

Demographics of Atlanta, Georgia:

A Visual Analysis of the 2000 and 2010 Census Data

36-315 Final Project

Rachel Cohen, Kathryn McKeough, Minnar Xie & David Zimmerman

Ethnicities of Atlanta

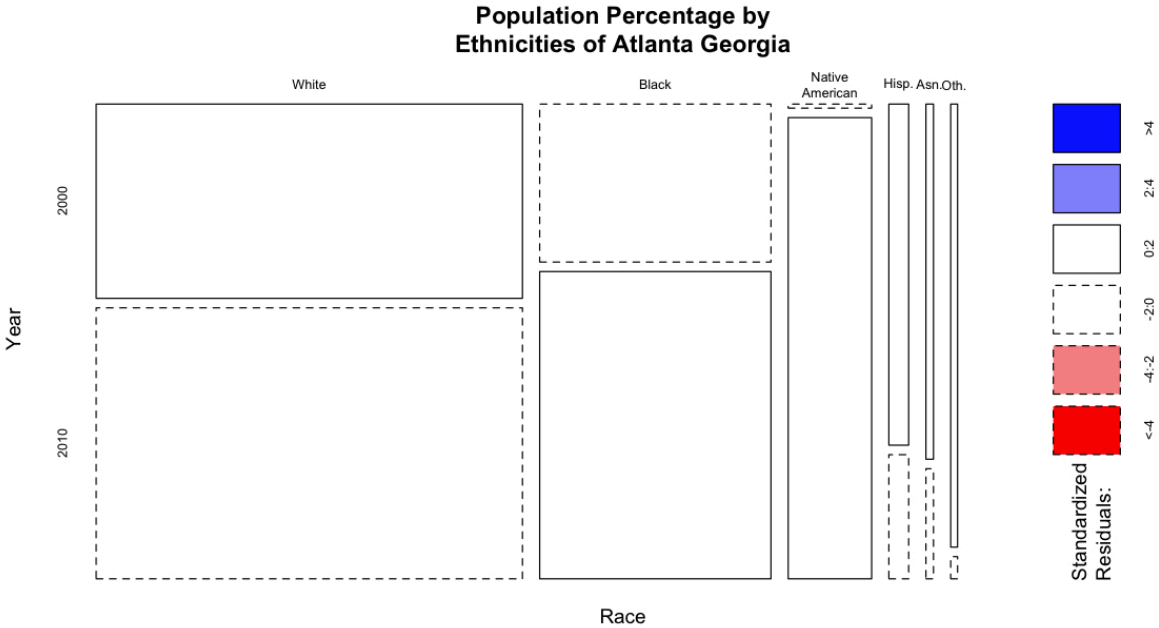


Figure 1: From 2000 to 2010, the white population remained roughly the same; the black population slightly increased. The percentage of Native Americans increased dramatically; Hispanic and Asian populations dropped significantly.

The mosaic plot shows a comparison of the populations of different ethnicities between 2000 and 2010. It shows information about ethnicity proportions conditioned on which year the census is taken. We may compare across ethnicities as well as compare the changes between years.

The width of each bar corresponds to that ethnicity’s proportion in the overall population; the white group is the dominant group, followed by blacks, which also represent a significant chunk of the population. Native Americans are a smaller but significant minority, followed by the much smaller Hispanic and Asian groups. The color scheme corresponds to the standardized residuals that indicate how different each conditioned group is.

In terms of the changes between years, the white and black populations appear relatively unchanged between 2000 and 2010. By contrast, the Native American population experienced more of an increase, whereas Hispanic and Asian populations dropped more.

The advantage of the mosaic plot is that we can condense a lot of information into one graph: This graph presents a proportional division of the entire population by ethnicity and shows the changes in population between years within those divisions. The disadvantage to this plot is that it is more difficult to directly compare across ethnicities, and since change is also measured in proportion to race, it is difficult to tell exactly by how much a population may have changed.

Minorities of Atlanta

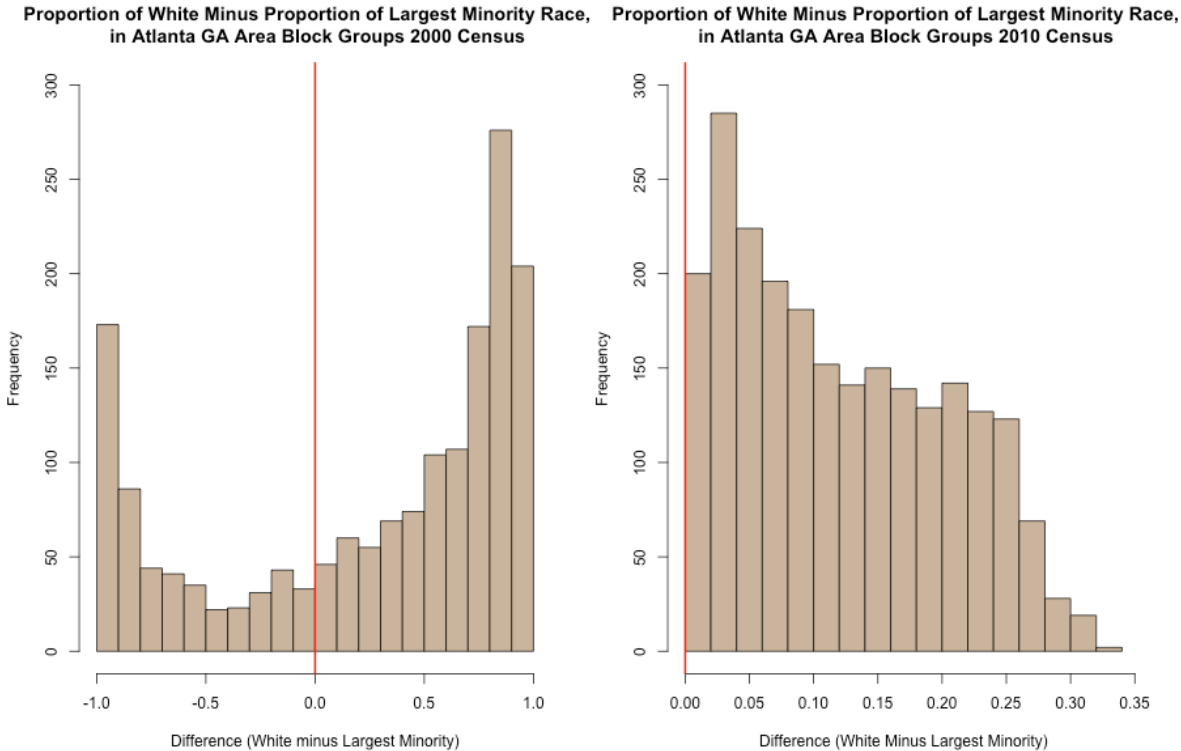


Figure 2: In 2000, some minority block groups were larger than white block groups; in 2010, the white block groups were larger than any minority block group.

We used two frequency histograms to illustrate the difference in proportion between the white population and the largest minority population within each block group in 2000 compared to 2010. We chose to look at the difference between whites and the largest minority because we were interested in the frequency by which whites were the largest group within individual block groups. One important note: the census questions on the census considers Hispanic and Latino as origins and not races, thus they are not included in the race demographics. The difference in proportion is on the x-axis; any data below zero on the x-axis shows block groups in which whites were not the dominant group, whereas positive data shows block groups that are mostly populated by whites. For 2000 there are a number of block groups that are almost entirely white and a number that have almost no whites at all. Overall this plot suggests that a lot of the block groups are quite segregated because the peaks of the data are at the extremes.

There is quite a reversal for 2010. Here every single block groups has a plurality of whites. It also is not as extreme since at maximum the proportion of whites is only .35 greater than the next most common race. Most of the data falls between zero and .25 difference in proportion. Based on maps of the distribution of whites it would appear that there was a massive dispersal of the white population throughout the Atlanta area, which caused each the whites to be much more homogeneously distributed.

Alternative graphs would include the ASH and a density plot. An ASH graph would have given a much smoother distribution across the proportions. It however greatly increases the number of observations in each bin and would not be an accurate representation of sample size. A density plot would have a very similar effect where it would smooth over the proportions as well but it does not show counts. In this particular plot it is helpful to get some sense of scale between the two years because they have the same frequency label.

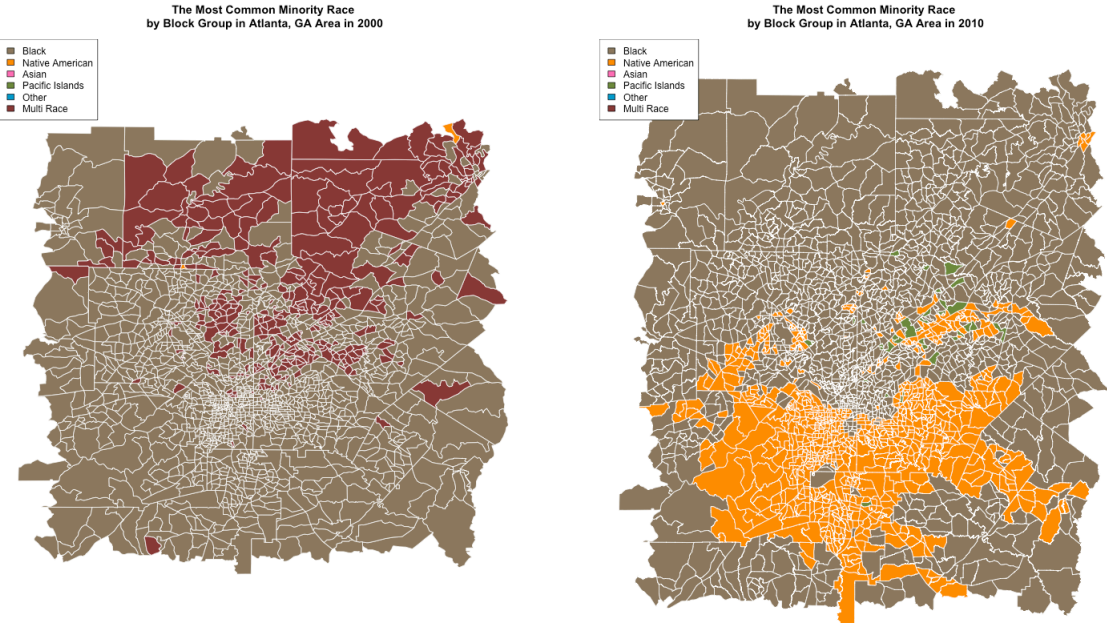


Figure 3: In 2000, the dominant minorities were black and multiracial; in 2010, the dominant minorities were black and Native American.

These maps show the most common minority race in the Atlanta area for each block group. Note again that Hispanic is not included because of structure of the US census. In 2000 black is the most common minority in most of the block groups, which is to be expected in the South. In the northern part of Atlanta multi-race is the most common minority race in many of the block groups. There is only one block group that is not multi-race or black. In this block group Native Americans are the most common minority. Not surprisingly Asian, Pacific Islands, and Other do not appear much in either year.

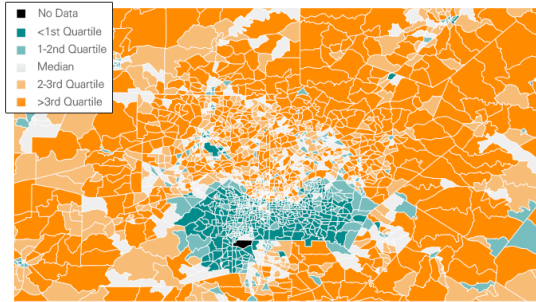
The transition to 2010 is quite surprising. All of the multi-race block groups changed and now the blacks are the most common minority in almost all of the northern block groups. Right outside of the city center extending to the south Native Americans are the most common minority. There is one notable spot where Native Americans are not the largest minority race, just south west of the city center, however this block group contains the Atlanta airport. To the northeast of the city center there are just a few block groups where Pacific islands is the most common. Unfortunately there is not an easy explanation for why Native Americans suddenly all moved the southern Atlanta, but the trend is quite surprising.

Another plot that could show similar information would be a rose diagram or donut chart. Here the area for each slice could correspond to the size of the block group and the color could represent the largest minority race. While this plot might work for the states in the US, the block groups of Atlanta would be more confusing. First it would be difficult to order the block groups in some intuitive manner. Additionally the geographic information would be lost, all that would be shown is the relative size of each block group and which minority was largest.

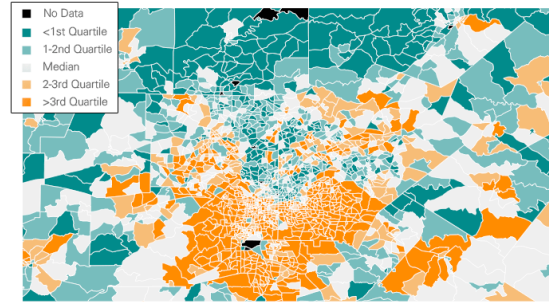
Shift in Population Densities of Whites and Blacks

Shift in Population Densities of Whites and Blacks

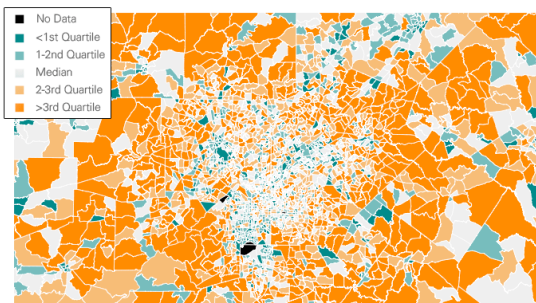
Density of Whites in Atlanta (2000)



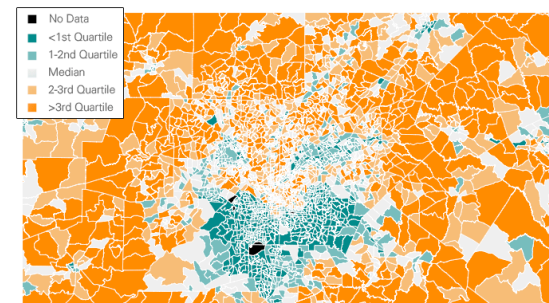
Density of Blacks in Atlanta (2000)



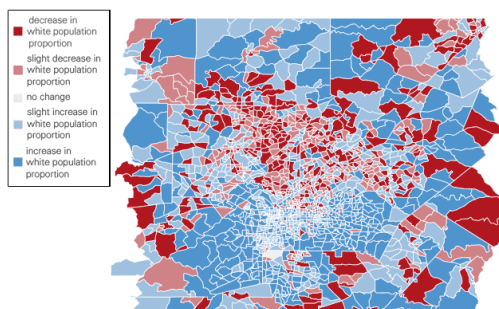
Density of Whites in Atlanta (2010)



Density of Blacks in Atlanta (2010)



Change in Proportion of Whites in Atlanta



Change in Proportion of Blacks in Atlanta

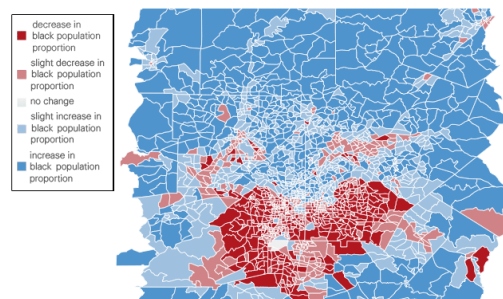


Figure 4: From 2000 to 2010, the black population has moved from the city to surrounding areas; the white population has become more urban, moving into the southern part of the city.

These maps are color-coded to illustrate the shift in population densities of whites and blacks in Atlanta from 2000 to 2010. The upper four maps plot the distribution of the densities of whites and blacks for 2000 and 2010 while the bottom two maps plot the change in the proportion of whites and blacks in Atlanta. To create the maps focusing on change, we wrote code to merge the block group data of 2010 (in which there were 500+ more block groups than the 2000 dataset) into the 2000 block groups so that the change in proportion could be plotted on the 2000 block group divisions.

In looking just at the shift of the white population, we see that in 2000 whites primarily live in the regions surrounding the densest southern portion of Atlanta. In 2010, whites have moved into the densest southern portion of Atlanta, as well as into particular block groups throughout Atlanta. This reflects the ongoing gentrification of particular neighborhoods in Atlanta for the past 10 years, and the movement of whites into particular block groups (as illustrated especially by the bottom map visualizing change) that have been gentrified.

Gentrification of particular neighborhoods in Atlanta is also reflected in the shift of Atlanta's black population. In 2000, the highest density of Atlanta's black population resided in the southern region and a few particular block groups in the central region. However, in 2010, the lowest density of Atlanta's black population resides in the same southern region and central block groups that in 2000 had the highest density of blacks. This stark change shown in the bottommost graph showing change reiterates that the areas that corresponded to a high density of blacks are the also the ones with the largest decrease in the proportion of blacks. When assessing the two graphs illustrating just the change of black and white population proportion in Atlanta, we see that the southern region that decreased very much in the black population proportion also increased a lot in the white population proportion. Again, the influence of gentrification is shown when comparing these plots: as certain neighborhoods become gentrified, whites are more likely to move into the neighborhood causing housing prices to move up, and blacks either move out or blacks moving into the city simply choose not to move into those neighborhoods, effectively decreasing the proportion of blacks in regions where they previously resided.

Map plots have the disadvantage of data being constrained to represent geographic space and not actual population counts. Larger visual space draws more attention, yet a map's largest block groups do not necessarily correspond to the highest population. The data could also be represented in a pie or donut chart, in which the amount of visual space would correspond to the size of the population. However, using a pie or donut chart in the case of block groups of a city is virtually useless because it is not practical to visualize where a block group is within a city.

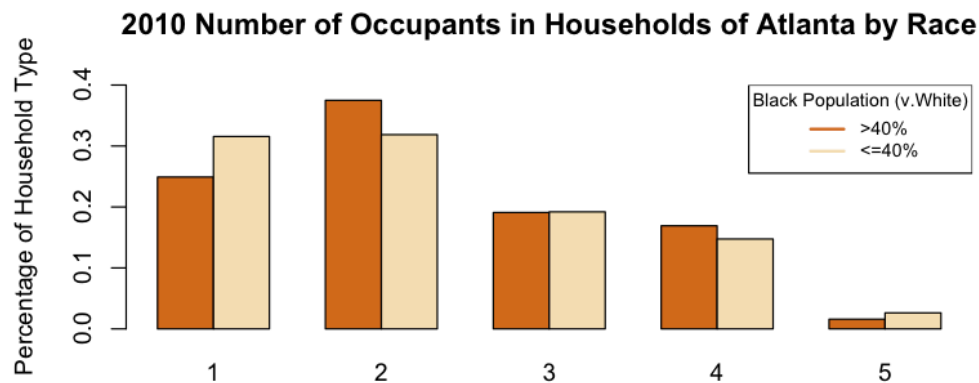
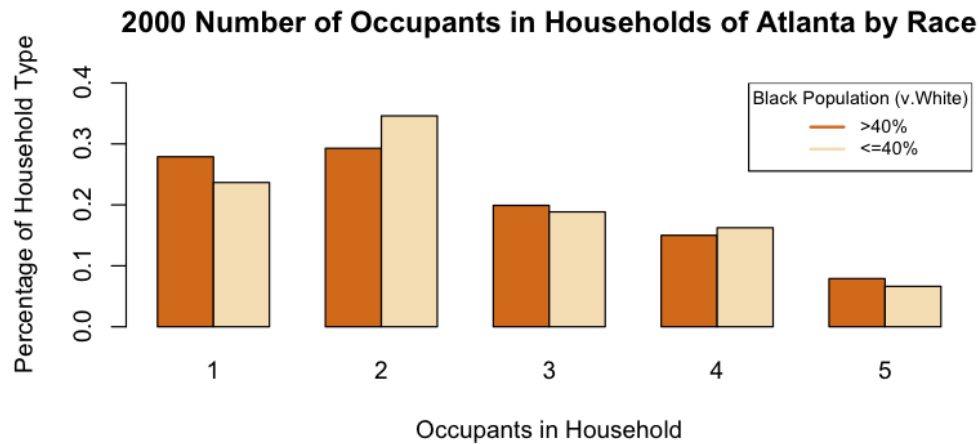


Figure 5: One-person households became more common for whites (25% to 31%); the percentage of black two-person households rose (30% to 38%). Overall the percentage of five-person households decreased, indicating a general shift toward smaller households.

It is evident that there is a large rural/ urban shift in population density between whites and blacks. To further explore changes that may have resulted from this, we looked into how the number of occupants in households may have also changed between the two ethnicities between 2000 and 2010. The block groups were divided into two different categories: those with a black population of greater than 40% compared to the white population and those with a black population less than 40% when compared to just the white population. Originally we wanted to look at block groups with more blacks than whites v. block groups with more whites than blacks; however, in 2010 there were not block groups that held a majority of black population relative to white.

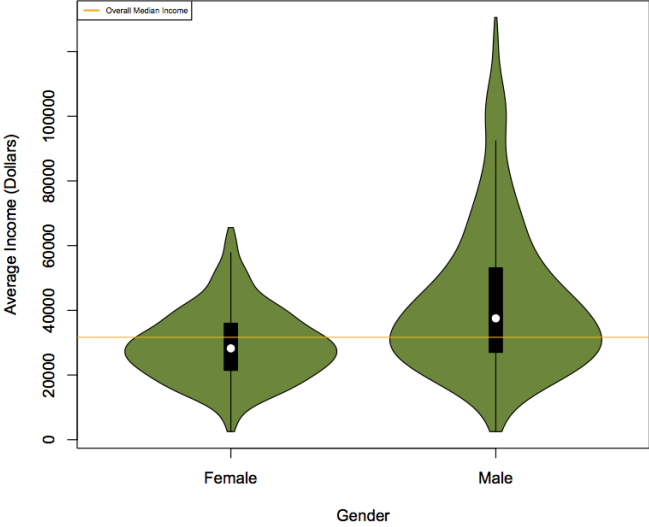
The bar plots show that there was a change between the number of occupants in households when comparing across race and year. The most interesting and significant change is between 1 occupant and 2 occupant households. In 2000, the black block groups had about 30% of each 1 and 2 occupant households. In 2010 the number of 1-occupant households dropped to about 25% and the 2 occupant households rose to about 40%. This is consistent with our hypothesis that there was a shift from urban to rural living for the black population because it is more likely to have 1-occupant households in the downtown of a city than in the suburbs. For the white block groups the percentage of 1-occupant households went from 25% to 30% and 2 occupant households went from 40% to 30% between 2000 and 2010.

Overall, it appears there are the highest percentage of two occupant households followed by one than three occupants. All of the occupant percentages remain relatively the same except for 5 occupant households, which decrease in percentage between 2000 and 2010.

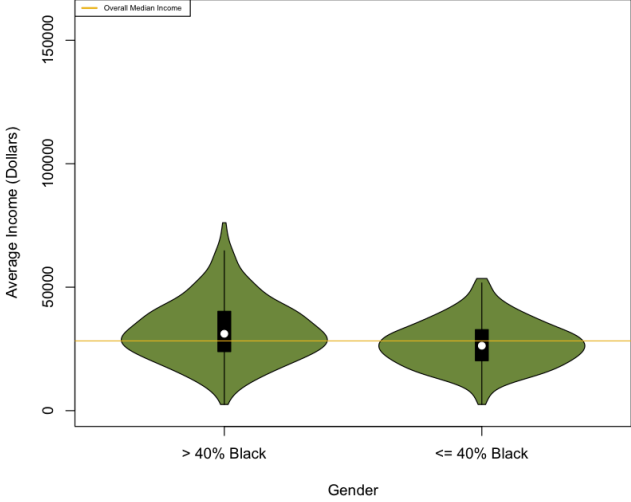
The advantages of using a side-by-side bar plot are that it is easy to compare across two dimensions of categorical data. It made it easy to compare between ethnicity for 5 different classifications of households. A bar plot is better than a pie or spine plot because we are able to compare the heights much easier than the proportions of a circle or rectangle. Quantitative data is easier to display on a bar plot in the form of the heights of the bars. The disadvantages are that is hard to compare between the two years. Because they are on top of each other we cannot compare heights. There is no cohesive way to compare 3 categorical variables with a side-by-side bar plot.

Income v. Gender & Race

Average Annual Income of Atlanta Block Groups of Bottom 99% by Gender



Average Annual Female Income of Atlanta Block Groups of Bottom 99% by Race



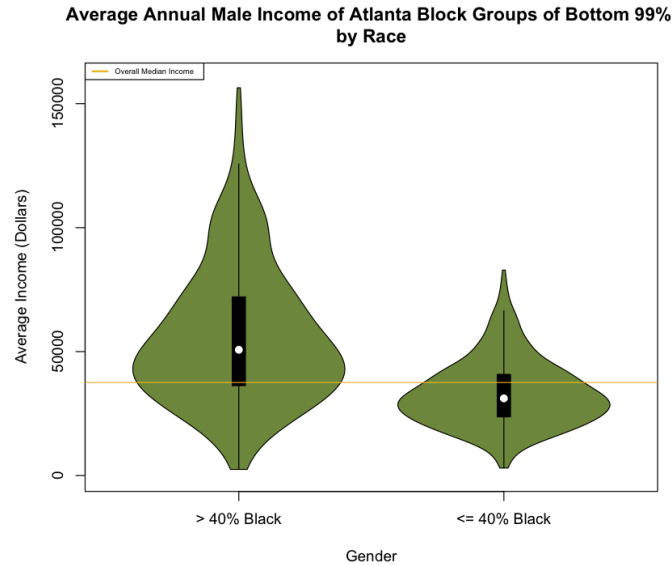


Figure 6: Women make less money than men (medians of ~\$30,000/year versus ~\$40,000 respectively). Black women make slightly more than white women; black men make significantly more than white men (by ~\$10,000/year).

The violin plots express the difference in income distributions across gender and race of the bottom 99% of block groups. The reason we chose to use the bottom 99% was to remove potential outliers. We did not care about the most wealthy block groups, and only wanted to look at the distributions of the commonly referenced 99%. We chose this over taking the log of the data, because it does not skew our data and in this case we do not care to plot the highest incomes. The golden line represents the overall median between the two groups. We chose the default kernel density because it best portrayed the uni-modal data without showing too much noise. The first graph shows the difference in income between males and females in 2010. It is evident that the center of male income is further above the overall median than the female income is below. The male distribution also extends higher than that of females. The next two graphs looked at black and white block groups. We defined a black ‘dominated’ block group by having over %40 of the population be black. The reason for this was that in 2010, the white population was the highest in every block group. For both males and females it is apparent that the income of the black dominant block groups is higher than that of the white dominant block groups. The effect was far less obvious for females than males. Overall the group with the highest median average income was black, male incomes and the lowest was white, female incomes.

There are many plots we may have used in order to show the center and spread of the data. Box and whisker plots are great, but they do not show the density of the data. We would not have been able to see its uni-modal spread with a box and whiskers plot. A density function or side-by-side histogram would have also been an acceptable choice, but two density plots on top of each other do not give us information about the median or quartiles. We wanted to show this in order to compare the two groups. The violin plot was a way to show both the information provided by a boxplot and a density plot.

We also could have used a bean plot or a box-percentile plot. We chose against the bean plot because we wanted to display the median of the data, which a bean plot does not show. A box-

percentile plot, on the other hand, shows density as a function of change in slope, and we wanted to use a graph that directly displayed density.

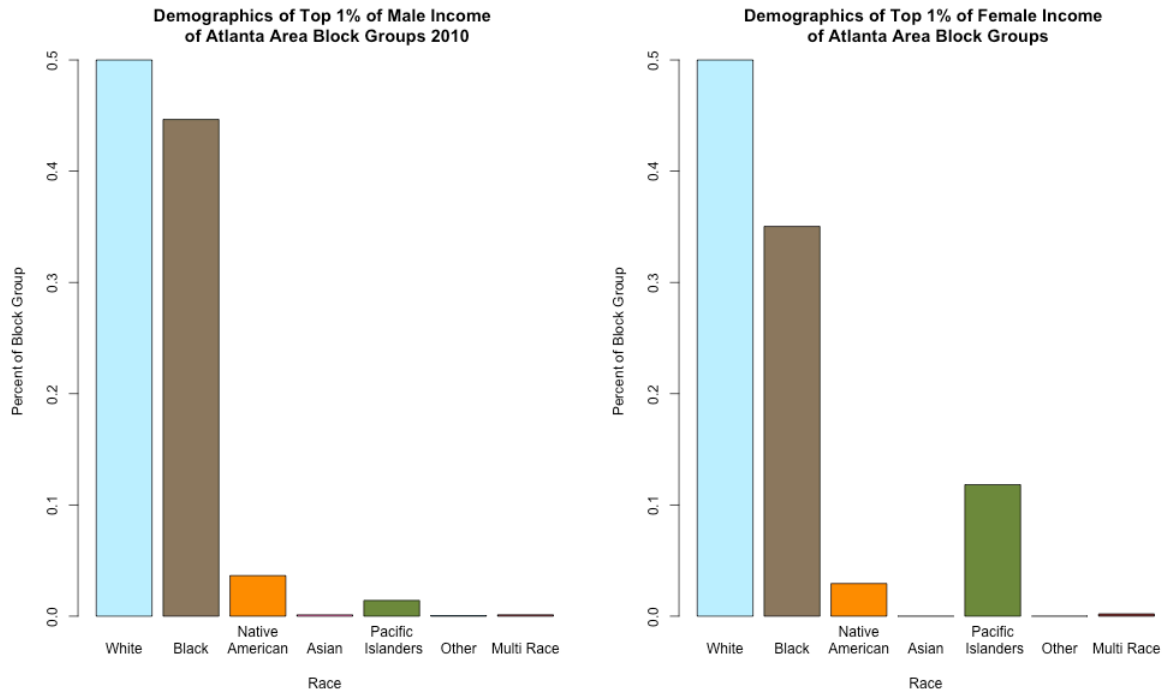


Figure 7: Each plot only considers the top one percent of block group income for males and females. Two block groups had average block group male incomes in the top 1% and one block group had average female in the top 1%, for each respective gender. There is a large shift from black to pacific islander between the groups however, Indians are considered pacific islanders which may contribute to this difference.

Another economic group that is interesting to analyze is the “top one percent”. In this case, the top one percent was classified as the block groups, which were in the top quartile of average income. The bar plots above show the percentage ethnicity populations that were in the top one percent in 2010 Atlanta Georgia. The left side shows the top average male incomes and the right side shows the top average female incomes. In both cases the dominating population was whites. The next demographic was blacks, however there was a higher percent of black males than females. Other demographics with a substantial proportion are Native Americans and Pacific Islanders. The proportion of native Americans remain fairly consistent between the two genders, however there is clearly more of a pacific islander proportion in the top female one percent than the top male one percent. Overall, the purpose of the bar plots is to grasp the ethnic break down of the block groups with the top one percent of average income and compare the distribution between genders.

Bar plots are good ways to compare across categorical variables. The percentages of each ethnicity can be easily compared by the heights of the bars. Since we are looking at percentages of a population, another option may have been a spine or a pie chart. These look cleaner and show the same data, however it would have been harder to compare ethnicity and gender. Also, neither of those plots show exact values without axis-labeling as the bar plot does.

Effects of Age

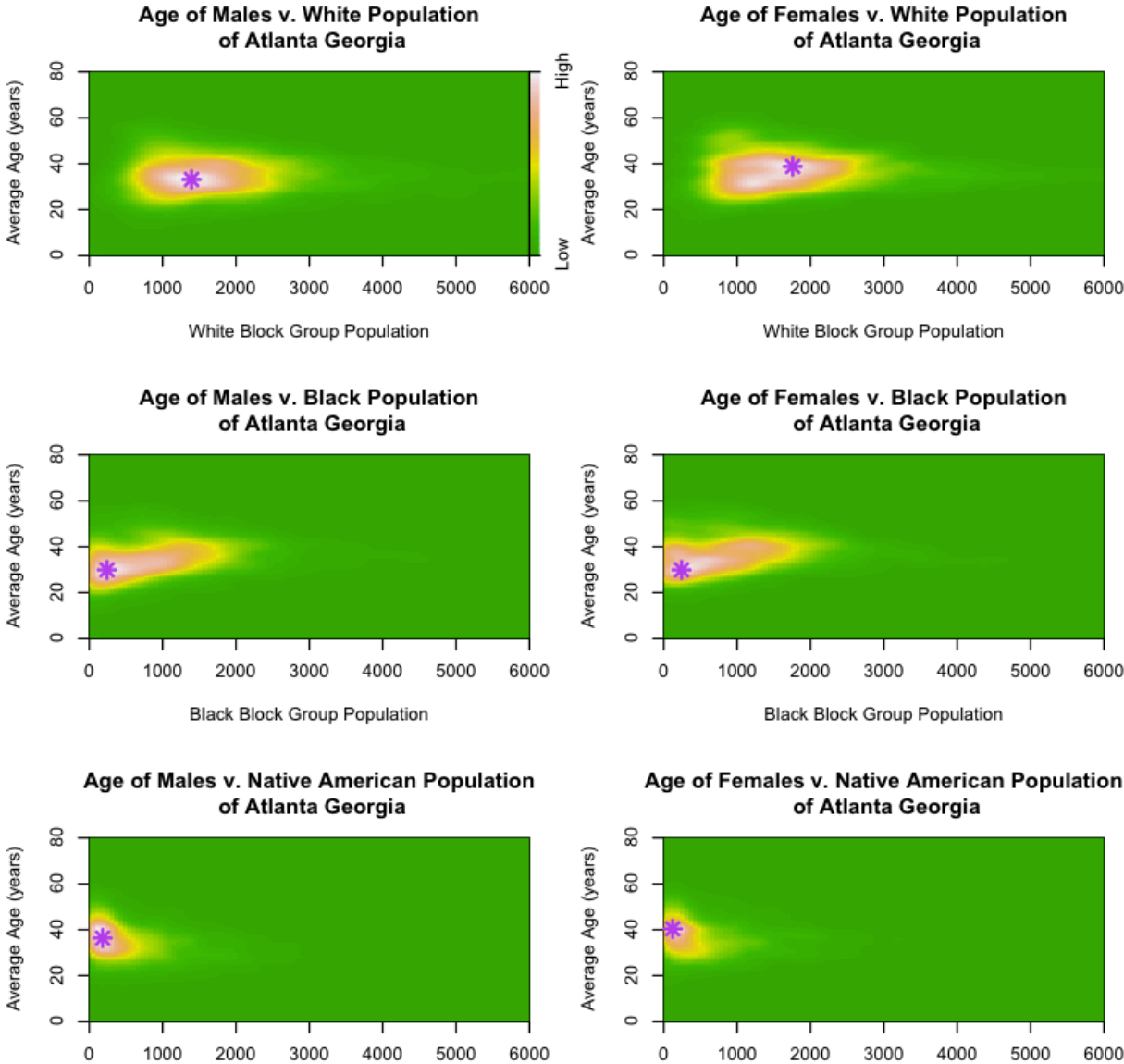


Figure 8: Population and age do not differ by gender. Whites have the highest of both (pop=1500, age=35-40); Native Americans and blacks are fewer (<500) and slightly younger (25-35), except for Native American women, who are older (~40).

We used a series of heat maps in order to explore differences in age and population across gender for the white population — the largest group — and blacks and Native Americans, the two largest minority groups in the population. We plotted population on the x axis and age on the y axis and produced side-by-side maps for males and females so we can compare across both gender and ethnicity. We concluded that population and age do not differ across gender lines, with the exception of Native American women, who are significantly older than Native American men. However, across ethnicity, we found that whites have the largest population, with around 1500 per block group, while blacks and Native Americans are significantly fewer in number, at less than 500 per block group; we also found that age for whites tends to center

around 35-40 years, whereas Native Americans and blacks tend to be younger, averaging around 25-35 years in age.

We could have represented the data in a variety of different ways. We could have used separate graphs to compare age and population by gender, but we felt that plotting a density estimate of the two variables was a more concise way of representing the data. We could have used either a perspective plot or a contour plot to display the density estimate, but we felt that a heat map most clearly showed the gradient in density. A contour plot would have more clearly shown change at different levels, but this information was not necessary for our purposes.

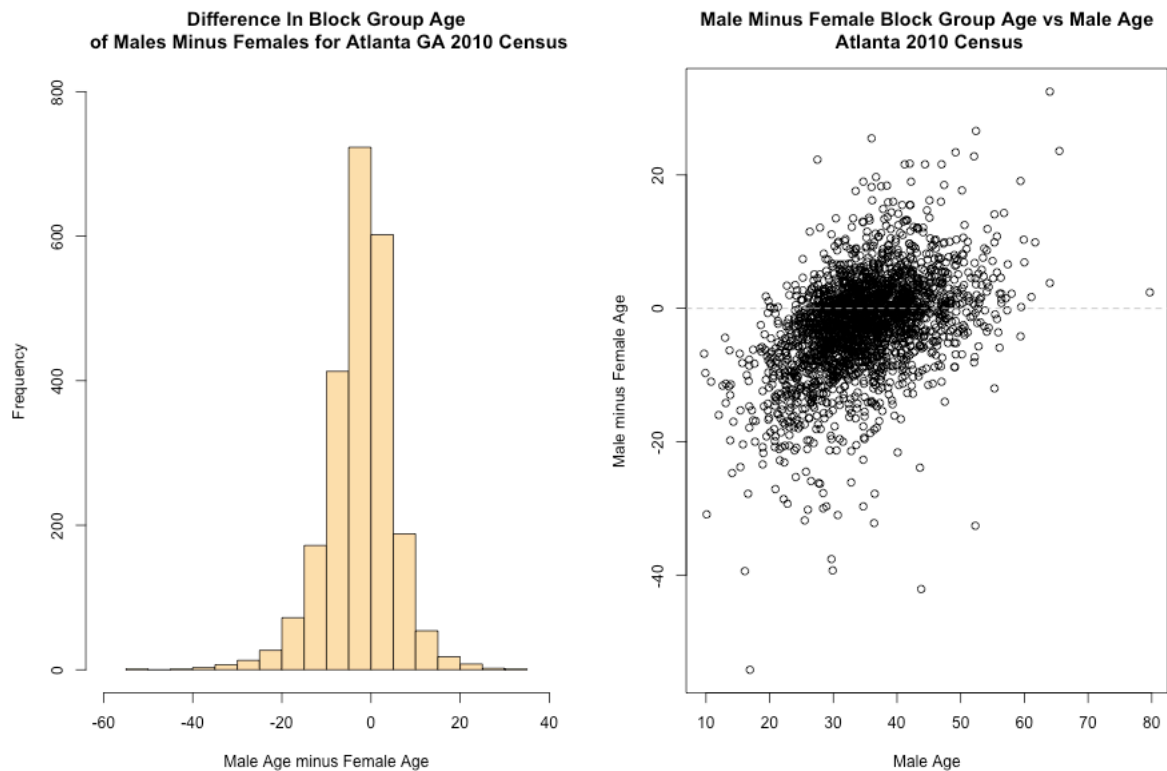


Figure 9: Within the block groups the median age difference is 1.6 years, with women being older.

The above plots show different aspects of the age gap between men and women within block groups. It is well known that women live longer than men and this plot confirms this. It is clear that the histogram is skewed to the left because female ages are higher than male ages. An unusual find is that the average age in a number of block groups differs by more than 20 years. A very intuitive explanation would involve single mothers with young children living in the same areas. However, since male and female children are born at very similar rates and single mothers tend to have children younger, it makes less sense that there would be a 20-year age difference.

In the scatter plot we looked at the difference in average block group age against the average male age in a block group. The gray dotted line shows where male and female ages would be equal. The most interesting trend is that as male age increases it appear that the difference in

block group ages decreases and eventually flips directions. When males are younger females tend to be much older than males. This would suggest that some block groups have high concentrations of younger males, ages 20-30, and females that are about 15 years older. When the average age in a block group for males is 35-40 a lot of the differences in average age attenuate to zero. Above 45 males appear to have a higher average age than females in block groups.

A density plot or a strip chart could replace the histogram. In this case the number of observations would overwhelm a strip chart, even though it would show the actual data points. A density plot would not show the actual counts for each age difference. The scatter plot could be replaced with a perspective plot, a contour plot, or a heat-map. In all three of these the densities would be shown instead of the actual data points. In this case there are a number of extreme points that would be covered up by the smoothing and thus be lost. The advantage would be that the large cluster of points would be more meaningful because you could see how density within the cluster changes.