

**Analysis of the Philadelphia, PA and Camden, NJ
Block Groups with emphasis of Census 2010 Children Data**

GROUP 12

**Zhuling Li
Olga Zubashko
Johnny Bae
Bernard Yuan
Asad Meghani**

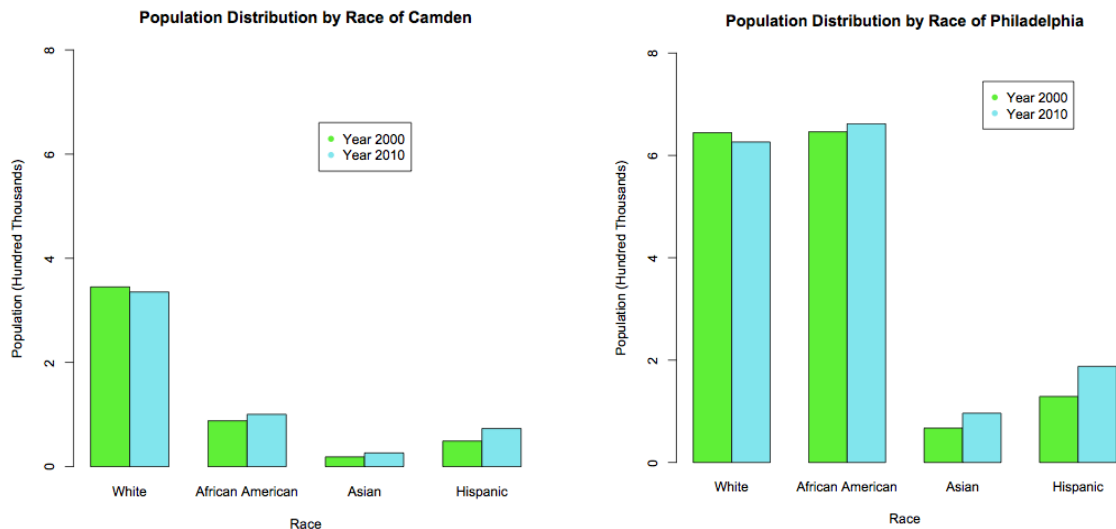
Abstract

For our final analysis and presentation, we are working with the US Census 2000 and 2010 data. Specifically, we are using the UScensus2000 and UScensus2010 R packages, as well as the American Community Survey (ACS) data from 2010, which surveyed a small portion of the population in 2010 to obtain census-related information. We focus our analysis on two adjacent cities: Camden, NJ and Philadelphia, PA.

We are given demographic data of two different cities, Philadelphia, PA and Camden, NJ. We want to draw conclusions using this information and display that information using graphs and written analysis. In this report, we will examine the relationship between two cities across multiple years and some of the demographics associated with them. Some of the specific aspects we examine include how many households contain a certain number of people, what portion of households in each city contain children, what the income of each household is in relation to the number of children they have, and the distribution of race in the household and cities. Using the knowledge acquired throughout the semester in this course, we were able to draw multiple conclusions, the results of which are included in depth.

Population Distribution by Race of Camden

Since our data set is large and filled with various variables, we begin our analysis by exploring a few basic variables to understand our data and to decide how we want to gear our analysis. We start by examining race distributions in 2000 and 2010 in each city.



We chose to display the population distribution by race of Camden and Philly through two bar charts. Using barplots, we easily compare the heights of each category of data and determine descriptive statistics, such as maximums and minimums. It is also easy to determine an approximation of the exact count for each category of data. We could have used a spine chart or a pie chart to illustrate the data, but neither would not been as easy to understand right off the bat and would require more work by the audience. It is easier to compare heights of a bar chart than the widths of bars in a spine chart and the sections of a pie chart. Furthermore, spine charts and pie charts do not illustrate the exact counts in each category. We could have also geographically shown the distribution of each race by plotting the points on a map, but we wanted to explore the data numerically.

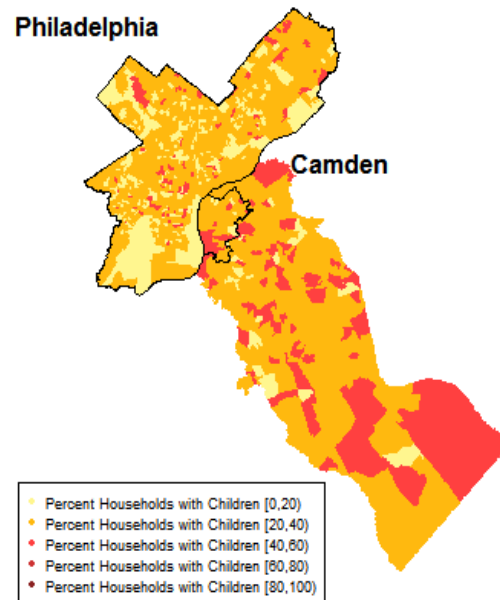
Analysis: The left bar in each pair represents the population data from the year 2000, and the right bar represents the population data from the year 2010. Looking at the white population, we see that it slightly decreased from the year 2000 to 2010. As expected, the other three races (African American, Asian, and Hispanic) increased from 2000 to 2010. Nationwide, the fastest growing minority population is Hispanic, and this is also evident in Camden and Philadelphia, as the Hispanic population experiences the largest population growth in our bar chart. The only difference between the two cities is the overall population. Looking at the y-axis and reading across the two graphs, we see that Philadelphia's overall population is greater than Camden's overall population, while the trends of growth and decline and each race remain homogenous across the two cities.

Children Under 18

After examining race in each city, we want to examine distribution by age. We assume that each city has a normal distribution of age, we numbers peaking between the ages of 30 and 50 and falling lower at the tails. We therefore choose to examine the distribution of households with children under the age of 18 – we want to see the number of households in Philadelphia and Camden choosing to have children.

Unfortunately, the 2000 and 2010 data is different – in 2000 we can see the total number of children of various ages in each blockgroup of each city, while in 2010, we see the number of households with children under the age of 18, regardless of the number of children in each household. We therefore do not compare the distributions of 2000 and 2010, but solely focus on the 2010 data because we are interested in the choices of households as opposed to individuals.

Percent of Households with Children Under 18 Philadelphia vs. Camden



We decide to focus on a specific age group, namely children under the age of 18, instead of analyzing the distribution of age as a whole. We create a map focusing on the distribution of children in households across Philadelphia and Camden block groups. Instead of showing how many children there are in each block group, we color the block groups according to the percentage of households in that block group that have at least one child under the age of 18. We are interested in seeing which block groups had high or low percentages of households with children and what general areas of each county (Philadelphia is both a city and county; Camden City is a subset of Camden County, shown with border on map) had more or less children in households; that's why we decided that a map would have been the best choice. To make the map easier to read, we use the built-in county function under the USCensus2010 package to show Philadelphia and Camden, without showing information about surrounding block groups outside of the two counties. A disadvantage of our map is that it is difficult to see what color the smaller block groups are and therefore it is not completely clear whether or not there is a pattern in the distribution. Another option would have been a density plot but the reason it was not used was because we would lose the ability to tell whether or not the location of the block group, if it was closer rural to areas, affected the percentage of households with children.

The way we set up the percentages and colors was by assigning a color to every 20th interval of percentages (i.e. 0-20%, 20-40%, 40-60%, etc). The lower percentages are assigned a more yellow color, with the hue increasing in darkness as the percentages increased. The higher

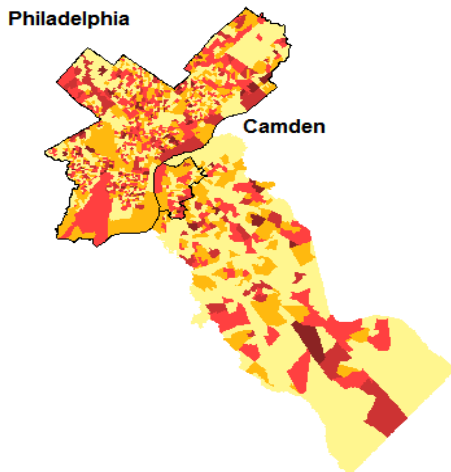
percentages are assigned a more red color, with the hue increasing in darkness as the percentages increased.

The majority of block groups with higher percentages of households with children are concentrated outside of the city of Camden, with the exception of one or two areas. We believe a reason for this might be because parents prefer to raise their children outside of the city because it is safer. Because cities are usually more densely populated, crime rates are also higher; these statistics show that cities might not be the best place to raise young children in. The distribution of households with children in Philadelphia also follows this pattern. As mentioned above, Philadelphia is a city, as well as county. There are very miniscule dark red areas (small specs), with the majority of block groups in Philadelphia being colored in with some hue of yellow. This is consistent with the idea that cities might not provide the best environment to raise children in. For the most part, 20-40% of households per block group have children less than 18 years of age. The region of South Philadelphia has only 0-20% of households with children. South Philadelphia also happens to be a very dangerous part of the city, so families may choose not to live there, which would explain the low percentage of households with children.

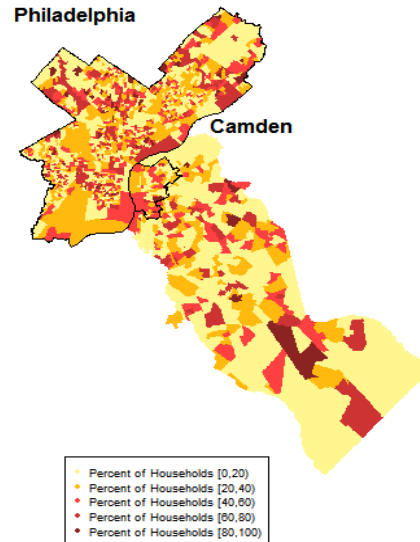
Renting vs. Owning

Having examined Household's choices to have or not have children in Camden and Philadelphia, we want to delve into the analysis of why households make these choices. We assert that having children is a choice households make based on a variety of factors: income, marriage status, etc. We therefore continue with our analysis of households with children under the age of 18 by examining how households with children are affected in a variety of other factors. We decided to first look into whether or not having a child under the age of 18 can affect whether parents decide to rent or own homes.

**Percent of Households Rented
Philadelphia vs. Camden**



**Percent of Households with Children, Rented
Philadelphia vs. Camden**

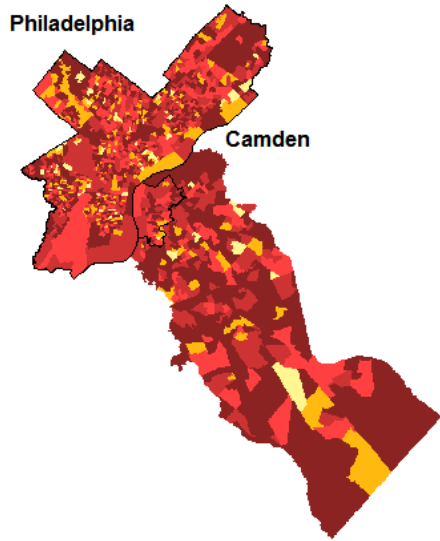


First, we focus on the distribution of rented homes and how it compares to the distribution of homes with children that are rented. By doing this, we examine first a household's choice to rent a home when it does not have children and then compare it to a household facing the same decision when it does contain children. In this case, we felt that using a map would give us the most information about the dataset (i.e. where higher concentrations of rented homes are and whether these concentrations change with homes that have children and are rented). For the left map, we showed the percentage of the total number of households that are rented (in each block group). For the map on the right, we showed the percentage of households that are rented and have children under the age of 18 divided by the total number of households with children (in each block group). The way we assigned colors and percentage ranges was the same as the previous map of the distribution of children under 18 for reasons of consistency.

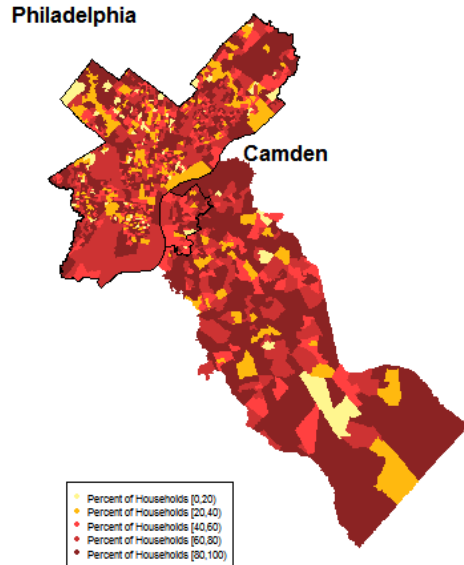
At first glance, the two maps actually look similar. However, there are a few interesting points that are worth investigating further. Near South Philadelphia, there is a big area where there is a relatively high percentage of households that are rented (40-60%), but that same area shows a lower percentage of households with children that are rented (20-40%). The pattern shows that a high percentage of rented households in general does not necessarily correlate with a high percentage of rented households with children. There are also some block groups that show low percentages of rented households and high percentage of rented households with children. This furthers our assertion that when facing the same choice of renting a home, households make different decisions based on whether they have children, though we cannot assert how a household will act simply by knowing whether there are children.

Having examined renting of housing, we want to see if the same pattern of difference based on children exists with houses that are owned.

**Percent of Households Owned
Philadelphia vs. Camden**



**Percent of Households with Children, Owned
Philadelphia vs. Camden**



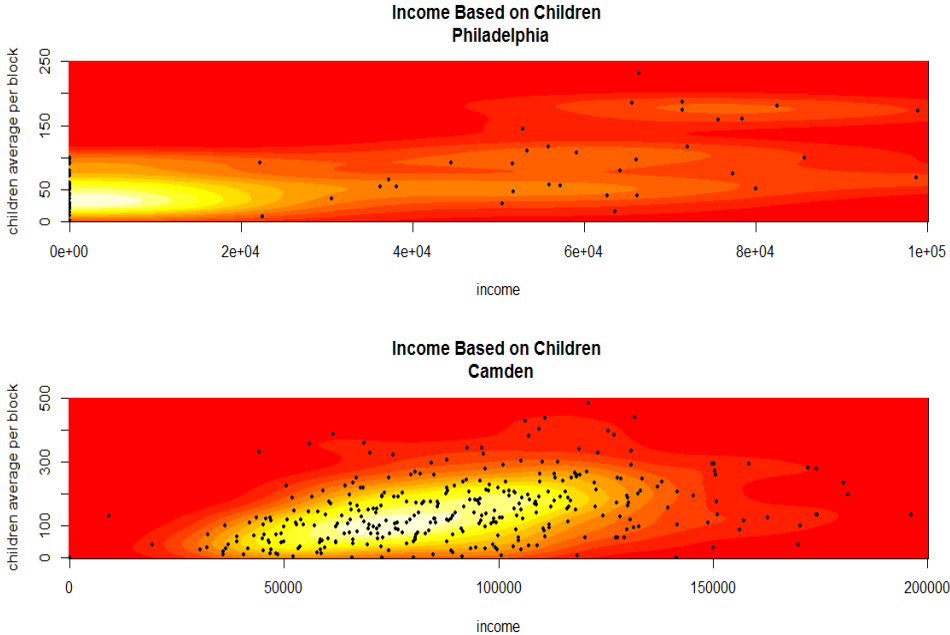
We therefore look into the distribution of owned homes and owned homes with children. We calculate the percentages of owned households by dividing the total number of owned homes by the total number households. We calculate the percentages of owned households with children by dividing the total number of owned households with children by the total number of households with children. We want to see if high percentage of owned household block groups also corresponded to a high percentage of owned households with children. Again, the assignment for colors and range of percentages remain the same for consistency in analysis. Similar to the rented household maps, it's difficult to see any correlation at first glance. We see that in many block groups, the percentage of owned houses remains the same regardless of children status. We can, however, note that in Philadelphia, the farthest block groups from Camden, are less likely to be owned when children are present than when children are not present. We also see this pattern right in the center of Philadelphia – the percentages of owned houses decrease with the presence of children. We see the opposite emerge in Camden County, outside of the city – more households with children are owned than households owned in general. We assert that because having children and owning a house are both expensive, having a child within the city takes money away from owning a house, especially since owning a home in the city is more expensive than in the suburbs. We believe the opposite happens in the suburbs – when people move to the

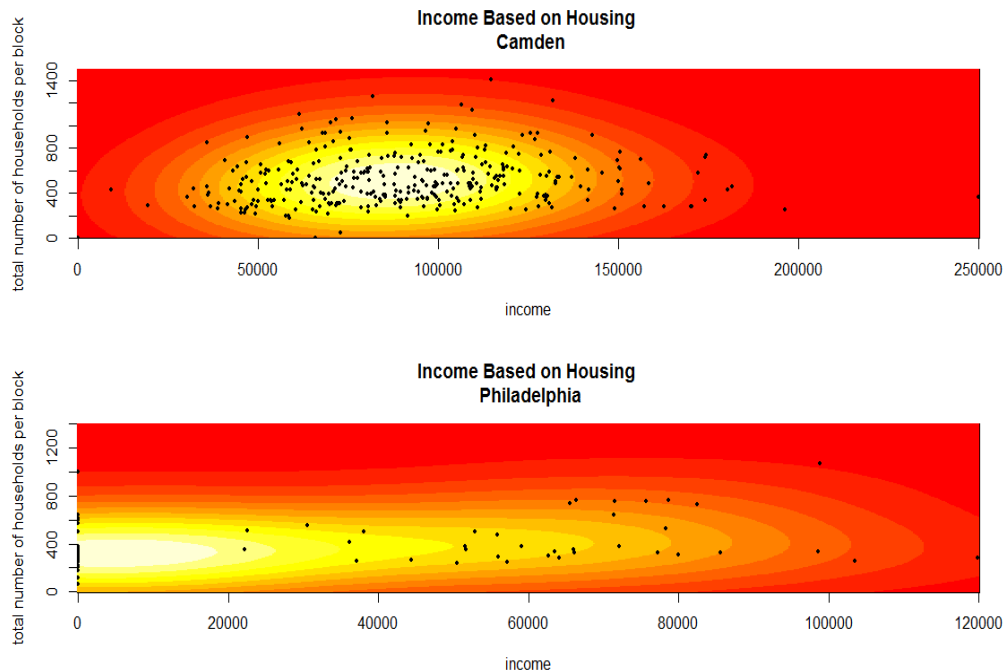
suburbs and choose to create a family, they also choose to buy homes and budget their money equally amongst various family values.

With the given analysis of renting and owning homes, we believed maps would give a valid representation of the locations of rented and owned homes. This would allow us to analyze the effects of not only children but also the city and the suburbs. If we wanted to examine pure numbers, we could have created an association plot to demonstrate the presence of children and owning/renting homes. Here, we would have to recode the variables accordingly to make them categorical instead of working with percentages.

Income Analysis

From our previous analyses, we asserted that many lifestyles changes because of the presence of children could also be attributed to income differences and the monetary effects of having children. We therefore choose to examine how income changes with the presence of children.





For Philadelphia, we realized that a heavy portion of the income data from the American Community Survey was simply missing. Because of this, the data could not be accurately represented, but we still felt that it was worth inserting in the graph, for otherwise it would be a heat map with only a few points. This is the main reason that the heat map looks to be centered around an income of 0.

However, Camden proved to be a reliable data set and displayed a strong heat center around the \$50,000 - \$100,000 income area. It is important to remember though that we worked heavily with the dataset of Camden county and not Camden city. When doing some inspection with the data to see just Camden city we learned that its income range is lower and focused more around the \$50,000 – \$75,000 range. The same general pattern applied with the housing heat maps.

Because of this, we assert that a higher income leads to a smaller probability of the existence of children. Although this often seems counterintuitive, we believe this makes sense. Literature has shown that as households become more financially literate, they also choose to have fewer children. We believe this same phenomenon happens in Camden, NJ and Philadelphia, PA – households that bring in more income also choose to have children at a later age, or when they are more financially stable.

When choosing the best way to analyze income based on children and income based on housing, we responded to two key factors: 1) we had two different non-categorical variables, and 2) our primary objective was to examine common features.

From the first factor, we understood that all one-dimensional plots would be inefficient in displaying this data because our objective was to examine the existence of a relationship between each. Because of this, we narrowed it down to many different two-dimensional plots ranging from sunflower plot, strip chart, perspective, contour, etc.

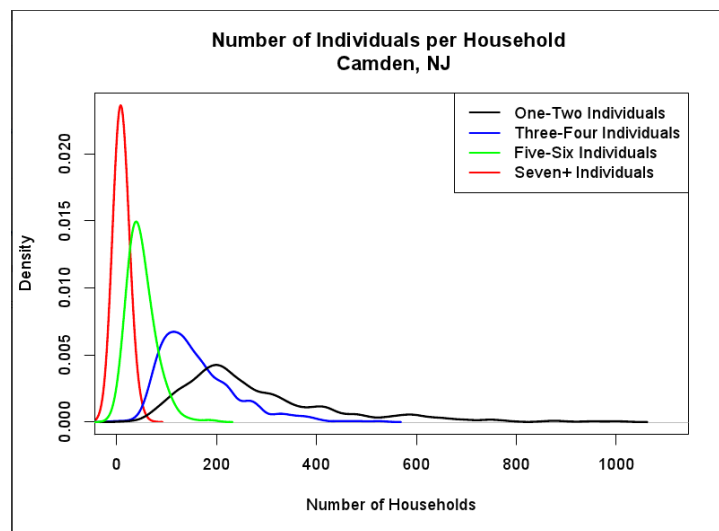
The heat map with points proved to be the optimal graph to display this data for a couple of reasons:

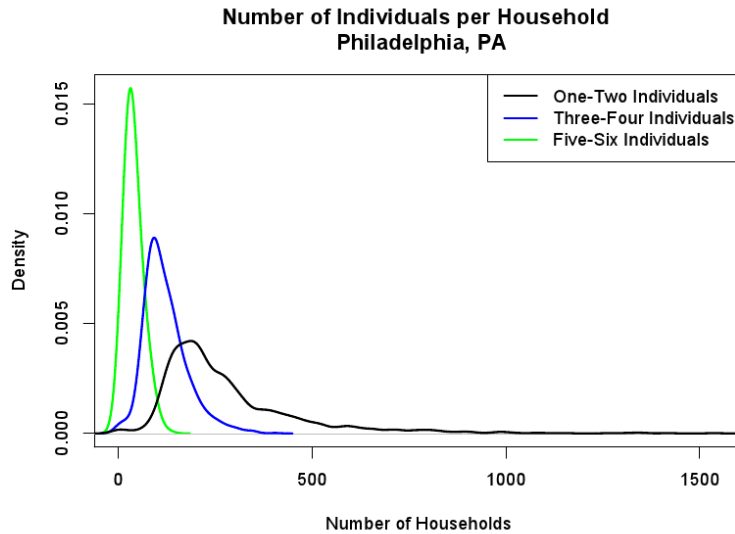
1. The data itself did not have repeating points diminishing the purpose of an alternative plot such as a sunflower plot.
2. A perspective plot would allow for a clean outlook, but it would not display the data in an easy to read manner.
3. A heat map would combine the positive effects of both a sunflower plot and a perspective plot by providing simplicity in understanding and allowing the audience to understand exactly where each point lies.

We also examined the option of overlaying a contour plot over the heat map. Although this would be a logical decision to make, we felt that a contour plot tended to create a more complicated graph than necessary. The points displayed clusters, and the heat map displayed the density. If there were multiple different densities in three or more different places, a contour map could show a percentage of each of the datasets lies in each cluster. For our data, that did not occur and we could solve our problem relatively easily with the above graphs.

Household Numbers

Having examined income and its relation to children, we believe that there is also a relationship between each city's household distribution and number of people living in the household.





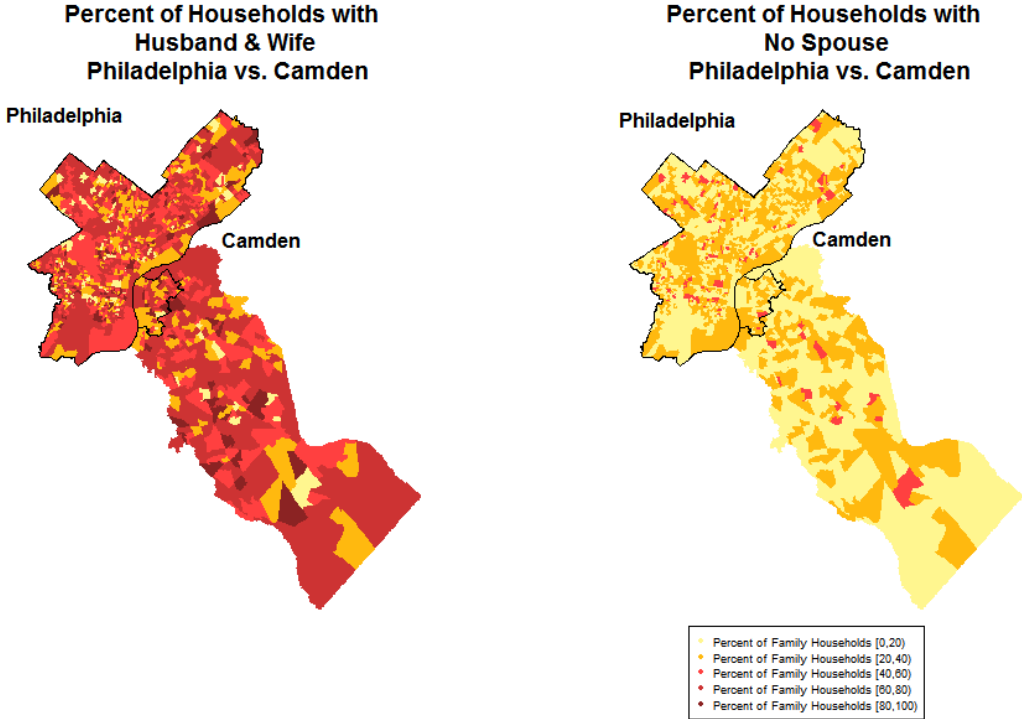
The lines represent how dense or, in layman's terms, how much data is located in a certain area, in this case, the number of households that contains either one-two, three-four, five-six, or seven-plus individuals. The first graph is for Camden, NJ and the second one is for Philadelphia, PA. The most prominent feature is that as the number of individuals in each household increases, there are less households that fit in that category. This makes sense, because it is much more likely for a household to have say, 3 people than 5, considering the national average of number of people in the household is around two and a half individuals. Additionally, more people per household suggest a higher income per household, as more individuals are harder to support. Our previous income analysis shows an income centered between \$50,000-\$100,000, numbers that may support 1-4 individuals comfortably. So as expected, a one-two individual has the most number of households and the widest distribution while seven-plus individuals, in Camden, or five-six individuals in Philadelphia, has the smallest number of households and shortest distribution. We attribute there being no 7+ individual households in Philadelphia because of congestion and less space for families, whereas Camden is less city-like. In conclusion, the trend that number of individuals in a household and the number of households that fit that description is negatively associated holds for both cities.

We chose to use a density line plot to indicate the number of people in each household for Philadelphia and Camden. We can easily quantify where each number lies in the total "Number of Individuals Per Household" distribution, for example, in households with 7+ individuals there are less than 100 households and in households with 1-2 individuals, it exceeds 1000. It is easy to glance at the graph and get a good overall idea of how many households contain how many individuals, because we put all the lines on the same plot and color coded them with the legend. This could not be done in our alternative choice of graph, which was originally a stacked strip chart. We had to have individual strip charts for each of the household classifications (1-2, 3-4, 5-6, 7+). In addition, the ylim cannot be controlled on a strip chart so many data points were cut off. We could have used a histogram, but we feel that since there is a large sample size, a smooth more aesthetic representation could be used. In addition, we would need multiple histograms. We could have used a violin plot, which does show density, however, the center and

IQR of this specific set of data doesn't really provide us necessary information and it would be better to choose a plot that shows only what we need and not excessive information. Thus, the best option is to choose a density line plot, because it has a solid combination of displaying relevant information in a easy to read and visually appealing way.

Households with Spouses vs. No Spouses

As we continue to examine the choice to have children, we believe that the choice to have children also depends on whether the household owner has a spouse or not.



We therefore create two maps depicting the percentages of family households with an owner and a spouse and an owner and no spouse. As would be expected, most of the households that identify themselves as “family households,” have a household head and a spouse. However, it does seem that there are quite a few block groups in the center of Philadelphia city with a high proportion of family households with no spouse. There appears to be fewer of these block groups within Camden City, though several exist in the county (proportionally a small number of block groups). We can also again see a trend of “family values” as the block groups move farther outside Camden city – there seem to be smaller percentages of family households with no spouse than within the city, where they range between 20-40% vs. 0-20% outside of the city.

We chose to again create a map to visualize the location of the data. We could have also examined the data numerically with weighted violin plots, but we believe that would have been more useful in the case of comparisons and including a stipulation for having children. Overall,

we believe the visual map provides a good representation of the data, and it is also easy to compare with previous maps that show the percentages of owning homes and having children.

Conclusion

Although we began the analysis by examining race in the cities, we focused most of our project on how children change the choices of households. After a thorough examination of how children affect the choice of renting homes vs. owning them, how income correlates with the presence of children, and whether households that identify themselves as family households have spouses, we believe that a household that has children makes its decisions differently than a household without children. Children significantly impact the location of resident, how money is allocated, and how many people live in each household. We believe further statistical analysis of the data would answer key questions of not only the location of households with these choices, as we examined here, but also how and why people make these decisions.