A Close Look at Detroit, Michigan

36-315: Statistical Graphics and Visualization

Spring 2013 Course Project

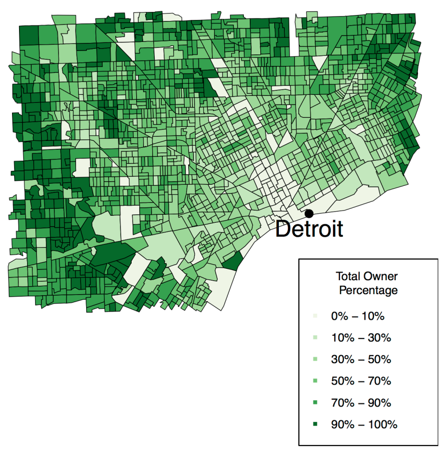
Alex Arkhangelskaya, Shivika Dhar, Wendy Mei, Jalen Poteat, Stephanie Stern

Introduction

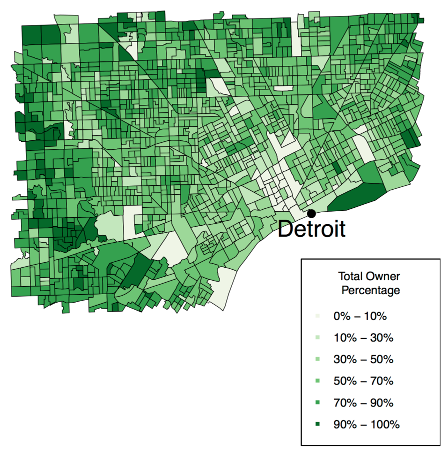
Combining the 2000 and 2010 United States Census data, as well as the 2011 American Community Survey, we were able to examine the distribution of several population characteristics across Detroit, Michigan at a block group level. Throughout the following series of graphs, we were able to arrive at several conclusions about how the population distribution compares over time from 2000 to 2010, with a focus on race, age, and income.

Proportion of House Owners and Renters

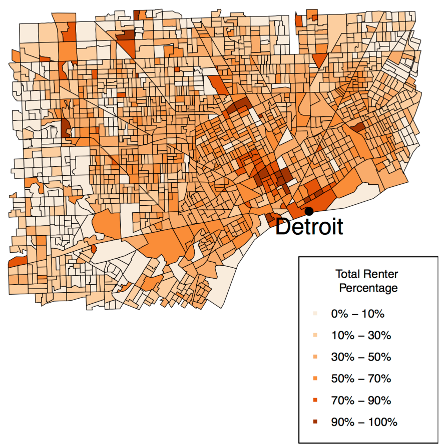
Proportion of House Owners in Detroit in 2000



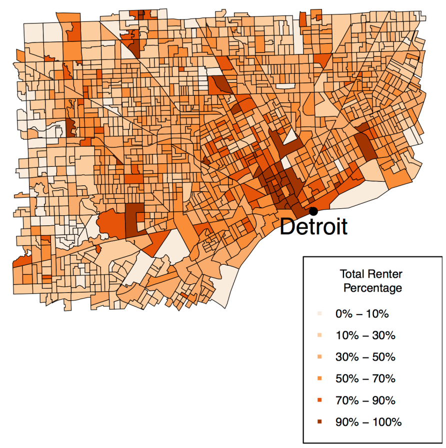
Proportion of House Owners in Detroit in 2010



Proportion of House Renters in Detroit in 2000



Proportion of House Renters in Detroit in 2010

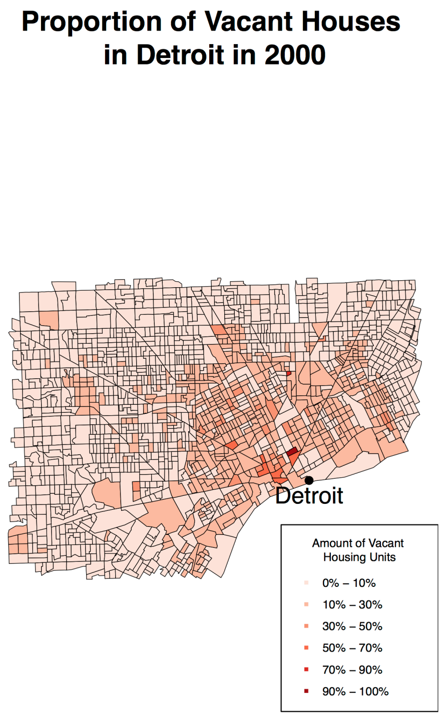


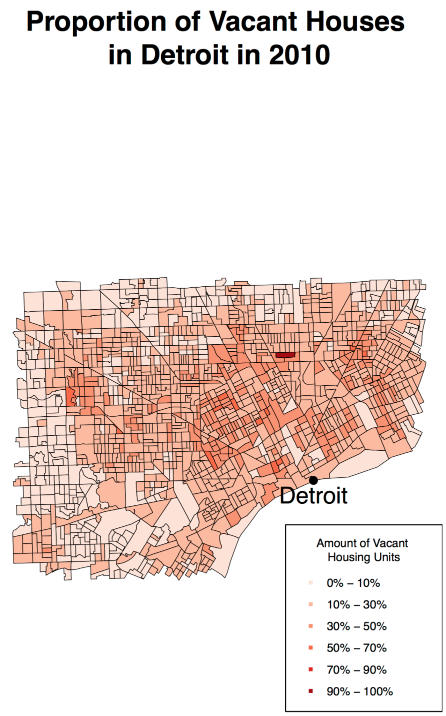
Our first set of graphs explores the changes in the proportions of house owners (number of house owners / total number of units) in Detroit from 2000 to 2010. The graphs show that the further away from the inner city one goes, the greater the proportion of owned houses in a block group becomes. Moreover, we noticed that very few of the housing units in inner city Detroit are owned. Looking at variations between 2000 and 2010, we see that in the outer city areas, fewer houses were owned in 2010. Regardless, a greater percentage of the houses in this area were owned if you compare the proportion to areas in the inner city. This may suggest that some people in greater Detroit were able to move to a different city between 2000 and 2010.

These two maps, of the proportion of house renters in 2000 and the proportion of house renters in 2010, are mirror images of the two shown above, indicating the proportion of house owners in 2000 and the proportion of house owners in 2010. Again, they show that in both 2000 and 2010 it was more likely to that a specific house was rented if it was selected from a block group closer to Detroit proper.

Using the choropleths provides an easy way to see the block groups that have a higher percentage of house ownership. However, using proportions means that we have no indication as to which block groups have the highest number of housing units. The choropleth gives no indication as to which block groups have to highest total number of housing units and a block group with 10 vacancies out of 40 housing units would be shaded the same color as a block group with 125 vacancies out of 315 housing units.

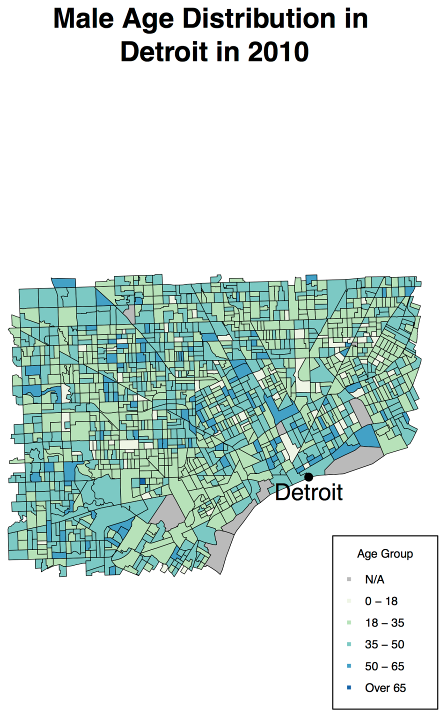
Proportion of Vacant Housing Units

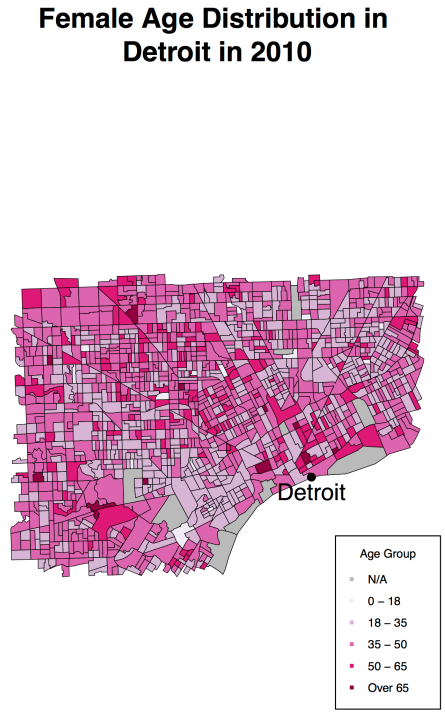




These choropleths examine the change in the proportion of vacant housing units (per block group) from 2000 to 2010. In 2000, we see that the block groups with the largest proportion of vacant houses, compared to occupied houses, are located in the inner city of Detroit with few heavily vacant housing units on the Detroit perimeter. The 2010 choropleth shows that not only did inner city Detroit remain an area with a high vacancy rates, but also that the number of block groups with a relatively high vacancy rate, compared to occupancy rate, increased significantly. This escalation of vacancies even extends to some block groups sitting on the perimeter of Detroit. As our next graphs will point out, it is also interesting to note that the area with the most housing vacancies is also the area that is dominated by an African American population.

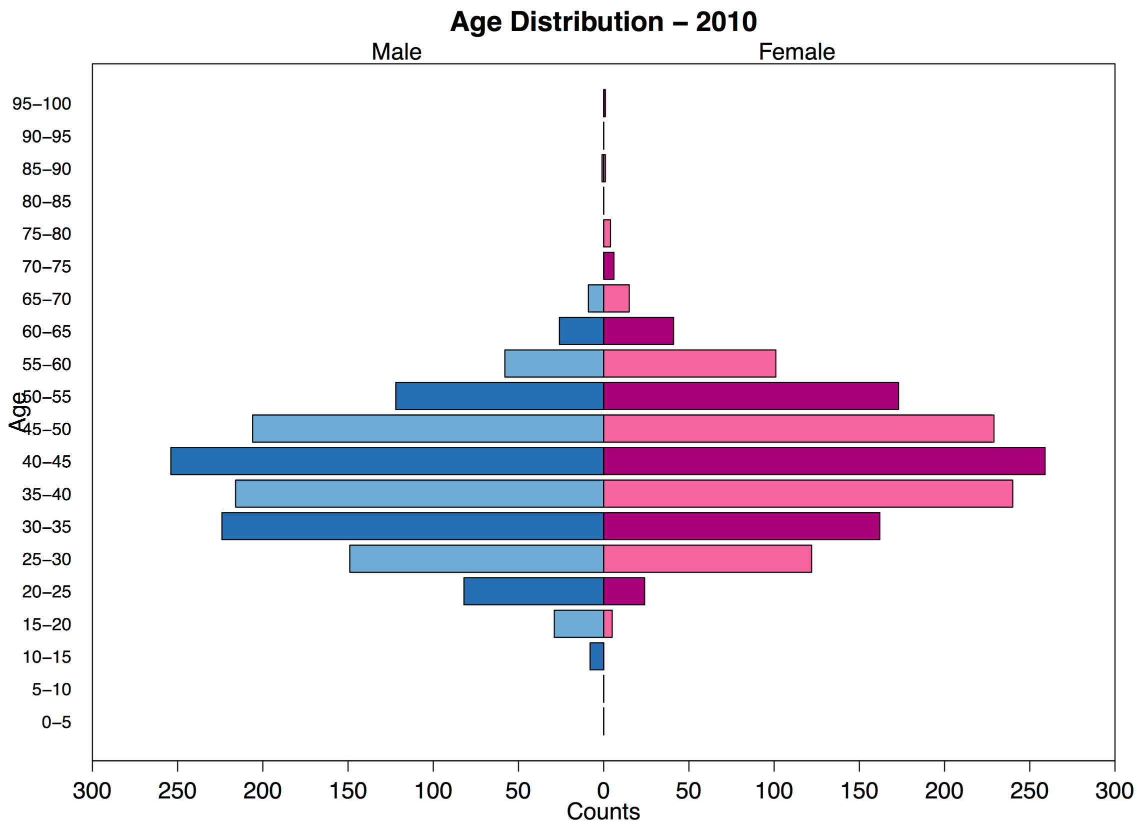
Male and Female Age Distribution





From the map above, we can see that in the block groups in the south central area in 2010 mainly includes people in the age range of 18-35. We did not have enough information to plot the age distributions for each gender in 2000. The south central area of Detroit may have a large proportion of young males and females because it is a fluid area as indicated by the renter map earlier. Young people are more likely to move around than older people. The rest of the age distribution is scattered with various pockets of older and younger generations. Regarding gender, we can see that in each block group, the age group for males and females is about the same. This means that young males and females live the same area and older males and females live in the same area.

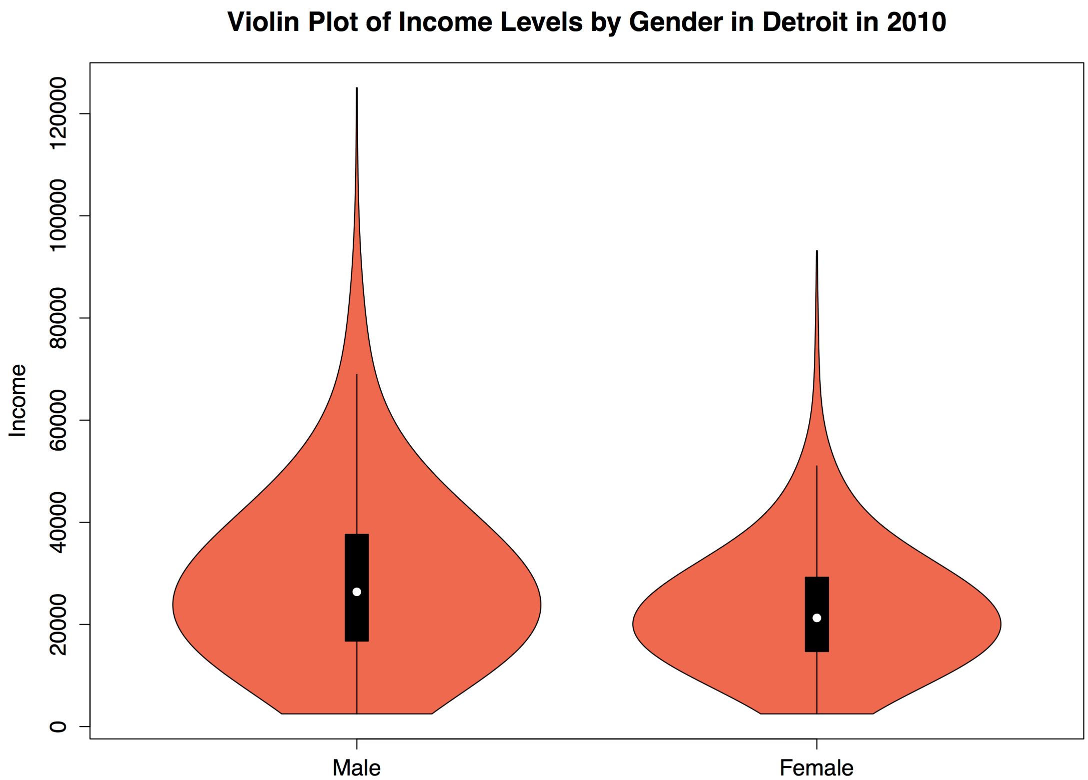
We chose a choropleth because the map provides a good indication of how people are distributed in Detroit by age. We could have used contour lines on a map instead, but the large number of block groups would cloud the impact of the contour lines. Additionally, since age is not entirely concentrated in one area, there would be many modes and the graphs would not be very conclusive.



The side-by-side histograms display information about the age distribution for both males and females in Detroit in 2010. Again, we did not have enough age information from 2000 to plot a histogram of the age distributions. The graph is unimodal for males at around 35 and at around 40 for females. As we can see, even though males and females have around the same average age in Detroit, there are more elder women than elder men. Additionally, there are also more young males than young females.

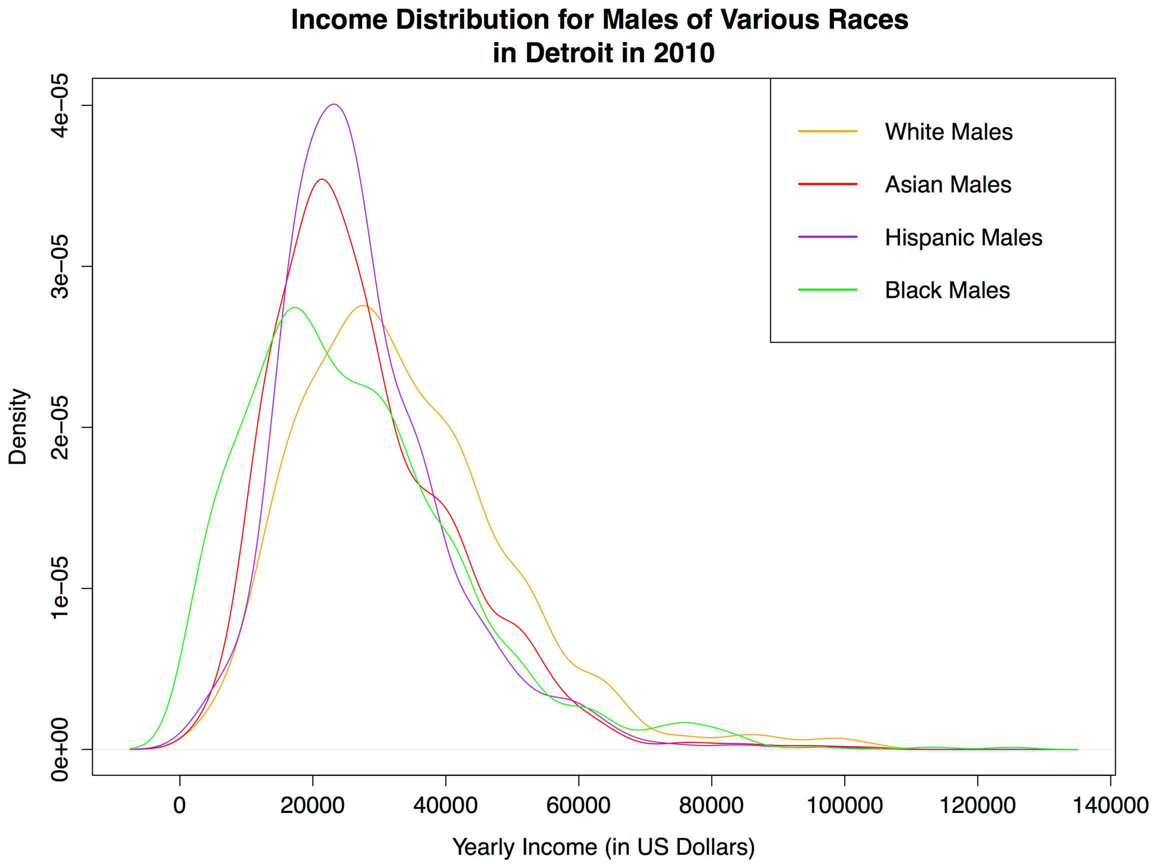
We chose to represent this information with a histogram, because histograms are generally used to display information over a continuous one-dimensional variable. In this case, the variable is age. We chose to place the two histograms side-by-side because the format allows one to compare each age category between the two genders. The side-by-side format is particularly useful at comparing age at the extremes, where the graph tapers off. We could have also used the ASH, because it gives a smoother estimate of the distribution. However, the ASH is dependent on the two parameters, the number of shifts and the number of bins. Therefore, different parameter choices may give different results. We chose to use a histogram because it is less dependent on the choice of bins than the ASH. We were focused on displaying the data truthfully without masking certain aspects of the graph, such as possible modes or the sample size. We could have also used a dotchart, a stripchart, or a density estimate. However, it is more difficult to gauge modes on a dotchart since it induces two-dimensional information by adding the index of the observation on the y-axis. A stripchart is only one-dimensional but may be difficult to see how big the different modes are, unless the chart is jittered or stacked. A density estimate would have given the same information as a histogram or an ASH, because it is dependent on bandwidth choice and does not contain sample size information. Furthermore, we chose to use alternating shades of each color because it created a more visually appealing graph. The alternating colors also made the bars easier to connect to the bins on the y-axis. We could have used only one shade of blue and only one shade of pink, but we felt that the graph would not have been as sophisticated as possible.

Income Distribution by Gender and Race



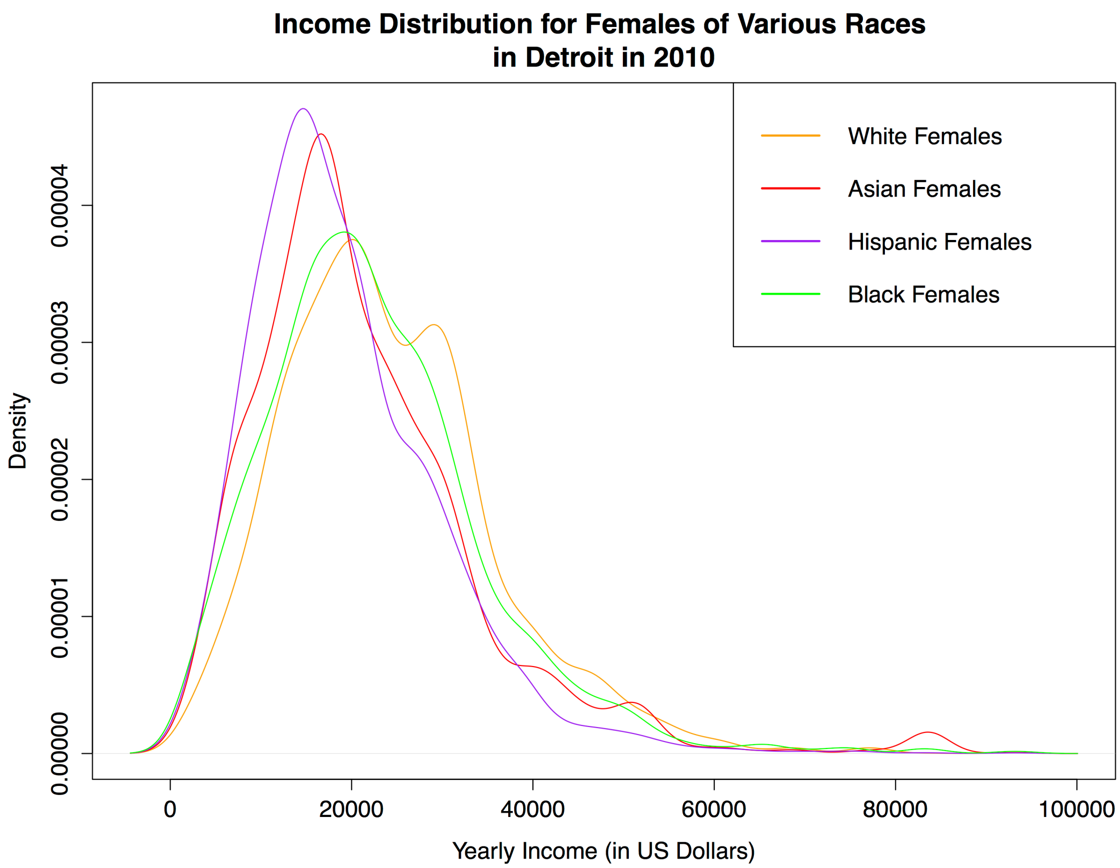
Male and female income distributions are similar in 2010; however, the male income range is larger than the female range. Additionally, from the graph we can see that the male income outlier is around an income of $125,000 and the female income outlier is around an income of $95,000. Male income has median around $25,000 and female income has median around $20,000. The female income distribution is denser around the median compare to the male income distribution.

We chose to use violin plots because we are comparing a categorical variable (gender) that has multiple groups with a quantitative variable (income). A violin plot shows descriptive information such as the minimum, median, maximum, minimum, and the inter quartile range. On top of this, a violin plot illustrates the distribution of voting for each genre through kernel density estimates, which a boxplot is not capable of. One disadvantage of the violin plot is that it is difficult to see what specific outliers there might be. Overall, though, we can estimate where outliers lie based on the fact that we have the kernel density estimates. Other plots that would have worked would be box percentile plots, bean plots, or even boxplots (but less so than the others). We chose to use to keep the default parameter values because we did not think that altering the parameters would be important nor make much of a difference in this case.



In the census data there are six races: White, Asian, Hispanic, Black, Native American, and Hawaiian. We chose to leave out the Native American and Hawaiian populations because there was very little information on them in the city of Detroit. Additionally, we only focused on 2010 data, as we did not have income information for 2000. From the density plot above, we can see that the majority of Asian males have an income around $21,260. This is a similar income level for Hispanic males ($23,218) except that Hispanic males have a higher density at the about $20,000 level, which means that a larger proportion of Hispanic males have an income of $23,218 than Asian males. We can also identify that white males have a slightly higher yearly income mode than all other races at around $27,680. However, the density is lower than that of Asians and Hispanics suggesting that there is a wider range of incomes for white males. On the other hand, the most common yearly income for black males in Detroit is about $17,365, which is lower than all other races, with a density level similar to white males.

We chose to use density estimates because we are interested in comparing the distributions of four different races. We could have used histograms but then run the risk of masking features by putting them in the same graph or having difficulty comparing them across graphs. We lose sample size information by using density estimates; however, since we are interested in the distribution, it is acceptable to lose the sample size. Other graphs such as violin plots, stripcharts and dot charts do not lend themselves easily to distribution comparison.

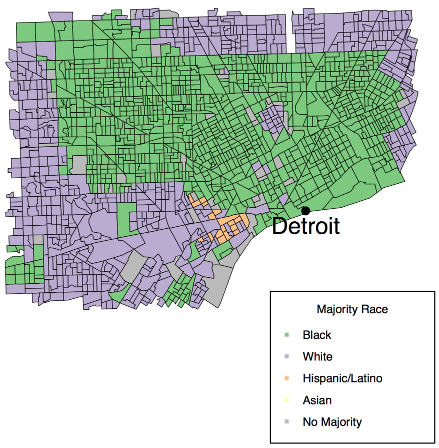


From the density plot above, we can see that the most common income for Asian females is around $16,655 in 2010. We only used 2010 data because as mentioned above, there was no income data available for 2000. The tail of the distribution of Asian females has a small local max density around an income level of $83,515. Hispanic females have an income level of around $15,000 with a slightly higher density than Asian females. We can also see that white and black females have similar distributions of yearly income of around $20,130 and $19,110, respectively. However, the distribution of white females has a second, smaller local maximum density around $30,150.

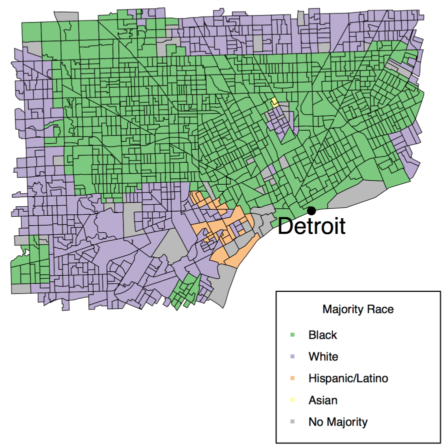
The reasoning for choosing density plots for incomes level of females by race is that same as that of the income distribution of males by race. The advantages and disadvantages are also the same.

Race Distribution over Time

Race Distribution – 2000



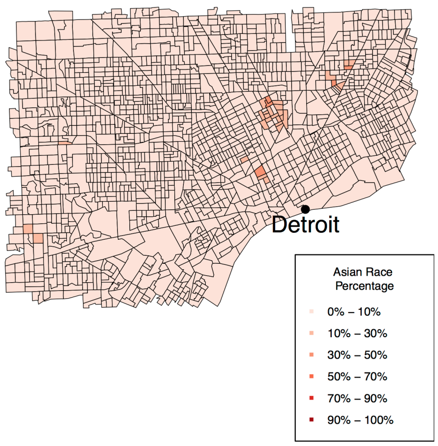
Race Distribution – 2010



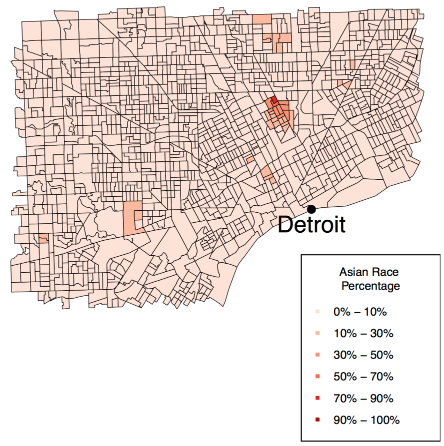
Our graph reveals that black people mainly populate Detroit proper while white people mainly populate the suburbs. There is a dense pocket of Hispanic population in the southeastern part of the greater Detroit area that grows from 2000 to 2010. There are no majority Asian block groups in 2000, but there is a small pocket of an Asian majority in 2010 in an area formerly dominated by white people. This area of a strong Asian population is a pocket in a larger black majority area.

We chose to use a choropleth map for race distribution because upon exploring the variables, we learned that Detroit is extremely divided geographically by race. We felt that showing this distribution would be vital to our project. Identifying the major race populations in each block group is necessary for explaining the demographic variations in Detroit. Advantages of this plot include that the racial distinctions are clear and the plot is not overwhelming with multiple graphs on top of one another. Disadvantages of this plot include the issue of only showing majority races, rather than the amounts of each race in each area, and making it seem like the Asian population in 2010 came from nowhere. Additionally, the many lines dividing the block groups provide a lot of ink.

Asian Race Distribution in Detroit in 2000



Asian Race Distribution in Detroit in 2010



We chose to use a choropleth map for the Asian race distribution because after plotting the majority races in each block group because we noticed that the Asian race has very little majority in Detroit. Additionally, we were curious if the small pocket of Asian majority block groups in 2010 appeared from nowhere or grew from a weaker Asian population in 2000. We plotted the 2000 and 2010 Asian race distribution maps side-by-side to examine the changes in distribution over time. The bins for each shade of red were determined by choosing extremes and dividing the rest into enough groups to show a gradient, but not too many groups because that would add noise. The graphs for Asian population distributions show that the Asian population in the northeastern section of greater Detroit in 2010, as evidenced in Graph 10, did not come from nowhere. In 2000, there was an Asian population in the same area, though it was not as distinct as it is 2010. Furthermore, there is a second cluster of Asian people in the southeastern section of Detroit that appears only in 2010. It is less distinct than the other cluster.

Advantages of this graph are that it shows a gradient for different levels of concentration of an Asian population, it shows the race distribution geographically, and the data is not overly confusing with many plots on top of one another (heat map, contour map, and points on top of a graph would be infinitely more confusing and unnecessary). Disadvantages of this graph are that it only shows one race at a time and that it only shows race and not other dimensions such as income.

Conclusion

In general, we can see that Detroit, Michigan went through a lot of changes between 2000 and 2010. The population of Detroit declined from 2000 to 2010, as evident in the choropleth maps displaying the vacant to occupied housing ratio. Poverty is an ongoing problem in the city proper. The overall mean income of Detroit is very low, even though Detroit is one of the larger cities in the United States. There were more vacant units present in Detroit in 2010 than in 2000. Detroit remains one of the most racially segregated cities in the United States, although segregation in Detroit has decreased from 2000 to 2010. As evident in our analysis, Detroit is a very fluid city and will continue to go through even more changes with respect to race, age, and income population distributions in the future.

R Code

```
##Final Project
##Greater Detroit Area

library(maps)
library(maptools)
library(UScensus2000blkgrp)
library(UScensus2010)
library(UScensus2010cdp)
library(RColorBrewer)
library(Hmisc)
library(wvioplot)
library(beanplot)
library(RColorBrewer)

#############
##Load Data##
#############
data(michigan.blkgrp) # 2000 data
load("michigan-census2010-plus-acs.RData") # 2010 data

#######################
##Setting Up the Data##
#######################
##Restrict 2000 Census Data
vec <- vector()
# for loop restricts the census data to blkgrps in the detroit area
for (yy in 1:length(michigan.blkgrp)){ # steps through every blkgrp
  # bbox retrieves boundary box from spatial data
  x <- bbox(michigan.blkgrp[yy,])
  # adds the blkgrp to the vector if the blkgrp is within the given doundaries
  det.plot <-
    if (x[1,1]>=-83.329926 & x[1,2]<=-82.910451 & x[2,1]>=42.255192 & x[2,2]<=42.482226) {
      vec <- c(vec,yy)
    }
}
##Check the vector
plot(michigan.blkgrp[vec,])
# Save the vector
det.data.2000 <- michigan.blkgrp[vec,]
save(det.data.2000, file="detroit-census2000.RData")

##Restrict 2010 Census Data
```

```
vec <- vector()
# for loop restricts the census data to blkgrps in the detroit area
for (yy in 1:length(michigan.blkgrp10)){ # steps through every blkgrp
  # bbox retrieves boundary box from spatial data
  x <- bbox(michigan.blkgrp10[yy,])
  # adds the blkgrp to the vector if the blkgrp is within the given doundaries
  det.plot <-
    if (x[1,1]>=-83.329926 & x[1,2]<=-82.910451 & x[2,1]>=42.255192 & x[2,2]<=42.482226) {
      vec <- c(vec,yy)
    }
}
# Check the vector
plot(michigan.blkgrp10[vec,])
# Save the vector
det.data.2010 <- michigan.blkgrp10[vec,]
save(det.data.2010, file="detroit-census2010-plus-acs.RData")

##########################
## Working with the Data##
##########################
load("detroit-census2000.RData")
load("detroit-census2010-plus-acs.RData")

#################
#### Detroit ####
#################
points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
text(x=-83.0245, y=42.319, labels="Detroit")

##########################################################################
## Question 1 - How does the population distribution compare over time, ##
################from 2000 to 2010, particularly by race?##################
##########################################################################

##########################################################################
#### Question 2 - How do characteristics like age and income correlate####
################# to population distribution and race in 2010?############
##########################################################################

##~ Graph 1 ~##
##Owner Occupied- 2000
par(mfrow=c(1,2))
race.oo.2000 <- det.data.2000$hh.owner/det.data.2000$hh.units
race.oo.2000[which(is.nan(race.oo.2000)==T)] <- 0
```

```
race.oo.2000.breaks <- c(0.1,0.3,0.5,0.7,0.9)
race.oo.2000.col <- brewer.pal(6,"Greens")
race.oo.2000.group <- apply(1*(outer (race.oo.2000, race.oo.2000.breaks, "-") > 0), 1, sum) + 1
plot(det.data.2010, col="white",border="white") #invisible
plot(det.data.2000, col=race.oo.2000.col[race.oo.2000.group],add=T,lwd=.5)
points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
text(x=-83.0245, y=42.319, labels="Detroit")
title(main="Proportion of House Owners in \nDetroit in 2000")
legend("bottomright", title="Total Owner \nPercentage",col=c(race.oo.2000.col), pch=15,
    legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
    xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)
# Owner Occupied- 2010
race.oo.2010 <- det.data.2010$H0140002/det.data.2010$H0140001
race.oo.2010[which(is.nan(race.oo.2010)==T)] <- 0
  race.oo.2010.breaks <- c(0.1,0.3,0.5,0.7,0.9)
  race.oo.2010.col <- brewer.pal(6,"Greens")
  race.oo.2010.group <- apply(1*(outer (race.oo.2010, race.oo.2010.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col=race.oo.2010.col[race.oo.2010.group],lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
title(main="Proportion of House Owners in \nDetroit in 2010")
legend("bottomright", title="Total Owner \nPercentage",col=c(race.oo.2010.col), pch=15,
    legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
    xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)

##~ Graph 2 ~##
##Renter Occupied - 2000
par(mfrow=c(1,2))
race.ro.2000 <- det.data.2000$hh.renter/det.data.2000$hh.units
  race.ro.2000[which(is.nan(race.ro.2000)==T)] <- 0
  race.ro.2000.breaks <- c(0.1,0.3,0.5,0.7,0.9)
  race.ro.2000.col <- brewer.pal(6,"Oranges")
  race.ro.2000.group <- apply(1*(outer (race.ro.2000, race.ro.2000.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col="white",border="white") #invisible
  plot(det.data.2000, col=race.ro.2000.col[race.ro.2000.group],add=T,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
  title(main="Proportion of House Renters in \nDetroit in 2000")
  legend("bottomright", title="Total Renter \nPercentage",col=c(race.ro.2000.col), pch=15,
    legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
    xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)
```

```
# Renter Occupied - 2010
race.ro.2010 <- det.data.2010$H0140010/det.data.2010$H0140001
  race.ro.2010[which(is.nan(race.ro.2010)==T)] <- 0
  race.ro.2010.breaks <- c(0.1,0.3,0.5,0.7,0.9)
  race.ro.2010.col <- brewer.pal(6,"Oranges")
  race.ro.2010.group <- apply(1*(outer (race.ro.2010, race.ro.2010.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col=race.ro.2010.col[race.ro.2010.group],lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
  title(main="Proportion of House Renters in \nDetroit in 2010")
  legend("bottomright", title="Total Renter \nPercentage",col=c(race.ro.2010.col), pch=15,
     legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
     xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)

##~ Graph 3 ~##
#Vacancy
#2000
par(mfrow=c(1,2))
vac.2000 <- det.data.2000$hh.vacant/det.data.2000$hh.units
  vac.2000[which(is.nan(vac.2000))] <- 0
  vac.2000.breaks <- c(0.1,0.3,0.5,0.7,0.9)
  vac.2000.col <- brewer.pal(6,"Reds")
  vac.2000.group <- apply(1*(outer (vac.2000,
                              vac.2000.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col="white",border="white") #invisible
  plot(det.data.2000, col=vac.2000.col[vac.2000.group],add=T,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
  title(main="Proportion of Vacant Houses \nin Detroit in 2000")
  legend("bottomright", title="Amount of Vacant \nHousing Units",col=vac.2000.col,pch=15,
     legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
xpd=TRUE, cex=0.5, bty="n")
  rect(-83.05277, 42.28678-0.1300292, -83.05277+0.1590747, 42.29678)
#2010
vac.2010 <- det.data.2010$H0030003/det.data.2010$H0030001
  vac.2010[which(is.nan(vac.2010))] <- 0
  vac.2010.breaks <- c(0.1,0.3,0.5,0.7,0.9)
  vac.2010.col <- brewer.pal(6,"Reds")
  vac.2010.group <- apply(1*(outer (vac.2010,
                              vac.2010.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col=vac.2010.col[vac.2010.group],lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
```

```
  title(main="Proportion of Vacant Houses \nin Detroit in 2010")
  legend("bottomright", title="Amount of Vacant \nHousing Units",col=vac.2010.col,pch=15,
     legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
xpd=TRUE, cex=0.5, bty="n")
  rect(-83.05277, 42.28678-0.1300292, -83.05277+0.1590747, 42.29678)

##~ Graph 4 ~##
# 2010 Age Distribution by Sex
par(mfrow=c(1,2))
age.male <- det.data.2010$age.male
  mean(age.male, na.rm=T) #34.46546
  min(age.male, na.rm=T) #6.4
  max(age.male, na.rm=T) #81.2
age.male.breaks <- c(18,35,50,65)
age.male.col <- brewer.pal(5,"GnBu")
age.male.group <- apply(1*(outer (age.male, age.male.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col=age.male.col[age.male.group],lwd=.5)
  plot(det.data.2010[which(is.na(age.male)),], col="gray", add=T,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
  title("Male Age Distribution in \nDetroit in 2010")
  legend("bottomright", title="Age Group",col=c("gray",age.male.col), pch=15,
      legend=c("N/A","0 - 18","18 - 35","35 - 50","50 - 65","Over 65"),xpd=TRUE,cex=.5)
age.female <- det.data.2010$age.female
  mean(age.female, na.rm=T) #38.05322
  min(age.female, na.rm=T) #10.9
  max(age.female, na.rm=T) #95
age.female.breaks <- c(18,35,50,65)
age.female.col <- brewer.pal(5,"PuRd")
age.female.group <- apply(1*(outer (age.female, age.female.breaks, "-") > 0), 1, sum) + 1
  plot(det.data.2010, col=age.female.col[age.female.group],lwd=.5)
  plot(det.data.2010[which(is.na(age.female)),], col="gray", add=T,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
  title("Female Age Distribution in \nDetroit in 2010")
  legend("bottomright", title="Age Group",col=c("gray",age.female.col), pch=15,
     legend=c("N/A","0 - 18","18 - 35","35 - 50","50 - 65","Over 65"), xpd=TRUE,cex=0.5)

##~ Graph 5 ~##
##Age Distribution by Sex - Bar Plots
age.male.breaks2 <- seq(5,95,5)
age.male.group2 <- apply(1*(outer (age.male, age.male.breaks2, "-") > 0), 1, sum) + 1
age.male.group2 <- table(factor(na.omit(age.male.group2), levels=0:19))
age.male.col2 <- rep(c("#2171B5","#6BAED6"),length.out=length(age.male.breaks2))
```

```
age.female.breaks2 <- seq(5,95,5)
age.female.group2 <- apply(1*(outer (age.female, age.female.breaks2, "-") > 0), 1, sum) + 1
age.female.group2 <- table(factor(na.omit(age.female.group2), levels=0:19))
age.female.col2 <- rep(c("#AE017E","#F768A1"),length.out=length(age.female.breaks2))
par(mfrow=c(1,1))
male.group.bp <- barplot(-1*age.male.group2, horiz=T,xlim=c(-300,300), col=age.male.col2,
cex.names=0.75,
    names.arg=c("0-5","5-10","10-15","15-20","20-25","25-30","30-35","35-40",
        "40-45","45-50","50-55","55-60","60-65","65-70","70-75","75-80",
        "80-85","85-90","90-95","95-100"), xaxt="n", las=2, ylab="Age")
female.group.bp <- barplot(age.female.group2, horiz=T, add=T, col=age.female.col2, xaxt="n",
yaxt="n")
axis(side=1, at=seq(-300,300,by=50),
labels=c("300","250","200","150","100","50","0","50","100","150","200","250","300"), cex=0.7)
mtext(side=1, text="Counts", line=2)
title("Age Distribution - 2010")
mtext(side=3, text="Male                                         Female")
box()

##~ Graph 6 ~##
#Side-by-side violin plots of income levels for owning and renting
wvioplot(det.data.2010$income.male, det.data.2010$income.female, names=c("Male",
"Female"), col="coral2")
title("Violin Plot of Income Levels by Gender in Detroit in 2010", ylab="Income")

##Races
his.dist.2000 <- det.data.2000$hispanic/det.data.2000$pop2000
  his.dist.2000[which(is.nan(his.dist.2000))] <- 0
his.dist.2010 <- det.data.2010$P0040003/det.data.2010$P0040001
  his.dist.2010[which(is.nan(his.dist.2010))] <- 0
white.dist.2000 <- det.data.2000$white/det.data.2000$pop2000
  white.dist.2000[which(is.nan(white.dist.2000))] <- 0
white.dist.2010 <- det.data.2010$P0030002/det.data.2010$P0030001
  white.dist.2010[which(is.nan(white.dist.2010))] <- 0
black.dist.2000 <- det.data.2000$black/det.data.2000$pop2000
  black.dist.2000[which(is.nan(black.dist.2000))] <- 0
black.dist.2010 <- det.data.2010$P0030003/det.data.2010$P0030001
  black.dist.2010[which(is.nan(black.dist.2010))] <- 0
asian.dist.2000 <- det.data.2000$asian/det.data.2000$pop2000
  asian.dist.2000[which(is.nan(asian.dist.2000))] <- 0
asian.dist.2010 <- det.data.2010$P0030005/det.data.2010$P0030001
  asian.dist.2010[which(is.nan(asian.dist.2010))] <- 0
####
```

```
##~ Graph 7 ~##
##Density plot of income variation over races
#2010 Male
par(mfrow=c(1,1))
#the densities use the weights argument. the densities use the income data from every
#blkgrp but adds a weight to it proportional to the dominance of a given race in that
#blkgrp
white.male.weights <- white.dist.2010/sum(white.dist.2010)
  white.male.weights <- white.male.weights[-which(is.na(det.data.2010$income.male))]
  white.income.male.2010.dens <- density(na.omit(det.data.2010$income.male),
                                  weights=white.male.weights/sum(white.male.weights))
asian.male.weights <- asian.dist.2010/sum(asian.dist.2010)
  asian.male.weights <- asian.male.weights[-which(is.na(det.data.2010$income.male))]
  asian.income.male.2010.dens <- density(na.omit(det.data.2010$income.male),
                                  weights=asian.male.weights/sum(asian.male.weights))
his.male.weights <- his.dist.2010/sum(his.dist.2010)
  his.male.weights <- his.male.weights[-which(is.na(det.data.2010$income.male))]
  his.income.male.2010.dens <- density(na.omit(det.data.2010$income.male),
                                weights=his.male.weights/sum(his.male.weights))
black.male.weights <- black.dist.2010/sum(black.dist.2010)
  black.male.weights <- black.male.weights[-which(is.na(det.data.2010$income.male))]
  black.income.male.2010.dens <- density(na.omit(det.data.2010$income.male),
                                  weights=black.male.weights/sum(black.male.weights))
plot(white.income.male.2010.dens, col="Orange",xlab="Yearly Income (in US Dollars)",
     ylim=c(0,max(c(white.income.male.2010.dens$y,asian.income.male.2010.dens$y,
           his.income.male.2010.dens$y,black.income.male.2010.dens$y))),
     main="Income Distribution for Males of Various Races \nin Detroit in 2010")
lines(asian.income.male.2010.dens,col="Red")
lines(his.income.male.2010.dens,col="Purple")
lines(black.income.male.2010.dens,col="Green")
legend("topright", legend=c("White Males", "Asian Males", "Hispanic Males", "Black Males"),
      col=c("Orange", "Red", "Purple", "Green"), lwd=2)

##~ Graph 8 ~##
##Density plot of income variation over races
#2010 female
par(mfrow=c(1,1))
#the densities use the weights argument. the densities use the income data from every
#blkgrp but adds a weight to it proportional to the dominance of a given race in that
#blkgrp
white.female.weights <- white.dist.2010/sum(white.dist.2010)
  white.female.weights <- white.female.weights[-which(is.na(det.data.2010$income.female))]
  white.income.female.2010.dens <- density(na.omit(det.data.2010$income.female),
                                    weights=white.female.weights/sum(white.female.weights))
```

```
asian.female.weights <- asian.dist.2010/sum(asian.dist.2010)
  asian.female.weights <- asian.female.weights[-which(is.na(det.data.2010$income.female))]
  asian.income.female.2010.dens <- density(na.omit(det.data.2010$income.female),
                                weights=asian.female.weights/sum(asian.female.weights))
his.female.weights <- his.dist.2010/sum(his.dist.2010)
  his.female.weights <- his.female.weights[-which(is.na(det.data.2010$income.female))]
  his.income.female.2010.dens <- density(na.omit(det.data.2010$income.female),
                              weights=his.female.weights/sum(his.female.weights))
black.female.weights <- black.dist.2010/sum(black.dist.2010)
  black.female.weights <- black.female.weights[-which(is.na(det.data.2010$income.female))]
  black.income.female.2010.dens <- density(na.omit(det.data.2010$income.female),
                                weights=black.female.weights/sum(black.female.weights))
options(scipen=10)
plot(white.income.female.2010.dens, col="Orange",xlab="Yearly Income (in US Dollars)",
    ylim=c(0,max(c(white.income.female.2010.dens$y,asian.income.female.2010.dens$y,
              his.income.female.2010.dens$y,black.income.female.2010.dens$y))),
    main="Income Distribution for Females of Various Races \nin Detroit in 2010")
lines(asian.income.female.2010.dens,col="Red")
lines(his.income.female.2010.dens,col="Purple")
lines(black.income.female.2010.dens,col="Green")
legend("topright", legend=c("White Females", "Asian Females", "Hispanic Females", "Black
Females"),
      col=c("Orange", "Red", "Purple", "Green"), lwd=2)

##~ Graph 9 ~##
##Most Dominant Race
#2000
par(mfrow=c(1,2))
#par(mar=par()$mar+c(0,0,0,5.2))
palette <- brewer.pal(4,"Accent")
col.vec.2000 <- vector()
for (yy in 1:length(det.data.2000)){
  if (black.dist.2000[yy] > 0.5)
    col.vec.2000 <- c(col.vec.2000,palette[1])
  else if (white.dist.2000[yy] > 0.5)
    col.vec.2000 <- c(col.vec.2000,palette[2])
  else if (his.dist.2000[yy] > 0.5)
    col.vec.2000 <- c(col.vec.2000,palette[3])
  else if (asian.dist.2000[yy] > 0.5)
    col.vec.2000 <- c(col.vec.2000,palette[4])
  else col.vec.2000 <- c(col.vec.2000,"gray")
}
plot(det.data.2010, col="white",border="white") #invisible
plot(det.data.2000, col=col.vec.2000,add=T,lwd=.5)
```

```
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
title("Race Distribution - 2000")
legend("bottomright", title="Majority Race",col=c(palette,"gray"), pch=15,
      legend=c("Black","White","Hispanic/Latino","Asian","No Majority"),xpd=T,cex=0.5)

##~ Graph 10 ~##
##2010
col.vec.2010 <- vector()
for (yy in 1:length(det.data.2010)){
  if (black.dist.2010[yy] > 0.5)
    col.vec.2010 <- c(col.vec.2010,palette[1])
  else if (white.dist.2010[yy] > 0.5)
    col.vec.2010 <- c(col.vec.2010,palette[2])
  else if (his.dist.2010[yy] > 0.5)
    col.vec.2010 <- c(col.vec.2010,palette[3])
  else if (asian.dist.2010[yy] > 0.5)
    col.vec.2010 <- c(col.vec.2010,palette[4])
  else col.vec.2010 <- c(col.vec.2010,"gray")
}
plot(det.data.2010, col=col.vec.2010,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
title("Race Distribution - 2010")
legend("bottomright", title="Majority Race",col=c(palette,"gray"), pch=15,
      legend=c("Black","White","Hispanic/Latino","Asian","No Majority"),xpd=TRUE,cex=0.5)

##~ Graph 11 ~##
##Asian / Total Pop
asian.dist.2000.breaks <- c(0.1,0.3,0.5,0.7,0.9)
asian.dist.2000.col <- brewer.pal(6,"Reds")
asian.dist.2000.group <- apply(1*(outer (asian.dist.2000, asian.dist.2000.breaks, "-") > 0), 1,
sum) + 1
plot(det.data.2010, col="white",border="white") #invisible
plot(det.data.2000, col=asian.dist.2000.col[asian.dist.2000.group],add=T,lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
title(main="Asian Race Distribution in \nDetroit in 2000")
legend("bottomright", title="Asian Race \nPercentage",col=c(asian.dist.2000.col,"white"),
pch=15,
      legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
      xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)
asian.dist.2010.breaks <- c(0.1,0.3,0.5,0.7,0.9)
```

```
asian.dist.2010.col <- brewer.pal(6,"Reds")
asian.dist.2010.group <- apply(1*(outer (asian.dist.2010, asian.dist.2010.breaks, "-") > 0), 1,
sum) + 1
plot(det.data.2010, col=asian.dist.2010.col[asian.dist.2010.group],lwd=.5)
  points(x=-83.0245, y=42.33, col="black", pch=19, cex=1)
  text(x=-83.0245, y=42.319, labels="Detroit")
title(main="Asian Race Distribution in \nDetroit in 2010")
legend("bottomright", title="Asian Race \nPercentage",col=c(asian.dist.2000.col,"white"),
pch=15,
      legend=c("0% - 10%", "10% - 30%", "30% - 50%", "50% - 70%", "70% - 90%", "90% - 100%"),
      xpd=TRUE, cex=0.5,bty="n")
rect(-83.03481,42.28594-0.1300292,-83.03481+0.1411214,42.29594)
```