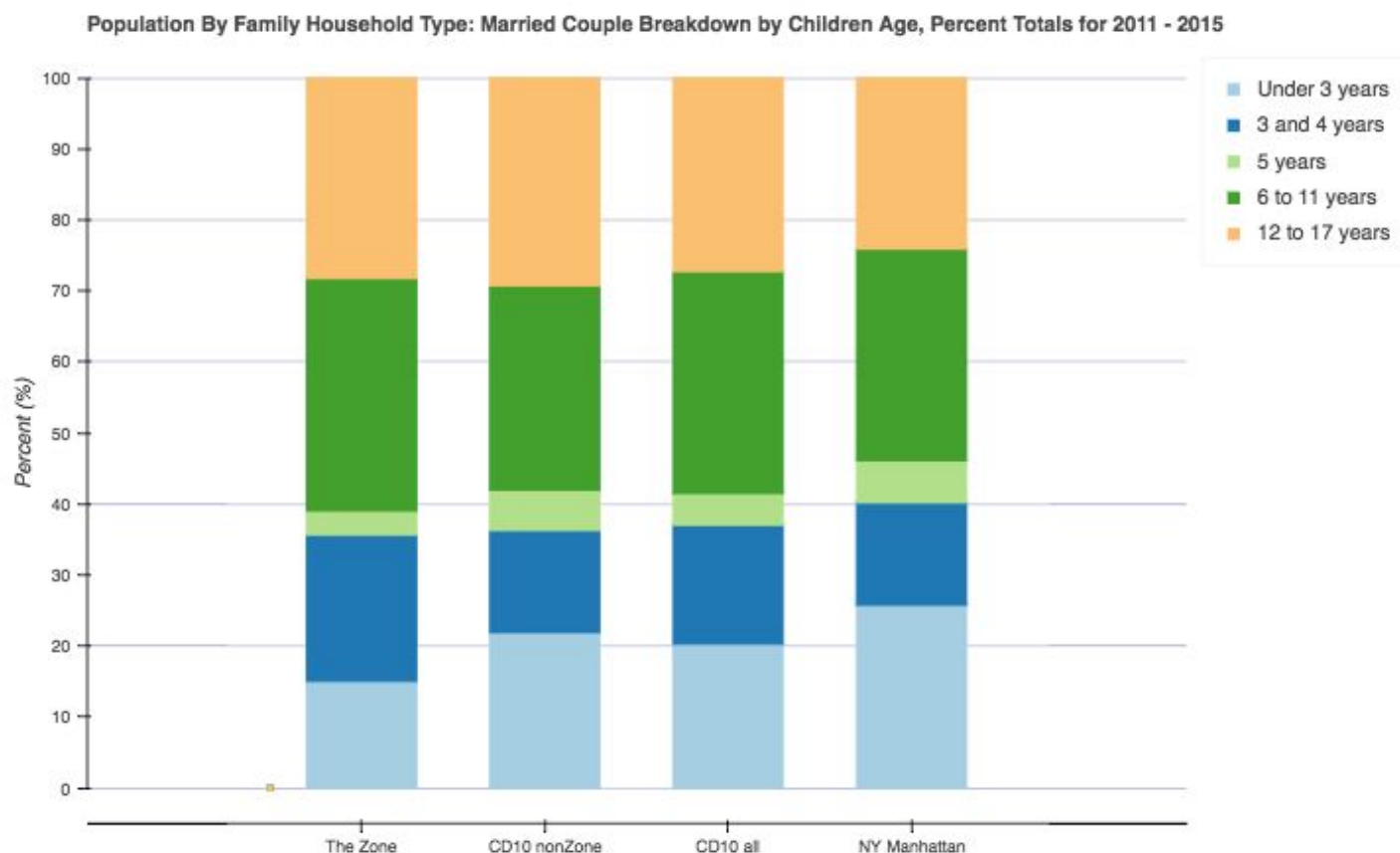
In *Figure 14*, q-q plots are displayed for five sample population comparisons (for married-couple families by age of children): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_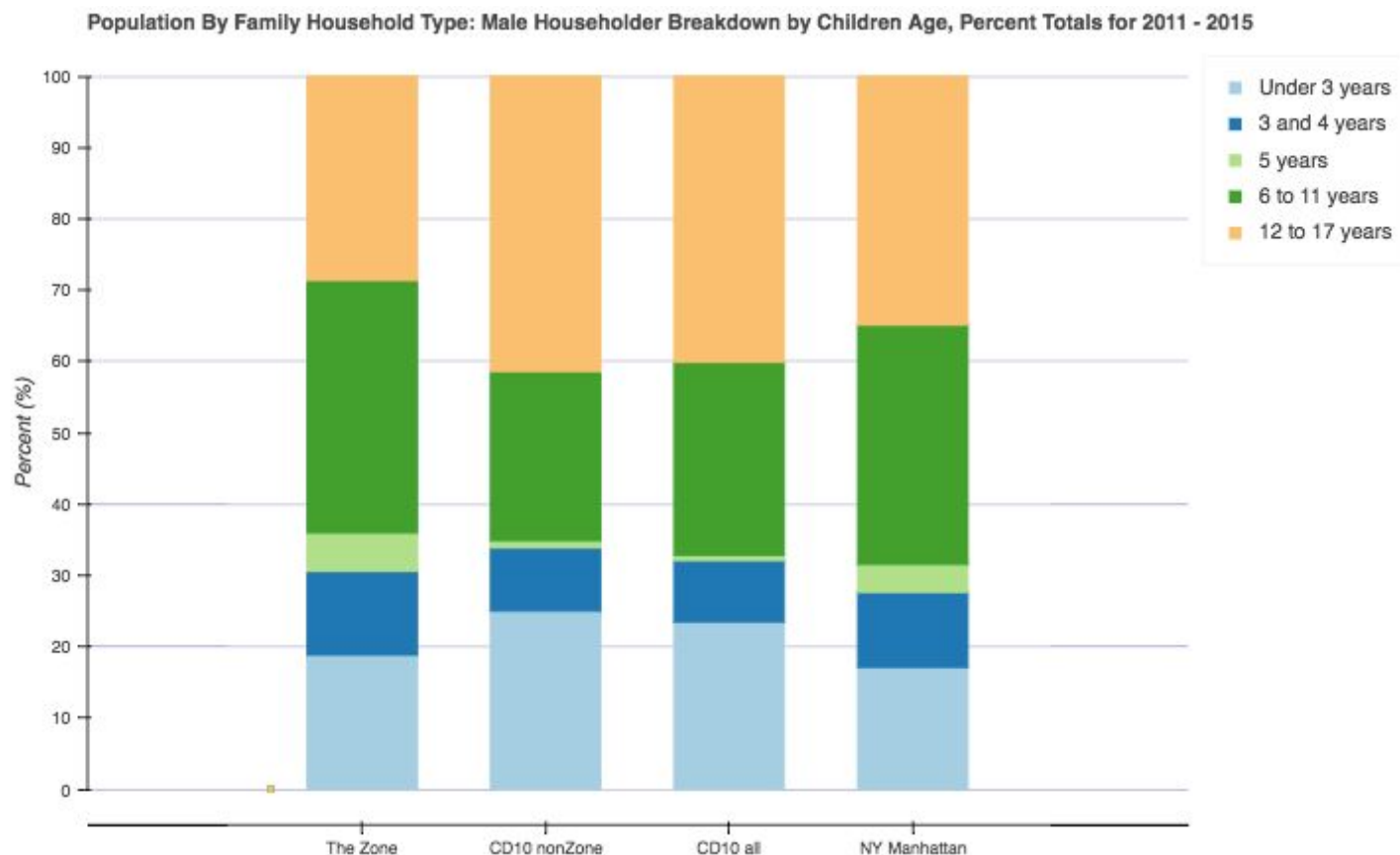all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for married-couple families with Community District 10 as a whole. Comparing Figures 14-A and 14-B with 14-D and 14-E confirms that the Zone has a slightly similar distribution for married-couple families with CD10 than with NYC Manhattan. Comparing Figures 14-A and 14-B with Figure 14-C confirms that population distributions for married-couple families are slightly similar between the Zone and Community District 10.

- Observation: Slightly similar distribution of married-couple families by age of children between the Zone and Community District 10.

Figure 14: Q-Q Plot of Married-Couple Families by Age of Children for 2011 - 2015
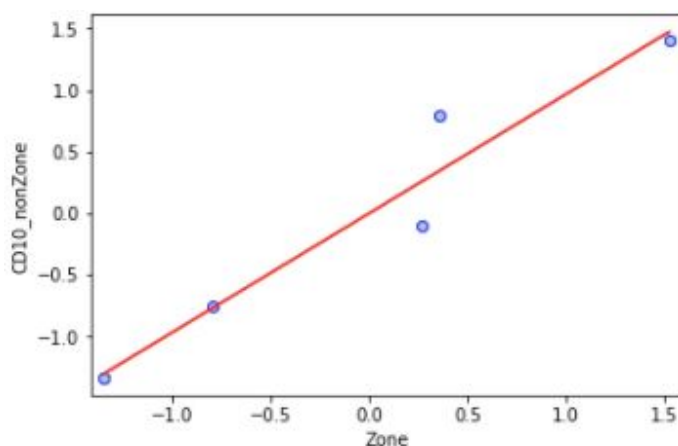

Figure 14-A: Zone vs. CD10 Non-Zone
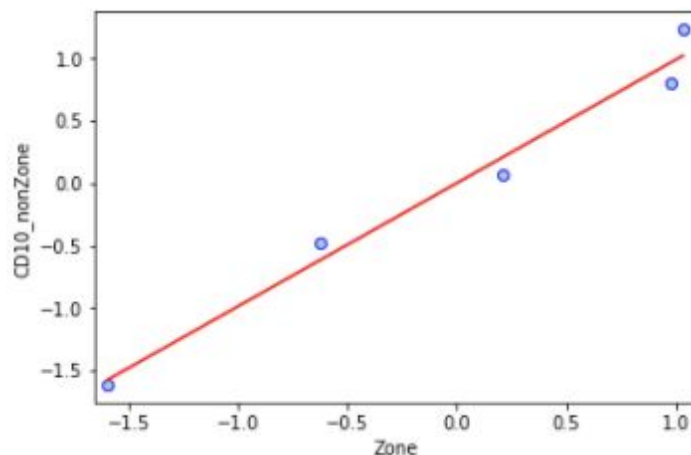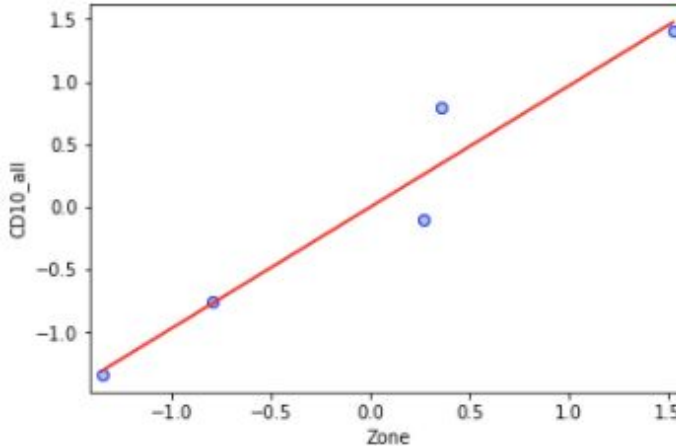
Figure 14-B: Zone vs. CD10 All
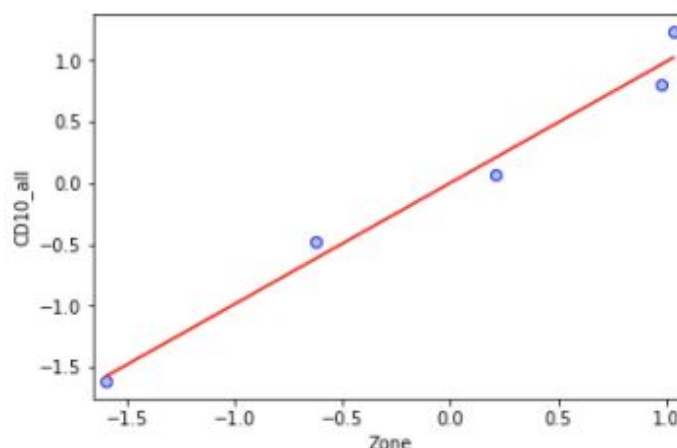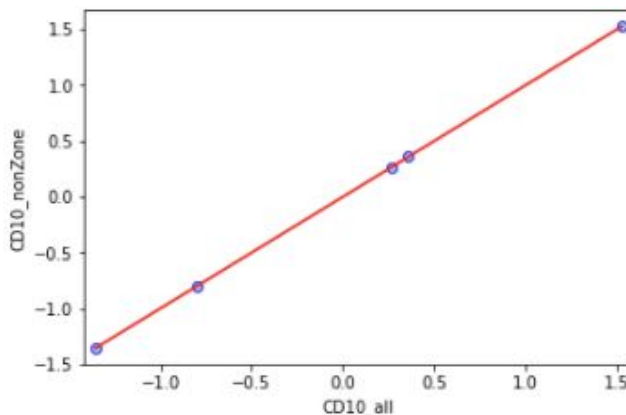
Figure 14-C: CD10_all vs. CD10 Non-Zone
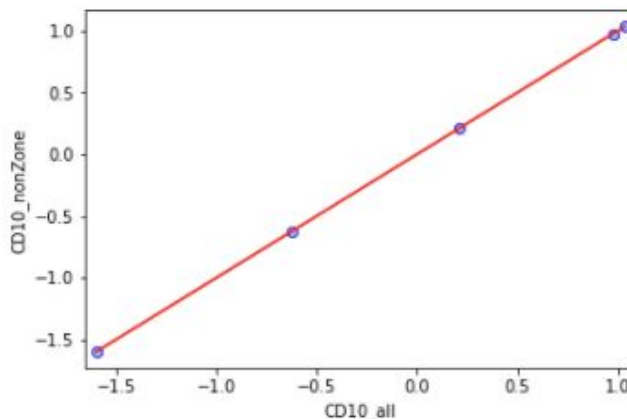
Figure 14-D: Zone vs. NY_Manhattan
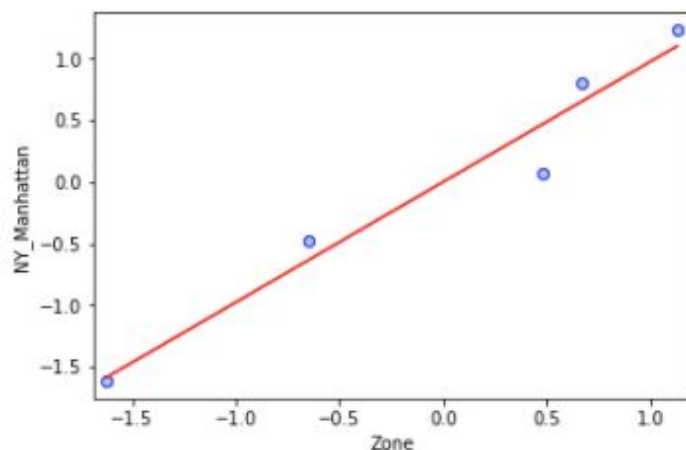
Figure 14-E: CD10_all vs. NY_Manhattan
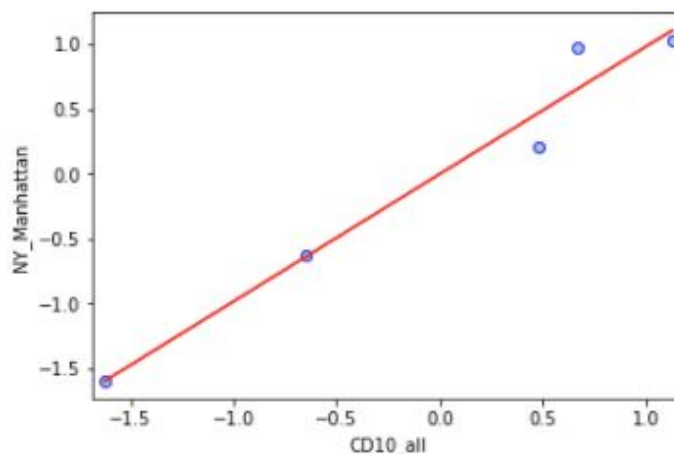
Findings -- Male-Householder (MHH) Family Population by Children Age

Focusing on only the population of male-householder families (no wife present), *Figure 15* visualizes the breakdown (in percentages) of the sample populations by age groups of the children in the family. In other words, out of the male-householder families in each area of interest, what percentage of them have children for the listed age groups. The denominator used to obtain percent totals was the sum of male-householder family household counts (5-yr estimates) for each area of interest.

● Observation: Different male-householder family distribution among the three populations of interest.

Table 19: Male-Householder Family Population, Breakdown by Age of Children, Percent Totals for 2011 - 2015

Family_Household	The Zone	CD10 nonZone	CD10 all	NY Manhattan
(MHH) Under 3 years	18.865598	25.017042	23.505572	17.147708
(MHH) 3 and 4 years	11.713933	8.861622	8.611955	10.518599
(MHH) 5 years	5.425401	1.022495	0.759878	3.920358
(MHH) 6 to 11 years	35.388409	23.721881	27.001013	33.515975
(MHH) 12 to 17 years	28.606658	41.37696	40.121581	34.897361

Figure 15: Male-Householder Family Population, Breakdown by Age of Children, Stacked Bar Graph of Percent Totals for 2011 - 2015

Population By Family Household Type: Male Householder Breakdown by Children Age, Percent Totals for 2011 - 2015

In *Figure 16*, q-q plots are displayed for five sample population comparisons (for male-householder families by age of children): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for male-householder families with Community District 10 as a whole. Comparing Figures 16-A and 16-B with 16-D and 16-E confirms that the Zone has a slightly similar distribution for male-householder families with NYC Manhattan than with CD10. Comparing Figures 16-A and 16-B with Figure 16-C confirms that population distributions for male-householder families are different between the Zone and Community District 10.

- Observation: Different distribution of male-householder families by age of children between the Zone and Community District 10.

Figure 16: Q-Q Plot of Male-Householder Families by Age of Children for 2011 - 2015

Figure 16-A: Zone vs. CD10 Non-Zone

Figure 16-B: Zone vs. CD10 All

Figure 16-C: CD10_all vs. CD10 Non-Zone
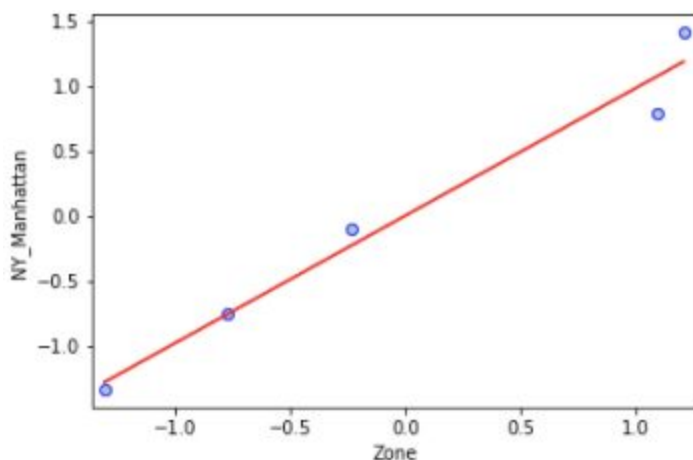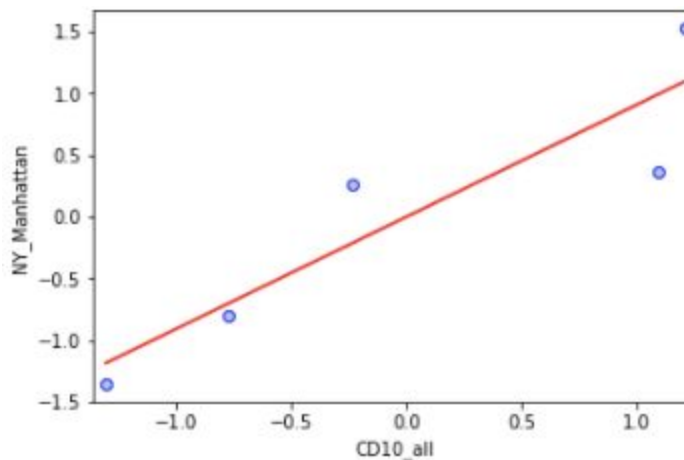
Figure 16-D: Zone vs. NY_Manhattan

Figure 16-E: CD10_all vs. NY_Manhattan

**Findings -- Female-Householder (FHH) Family Population by Children Age**

Focusing on only the population of female-householder families (no husband present), *Figure 17* visualizes the breakdown (in percentages) of the sample populations by age groups of the children in the family. In other words, out of the female-householder families in each area of interest, what percentage of them have children for the listed age groups. The denominator used to obtain percent totals was the sum of female-householder family household counts (5-yr estimates) for each area of interest.

- Observation: Similar female-householder family distribution among the three populations of interest.

Table 20: Female-Householder Family Population, Breakdown by Age of Children, Percent Totals for 2011 - 2015

Family_Household	The Zone	CD10 nonZone	CD10 all	NY Manhattan
(FHH) Under 3 years	11.031452	14.332925	12.853333	13.030133
(FHH) 3 and 4 years	14.73173	13.586291	14.195556	10.26456
(FHH) 5 years	5.712303	5.887393	5.742222	5.241345
(FHH) 6 to 11 years	35.360777	32.350061	33.884444	32.012091
(FHH) 12 to 17 years	33.163737	33.843329	33.324444	39.451871

Figure 17: Female-Householder Family Population, Breakdown by Age of Children, Stacked Bar Graph of Percent Totals for 2011 - 2015
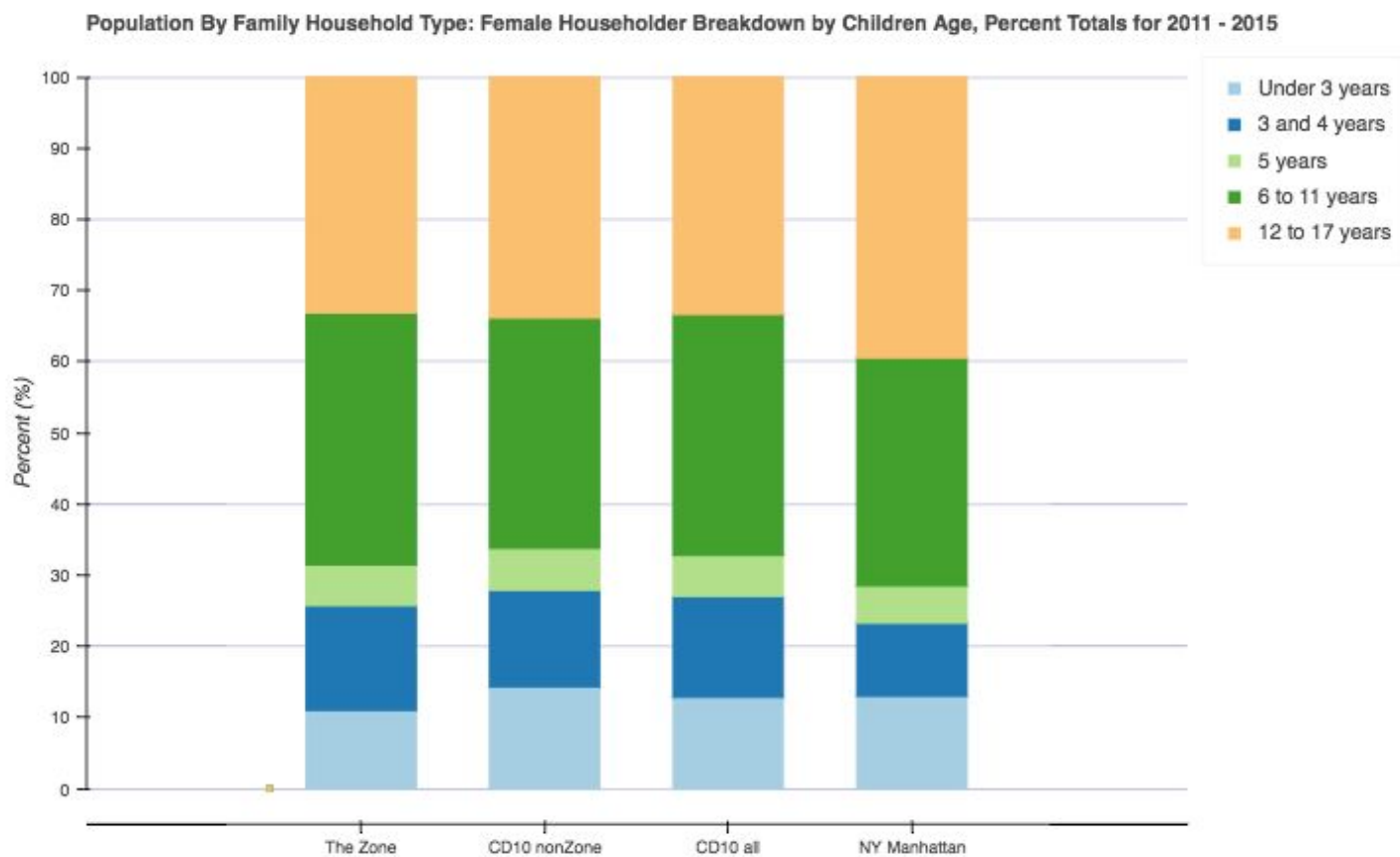
In *Figure 18*, q-q plots are displayed for five sample population comparisons (for female-householder families by age of children): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for female-householder families with Community District 10 as a whole. Comparing Figures 18-A and 18-B with 18-D and 18-E confirms that the Zone has a slightly similar distribution for female-householder families with CD10 than with NYC Manhattan. Comparing Figures 18-A and 18-B with Figure 18-C confirms that population distributions for female-householder families are similar between the Zone and Community District 10.

- Observation: Similar distribution of female-householder families by age of children between the Zone and Community District 10.

Figure 18: Q-Q Plot of Female-Householder Families by Age of Children for 2011 - 2015
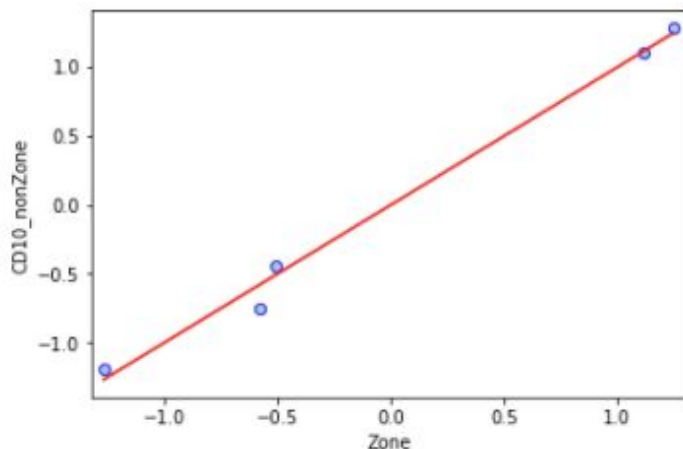

Figure 18-A: Zone vs. CD10 Non-Zone
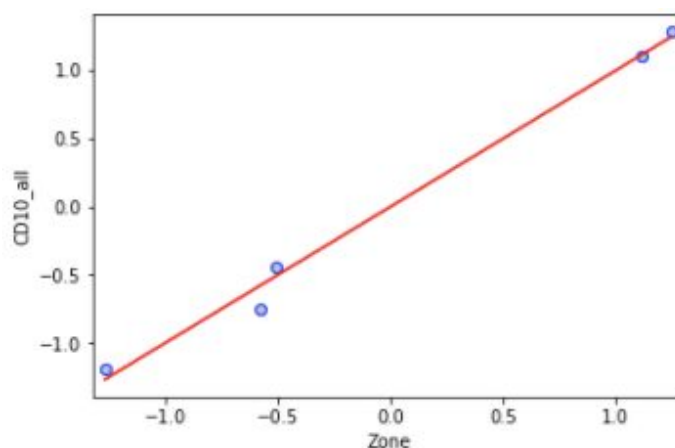

Figure 18-B: Zone vs. CD10 All
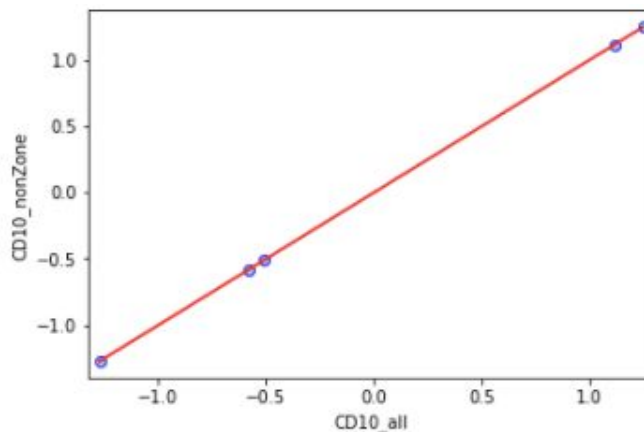

Figure 18-C: CD10_all vs. CD10 Non-Zone

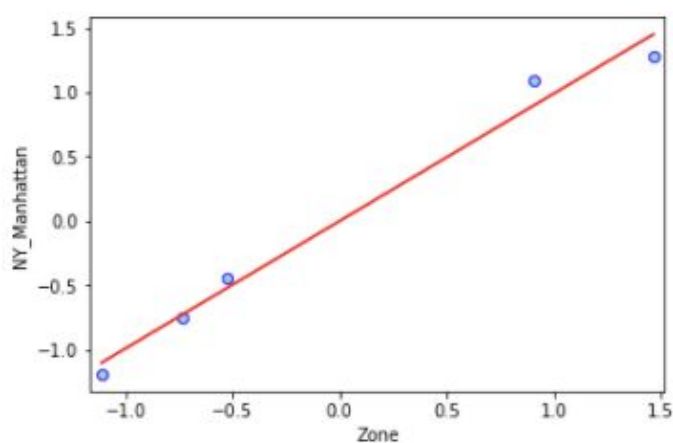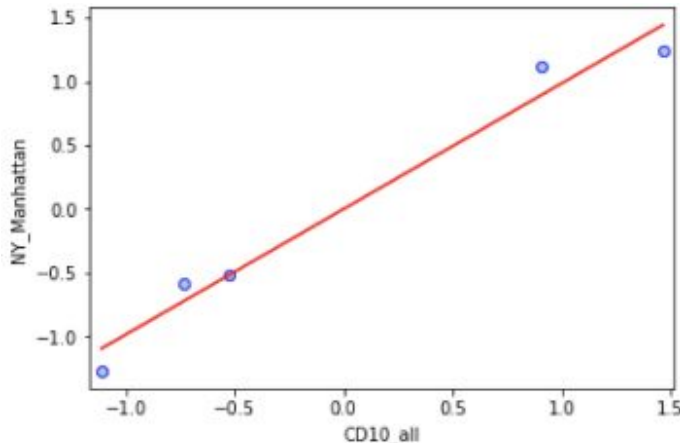Figure 18-D: Zone vs. NY_Manhattan

Figure 18-E: CD10_all vs. NY_Manhattan

Findings -- Population by Household Income

In *Figure 19*, we can see the breakdown (in percentages) of the sample populations by household income. The denominator used to obtain percent totals was the sum of household counts (5-yr estimates) for each area of interest.

- Observation: Similar household income distribution among the three populations of interest.

Table 21: Population by Household Income, Household Count Totals for 2011 - 2015

Household_Income	Zone	CD10_nonZone	CD10_all	NY_Manhattan
Less than $10,000	3143	5365	7896	71430
$10,000 to $14,999	1791	2660	4212	39773
$15,000 to $19,999	1382	2095	3213	33103
$20,000 to $24,999	1027	1796	2602	28845
$25,000 to $29,999	879	1833	2507	24429
$30,000 to $34,999	963	1977	2714	25933
$35,000 to $39,999	1055	1601	2487	22258
$40,000 to $44,999	655	1291	1852	21844
$45,000 to $49,999	731	1150	1789	18969
$50,000 to $59,999	1529	2180	3576	39737
$60,000 to $74,999	1376	3069	4226	54769
$75,000 to $99,999	2090	2175	3942	72568
$100,000 to $124,999	1052	1922	2894	59231
$125,000 to $149,999	646	956	1533	41714
$150,000 to $199,999	835	1230	1997	58474
$200,000 or more	1042	1392	2381	137342

Table 22: Population by Household Income, Household Percent Totals for 2011 - 2015

Household_Income	Zone	CD10 nonZone	CD10 all	NY Manhattan
Less than $10,000	15.562488	16.410743	15.848738	9.518682
$10,000 to $14,999	8.868093	8.136547	8.454266	5.300106
$15,000 to $19,999	6.842939	6.408296	6.449088	4.411269
$20,000 to $24,999	5.085165	5.493699	5.222697	3.843853
$25,000 to $29,999	4.352347	5.606876	5.032015	3.255381
$30,000 to $34,999	4.768271	6.047351	5.447502	3.455803
$35,000 to $39,999	5.223807	4.897223	4.991871	2.966076
$40,000 to $44,999	3.243216	3.948978	3.717308	2.910907
$45,000 to $49,999	3.619529	3.51768	3.590855	2.527788
$50,000 to $59,999	7.570806	6.668298	7.177696	5.295308
$60,000 to $74,999	6.81323	9.387618	8.482367	7.298456
$75,000 to $99,999	10.348584	6.653004	7.912326	9.670331
$100,000 to $124,999	5.208952	5.879114	5.808795	7.893057
$125,000 to $149,999	3.198653	2.924263	3.077016	5.558761
$150,000 to $199,999	4.134482	3.762388	4.00835	7.79218
$200,000 or more	5.159438	4.257922	4.779109	18.302042

Figure 19: Population by Household Income, Stacked Bar Graph of Percent Totals for 2011 - 2015

Population By Income, Percent Totals for 2011 - 2015
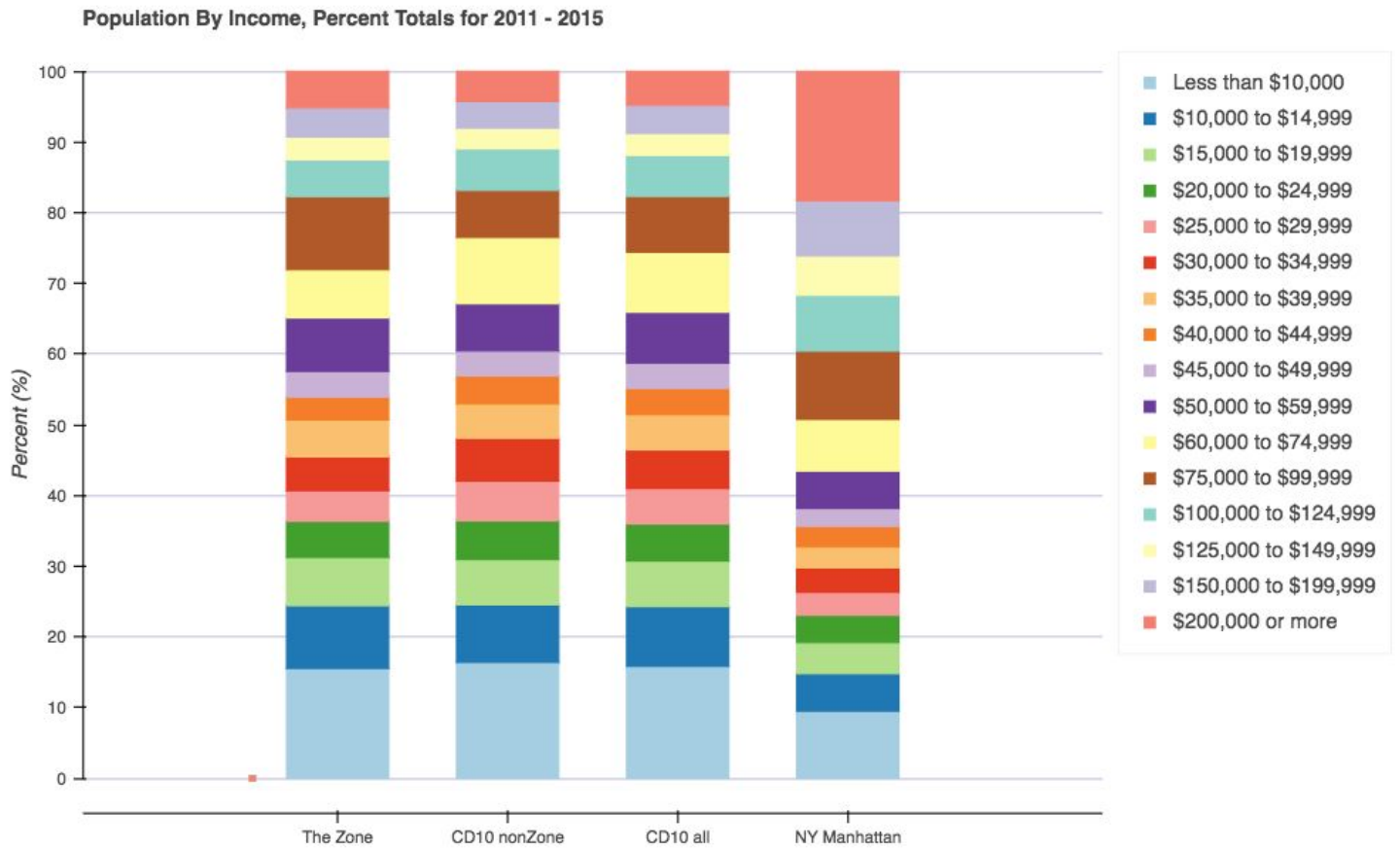
In *Figure 20*, q-q plots are displayed for five sample population comparisons (for household income distribution): (A) Zone vs. CD10 Non-Zone; (B) Zone vs. CD10 All; (C) CD10_all vs. CD10 Non-Zone; (D) Zone vs. NY_Manhattan; and (E) CD10_all vs. NY_Manhattan. Figure C displays an ideal fit, which confirms that non-Zone areas in Community District 10 have a common distribution for household income with Community District 10 as a whole. Comparing Figures 20-A and 20-B with 20-D and 20-E confirms that the Zone has a similar distribution for household income with CD10 than with NYC Manhattan. Comparing Figures 20-A and 20-B with Figure 20-C confirms that population distributions for household income are slightly similar between the Zone and Community District 10.

- Observation: Slightly similar distribution of household income between the Zone and Community District 10.

Figure 20: Q-Q Plot of Household Income for 2011 - 2015

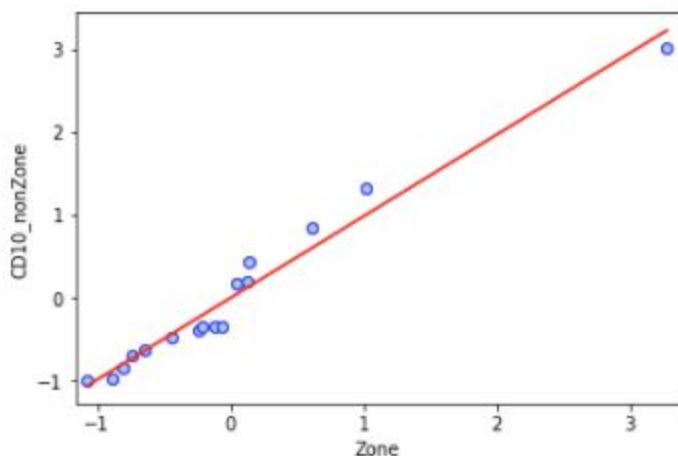Figure 20-A: Zone vs. CD10 Non-Zone
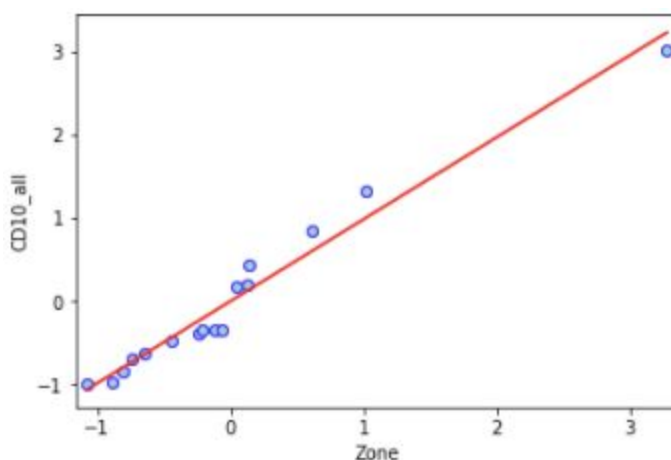
Figure 20-B: Zone vs. CD10 All
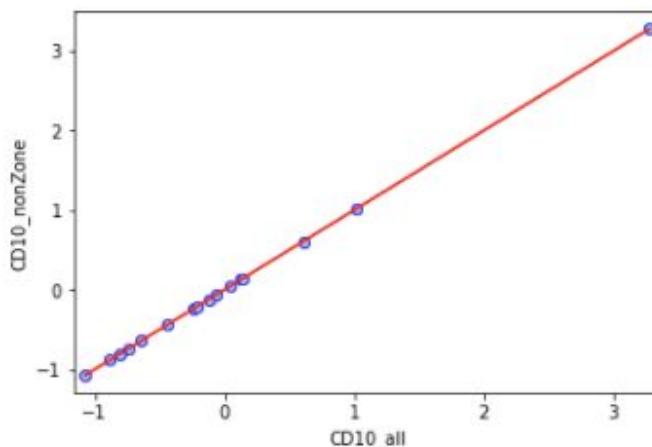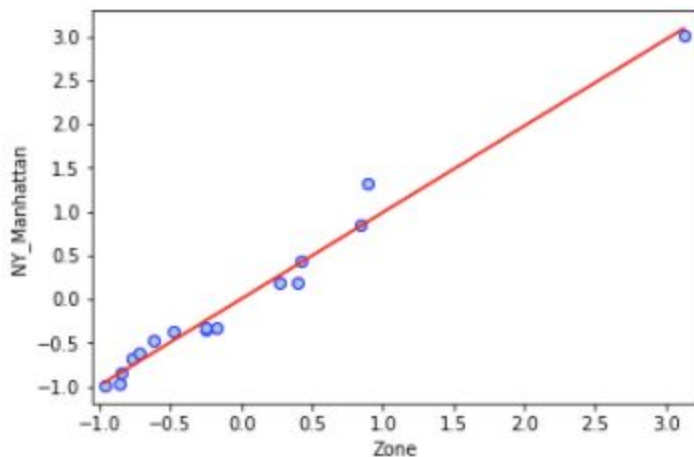
Figure 20-C: CD10_all vs. CD10 Non-Zone

Figure 20-D: Zone vs. NY_Manhattan

Figure 20-E: CD10_all vs. NY_Manhattan