

Final Project

05/09/2013

Team #6

Jason Ha, Zhijun Huang, Terence Kwak, Stephany Park

Statistical Findings on Miami-Dade County
(US Consensus 2000 and 2010)

Abstract

We explore the combined data of US Census 2010 and ACS for the Miami Florida. We saw that Miami is a growing city population wise since our data doesn't explore anything about the area or the size of the city. The city of Miami is very dominated by White population with large of black, and asian population. Age and income has a negative correlation with total population of block groups. Moreover, age is positively correlated with only white population percentage, and income is positively correlated with white and Asian population percentages. We used geo-graphical plotting in order to find connection between race and income based on their location. We learned that population of higher income reside in suburbs. The population of highest income reside towards the ocean. We concluded that the trend speaks for white population since Miami Metropolitan Area is consisted of 76% white population. Lastly, the analysis of homeowners/renters and how they own their housing is a good indication of income. From our analysis we saw that different race had different housing preference and this may be attributed to the preference or characteristics of each race groups. Overall, Miami is a very diverse city with a large metropolitan area.

Introduction

Our group researched Miami Metropolitan Area because we are assigned with City of Miami. The city is part of the larger metropolitan area. The majority of population in the area resides within the city limits of Miami. The population choropleth map of Miami Metropolitan Area(Figure 7 a) is to give an overview of the county. We chose the color gradient to show that the darker color is associated with the higher population count of the block group. Note that the lightest pink area on the very right is the ocean. The big vast areas on the left and the bottom are not part of the county, therefore it is not divided into block groups. Also, some white spaces are farmlands. In the coastline there is a slit, which is an island off of Miami connected by the bridge. The coastal area is the tourist area, where Miami is very famous for its beach. North east of the map is the city center. As we move towards the south and the west we move into Miami suburb, where most people reside. The city center is not a large part of the city because of the unique characteristic that is Miami. The vast area on the left is not part of Miami

Metropolitan Area. Miami is an amalgamation of farm, tourist area, downtown, and suburb. Overall, when we see the map as a whole we do see that the most population is in the suburbs.

Analysis of Hispanic Origin

We started our data analysis by looking into the population distribution by race. We notice that the population of Hispanic and Latino origin (Hispanic) is set aside from the general race category.

First, we look at the Miami Population by Hispanic Origin in 1a through a pie chart. We used the pie chart to easily differentiate the difference between each region of the pie. We could have used a histogram of hispanic and non hispanic in Miami. This would have given a better absolute information of population. However, it may be difficult to visualize the difference. From the plot, we notice there is a large percentage Hispanic population. This is not surprising due to Miami's history and geographic location. In 1566, Pedro Menéndez de Avilés from Spain led the first European group to visit Miami. They claimed the Miami area for Spain and established Spanish mission one year later. Moreover, Florida is located very close to Latin America geographically. We can see a further breakdown of Hispanic population in Figure 1b, in which almost 90% are White. Also, the survey questions for hispanic population and race were separate. Hispanic population wasn't measured as part of the race variable. It was measured indepently from the other races. The survey measures the hispanic origin of the people.

After some calculation, we know that in 2010, there is only 15% Non-Hispanic White population. It decreased by 75%, compared to the 90% in 1960. This decrease could be associated with the Mariel boatlift in 1980. It is a mass immigration of 150,000 Cubans to the U.S from the Mariel Harbor, located at the South of Florida. After this, many middle class Non-Hispanic Whites left Miami due to the increase in the minority population. This is referred as the term "white flight".(Clary)

Although there is a large percentage of Hispanic population, we decided to not include Hispanic in the race category as we move forth with our data analysis. The main reason is that US census has some variables (such as tenure of the household) missing for Hispanic population. Also, when we examined the hispanic population and broke it down into different race, we saw that the numbers and information overlapped with the whole Miami Population distribution. We acknowledge that we might be masking certain features by not looking into the people of Hispanic origin. But for consistency, we decide to focus on the race categories White, Black, Asian, and Other.

Finally, we have the total population distribution by race in 1c. The majority is White (73%), followed by Black (19%). There is a small Asian population and other population (a variable combining

American Indian, Native Hawaiian, other race alone, and 2 or more races).

We used pie charts for Figure 1 a,b,c to show the percentage distribution by racial categories. The size of the area indicates the percentage of the category. One may argue that rose diagram plot is a better display since it is easier to compare the categories by comparing the length of the radii. However, due to the huge difference between each category for this specific set of data, rose diagram plot does not have as great an advantage.

Overview of the Race Information in Miami

In order to better visualize the population breakdown by race, and how the different race groups' size in Miami change overtime we used the barplot with different categories. The bar plot (Figure 2 a,b) seems to be the most logical choice of visualization because we are able to clearly distinguish between groups. Also, we can see which group is the majority in the city of Miami. However, from the bar plot of population we cannot discern exact proportional information of the change in the overall population. However, we see that overall there is an increase in population from 2000 to 2010 in Miami in absolute terms. Miami is a growing city. When we examine the overall population distribution by race in proportional terms, we do see that white population is the majority, then it is followed by the Black population, and Asian. This trend is still true in 2010. We used the proportional values for the examination of individual race because we wanted to see if the population increase for each race is due to the increase in the overall population or if there was an actual increase in each of the groups. There is a greatest increase in the white population out of the four race groups. This means that size wise the white population has grown in number not just because the population has increased but because there are more white people who came into the city than other race. Black population shows a proportional drop. The increase in black population is just due to the population increasing from year 2000 to 2010. The biggest drop is from the other groups, which include the other minority groups. The population of these groups are not increasing relative to the other groups.

After examining the overall population distribution, it would be best to examine the block group information of race (Figure 3 a,b), as this would be a useful information on where each race prefers to live. We examined different bandwidth to see which bandwidth value will give the most interpretable information. The conclusion was that the default bandwidth gave the most useful information. The larger the block group means that the location is closer to the city center or larger in size, thus more densely populated. We have, through this density plot examined the distribution change over time to see how the block groups that each race reside in differ over time. We figured that density plot was the best plot to use for this variable because it gives a snapshot on where the biggest concentration of people are and which groups of people live in bigger or more densely populated block groups. We are able to obtain a

proportional information and an overview. However, if one were to know more about the exact number of block groups or any concrete numbers, this plot is unable to give that information. However, this plot is sufficient and covers all the information that we intend to extract. From the density plot we see that white population tends to live in smaller block groups. This probably means that the white population is concentrated in the suburbs. The black race lives in bigger and more populated block groups. This may indicate that black people are more concentrated in the city center, then the suburbs. The asian population has the similar trend as the white people except we see that the concentration is more towards the higher block group than that of the white population. The trend continues for the 2010 density plot. However, overall, we see a shift towards the right, which indicates bigger block groups. This increase may be due to two possible reasons, which will be better highlighted in the choropleth map. First reason may be that the block group that each race groups are dwelling in is getting bigger as the population increases. This will indicate the city getting bigger everywhere. The second reason may be that we are seeing a migration of each group towards bigger and more populated block groups. This will mean that people are tending towards the city center. This plot is there to give a brief insight into what we will cover in detail with a choropleth map.

Population and Race Characteristics

Figure 5.a is the linear regression plot between income and population of each block group. This graph shows a negative slope between income and population, suggesting that two variables are negatively correlated. The actual model is $\text{income} = 38594.80 - 3.84 * \text{population}$. Because the slope is significant, we can say that there is really a relationship between the variables. However, the correlation for these two variables is only -.17. Because the correlation coefficient is close to 0, we can infer that population and age do not have strong linear correlation. Weak correlation can also be inferred from the graph, as there is no clear pattern for positive or negative relationships.

Similar to the previous pair, age and population are not strongly correlated. This is evident in the fact that the correlation coefficient is -0.13. The linear regression is portrayed in Figure 5.b. The linear model is $\text{average age} = -0.002 * \text{population} + 42.68$. One thing to note is that the coefficient for population is significant, thus there is a clear relationship between age and population.

For the previous two graphs, we utilized linear regression models, because we were more interested in obtaining correlation than obtaining distribution. Linear regression models allow us to quantify the associations between the variables. However, linear regression models are dependent on linear model assumptions such as normality of residuals. Moreover, they do not explicitly show what the coefficients are unless we use a summary method. Heat maps and contour plots are better used for displaying distributions than correlations. Perspective plots also show density and distribution information, but it is

harder to interpret and visualize the relationship between the variables, because they are mostly three-dimensional.

Figure 6 displays a pairs plot between age, income and percentages of race populations. The pairs plot is a group of scatterplots with correlation coefficients in the bottom half of the graph. Although we did not explicitly go over this concept, we did learn about scatterplots and correlations, and this was the cleanest way to display correlations between multiple variables. We used the percentages instead of race population, because it allows us to isolate the correlations between race and variables of interest. Age is only positively correlated with white percentage, suggesting that block groups with higher white percentage generally have higher average age. Block groups with relatively high population percentage of black and other races seem to have lower average age.

Income shows a bit more interesting characteristic, as correlations between income and race population are positive for white and Asians, and negative for black and other races. This suggests that block groups with relatively high Asian and white population percentage seem to have higher average income, while block groups with relatively high black percentage and population percentage of other races seem to have lower average income.

We utilized a pairs plot, because it is simple and easier to interpret correlation between variables. Although it might be prone to fitting too much information in one graph, it is the best graph to portray correlation coefficients, scatterplots, and smoothers with the minimum number of graphs. Instead of the pairs plot, multiple heat maps, contour plots, or perspective plots could have been used. They would solve the problem of fitting too much information in one graph, but they do not portray quantitative information about the correlation coefficients. Also, they are harder to visualize relationships, as readers have to switch back and forth between graphs to really grasp the relationships between the variables.

Geographic Income Distribution and its Relation to Races

For all of Figure 7 plots, we used choropleth map in order to visualize the data distribution by its location. Figure 7 a is a choropleth map for total population count per blockgroup. Figure 7 b is a choropleth map for total income in dollars per blockgroup. Figure 7 c-f are population proportion per blockgroup for each race (white, black, asian, others). Also, for Figure 7 c-f, we decided to use proportion instead of count because proportion would be more legitimate to explain how racial population contributes to the blockgroup's income. The color gradient shows that the darker color is associated with higher numerical value. For example, if the region has darker color, there is a higher population count. The same goes for income and proportion.

The reason why we chose choropleth map is because we wanted to make connection between income and population data with geographic distribution. Then we can get some interesting explanation on people's preference of location based on their income and race. We also anticipated on drawing sociological information based on the landscape. Another advantage of this plot is that anomalies can easily be identified. A disadvantage of this plot is that we could not plot both income and race information in one plot. There is no way to put them together without making them look too cluttered. An alternate plot of choice could be boxplot. With such choice, we can use both variables, income and race, in one plot. However, boxplot does not convey any information about geography.

Although stated in the introduction, here is a brief observation from total population choropleth map (Figure 7 a). Downtown is located in the north and does not have the highest population. The population is higher in the suburbs and towards the inland. The lightest pink area that is stretched on the very right is the ocean. Some inland spaces that are in lightest pink are farmlands. The big vast areas on the left and the bottom are not part of the county, therefore it is not divided into block groups.

According to our research, there is a supporting sociological fact on why lower population is associated with higher income. The small peninsula on the right is an area called Key Biscayne. Key Biscayne is an area where multi million dollar homes are owned by Fortune 500 Executives and celebrities. Interestingly enough, the total population and income choropleth maps (Figure 7 a, b) show that the area is not densely populated (12,344 in 2010 Census), but scored \$1,054,213 per capita income. Similar trend is exhibited at the shores as higher income is distributed towards the shores where there are more expensive houses.

Another interesting observation is a comparison between white and black population. A strong sign of negative correlation (-0.99) between the two races is also prominent in white and black population choropleth maps (Figure 7 c, d). Blockgroups of higher percentage of white people are located in the suburbs and the shores. This could be perhaps due to white flight* and the association of white race and higher income. On the other hand, blockgroups of higher percentage of black people are located in the downtown area. This distribution supports the association of black race and lower income.

Although the population of Asian and other races are very low, the choropleth maps (Figure 7 e, f) show some trend. Asian population resemble white population in the way that they tend to prefer living in the suburbs and places of relatively higher income. Population of other races resemble black population and is more likely to be distributed in downtown and other places of relatively lower income.

*Some working-class and middle-class white families moved out to the suburbs because they felt pressure from increases in minority populations and overcrowding in cities.

Income and Race Relationship Through Housing Information

The possession of houses could be an indication of income and values (whether they are more family/home-oriented). Therefore, we want to see how it varies by race.

We use mosaic plot to examine the relationship between two categorical variables Tenure (Owner, Renter) and Race (White, Black, Asian, Other) in Figure 8a and the relationship between Mortgage Payment (Owned free and clear, Owned with Mortgage or Loan) and Race (White, Black, Asian, Other) in Figure 8b. The width gives the marginal distribution of Race and the height gives the conditional distribution of Tenure and Mortgage Payment given race in Figure 8a and Figure 8b, respectively. We could have used a rose diagram as an alternative plot. This would have made the plot simple and we can compare the size difference of each sector as the rose petal sizes differ. However, the mosaic plot is able to give more information of such as expected and actual value information and enable comparison side by side. The box shaded in blue means the value of a particular (Tenure/Mortgage, Race) is greater than the expected value and the box shaded in red means less than expected value.

In both graphs, every box shows substantial deviation from expected counts.

In Figure 8a, the boxes correspond to White and Asian homeowners and the boxes for Black and Other renters have positive deviation. The rest of the boxes have negative deviation, indicated in red. This shows that there are more homeowners and fewer renters for White and Asian than expected by chance. It is the reverse for Black and Other.

In the subsequent mosaic plot, we can see who really owned the house in full. This mosaic plot has the same pattern as the previous one. Every box shows substantial deviation from expected counts. Again, we see more White and Asian homeowners owned the house free and clear and fewer owned the house with mortgage or loan than expected. The reverse is true for Black and Other. Another interesting observation is that the majority of homeowners own the house with mortgage or loan regardless of their races.

One could have chosen association plot to display the data. This would give us a ranking based on standard deviation but the marginal and conditional information would have been lost. A double decker plot would also work but would not contain information regarding standard deviation.

Conclusion

Overall, Miami is a very dynamic and diverse city. It is a growing city with a large metropolitan area.

We saw this from the population difference from 2000 to 2010. It doesn't necessarily have a big city center but the city borders are more spread out. The tourist city that Miami is well known for explains the large metropolitan area. Overall, in terms of race, the city of Miami is very dominated by White population with large shares of black, and asian population. There is a relationship between Age and income, and other variables such as white population. Amongst the different race groups we see that white and asian population shows similar traits in terms of age. From the map information of Miami, we learned that population of higher income reside in suburbs. The population of highest income reside towards the ocean, as it is the most beautiful part of the city. We concluded that the trend speaks louder for white population since Miami Metropolitan Area is consisted of 76% white population. Lastly, the analysis of homeowners/renters and shows that on average White and Asian population in Miami may be the more wealthier people. These groups also tend to be the home owners, who doesn't have debt on their housing. It was very interesting to worth with a city that is so multicultural and is historically very diverse.

Appendix

Bibliography

1. Clary, Mike. "The Melding Americas : Society : Miami: A Laboratory of Social Change : The Migrant Stream from Latin America and the Caribbean, Coupled with 30 Years of White Flight, Has Pushed the City to the Frontier of the Urban Future." *Los Angeles Times*. Los Angeles Times, 27 Sept. 1994. Web. 09 May 2013.

Graphs

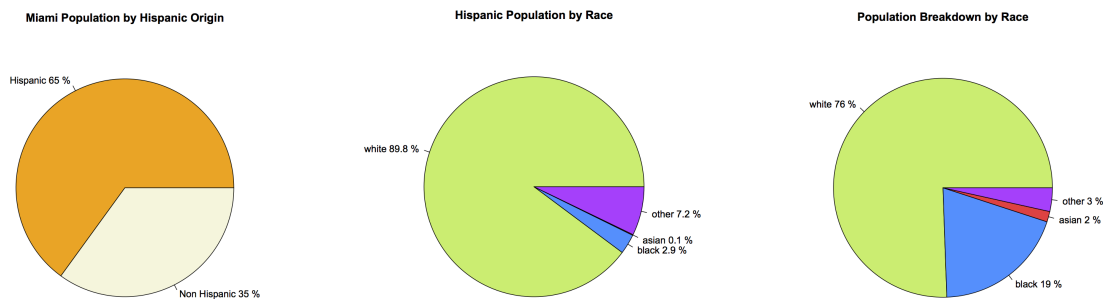


Figure 1 a,b,c

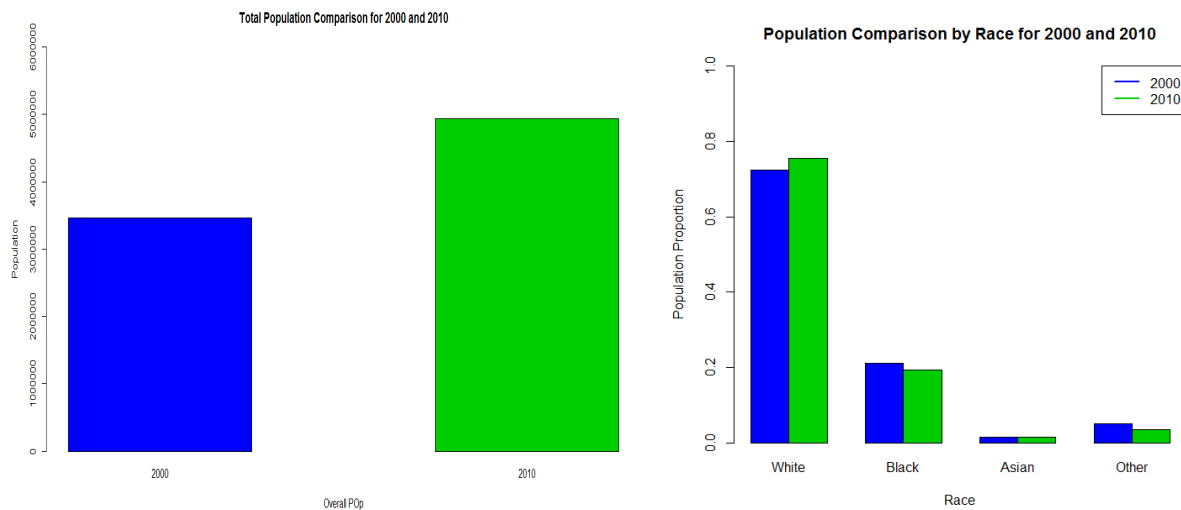


Figure 2 a, b.

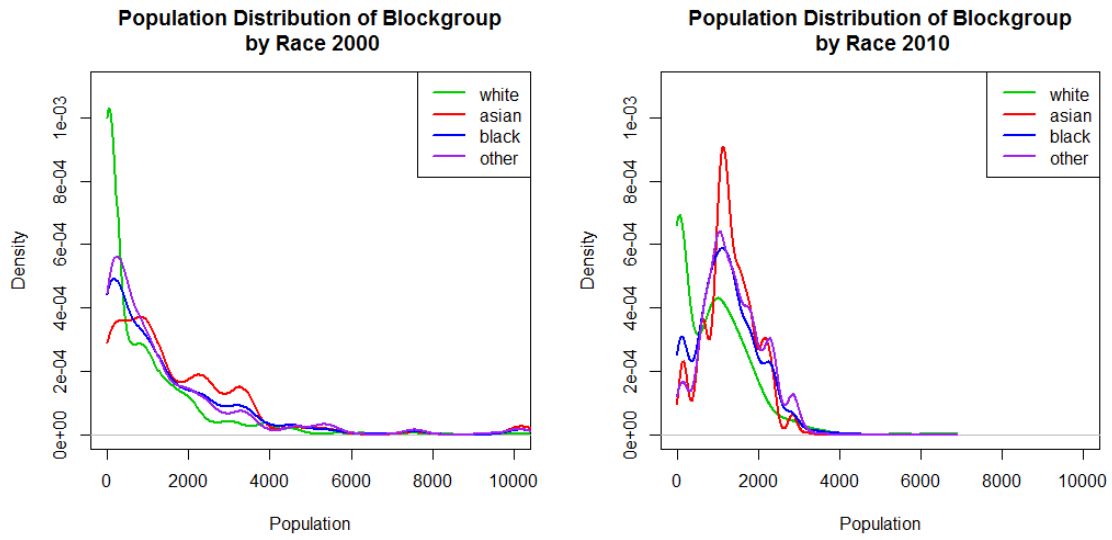


Figure 3 a,b.

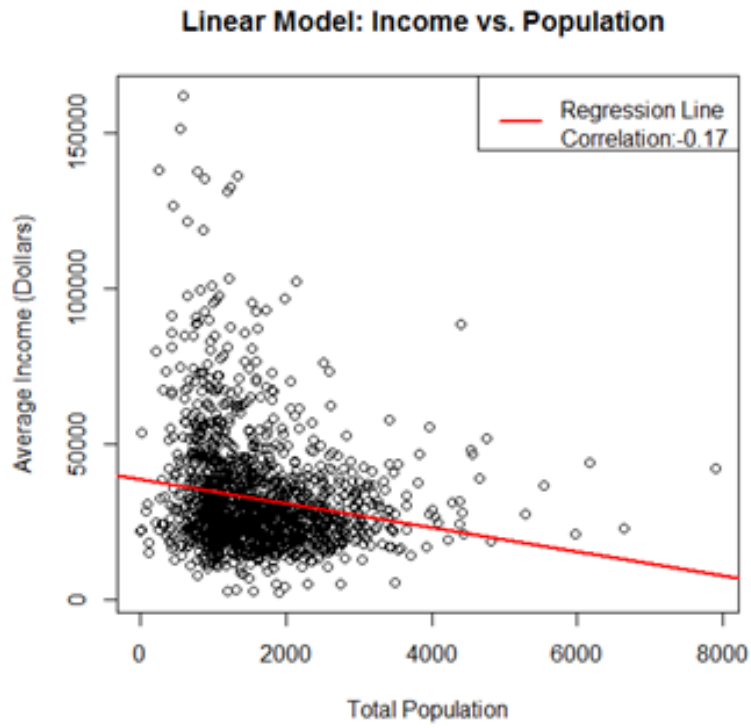


Figure 5 a.

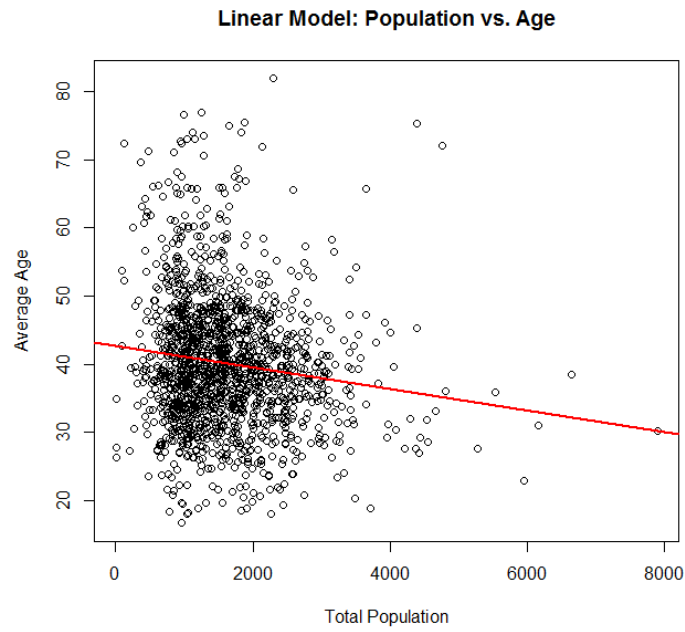


Figure 5 b.

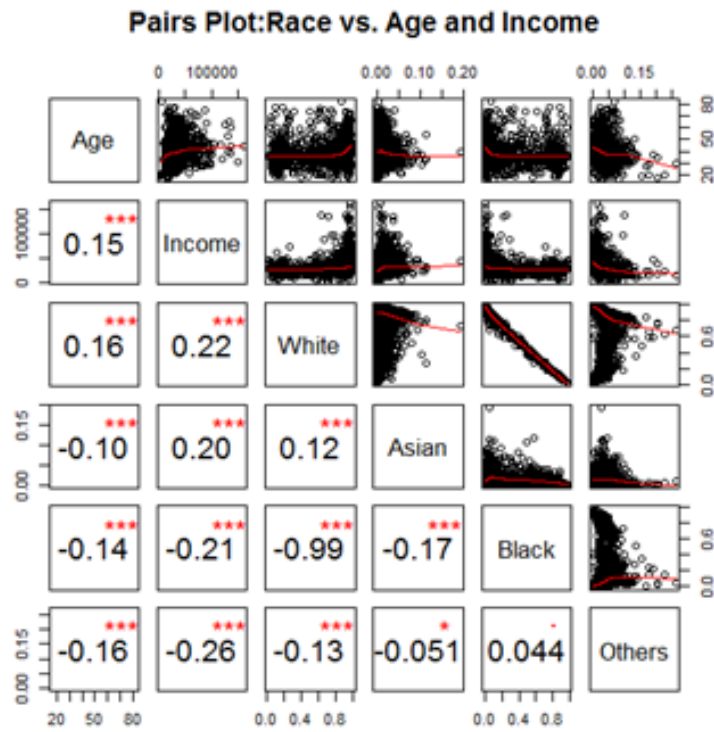


Figure 6.

Total Population Choropleth Map of Miami-Dade County, Florida 2010

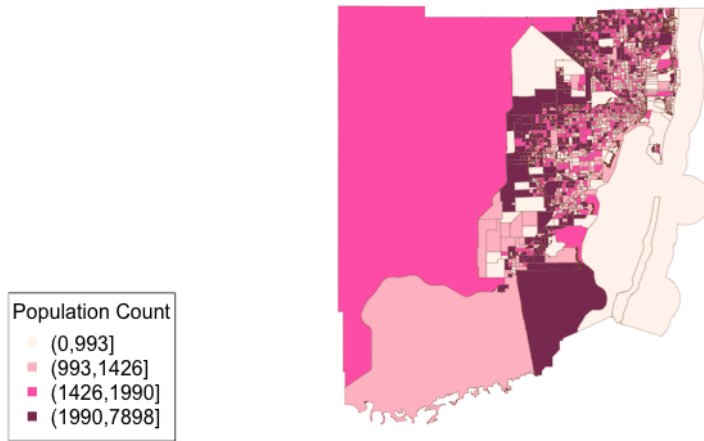


Figure 7 a.



Figure 7 a (Zoom In).

Income Choropleth Map of Miami-Dade County, Florida 2010

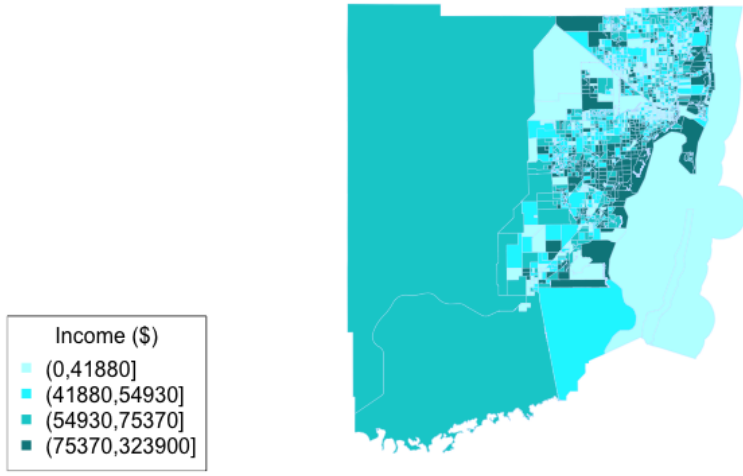


Figure 7 b.

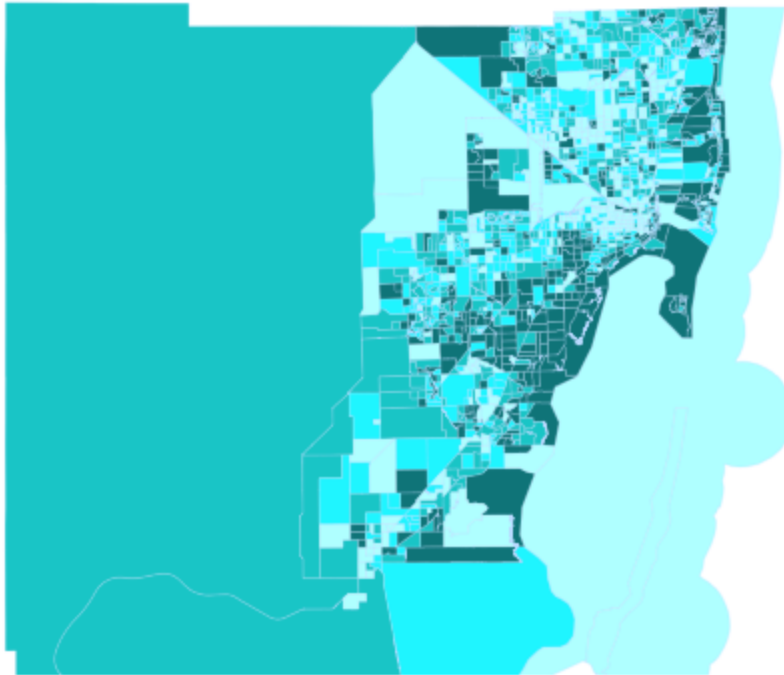


Figure 7 b (Zoom In).

**Population Proportion Choropleth Map of Miami-Dade County, Florida 2010
Race: White Alone**

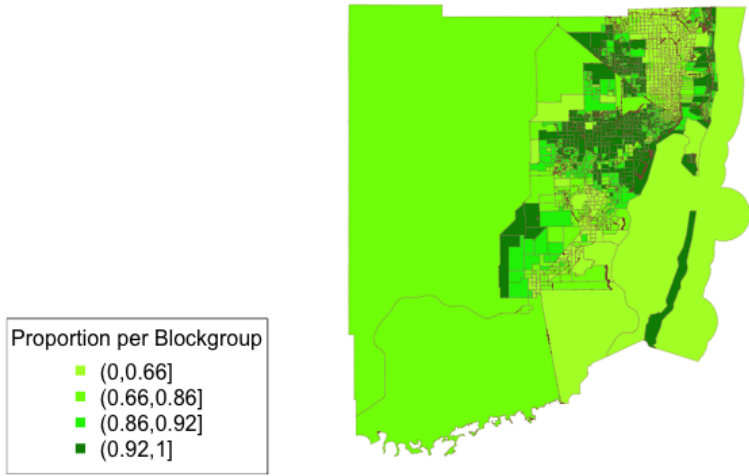


Figure 7 c.



Figure 7 c (Zoom In).

**Population Proportion Choropleth Map of Miami-Dade County, Florida 2010
Race: Black or African American Alone**

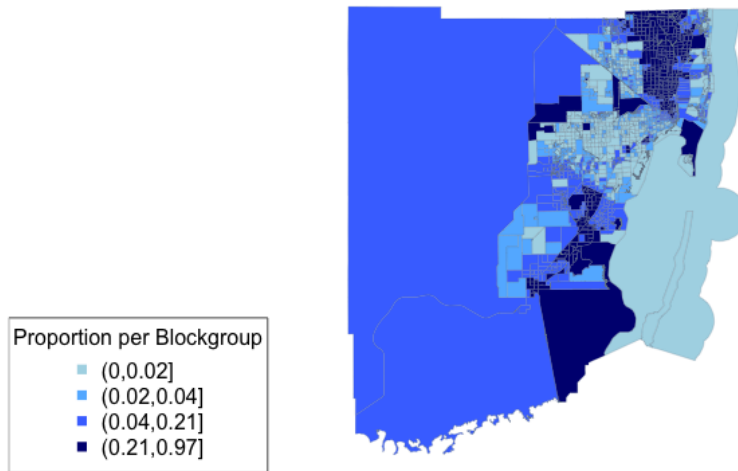


Figure 7 d.

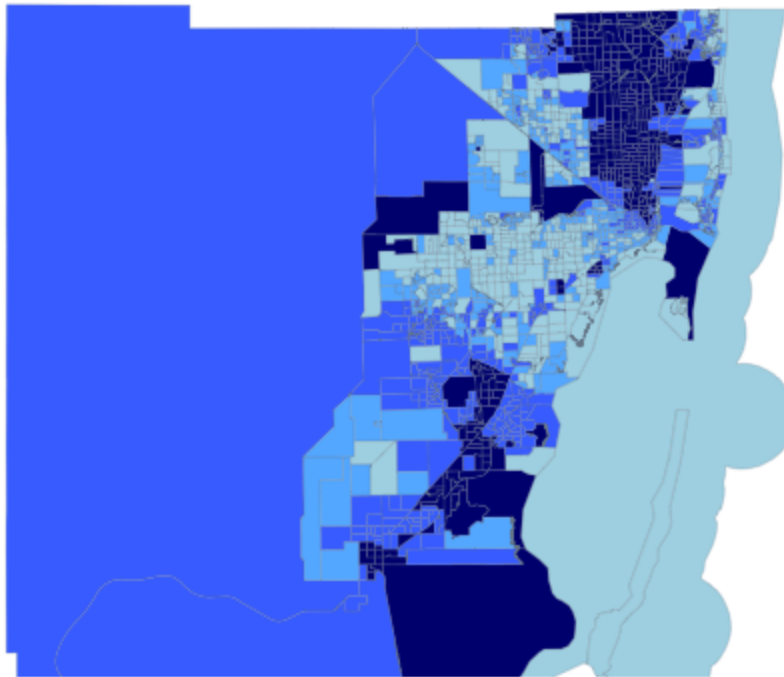


Figure 7 d (Zoom In).

**Population Proportion Choropleth Map of Miami-Dade County, Florida 2010
Race: Asian Alone**

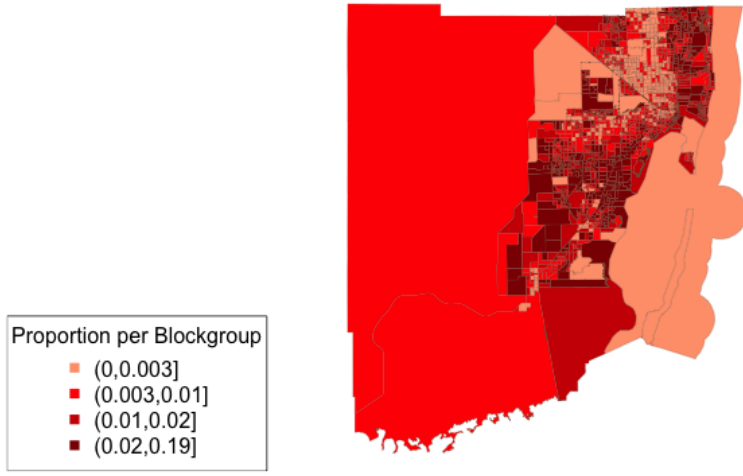
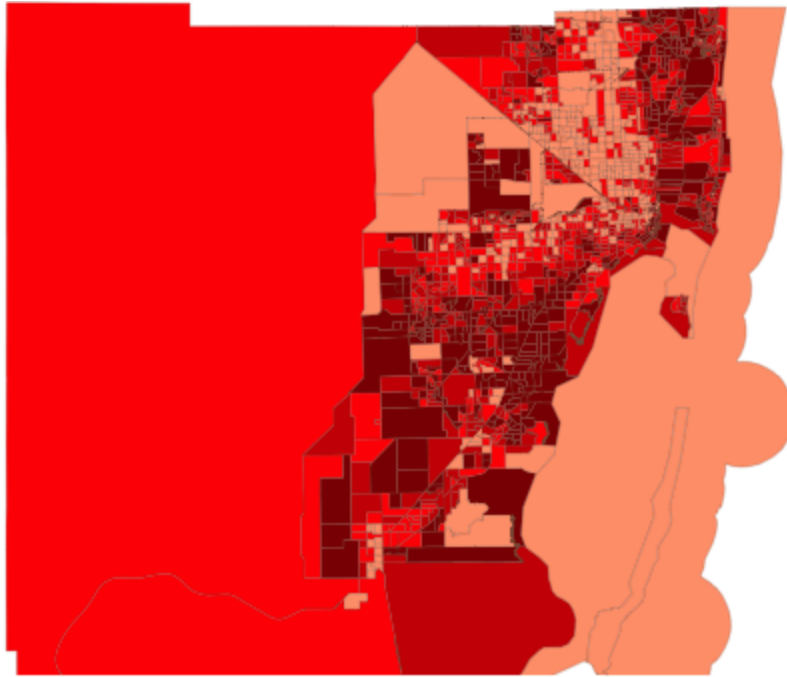


Figure 7 e.



**Population Proportion Choropleth Map of Miami-Dade County, Florida 2010
Race: Others**

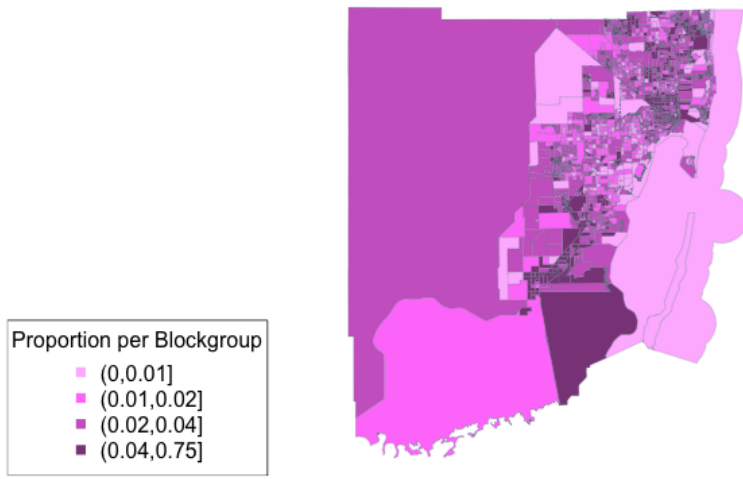


Figure 7 f.



Figure 7 f (Zoom In).

Relationship Between Tenure of a Housing Unit and Race

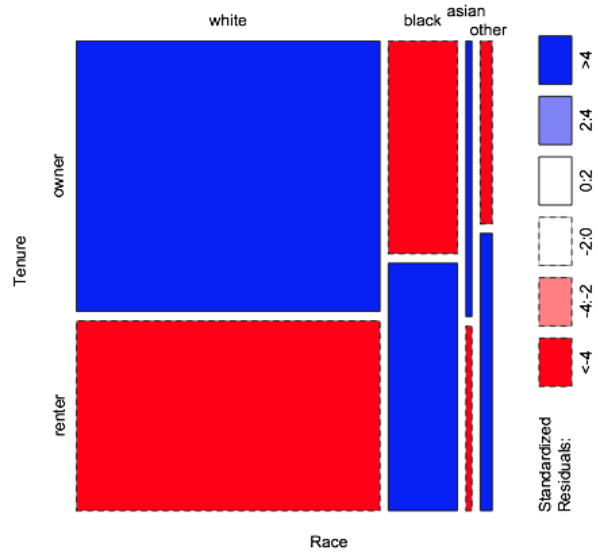


Figure 8 a.

Relationship between Mortgage Payment and Race

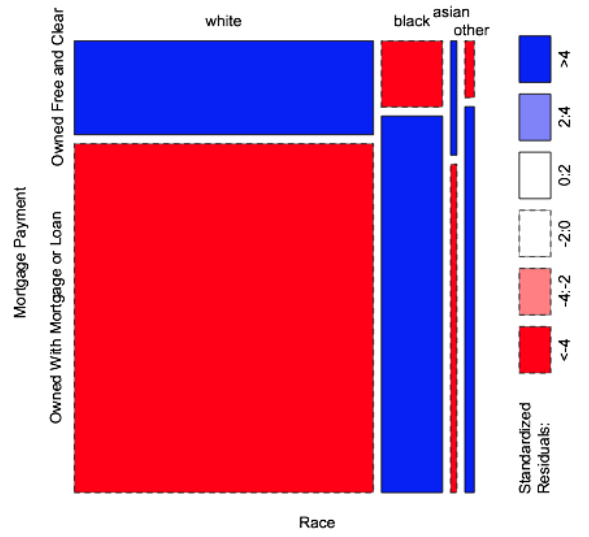


Figure 8 b.