36-315 Final Report

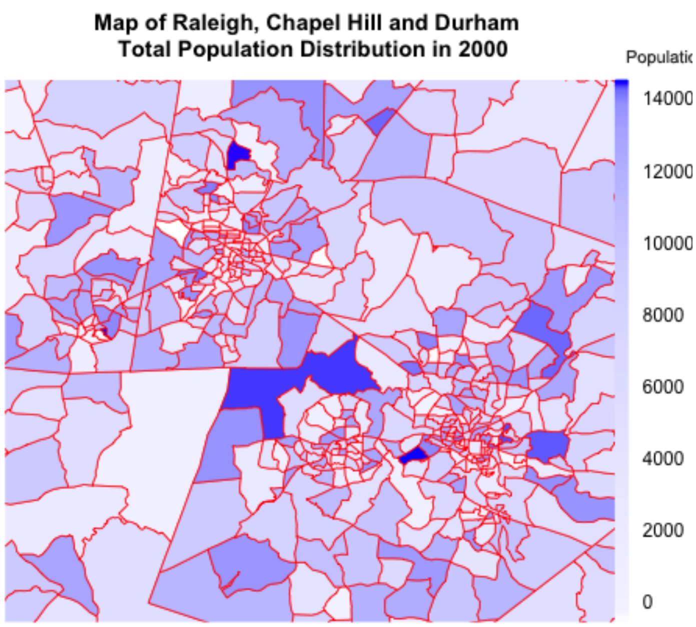
Group 3
Nicole Anderson, Kyongche Kang, Henry Liu, Angie Ng, Stephy Wong

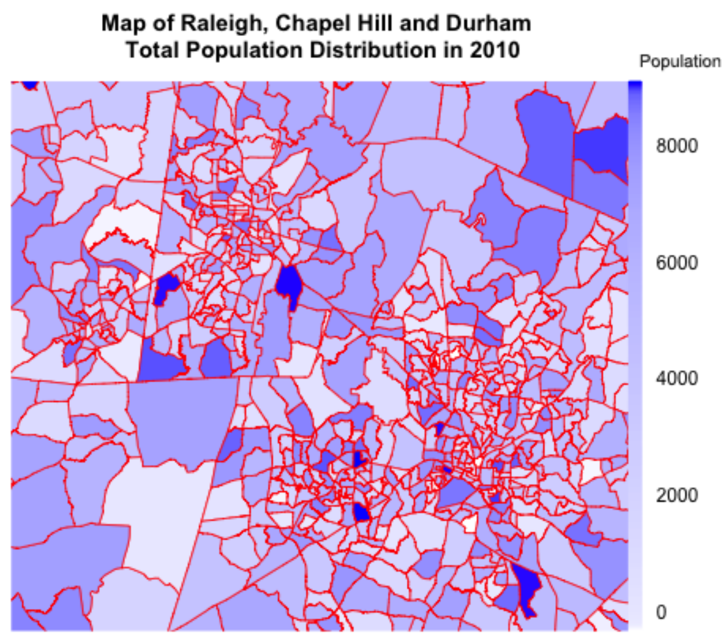
Introduction

Our group analyzes the population distribution by race and how they are related with characteristics like income and age of Durham, Raleigh and Chapel HIll in North Carolina. We are interested in population distribution of White, Black and Hispanic groups, as they constitute majority of residents in three cities of North Carolina.

Change in total population distribution from 2000 to 2010

Image 1: Choropleth Map of Raleigh, Chapel Hill and Durham on Total Population Distribution in 2000 and 2010



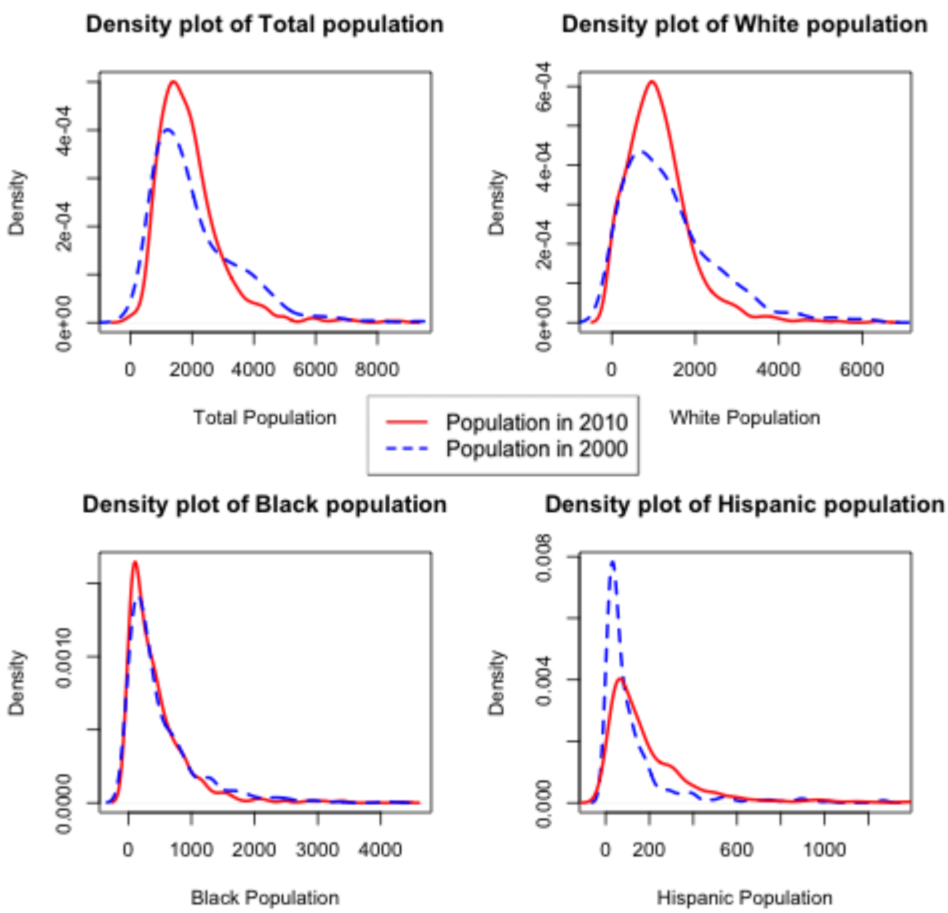


We chose to use choropleth map to illustrate the change in total population from 2000 to 2010. The advantage of choropleth maps is that it shows the movement of population with respect to geographical settings. It is easy to see the all the blockgroups and the borders. We can see and compare the population of each blockgroup and the geographic location of the blockgroup. The disadvantage is that some of the blockgroups changed from sometime between

2000 and 2010 so the maps do not look exactly the same and it may be hard to compare certain blockgroups. Furthermore, there is only so much a choropleth map can tell you. We couldn't break down the population by race, income, gender, etc. without making even more choropleth maps.

The maps contains the three cities: Durham, Raleigh, and Chapel Hill.From the maps above, we can see that the population changes from 2000 to 2010. There are more areas of high population in 2010. To further analyze the blockgroup data, we have to use different plots to analyze different parts of the data. Other than using the maps to show population change, it would've also been practical to use a density plot, scatterplot, or maybe even a bar plot, but they do not display geographic settings.

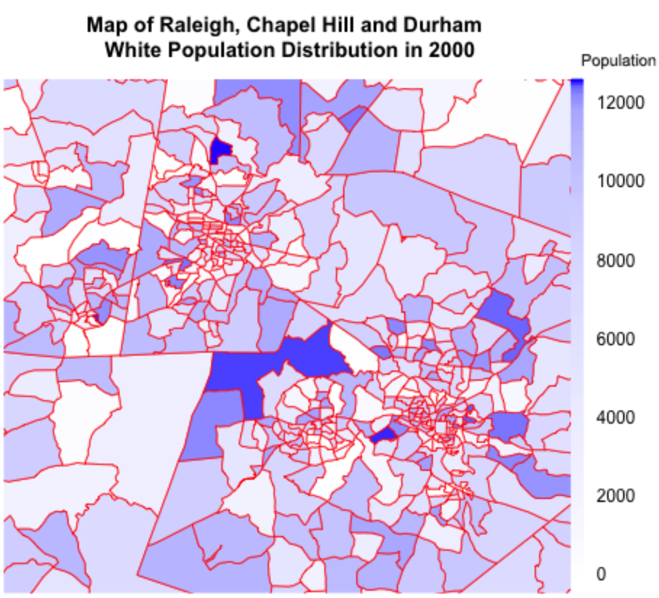
Image 2: Density plots of population distribution by race



The next aspect of the data that we wanted to look at was the change in population broken down by the three races that we are looking at. we decided that it will be easier to just look at population change and ignore geographic settings. Using a density line plot, we can easily observe the change in population between 2000 and 2010 (advantage). Density line plots are pretty easy for everyone to interpret and they show the density of data at certain population sizes. We don't have to deal with the change in blockgroups and we can look at individual races. However, the disadvantage of density plots is that it is hard to tell the exact number of blockgroups at each population level.

The density plot shows the change in concentration of population sizes for blockgroups. From the top left plot, we can see that the total population increased from 2000 to 2010. However, by looking at the other three plots, we can see that only the white population showed a significant increase. This makes sense, because the population of North Carolina is 68.5% white. Looking at the density plot for blacks, we see that there is little change from 2000 to 2010. The population for hispanics seems to have changed at a glance, but upon closer observation we can see that there are less blockgroups of low hispanic populations and more blockgroups of high hispanic populations. This means that the hispanic population became more concentrated in certain areas leaving the other areas with lowered hispanic populations. The most interesting pattern observed in this cluster of plots is the population movement of hispanics. The trend is different from the general population, whites, and blacks. While the other three plots display a less concentrated population, the hispanic plot showed something completely opposite. Possible alternate choices are scatterplot, boxplot, violin plot, beanplot, and mosaic plot.

Image 3: Choropleth Map of Raleigh, Chapel Hill and Durham on White Population Distribution in 2000 and 2010



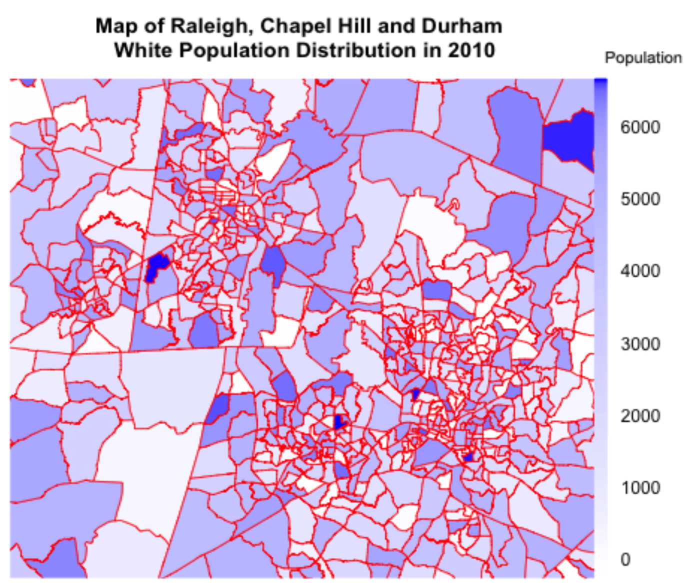


Image 4: Choropleth Map of Raleigh, Chapel Hill and Durham on Black Population Distribution in 2000 and 2010

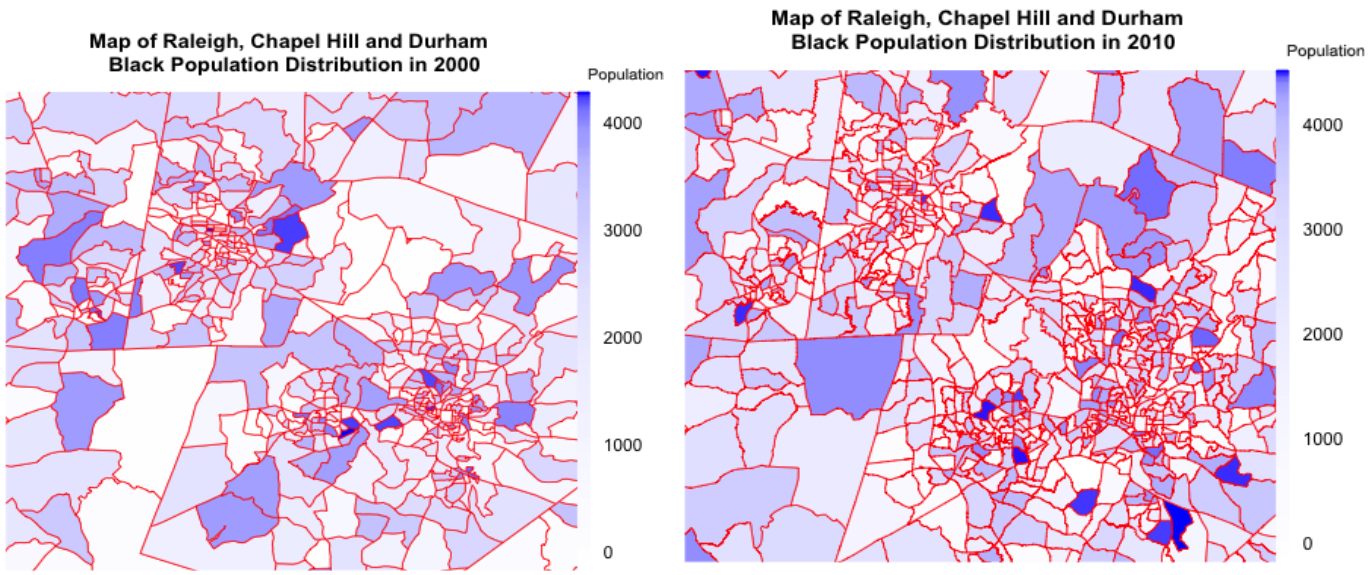
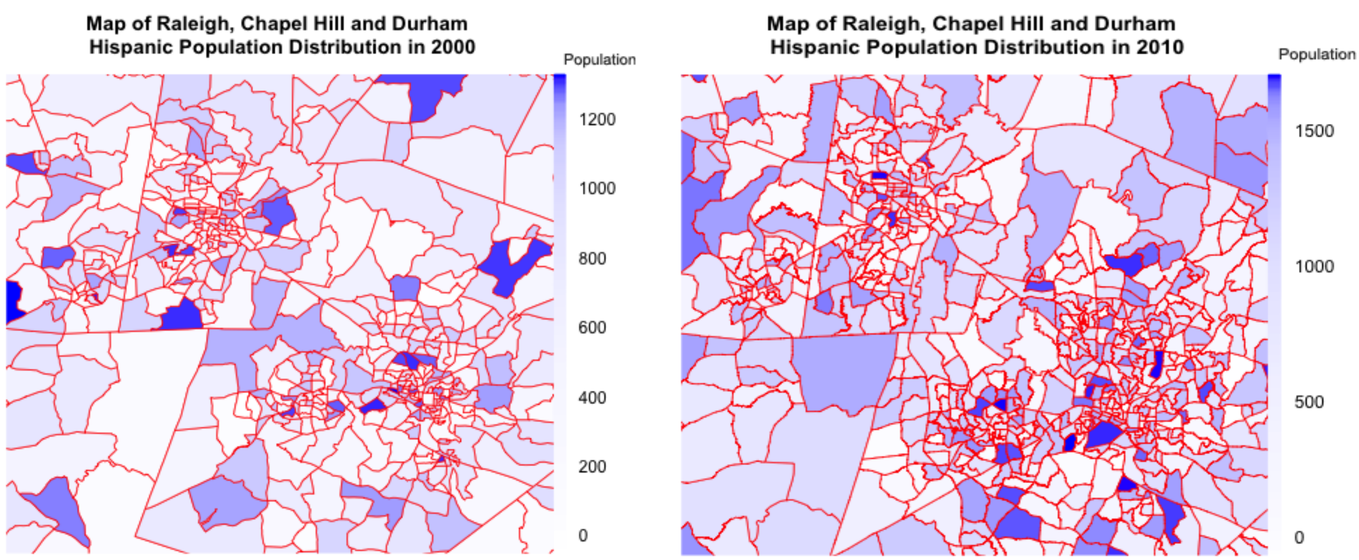


Image 5: Choropleth Map of Raleigh, Chapel Hill and Durham on Hispanic Population Distribution in 2000 and 2010



Here we chose to use a choropleth maps to show the different race populations of White, Black and Hispanics of the block groups for Durham, Raleigh and Chapel Hill in both 2000 and 2010. The darker the shade of blue of the blockgroup, the greater the population for that area, which allows us to draw conclusions about population density of each race in each year, and by comparing them, understand the migratory patterns of the white population over the course of the 10 years.

The advantages of this plot are that it gives very clear information about the population distribution in relation to geography, and can easily visualize the change in distribution between the time periods. The primary disadvantage of these plots are that the blockgroups do not match between 2000 to 2010 so it can be challenging to directly compare the blocks as the boundaries are different in the two data sets. In addition the color gradient is calculated by determining the ratio of that races population to the maximum blockgroup population for that year. However because the maximum blockgroup population is different for 2000 and 2010, the gradient scales differ slightly. Though you can still visually compare relative color within one map, and can compare the distribution of population density between the two maps, visually the color of one blockgroup in 2000 can not necessarily be visually compared to the color of that same group in 2010 to determine if the population of that blockgroup increased or decreased in the ten years. In addition, choropleths can only display one variable at a time, and these maps only show population density and cannot be used to address other variables such as income or age.

From the population distribution by race in the above maps, we see that for White and Hispanic populations, we see some shift in population density of blockgroups away from the center of the cities to outlying areas. In the first plot we mapped the distribution of the white population in 2000 and 2010. There is some variation in population distribution within individual blockgroups, in particular the dispersal of whites from one group to the northwest of Raleigh (though this may appear this way because of the redistribution of blockgroups in 2010), and a relatively stronger concentration or whites to the north east of Raleigh in 2010, but for the most part, there was not noticible pattern in the redistribution of the white population from 2000 to 2010. By examining the legend, you can tell that the maximum blockgroup population was much smaller in 2010 in comparison to 2000 which can probably attributed to the fact that the census data made more smaller blockgroups in 2010. For the most part, there seems to be less concentrated white populations in the center of the cities, with most of the white population distributed to the surrounding suburbs.

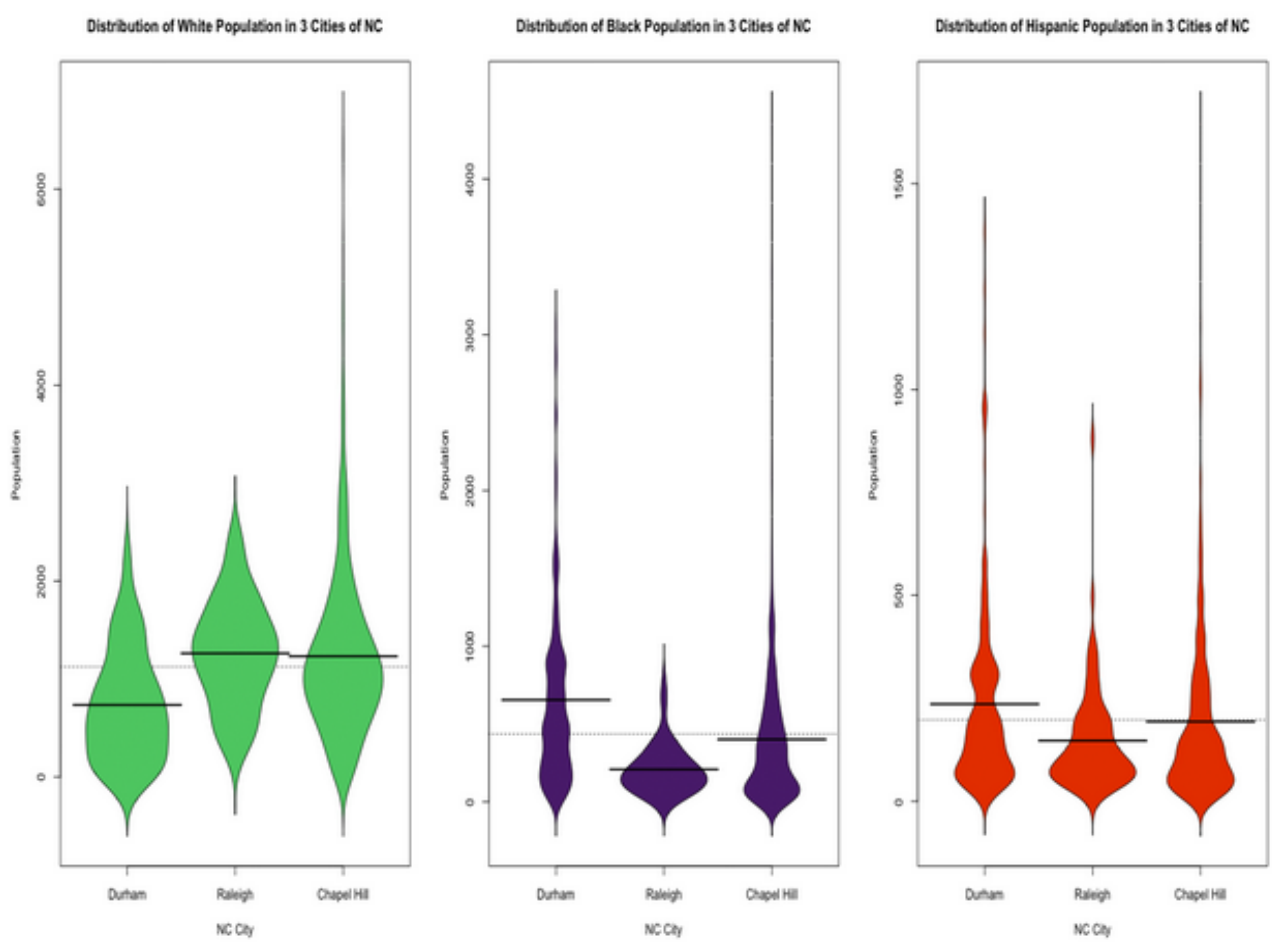
For Black population, the population distribution among blockgroups remained relatively similar. In the second plot we made two choropleths of the distribution of the black population in 2000 and 2010. These plots are much lighter in general than the plots for the white population, and also have a smaller range in the legend, indicating that there are fewer blacks than whites. The black population tends to be very low in the suburban areas where the whites were shown to be settled, and instead live in a few high density pockets closer to the urban areas. This pattern did not change between 2000 and 2010, but the location of these highly populated blockgroups did shift, in particular with more blacks settling to the south of Raleigh. There also seems to be more blacks distributed to the suburbs to the north of Raleigh in 2010 than in 2000.

For Hispanic population, the range in the legend for the 2010 plot is larger, and the overall plot is much darker, than the 2000 plot. This reflects the density plots above which showed the increase of the Hispanic population in these areas. In terms of geographic distribution, in 2000 there were only a few blockgroups with a high distribution of Hispanics, primarily located in the center of the urban areas. In 2010 there were many more blockgroups that had a Hispanic population of more than 1500. The Hispanic population appears to have migrated south towars Raleigh, with significantly more blockgroups of high distribution in Raleigh in 2010, and fewer high distribution plot groups in Chapel Hill and Durham.

However, as mentioned, because the blockgroup boundaries changed from 2000 to 2010, it is difficult to see whether this change is caused by dividing them into smaller blockgroups. We could have used the density or boxplots to show distribution of population by race, but it wouldn't display the geographic information that we want.

Relationships between age and population by race

Image 6: Comparison of distributions of Population by Race by Cities

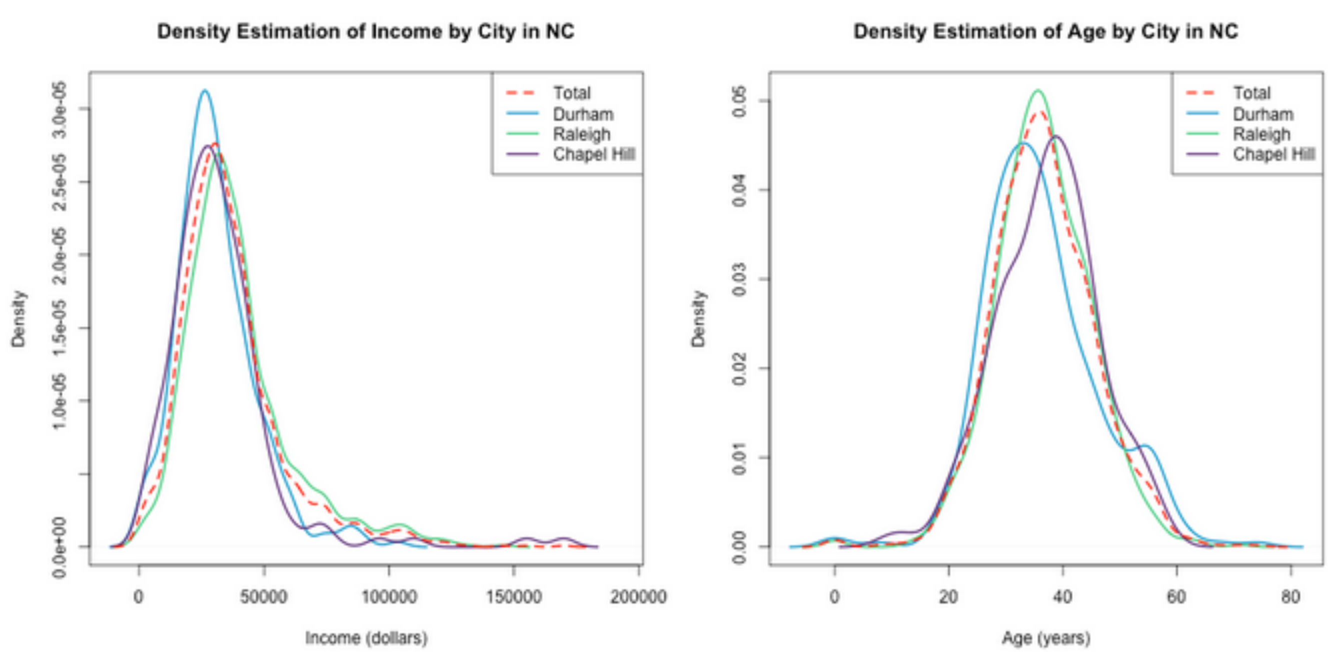


These graphs are bean plots of the distributions of population by race separated by the 3 cities we were assigned in North Carolina, Durham, Raleigh and Chapel Hill. The advantage of bean plots is that it is easy to compare distributions by categories in one plot of graphs and within the plot we can see statistical information such as median value and average, and also unlike boxplots, we get visually appealing shapes even with observations that could be possible outliers. The disadvantage is that this plot does not allow different coloring of the bean boxes within the same plot. It would have been better to use different colors for cities, but bean plot only takes one color argument.

The first graph in green shows the distribution of the white population. The median value of Durham is much lower than the medians of Raleigh and Chapel Hill therefore demonstrating that Durham has the lowest amount of white people. Looking at the purple bean plots in the middle, you can see that the median for black people is the lowest in Raleigh and the highest in Durham, therefore, Durham has the largest population of white people and Raleigh has the lowest population of black people. The last graph of orange bean plots show that similar to the black population in the cities, Durham has the largest population of Hispanic people and Raleigh has the smallest population of Hispanic people. Also by looking at the y-axises on the 3 separate bean plots, you can clearly tell that the white population is the largest overall and the Hispanic population is the smallest overall. The most important information and differences that we can visualize from these bean plots is that the city of Durham has the least amount of white people and the most amount of black and Hispanic people, and the city of Raleigh has the most amount of white people and the least amounts of black and Hispanic people.

Other plots that we could have used were boxplots and violin plots. A violin plot would give you similar information to the bean plots but the box plots would lose out on the density information. The advantage of using bean plots is that we get a density estimate in addition to other information such as medians of our data.

Image 7: Density Estimation of Income and Age by Cities



This graph shows the density estimation of income and age by individual cities in North Carolina. We distinguished between the 3 cities that my group was assigned: Durham, Raleigh and Chapel Hill.

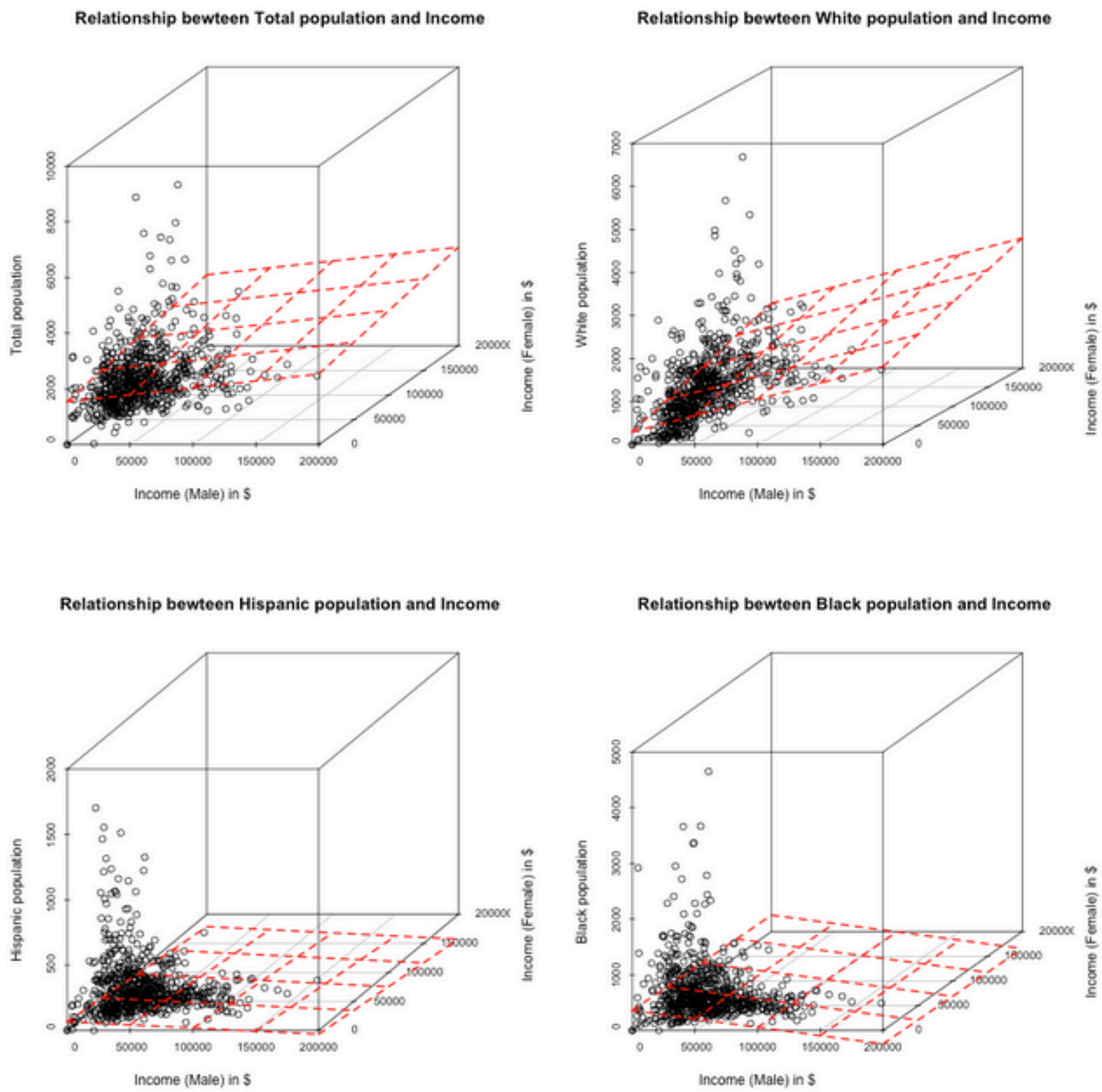
The advantage of density estimate plot is that it is one of the easiest methods for visualizing the particular information that we put into one plot since we are most interested in displaying the distribution comparisons. The disadvantage of using a density estimate is that sample size information is lost and estimate depends on the choice of bandwidth.

As shown in the legend, the blue line represents Durham, the green line represents Raleigh, the purple line represents Chapel Hill and the red dotted line represents the total population. On my first graph on the left, I have density estimation of income. Income in dollars is on the x-axis and the density estimate is on the y-axis. We can see that the city of Durham has the highest density in the lowest income of the 3 cities. Raleigh has the highest income density in the highest income of the 3 cities. Durham's density estimation of income cuts off around 100,000, Raleigh's density estimation of income cuts off around 125,000 and Chapel Hill's density estimation of income cuts off around 175,000. The distributions of densities of incomes for all the cities are positively skewed to the right which makes sense because in general many people are making an average amount of money and a much smaller group of people are making much more income than the average. The most important information that we displayed using the density estimates on the graph to the left is that the city of Durham has the lowest income and the city of Raleigh has the highest income. This is also related to the information that we discovered from our previous bean plots in that Durham has the lowest amount of the white population and Raleigh has the highest amount of the white population.

The density estimate on the right displays the estimate of age in the three cities. Generally, you can see that Chapel Hill has a greater density of older people, Durham has a greater density of younger people and Raleigh has the greatest density of middle-aged people. This can relate to the density estimate of income in the three cities because the Durham population has the highest density estimate of the lowest income and the city also has the highest density estimate of the youngest in age. It makes sense that younger people do not make as much money as older people. Also the city of Raleigh has the highest density estimate of the highest income and also has the highest density estimate of the most middle-aged people. This also makes sense because middle-aged people are more likely to have established careers and be making a stable amount of income.

Other graphs we could have used include strip charts, dot charts, histograms and boxplots. The advantage of using a density estimate is that we can easily visualize the comparison of data between the three cities and the total.

Image 8: Relationships between Income and Population by Race

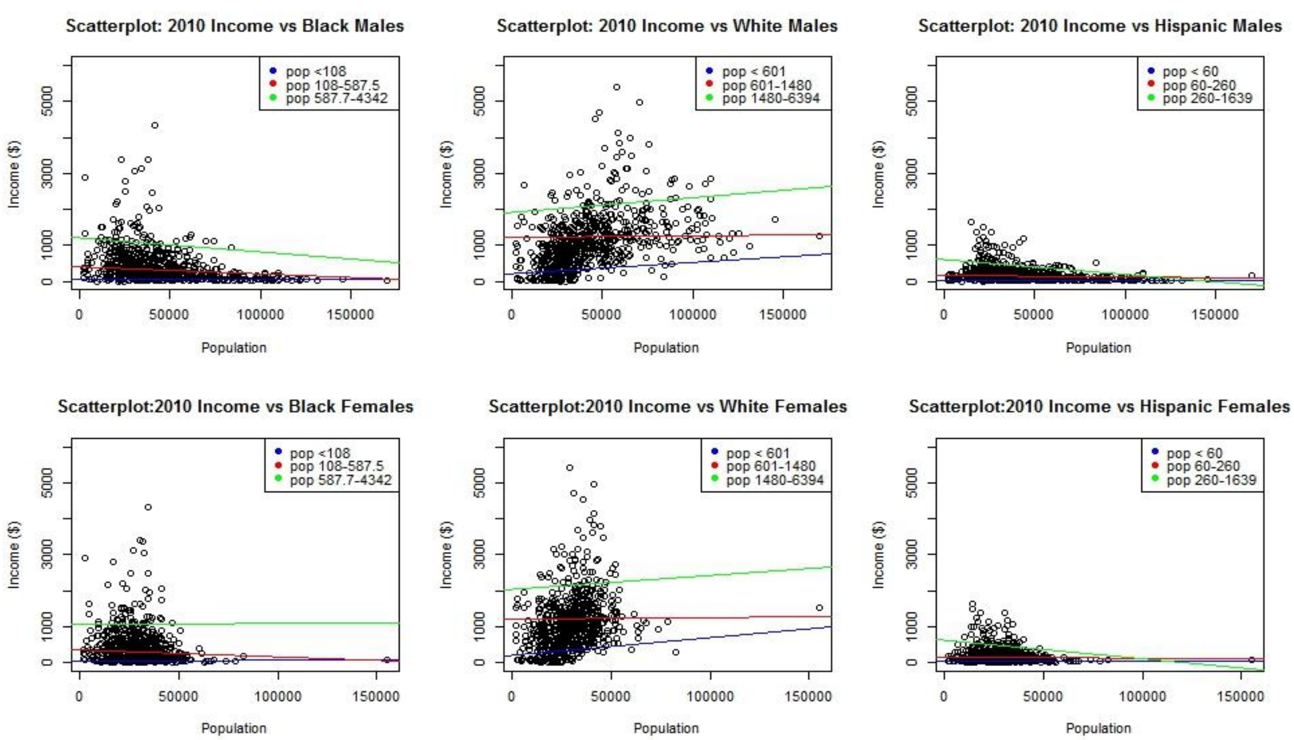


We chose three-dimensional scatterplots to display the relationships between income and population by race. The reason for our choice is that because we have two different variables representing income, male and female, and since using only one of them would result in loss of information, we choice to represent the relationships in 3D scatterplot with both income for male and female as x and y-axis. The z-axis represents the population by race. The dotted red plane represents the fitted regression line plane, regressed population by race on income of male and female, which gives an idea of how change in income by male and female are associated with change in population by race. The advantage is that we can incorporate three variables into scatterplot to display the relationships between income of male and female and population distribution by race. The disadvantage is that because it is three dimensional, it is difficult to see the relationships, without additional aid like the fitted regression plane.

From the top left plot, we see that income and total population exhibits positive association, by upward sloping of fitted regression plane. There seems to be positive relationships between income of male and female and total population. We further examine the relationships by looking at scatterplots of individual ethnic groups. On the top right plot, relationships between income and white population seems to be a positive one. As income of male and female increase, we generally see more white population. However, it is striking that in the bottom, scatterplots of income and black and Hispanics population, we see the opposite results. There seems to be almost no relationships, or rather, negative association, between income and Black/Hispanic population. We see that the positive relationships we saw earlier with total population is mostly driven by white population. Generally, white population is associated as higher income class, and minority like Black and Hispanic, as low income class. There seems to be a positive relationship with income for white population, and slightly negative relationship with Black and Hispanic population.

The alternative graph would be normal scatterplot or perspective plots. However, because we want to incorporate income of both male and female, because gender can be a confounding factor when examining the relationships.

Image 9: Relationships between Income and Race by Age

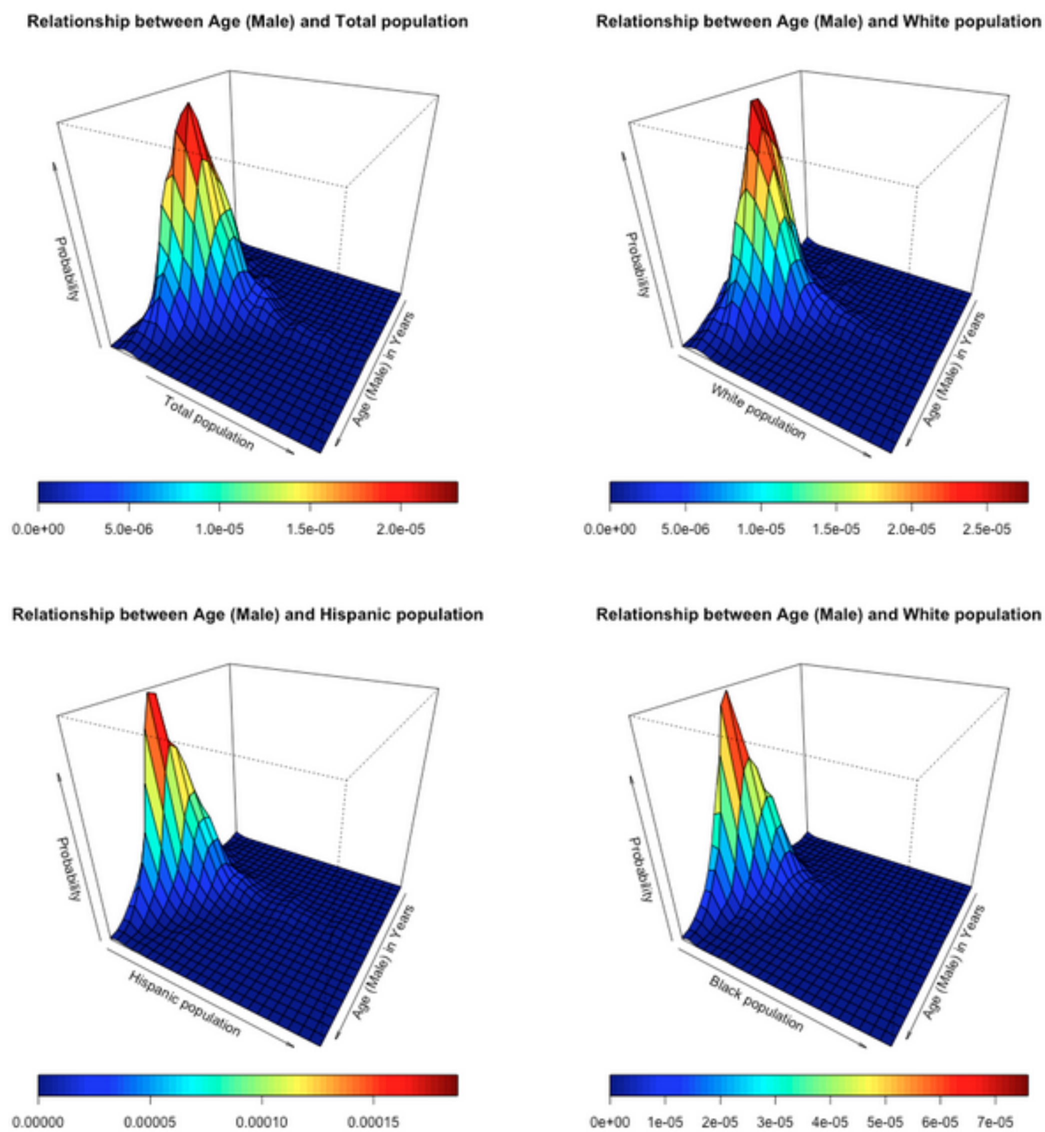


The above scatterplots explores the relationship between income and the 3 different races by gender. For each race, we divide the race population into 3 categories according to their population, low population areas which is below their first quartile population distribution (in blue), middle population areas which is between the first and third quartile (in red), and high population areas which is above the third quartile population distribution (in green). For the black population, there are 174 samples in the low population areas, 338 in the middle population areas and 171 in the high population areas. Similar for the white population, there are 171 samples in the low population areas, 341 in the middle population areas and 171 in the high population areas. And for the Hispanic population, there are 173 samples in the low population areas, 339 in the middle population areas, and 171 in the high population areas.

From the scatterplots, we can see that white population a relatively higher income in general while the areas with hispanic population has the lowest income among the three races. Looking at income by race, for both white and black population, higher population areas have higher incomes. Among the black population, male income actually decreases a little as population increases while female income increases a little as population increases in high population areas. For the white population, in all areas and across gender, there is a positive association between income and population size in each tract. It is also the only race among three with the most apparent positive correlation. In high Hispanic populated areas, we can see an apparent decrease in income as population increases, to a point lower than income in low Hispanic populated areas. There is a similar trend across both Hispanic genders. One of the implications we drew from this is that in high white populated areas, residents are associated with more white collar jobs where income accumulates and become higher, whereas for high Hispanic populated areas, residents are associated with more blue collar jobs and residents who are unemployed, therefore, incomes are relatively lower.

Other graphs choices are perspective plots, heat maps and 2d contour plots. Scatterplots are useful here because it is easier to compare the distribution between different race and gender with the regression lines visually. Also, it is easier to explore a positive/negative correlation between the two variables whereas it would be harder for a heat map or contour plot. One disadvantage of scatterplot is that it sometimes conceal overprinting, which might be significant for multimodal data in some cases.

Image 10: Relationships between age and population by race



To explore the relationships between age and population by race, we chose perspective plot draped with colors to represent the relationship between age and population by race. Specifically, we explored the relationship between age of male and population by race.

The advantage of this plot is that we can visually see the change in density of the two variables with colors, red being the highest and blue being the lowest, with help of legend. We can also visualize the relationship clearly, if change in age is correlated with change in population by race. The disadvantage is that there are no measurements shown on x and y axis and it is difficult to grasp how much the variables change.

From the above graphs, we do not observe any relationship, as change in age of males did not show apparent change in race population. There seems to be no change in density with population by race as age of male increases. From this graph, we can infer that there does not seem to be relationships between age and population distribution by race.

We could have used normal perspective plot or conditional density plot to display this kind of relationship. However, the normal perspective plot and conditional density plots do not give control over the labels of axis and when we plotted, the labels were overlapping each other. Also, the color scheme was difficult to implement.

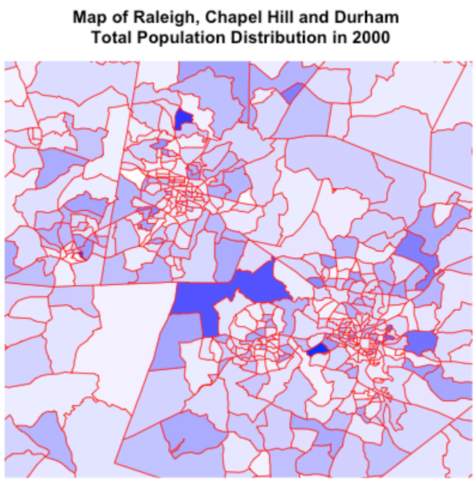
Conclusion

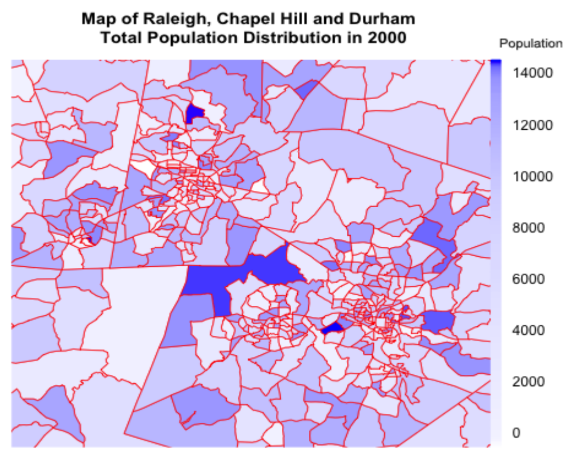
From our analysis, we showed that there were changes in distribution total population over time, from 2000 to 2010, as well as population by race with choropleth maps and density plots. We further analyzed how the change in distributions of population is related with income and age. With three-dimensional/two-dimensional scatterplots and perspective plots, we showed that there seemed to be a relationship between income and distribution of population by race, although this was strongly correlated with white population specifically. There did not seem to be relationships between age and distribution of population by race.

Appendix A: Revision

1. Choropleth Map: Images 1,3,4,5

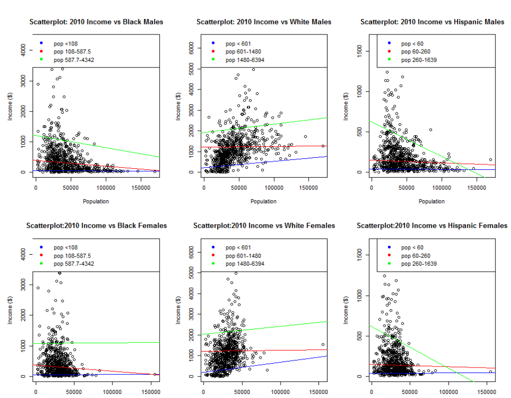
Our previous choropleth maps did not have a legend. It was difficult to see what the colors in blockgroups represented. We incorporated color gradient legend on the right axis by level of population.

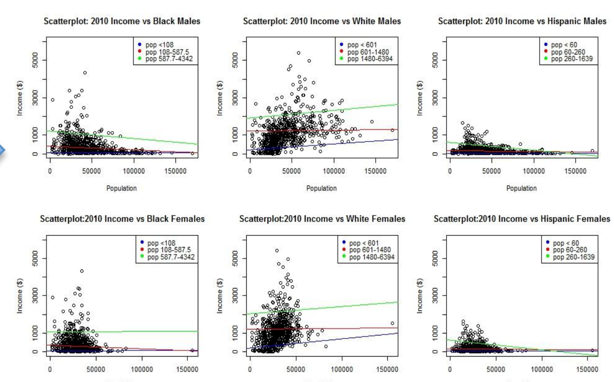




2. Scatterplot: Image 9

The range of y-axis was different across the plots. We made it the same across the plots, so it would be easier to see the difference in income levels. We also included the sample size for each group of population (low, medium, high) in the writing, so it would give better idea of how they are distributed.





Appendix B: R-code

```
library(UScensus2000blkgrp)
library(UScensus2010blkgrp)
library(MASS)
library(beanplot)
library(lattice)
library(graphics)
library(fields)
library(plotrix)
library(scatterplot3d)
library(scatterplot3d)
data(north_carolina.blkgrp)
data(north_carolina.blkgrp10)

color.grad <- function(input)
  sapply(input, function(kk)
  rgb(1-kk/max(input), 1-kk/max(input), 1)
)

col.total2010<-color.grad(nc2010$P0010001)
col.total2000<-color.grad(nc2000$pop2000)
col.white2010<-color.grad(nc2010$white)
col.white2000<-color.grad(nc2000$white)
col.black2010<-color.grad(nc2010$black)
col.black2000<-color.grad(nc2000$black)
col.hispanic2010<-color.grad(nc2010$hispanic)
col.hispanic2000<-color.grad(nc2000$hispanic)
cols.total2010<-find.color(col.total2010,nc2010$P0010001)
cols.hispanic2010<-find.color(col.hispanic2010,nc2000$hispanic)

png("1-total2010.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp10,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.total2010,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Total Population Distribution in 2010")
legend.col(col=sort(col.total2010,decreasing=T),nc2010$P0010001)
dev.off()

png("1-total2000.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.total2000,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Total Population Distribution in 2000")
```

```
legend.col(col=sort(col.total2000,decreasing=T),nc2000$pop2000)
dev.off()

png("1-white2010.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp10,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.white2010,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n White Population Distribution in 2010")
legend.col(col=sort(col.white2010,decreasing=T),nc2010$white)
dev.off()

png("1-white2000.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.white2000,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n White Population Distribution in 2000")
legend.col(col=sort(col.white2000,decreasing=T),nc2000$white)
dev.off()

png("1-black2010.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp10,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.black2010,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Black Population Distribution in 2010")
legend.col(col=sort(col.black2010,decreasing=T),nc2010$black)
dev.off()

png("1-black2000.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.black2000,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Black Population Distribution in 2000")
legend.col(col=sort(col.black2000,decreasing=T),nc2000$black)
dev.off()

png("1-hispanic2010.png")
par(mar=c(5, 2, 4, 4) + 0.1)
plot(north_carolina.blkgrp10,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.hispanic2010,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Hispanic Population Distribution in 2010")
legend.col(col=sort(col.hispanic2010,decreasing=T),nc2010$hispanic)
dev.off()

png("1-hispanic2000.png")
par(mar=c(5, 2, 4, 4) + 0.1)
```

```
plot(north_carolina.blkgrp,xlim=c(-79.12754,-
78.51098),ylim=c(35.68054,36.09091),col=col.hispanic2000,border=2)
title("Map of Raleigh, Chapel Hill and Durham \n Hispanic Population Distribution in 2000")
legend.col(col=sort(col.hispanic2000,decreasing=T),nc2000$hispanic)
dev.off()


png("1-density.png")
par(mfrow=c(2,2))
plot(density(nc2010$P0010001),main="Density plot of Total population",xlab="Total
Population",col="red",lwd=2)
lines(density(nc2000$pop2000),lwd=2,col="blue",lty=2)
plot(density(nc2010$white),main="Density plot of White population",xlab="White
Population",col="red",lwd=2)
lines(density(nc2000$white),lwd=2,col="blue",lty=2)
plot(density(nc2010$black),main="Density plot of Black population",xlab="Black
Population",col="red",lwd=2)
lines(density(nc2000$black),lwd=2,col="blue",lty=2)
plot(density(nc2000$hispanic),lwd=2,col="blue",lty=2,main="Density plot of Hispanic
population",xlab="Hispanic Population")
lines(density(nc2010$hispanic),col="red",lwd=2)
dev.off()
plot.new()
legend("topleft",legend=c("Population in 2010","Population in
2000"),col=c("red","blue"),lty=c(1,2),lwd=2)


#Getting rid of NAs
nc2010$income.female[which(is.na(nc2010$income.female))]<-0
nc2010$income.male[which(is.na(nc2010$income.male))]<-0
nc2010$age.male[which(is.na(nc2010$age.male))]<-0
nc2010$age.female[which(is.na(nc2010$age.female))]<-0


#Relationship between Total population and Income
png("2-total3d.png")
s3d<-
scatterplot3d(nc2010$income.male,nc2010$income.female,nc2010$P0010001,main="Relationshi
p bewteen Total population and Income",xlab="Income (Male) in $",ylab="Income (Female) in
$",zlab="Total population")
fit<-lm(P0010001~income.male+income.female,data=nc2010)
s3d$plane3d(fit,lwd=2,col="red")
dev.off()
```

```
png("2-white3d.png")
#Relationship between White and Income
s3d<scatterplot3d(nc2010$income.male,nc2010$income.female,nc2010$white,main="Relationshi
p bewteen White population and Income",xlab="Income (Male) in $",ylab="Income (Female) in
$",zlab="White population")
fit<-lm(white~income.male+income.female,data=nc2010)
s3d$plane3d(fit,lwd=2,col="red")
dev.off()

png("2-black3d.png")
#Relationship between Black and Income
s3d<scatterplot3d(nc2010$income.male,nc2010$income.female,nc2010$black,main="Relationshi
p bewteen Black population and Income",xlab="Income (Male) in $",ylab="Income (Female) in
$",zlab="Black population")
fit<-lm(black~income.male+income.female,data=nc2010)
s3d$plane3d(fit,lwd=2,col="red")
dev.off()

png("2-hispanic3d.png")
#Relationship between Hispanic and Income
s3d<scatterplot3d(nc2010$income.male,nc2010$income.female,nc2010$hispanic,main="Relation
ship bewteen Hispanic population and Income",xlab="Income (Male) in $",ylab="Income (Female)
in $",zlab="Hispanic population")
fit<-lm(hispanic~income.male+income.female,data=nc2010)
s3d$plane3d(fit,lwd=2,col="red")
dev.off()


#Relationship between Age and Population

png("2-totalage.png")
bw.total.age<-kde2d(nc2010$age.male,nc2010$P0010001)
drape.plot(bw.total.age,add.legend=F,theta=120,phi=30,main="Relationship between Age (Male)
and Total population",xlab="Age (Male) in Years",ylab="Total population",zlab="Probability")
image.plot(bw.total.age,legend.only =TRUE, horizontal =F)
dev.off()

png("2-whiteage.png")
par(mfrow=c(1,3),oma = c(0, 0, 0, 2))
bw.white.age<-kde2d(nc2010$age.male,nc2010$white)
drape.plot(bw.white.age,add.legend=F,theta=120,phi=30,main="Relationship between Age (Male)
and White population",xlab="Age (Male) in Years",ylab="White population",zlab="Probability")
image.plot(bw.white.age,legend.only =TRUE, horizontal =T)
dev.off()
```

```
png("2-blackage.png")
bw.black.age<-kde2d(nc2010$age.male,nc2010$black)
drape.plot(bw.black.age,add.legend=F,theta=120,phi=30,main="Relationship between Age (Male)
and White population",xlab="Age (Male) in Years",ylab="Black population",zlab="Probability")
image.plot(bw.black.age,legend.only =TRUE, horizontal =T)
dev.off()

png("2-hispanicage.png")
bw.hispanic.age<-kde2d(nc2010$age.male,nc2010$hispanic)
drape.plot(bw.hispanic.age,add.legend=F,theta=120,phi=30,main="Relationship between Age
(Male) and Hispanic population",xlab="Age (Male) in Years",ylab="Hispanic
population",zlab="Probability")
image.plot(bw.hispanic.age,legend.only =TRUE, horizontal =T)
dev.off()



income <- c(nc2010$income.female,nc2010$income.male)
income[is.na(income)] <- 0
durhamincome <- income[nc2010$county=="063"]
wakeincome <- income[nc2010$county=="183"]
orangeincome <- income[nc2010$county=="135"]

income <- c(nc2010$age.male,nc2010$age.female)
income[is.na(income)] <- 0
durhamincome <- income[nc2010$county=="063"]
wakeincome <- income[nc2010$county=="183"]
orangeincome <- income[nc2010$county=="135"]

d.total<-density(income)
d.durhami<-density(durhamincome)
d.wakei<-density(wakeincome)
d.orangei<-density(orangeincome)
x.min<-min(d.total$x,d.durhami$x,d.wakei$x,d.orangei$x)
x.max<-max(d.total$x,d.durhami$x,d.wakei$x,d.orangei$x)
y.min<-min(d.total$y,d.durhami$y,d.wakei$y,d.orangei$y)
y.max<-max(d.total$y,d.durhami$y,d.wakei$y,d.orangei$y)

png("2-age density.png")
plot(d.durhami, xlim=c(x.min,x.max),ylim=c(0,y.max), col=c("#0099CC"),lwd=2,xlab="Age
(years)",main="Density Estimation of Age by City in NC")
lines(d.wakei,col="#43CD80",lwd=2)
lines(d.orangei,col="#5E2D79",lwd=2)
lines(d.total,lty=2,lwd=2,col="red")
```

```
legend("topright",c("Total","Durham","Raleigh","Chapel
Hill"),col=c("red",'#0099CC','#43CD80','#5E2D79'),lty=c(2,1,1,1),lwd=2)
dev.off()


#Population distribution by race in cities
png("1-bean.png")
par(mfrow=c(1,3))
beanplot(nc2010$white~nc2010$county, col="#43CD80", main = "Distribution of White Population
in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
beanplot(nc2010$black~nc2010$county, col="#5E2D79", main = "Distribution of Black Population
in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
beanplot(nc2010$hispanic~nc2010$county, col="#EE4000", main = "Distribution of Hispanic
Population in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
dev.off()

beanplot(nc2000$white~nc2000$county, col="#43CD80", main = "Distribution of White Population
in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
beanplot(nc2000$black~nc2000$county, col="#5E2D79", main = "Distribution of Black Population
in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
beanplot(nc2000$hispanic~nc2000$county, col="#EE4000", main = "Distribution of Hispanic
Population in 3 Cities of NC", xlab = "NC City", ylab = "Population", what=c(1,1,1,0),
names=c("Durham","Raleigh","Chapel Hill"))
```