Maryland County Analysis

Nate Kaduk

ECON 42050

Spring 2024

Executive Summary

This report seeks to compare the differences across Maryland counties, with respect to age, income, education, race, and sex. The exact counties being studied are Anne Arundel County, Baltimore County, Cecil County, Charles County, and Montgomery County. For each Maryland county, the history for each county was researched and a brief history will be discussed in the introduction.

Next, a few subsample analyses comparing the overall distribution of income, age brackets, gender, and race distribution in Maryland counties will be performed. Then, a further subsample analysis will be performed in each county comparing the educational attainment, income, and age for minority groups in each Maryland county. To help get a further perspective of Maryland counties, this analysis will also be extended to sex.

There were several results for this report. First, race and education are very useful in predicting which county is which. Second, Anne Arundel and Baltimore counties are extremely similar when it comes to income and education, although Anne Arundel is slightly better. Third, each county has disparities in gender income, with Cecil being the worst on this front. Fourth, Charles County has a high minority population, and is the best out of all these counties when it comes to the gender income gap. This is most likely due to the migration of affluent African American females into the county. Fifth, Montgomery County has the richest and most-educated population whereas Cecil County has the worst in both categories.

The data used in this project came from 2022 Census data using IPUMS. IPUMS allows users to retrieve census data for particular years. The uncleaned data in this project came with numerically-coded attributes income, age, race, education, state/county FIPS (Federal Informational Processing Standards) codes, and sex for many observations across the US. The uncleaned data had 3,373,378 observations whereas the cleaned data only had 22,266 observations. Specifically, this data was cleaned to include only adults, only specific Maryland Counties, and positive/identifiable yearly incomes. Some numeric variables such as race, sex, and education were grouped into categories. More details on the cleaning are included in the introduction.

Introduction

The goal of this project is to compare different aspects of Maryland counties and see how those counties differ from one another. This will be done at the county-level, with the counties being compared directly, and at personal level, where different subsamples (race and gender) for each county are taken and analyzed.

The aspects being used for comparison will be income, age, sex, race, and education. The distribution of these county attributes will be compared across counties to answer questions such as:

- Which county is the richest when it comes to income? Which county is the poorest?
- Are there any counties that have a large minority population?
- Which counties are the best-educated? Which are the worst?
- Do smaller counties in Maryland fare better or worse in these areas than larger counties?
- How do the above answers stack up when breaking down a county into different sub-populations?

The five Maryland counties being studied are Baltimore, Montgomery, Anne Arundel, Charles, and Cecil. The first three are large counties representing major cities. Baltimore County surrounds Baltimore City, Anne Arundel is home to the city of Annapolis (capital of Maryland), while Mongomery is one of the counties bordering Washington DC (capital of the United States). Attributes from these cities may bleed into their respective counties, causing them to take on similar aspects. As such, a brief history of these cities and the present-day synopsis of each county is below.

Baltimore was created on the Patapsco River (Presently known as the Inner Harbor), a river touching the Chesapeake Bay [1]. The Baltimore and Ohio railroad became operational in 1830, which brought a wave of economic prosperity to Baltimore rivaling that of the Erie Canal. Baltimore has seen a lot of immigration and migration over the years. Many African Americans went to Baltimore post-civil war. An influx of Italians and Germans immigrated to Baltimore during the pre-world war era. Still, Baltimore has seen its share of problems. During the 1990s, Baltimore was one of the worst in the nation when it came to public health, violence, and

education. However, education programs have gotten much better with the addition of a CitiStat program which changes the previous local government into a results-based one.

Present-day Baltimore County is filled with different colleges. Several colleges are present around the Towson area alone, namely Towson University, Goucher University, Loyola University, and with a bit more travel, Johns Hopkins University. Baltimore today still has a large amount of crime, ranking 3rd most in the country for violent crime [2].

During the 1800s, the majority of Annapolis's economy came from fishing and shipping [3]. Annapolis has always been high on architecture. A report of George Washington mentioned that he wanted the capitol of the US to be Annapolis because of its architecture (It became the capitol of the US for about nine months). Furthermore, Annapolis has spent a good chunk of money on maintaining historic buildings [4].

Anne Arundel today still retains a solid fishing/shipping economy. This stems from the county being in the heart of Maryland with a side completely touching the Chesapeake Bay. Furthermore, the Bay Bridge, which connects Anne Arundel to the other side of Maryland, is a very commonly used bridge. This is because it allows for direct paths to both Delaware and Ocean City.

Washington DC was built with the premise of connecting the Eastern seaboard with westward frontiers [5]. The War of 1812 saw many federal buildings burnt down, crippling the architecture. It wasn't until 1830, when new railroads allowed tourists to enter the city, that the economy thrived again. After the Civil War, the city started to become a scenic place, when numerous infrastructure changes were made both in agriculture and in city projects. Not long after, many "affluent intellectuals" started moving to the city. Since the 1900s, Washington DC has grown in economic prosperity, with only a few hiccups along the way.

Modern day Montgomery County is home to several affluent cities such as Silver Spring. It is also the location for the National Institute of Health along with many government research facilities [6]. Their economy is primarily health, research, and business.

Both Charles and Cecil counties are small, with some minor advancements throughout the years. Charles County was founded in 1658 [7]. In 1903, the McDonough Institute opened,

which was the first high school in the county. The only college in the county is the College of Southern Maryland, and it was established in 2000. In 1674, Cecil County was created from Baltimore and Kent Counties [8]. During the 1800s, railroads linked Cecil County to areas in Delaware and Baltimore. Cecil Community College opened in 1968.

Data on these counties was created and obtained from United States census data using IPUMS [9]. The dataset obtained contains county and state FIPS codes for all recognizable counties in the US. Through selecting certain aspects of the census data, observations in the data contain information on surveyed individuals such as sex, age, income, race, and education. To focus on specific Maryland counties, all observations without Maryland's state FIPS code were dropped, then all observations with county FIPS codes not belonging to the counties specified above were dropped.   As IPUMS numerically hardcodes its data, the variables sex, race, and education were restricted to specific groups for this analysis, namely

- Less than high school, high school, bachelor's, master's and above for education
- Male/Female for sex
- Black, White, and Other for race.

The data was also restricted based upon age. In this report only those who are 18 or older (adults) are included. Observations with negative yearly incomes were removed, as those with observations containing system missings. As the data used from IPUMS was representative of the year 2022, all yearly incomes were adjusted for inflation by multiplying by 1.06 (value found by CPI calculator) in order to make the incomes more representative of 2024 dollars.

Section 0: Pre-Comparison Analysis

One important thing to note is that all counties are different. Each county has its own distribution of income, sex, race, education, and age. Taking this into account, figure 0.1. below illustrates how well these factors can be used in determining which county is which. This is done both on an individual level (mean accuracy) and an overall influence level (Mean Gini).

**Figure 0.1. Random Forest Pre-Analysis**
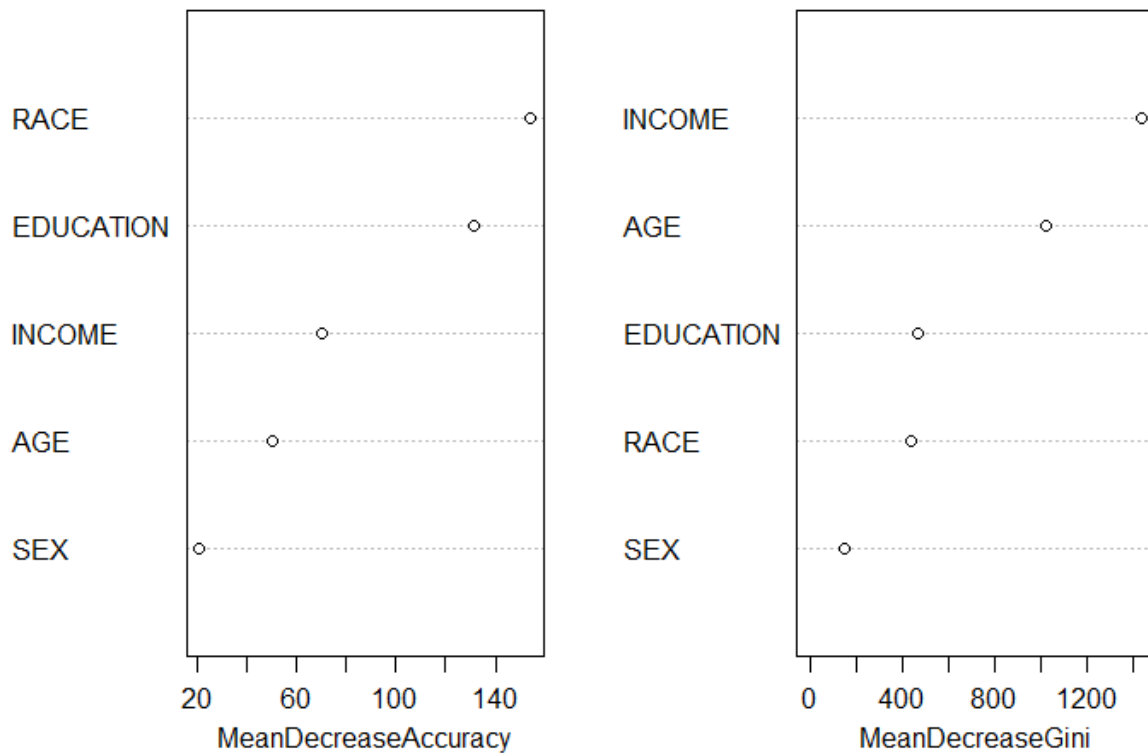
Mean Decrease Accuracy and Mean Decrease Gini chart for different variables in the data.

The MeanDecreaseGini corresponds to the importance of that variable in a decision-tree split to identify a county. Income is the most important as there are plenty of unique incomes across all counties. Education, Race, and Sex have the least number of possible options, and so they would result in a limited split for the decision tree.

The MeanDecreaseAccuracy measures how accuracy would be lost if the variable was removed from a prediction model. Race is the most irreplaceable variable here, meaning that if race was dropped, the prediction would not be as good. Similarly, as education is the second-best, there is likely enough education deviation between the counties that removing the education term would be detrimental. Sex is the worst, as each county can be assumed to be fifty percent male and fifty percent female. This leads to sex having a very small contribution to information gain when determining counties.

Section 1: Comparing Counties

Figure 1.1. Number of Observations in Maryland Counties
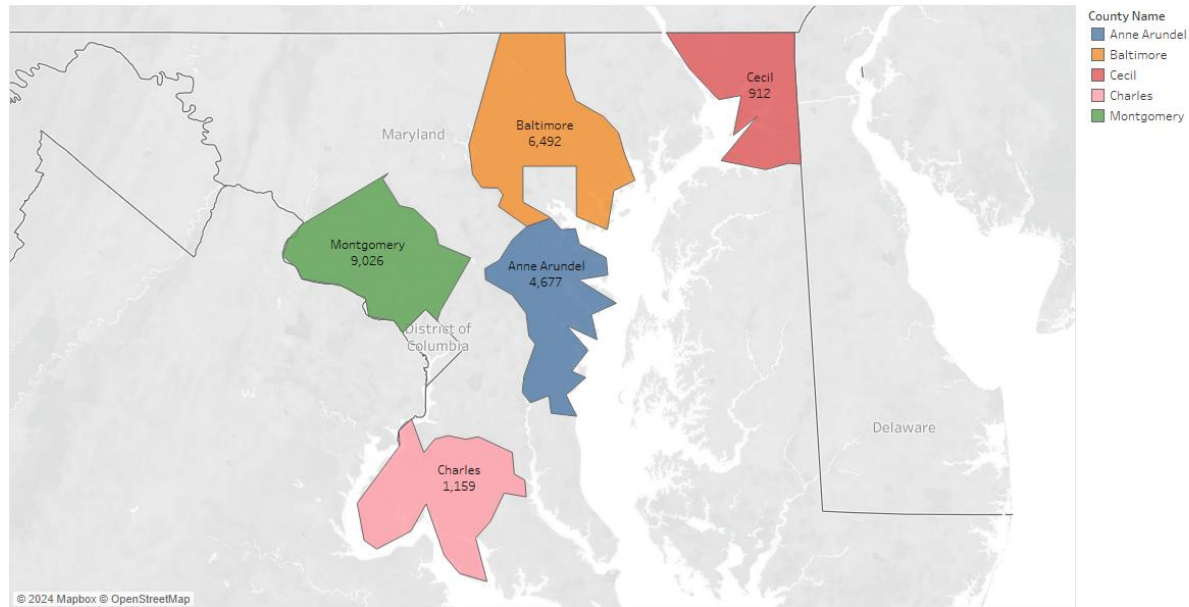

Map of the Maryland counties being studied. Below the name of each county is the number of observations (respondents) present for that county.

Each county has its own size. However, Montgomery is the largest county being studied with a population of 1,052,521 in 2022 [10] and has a total of 9,026 observations from IPUMS. The smallest county is Cecil with a 2022 population of 104,942 and 912 observations from IPUMS. For the purpose of this report the counties Anne Arundel, Baltimore, Cecil, Charles, and Montgomery are labeled with the colors blue, orange, red, pink and green respectively. The difference in populations are most likely due to the city sizes the counties represent. Washington DC has the largest population, Baltimore has the second-most , and Annapolis (Anne Arundel) has the third. The small counties have no big cities nearby and as such have a smaller population than the big three.

Like with population, each county each has its own degree of wealth. Consider figure 1.2. below.

**Figure 1.2. Income Distribution for Maryland Counties**
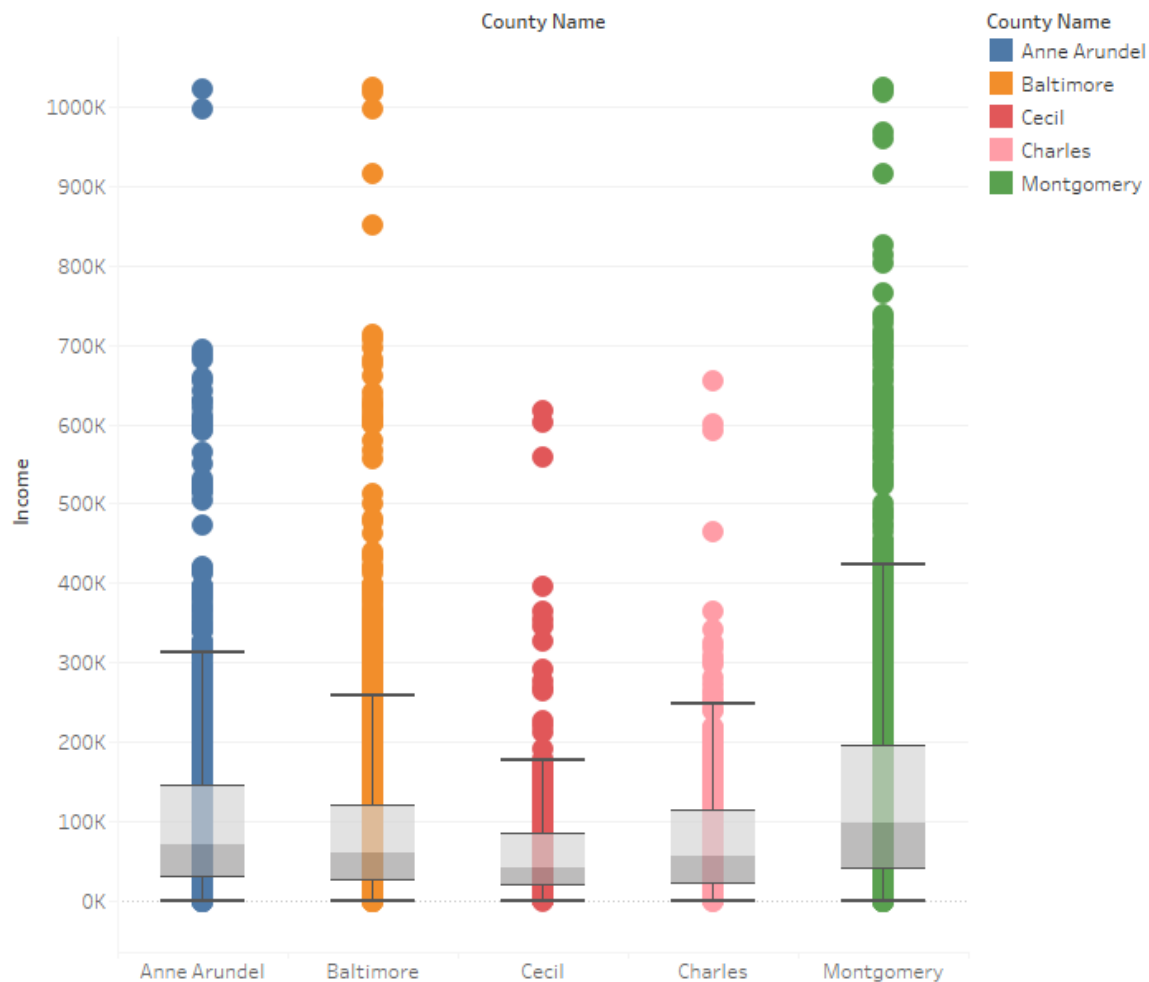
Box plot of yearly incomes representing a 99% confidence interval for each Maryland county.

Both Cecil County and Charles County have the smallest 99% maximum income and overall maximum income. The largest overall belongs to Baltimore, Anne, Arundel, and Montgomery counties, which all have maximum incomes (due to the top code being 999,998 and then taking into account inflation). Montgomery has the largest maximum for the 99% confidence interval. Anne Arundel has a larger median/maximum 99% confidence interval income than all of the counties except for Montgomery. As education can influence overall income, the figure below displays the education distribution for each county.

Figure 1.3. Education Distribution in Maryland Counties
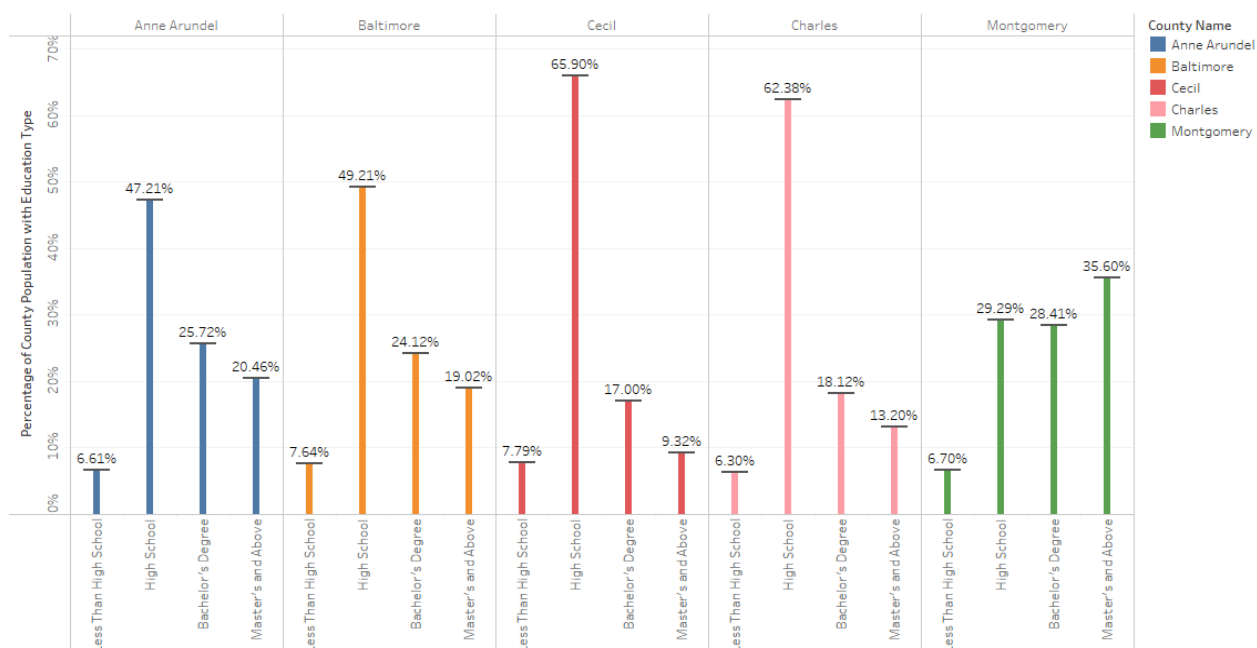
Distribution of education for each Maryland county based upon the percentage of a county's population with that degree.

Baltimore County and Anne Arundel County are very similar when it comes to education distribution, as are Cecil County and Charles County. Anne Arundel is only slightly better educated than Baltimore County, with a higher percentage of their population belonging to Bachelor's or above education types and a smaller percentage being without a high school diploma. The two smaller counties, Cecil and Charles, are less educated than the larger counties. An overwhelming majority of their populations only have a high school diploma. With this being said, Charles County is slightly more educated than Cecil. Charles even has the lower overall percentage without a degree at all.

Montgomery County is by far the best educated county here. Their master's and above population outweigh their high school diploma population. Furthermore, the percentage of those with a bachelor's degree is almost the same as the percentage of those with just a high school diploma. These two county characteristics contribute to Montgomery County having a better education than the other counties.

Some possible explanations of the above distributions include:
- Montgomery is filled with "affluent intellectuals" living around the DC area, leading to a higher education distribution than the other counties.
- Montgomery County is very research-based, and thus relies on people with degrees for their research.
- Cecil and Charles have the least amount of education because both counties only have a single college in them.
  - Plus, both colleges have only been open for less than 70 years.
- Cecil and Charles have the least amount of education because their colleges have only opened within the last century, giving less time for people to get a degree.
- Baltimore County and Anne Arundel County are very similar because they are so close to one another.

Each county also has their own distribution of race, as shown below.

Figure 1.4. Distribution of Race in Maryland Counties
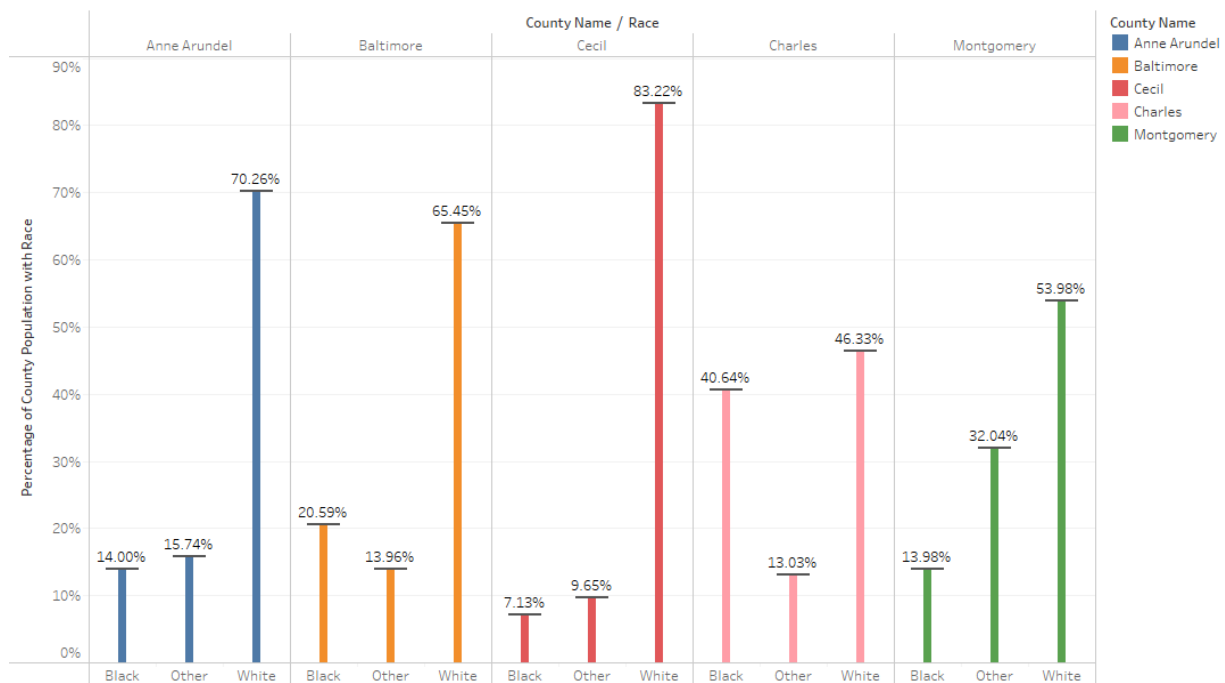
Distribution of race for each Maryland county based upon the percentage of a county's population belonging to a race.

Anne Arundel and Baltimore are very similar when it comes to race, only Baltimore's Black population is slightly more prominent. This might be because African American families moved to Baltimore after the Civil War. Cecil County is predominately white, more so than the other counties. Charles has a massive black population (possible reason for which is explained in the sex subsample analysis). Montgomery County has a large population of other races.

To compare these counties on a more individual level, two subsample analyses were performed: one for race, the other for sex.

Section 2: Race Group Subsample Analysis Based on Income

Each county has its own yearly income distribution when compared to race. The smaller counties have a lower maximum income for all races when compared to the larger counties. As a result, comparing a smaller county's distribution of income with a larger one is rough. However, when comparing the 99% confidence intervals for the races in a county, a fair comparison can be made.

Concerning just the small counties, Cecil has a much lower median and maximum income for its Black group than either the White or Other race groups. The White group has a higher median/maximum 99% confidence interval than the Other group. By contrast, the distribution of income by race in Charles County is roughly the same when comparing Black and Other groups. Like with Cecil, those who are white have the highest income, both when comparing 99% confidence interval maximum and median. Here, there exist a few outliers in the Other category, one of which surpasses the majority of white's income distribution.

Figure 2.1. Income Distribution in Cecil County
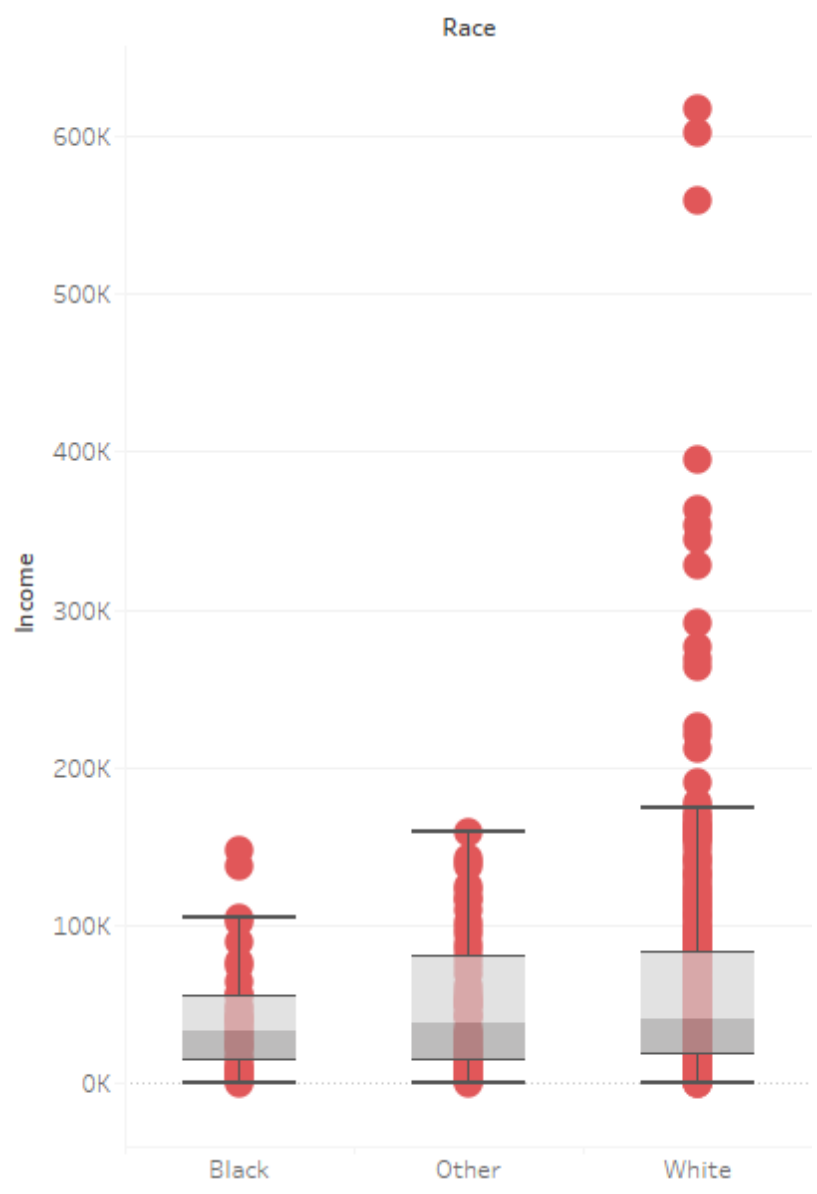
Boxplot represents a 99% confidence interval for income in Charles County, separated by race.
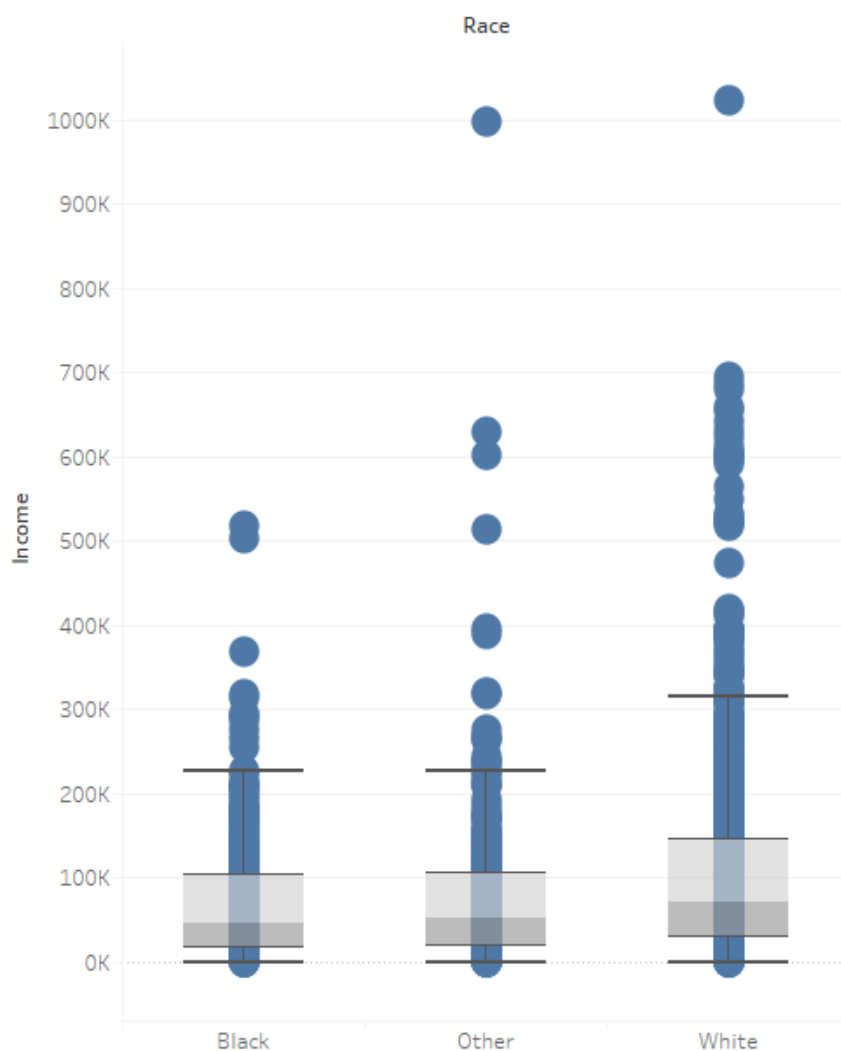
Figure 2.2. Distribution of Income by Race in Charles County
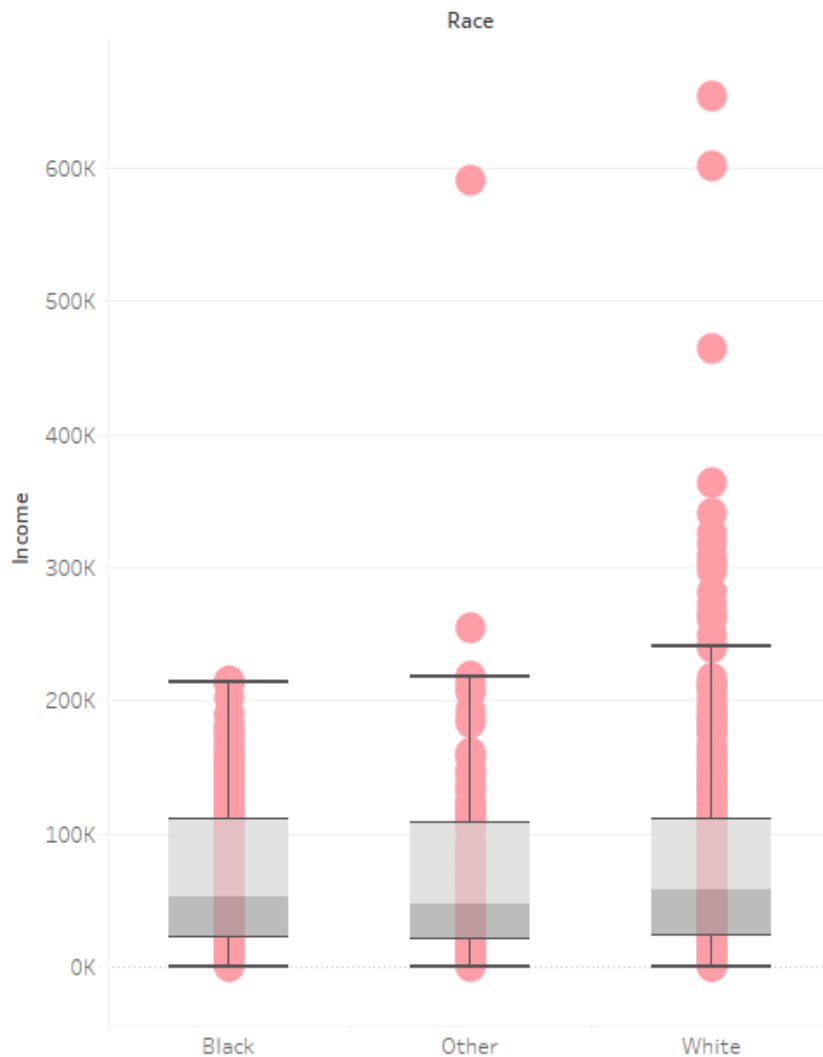
Boxplot represents a 99% confidence interval for income in Charles County, separated by race.

Moving to the larger counties, the income distributions between Anne Arundel County and Baltimore County are astronomically similar, as can be seen by figures 2.3. and 2.4. below.
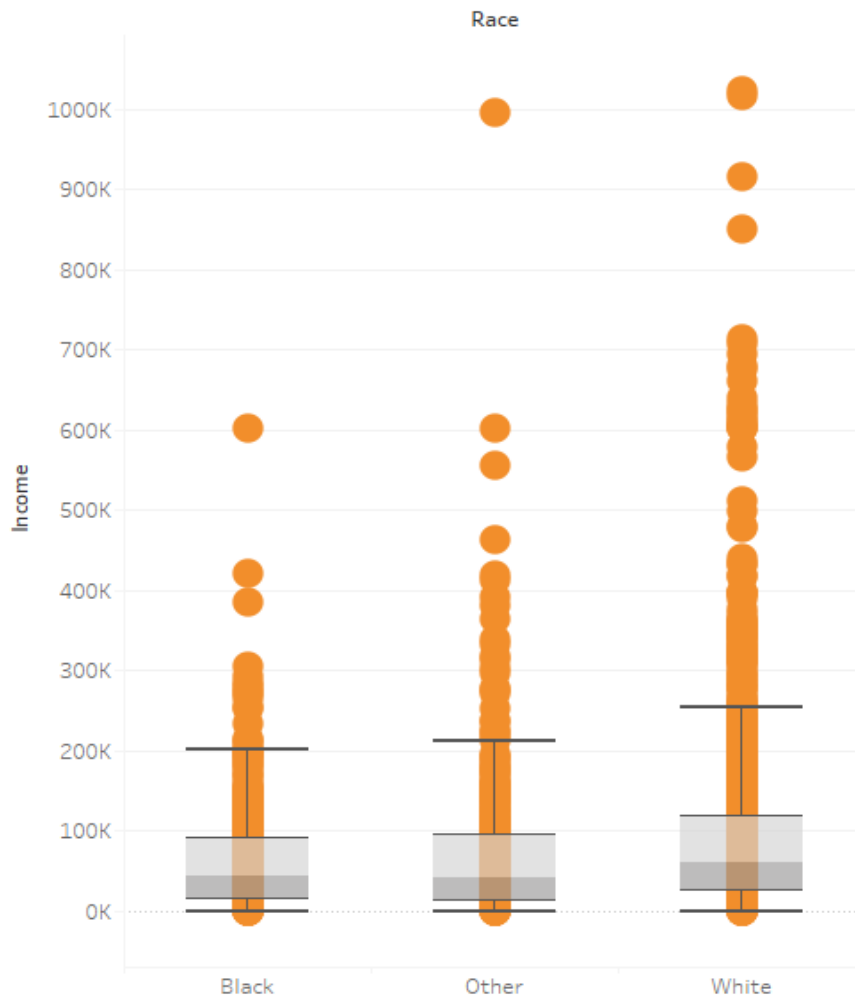
Figure 2.3. Distribution of Income by Race in Anne Arundel County

Boxplot represents a 99% confidence interval for income in Anne Arundel County, separated by race.

The confidence interval median incomes for those who are black or other races are roughly equivalent at around $50,000 a year, whereas this value jumps to $75,000 for those who are white. A similar argument holds for Baltimore County below.

Figure 2.4. Distribution of Income by Race in Baltimore County

Boxplot represents a 99% confidence interval for income in Baltimore County, separated by race.

Both counties have almost the same 99% confidence interval boxplot for the Black and Other groups. Similarly to the smaller counties, the median and maximum income for the White group is higher than both the Black or Other groups. Montgomery, on the other hand, is a far more interesting story.

Figure 2.5. Distribution of Income by Race in Montgomery County
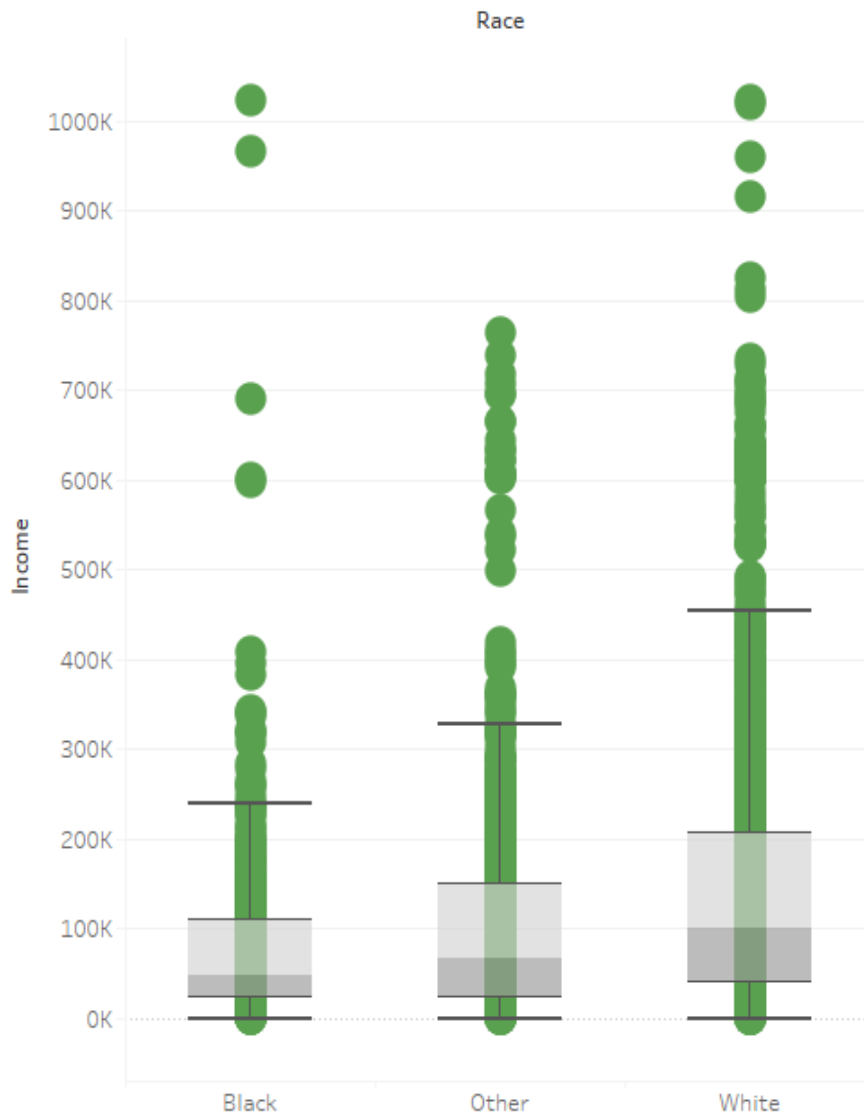
Boxplot represents a 99% confidence interval for income in Montgomery County, separated by race.

In Montgomery County, the 99% confidence interval for all categories is roughly the same as Cecil County in terms of comparison between the groups. However, the income for each group is much more. The median incomes for Montgomery County are approximately $50,000, $75,000, and $100,000 for the groups Black, Other and White respectively. The maximum of the 99% confidence interval is Above $200,000 for each group. Another interesting story on this graph is the outliers. For those of other races, the top incomes range between $500,000-800,000, with not a single income being above this range. Although a good chunk of top

incomes for those who are black lie within $200,000-400,000, there are a few earners who make more than $900,000. Those who are white have the majority of their top earners making within $500,000-$800,000. However, they also have a couple of people with incomes above $800,000.

There is disparity in income for Montgomery County when comparing Black-Other groups and the Other-White groups. There is hardly any disparity between the Black-Other groups when just comparing the 99% confidence interval for the other counties (except Cecil). Still, both the Black and Other groups are not making as much as the White group in all counties.

Section 3: Race Group Subsample Analysis Based on Age

**Figure 3.1. Median Age by Race in Maryland Counties**
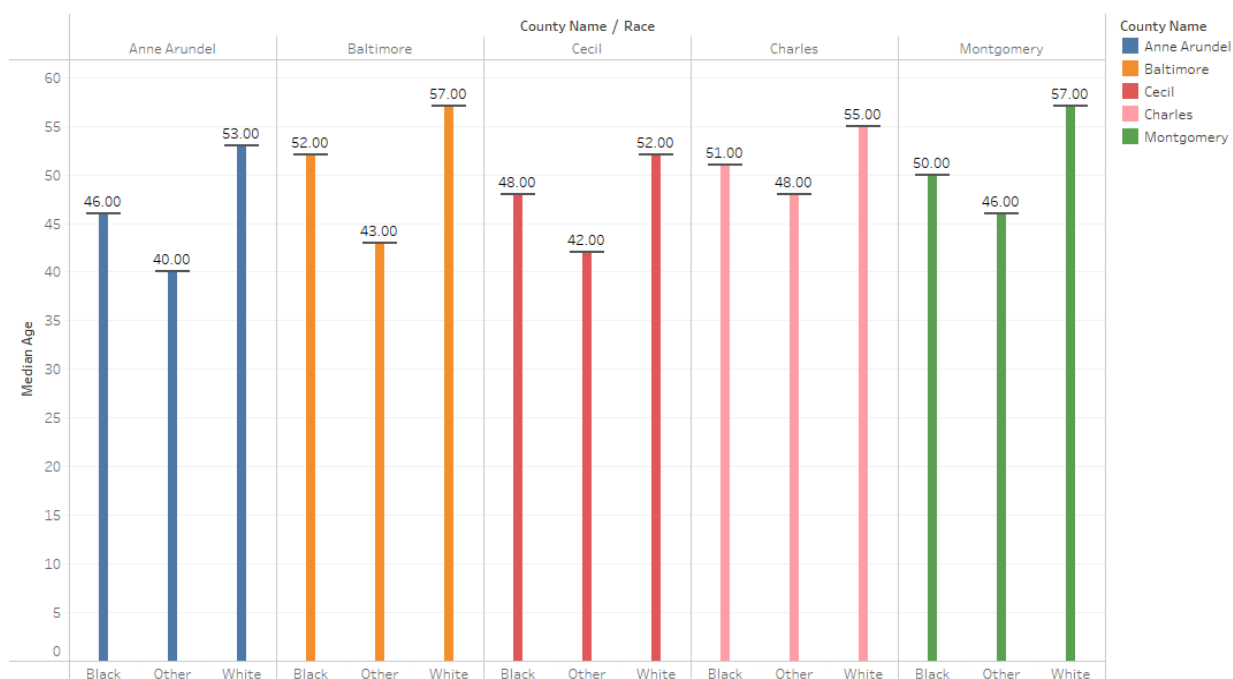
Median age by race group for each Maryland County.

In all counties, the median age for other races is less than the median age for the Black group, which is also less than the median age for the White group. With this being said, the people in Anne Arundel County and Cecil County are slightly younger in all races when compared to the other three counties. For Anne Arundel, the median age Black, Other, and White is 46, 40, and

53 respectively. The oldest county is Charles with a combined median age of 154 across all races.

Section 4: Race Group Subsample Analysis Based on Educational Attainment

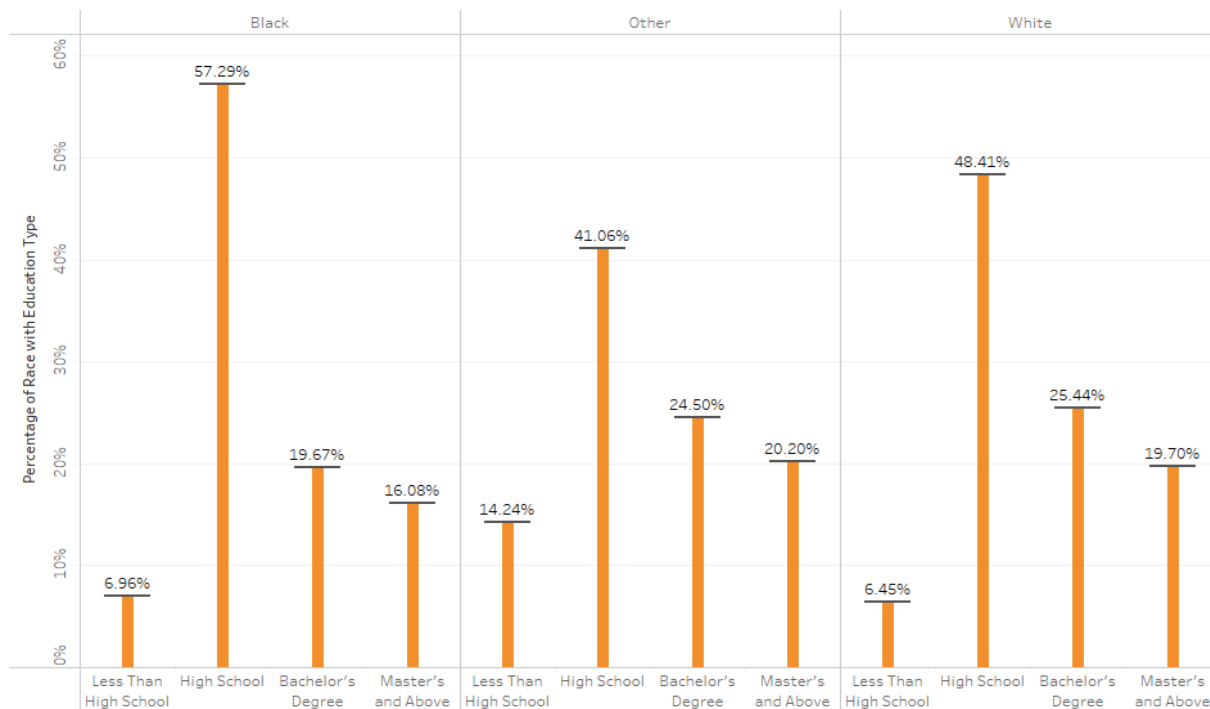**Figure 4.1. Education Distribution by Race in Anne Arundel County**
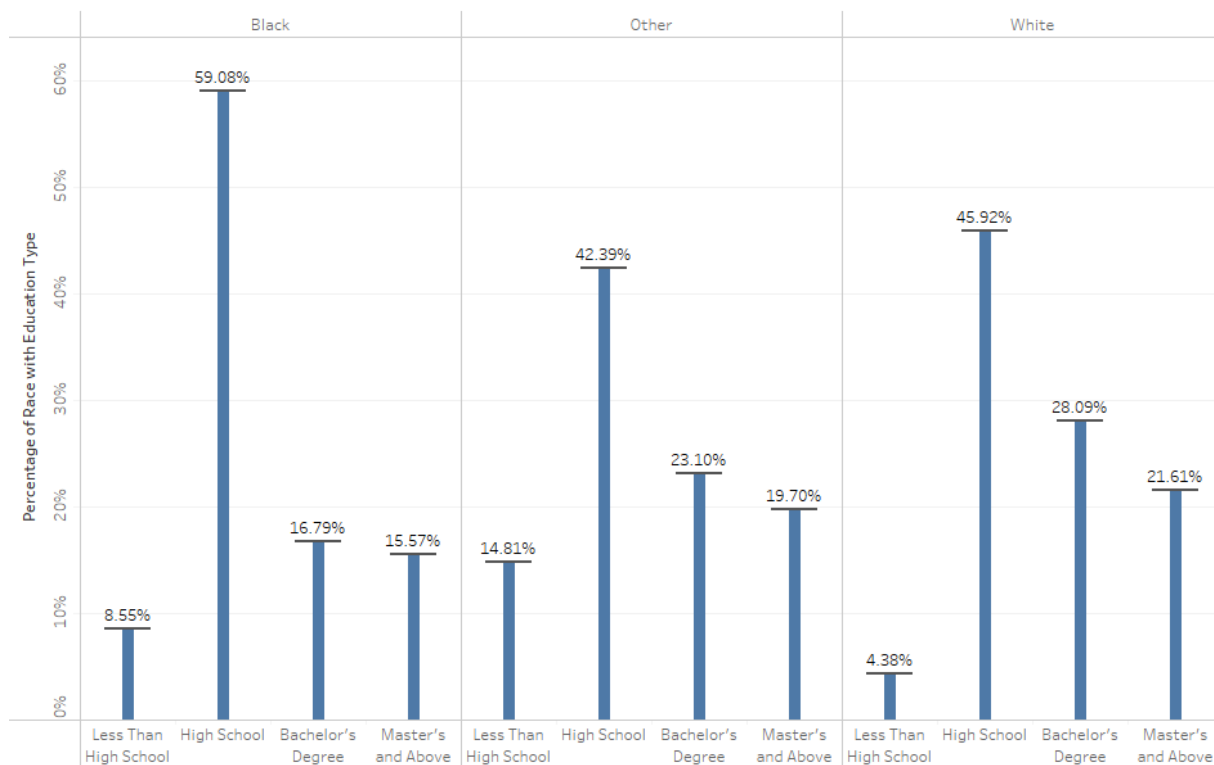
Education distribution bar charts separated by race for Anne Arundel County. Percentage indicates the percentage of race in Anne Arundel County with a particular degree.

In Anne Arundel County, those with races belonging to the Other group have more college degrees than those who are black (on average). However, those belonging to differing races also have an astronomically high percentage of their population with no degree at all (14.81%). Those who are white have a larger percentage with college degrees than the other groups. They also have the lowest percentage of those without a degree at all when comparing the

different races. Very similar trends exist when comparing the races in Baltimore County, as shown in figure 4.2 below.
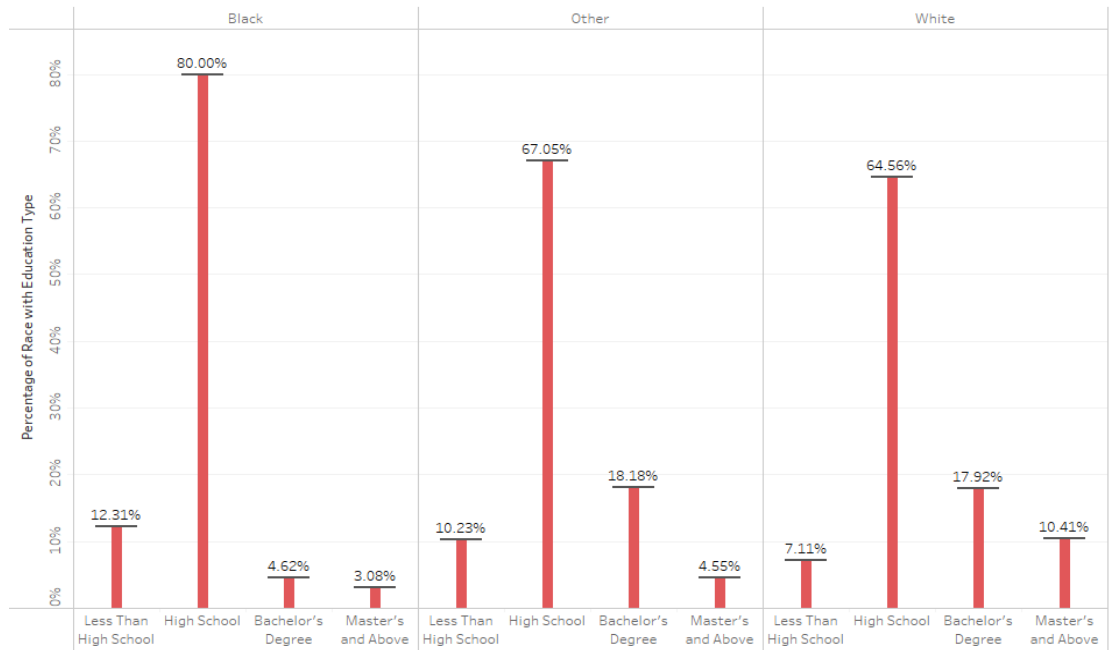
Figure 4.2. Education Distribution by Race in Baltimore County

Education distribution bar charts separated by race for Baltimore County. Percentage indicates the percentage of race in Baltimore County with a particular degree.

Both Baltimore and Anne Arundel are doing alright when it comes to education by race. The biggest disparity when it comes to education by race occurs in Cecil County (below). Each race has a large percentage of their population with a high school diploma or without a degree at all. However, the percentage of the Black population without a college degree is extremely significant compared to Other or White groups, as less than 10% of those who are black have a college degree in Cecil County. The White and Other groups are very similar, but there is a higher percentage of the white population with a master's degree, and less of the white population without a degree at all, than the Other group.

Figure 4.3. Education Distribution by Race in Cecil County

Education distribution bar charts separated by race for Cecil County. Percentage indicates the percentage of race in Cecil County with a particular degree.
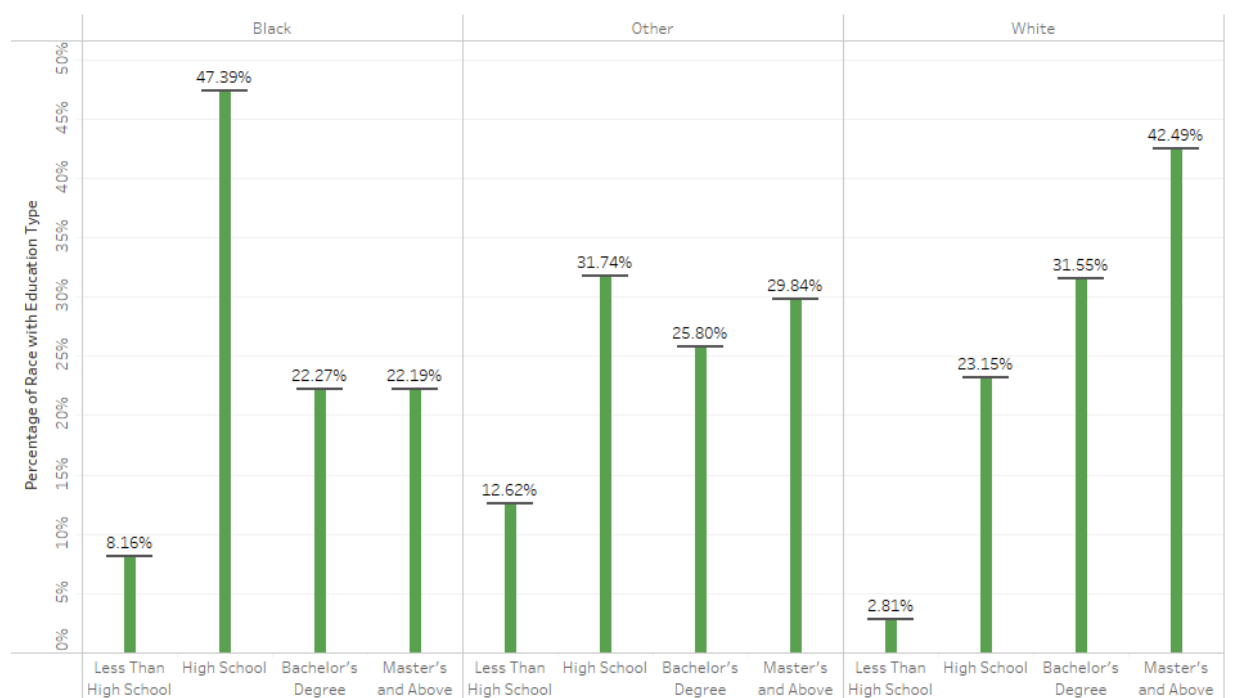
Figure 4.4. Education Distribution by Race in Charles County
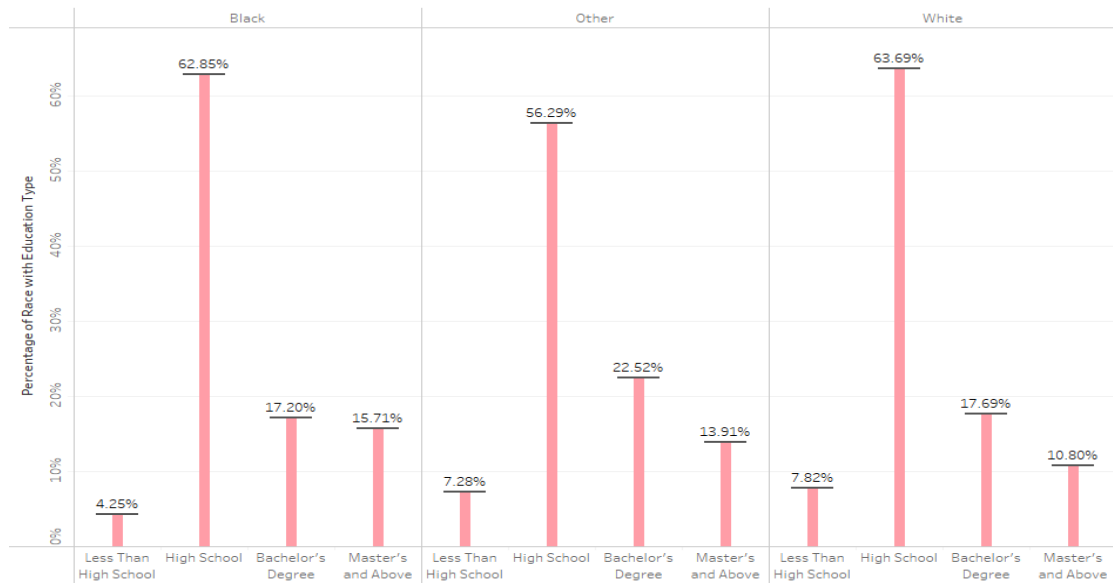
Education distribution bar charts separated by race for Charles County. Percentage indicates the percentage of race in Charles County with a particular degree.

In Charles County, the Other group is slightly more educated than Cecil County, with a higher percentage of their population having college degrees and a lower percentage being without a degree at all. In this county alone, the Black group is doing decently better than the White group in the sense that a higher percentage of the Black group has beyond a bachelor's (15.71%) and a much lower percentage has no degree at all (4.25%).

Figure 4.5. Education Distribution by Race in Montgomery County

Education distribution bar charts separated by race for Montgomery County. Percentage indicates the percentage of race in Montgomery County with a particular degree.

There is a fair amount of disparity in Montgomery County. Comparing Black and Other groups, there is a much higher percentage of the Other group population with a college degree than the percentage of the Black population with a college degree. Furthermore, the percentage of the Other group population with a bachelor's degree (25.80%) and the percentage with a master's or above (29.84%) are both almost surpassing the percentage of the population with only a high school diploma (31.74%). By contrast, the Black group has an extremely large percentage of their population with a high school diploma, enough that the previous claim cannot be made.

With this being said, there is a smaller percentage of the those who are black with no degree at all than those of other races. Comparing the Other group with the White group, there is a much lower percentage of the White group without a degree than the percentage of the Other group without a degree. Furthermore, the percentage population with only a high school diploma is less for the White group than it is for the Other group. These in conjunction lead to the White group having the percentage of their population with a bachelor's degree being higher than the percentage of their population with only a high school diploma. The same case can be made with the master's or above percentages.

With all of this being said, those in Montgomery County are extremely well-educated when comparing any race to the same race in other counties. Still, there is an education gap between races in Montgomery County.

Section 5: Sex Subsample Analysis Based on Income

Like with race, each county has its own distributions of income, education, and age when it comes to sexes. As such, it is beneficial to consider another subsample analysis in these categories: one of males and females. For the first part of this subsample analysis, each county will be studied based upon their income distribution by sex. As usual, the first county to be studied is Anne Arundel County. Consider the box and whisker plot of figure 5.1. below.
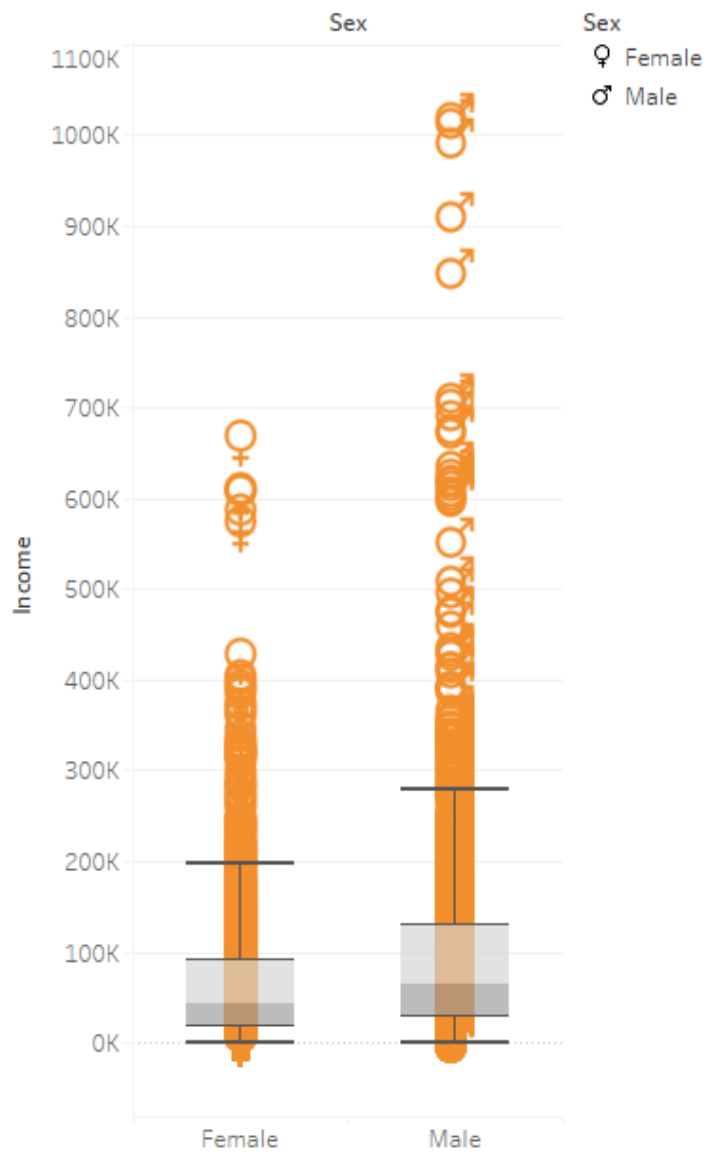
Figure 5.1. Distribution of Income by Sex in Anne Arundel County
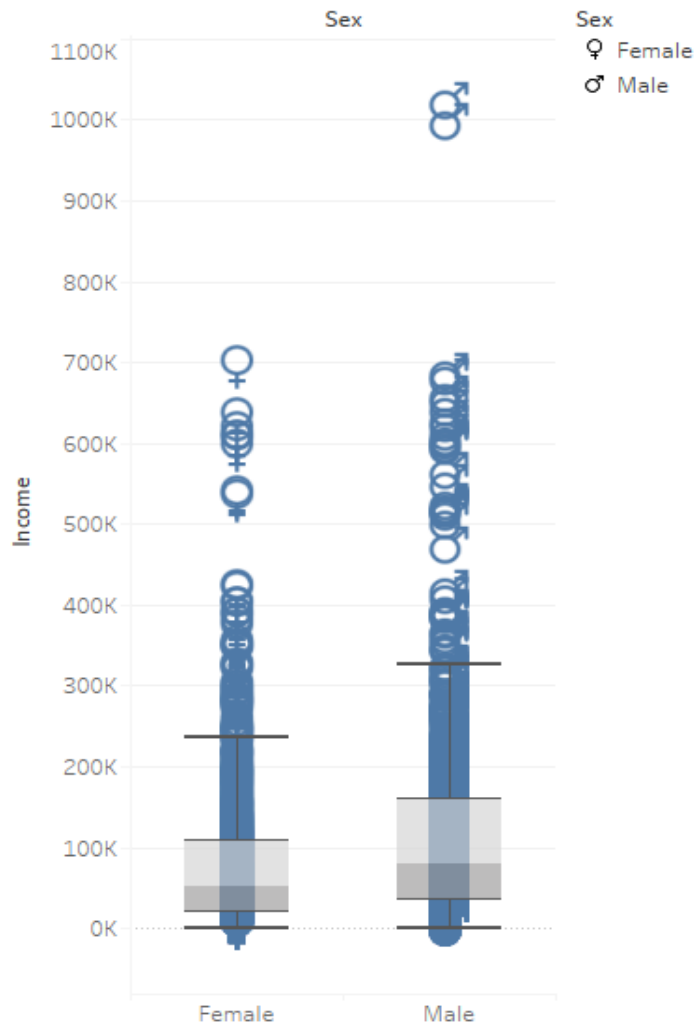
Boxplot represents a 99% confidence interval for income in Anne Arundel County, separated by Sex.

In Anne Arundel County, males have higher incomes than females   Males have a 99% confidence interval median around $75,000, but females only have a 99% confidence interval median of around $50,000. The maximum of the 99% confidence interval of the sexes is separated by $100,000. The overall maximum between the incomes is separated by nearly $350,000. The exact same observations can be made with Baltimore County.

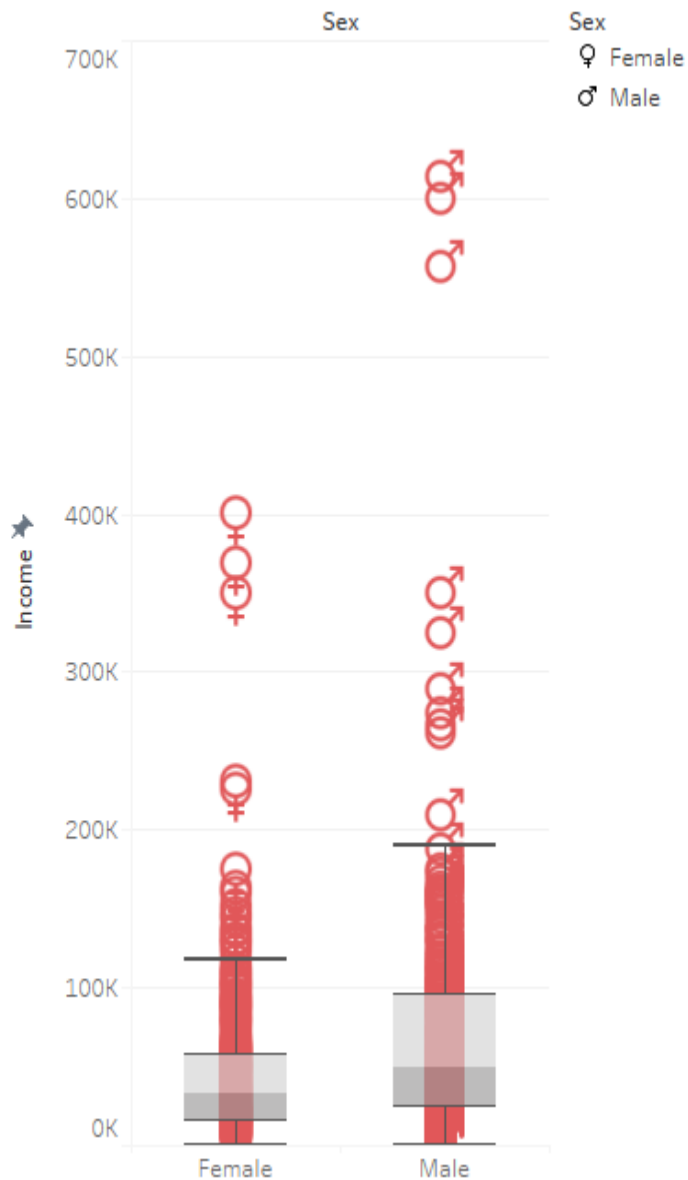**Figure 5.2. Distribution of Income by Sex in Baltimore County**

Boxplot represents a 99% confidence interval for income in Baltimore County, separated by Sex.

Cecil County, on the other hand, is slightly different with income by sex.

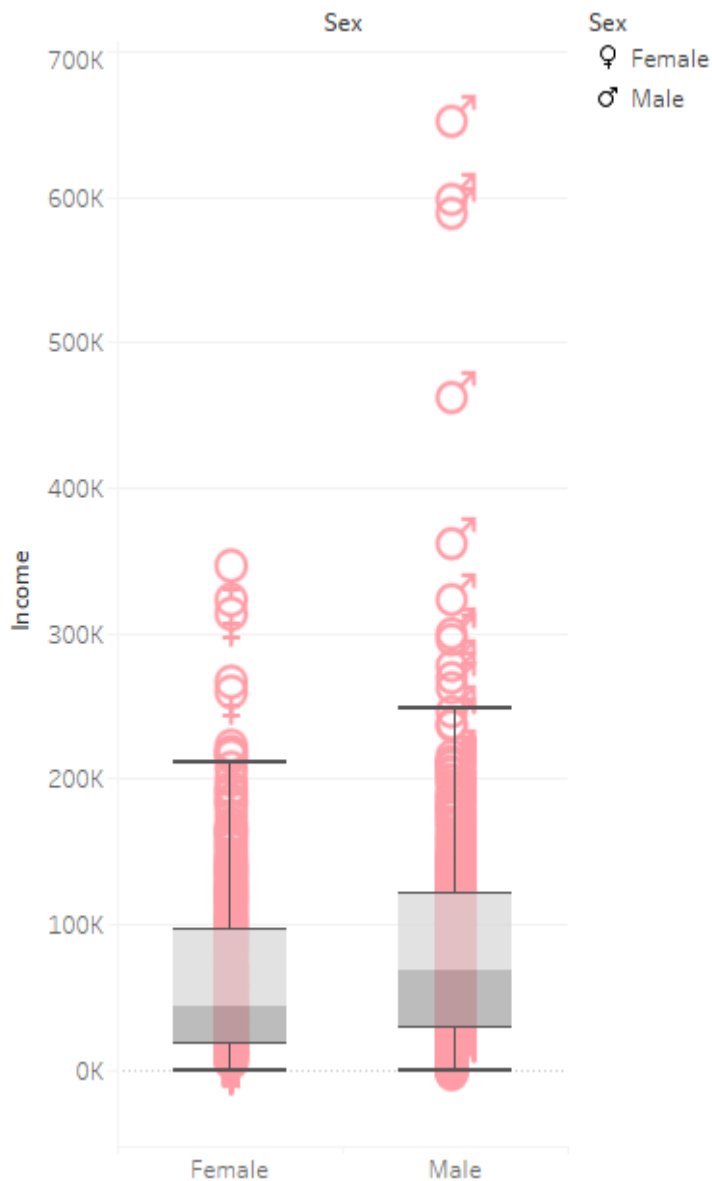**Figure 5.3. Distribution of Income by Sex in Cecil County**

Boxplot represents a 99% confidence interval for income in Cecil County, separated by Sex.

Cecil County does have an income gap between the sexes, like with Anne Arundel or Baltimore, and the differences between each sexes' boxplot graphs are similar to Anne Arundel or Baltimore. The median of their 99% confidence interval for income is around $25,000 for females and $50,000 for males (difference of $25,000). The maximum for the 99% confidence intervals is around $125,000 for females and $200,000 for males. By contrast with the bigger counties, the important cutoff points for females and males are smaller here than in Anne

Arundel or Baltimore. Furthermore, the maximum in general is around $400,000 for females and $600,000 for males (difference of $200,000).

Figure 5.4. Distribution of Income by Sex in Charles County
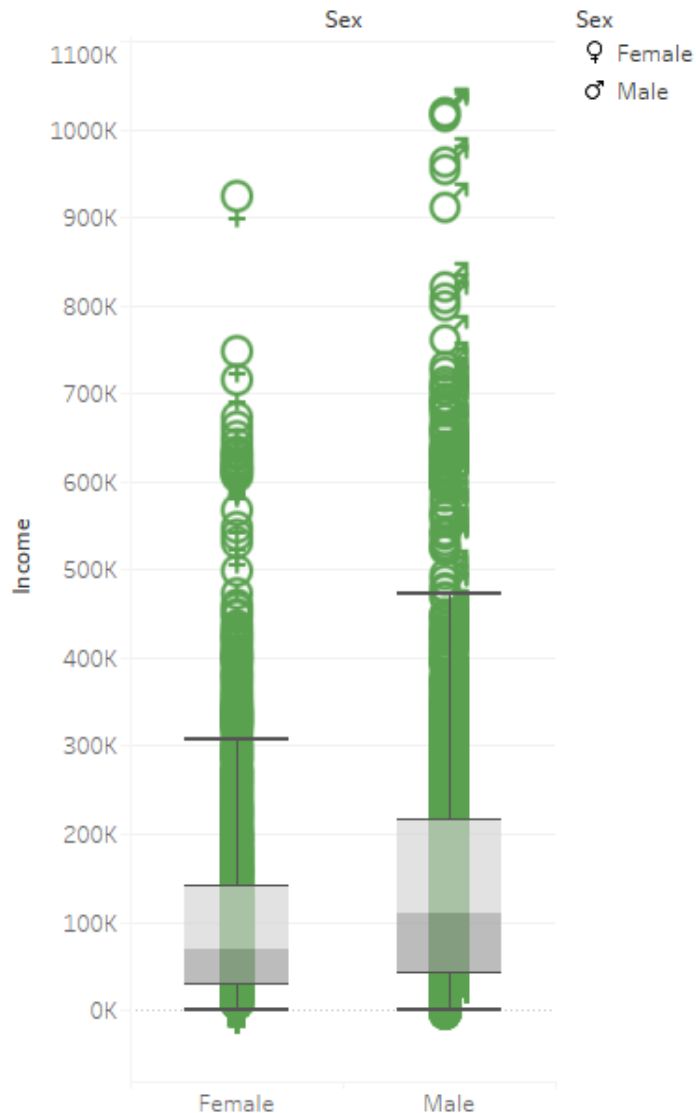

Boxplot represents a 99% confidence interval for income in Charles County, separated by Sex.

Like with the other counties, there is an income gap between males and females. The median income for females in Charles County is $50,000 while the median for males is $75,000. The maximum overall income for females is around $350,000, compared to the maximum income of

$650,000 for males. With this being said, Charles County is better with the 99% confidence interval maximum income than other counties. The 99% confidence maximum interval for females is $200,000 and $250,000 for males.

**Figure 5.5. Distribution of Income by Sex in Montgomery County**

Boxplot represents a 99% confidence interval for income in Montgomery County, separated by Sex.

The income disparity is present in Montgomery County as well. The median income of the 99% confidence interval has a $25,000 difference between males and females, similar to the other

counties. The income gap of the maximum of the 99% confidence interval is worse in Montgomery County than the other counties with a $200,000 gap. The maximum outliers, however, have less of a gap compared to other counties with only a $150,000 gap.

Figure 5.6. Percentiles of Income (75th, Median, 25th) by Sex in Maryland Counties
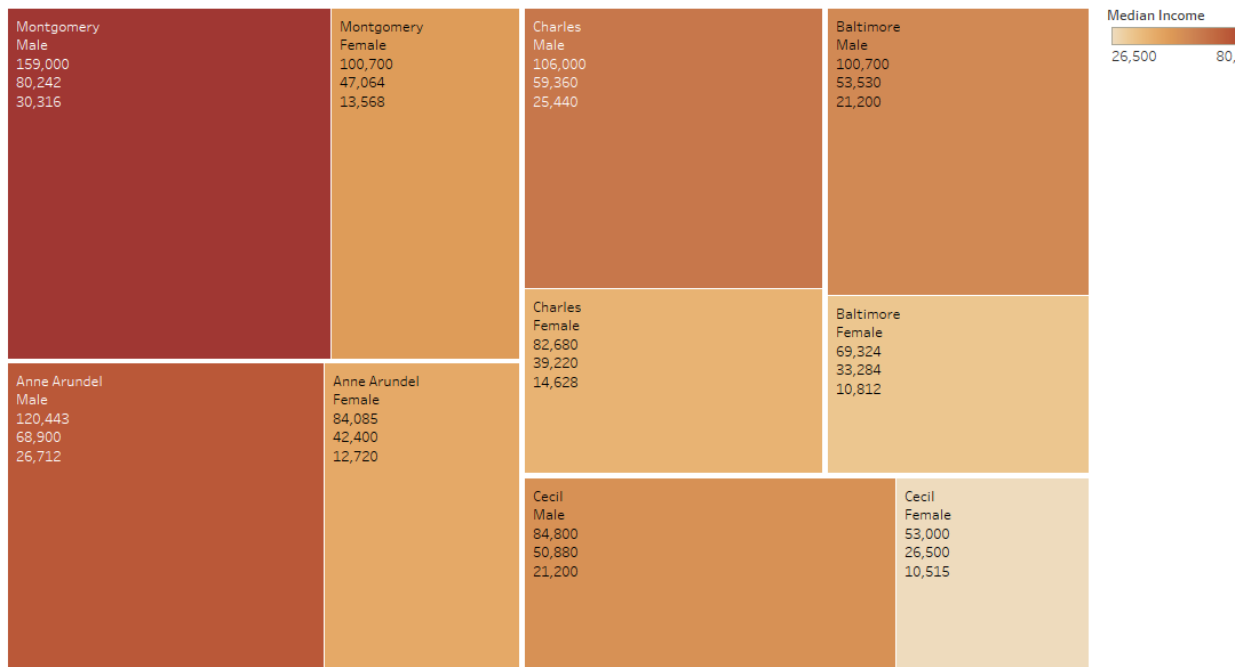
Blocks display shows county name, gender of individuals in that county, 75th percentile of their income, median, and 25th percentile, in that order. White-brown color scheme is based upon the median income in each county's gender population.

In all Maryland counties, males are making more than females on average. However, the average (median average) female in Montgomery County makes more than the average female in other counties. Comparing the males in the poorest country (Cecil) with the females in the richest county (Montgomery), males have a larger income on both the 25th percentile and median metrics while females win out on the 75th percentile.

Each county has their own pay gap. To specify how large each pay gap is, the median income for females will be divided by the median income for males. If the value is more than one, females have a larger income in that county. If the value is less than one, males have a larger

median income for that county. The lower/higher the value is from one, the bigger the disparity. This ratio was found for each county. The smallest ratio was 0.52 for Cecil County. The largest was 0.66 for Charles County. Thus, Cecil County has the largest income-ratio gap between males and females while Charles has the smallest gap.

Each county's pay gap is representative of the overall pay gap in the United States. This base reason can explain why there is a disparity in pay between the sexes here. Specifically, Cecil has the worst disparity. This could be a result of their lack of education. By contrast, Charles has the least pay gap. This could be for a variety of reasons. However, one possible reason is Prince George's County (county touching Charles, Montgomery and Anne Arundel that I did not study here). Prince George's County had started to fill with African Americans for decades since the Fair Housing Act was passed in 1968 [11], and according to a Maryland Matters article [12], this causes the median wage for Black/African American females to be higher than the median wage for Black/African American males. According to the same NBC graphic, since 1990, black families in Prince George County have been migrating into Charles County. These cascading events would lead to a massive black population in Charles County, and since richer African American females are also moving to Charles, would explain the small income gap between genders. In short, one possible reason the pay gap is small in Charles County is because families from Prince George's County with richer African American women have been moving to Charles County during the past few decades.

Section 6: Sex Subsample Analysis Based on Educational Attainment

Like with race, the distribution of education also differs when comparing the different sexes.

Figure 6.1. Distribution of Education by Sex in Anne Arundel and Baltimore Counties

Distribution Of Education By Sex In Anne Arundel County
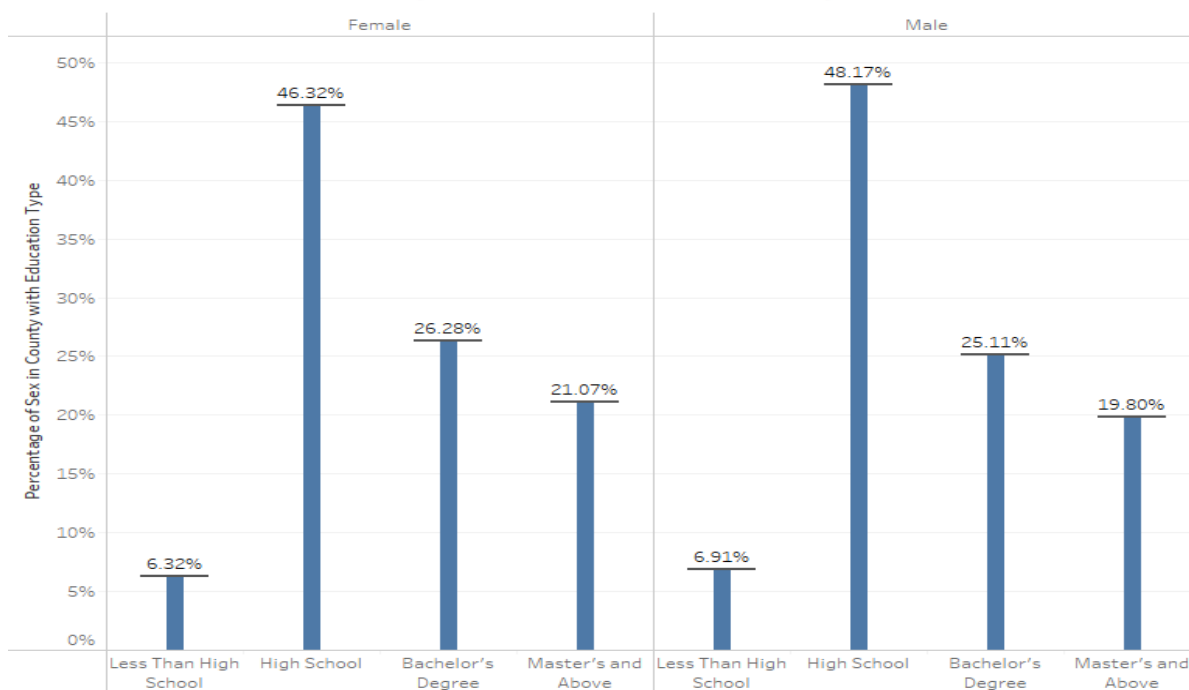
Distribution Of Education By Sex In Baltimore County
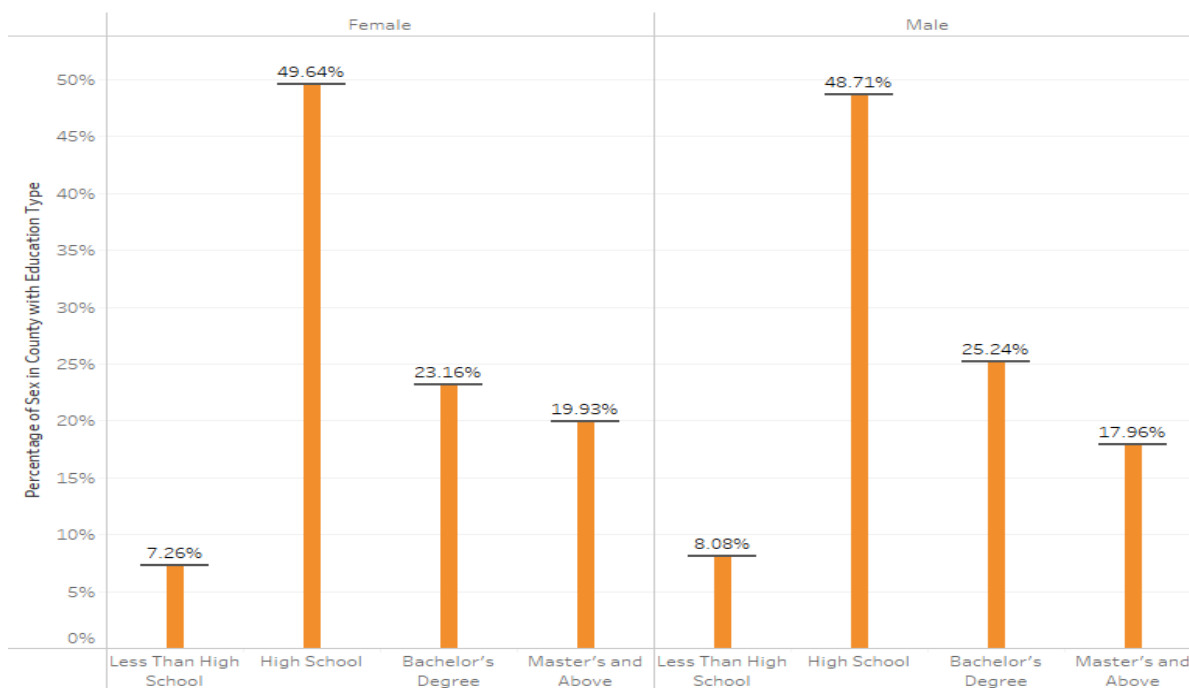
Education distribution bar charts separated by sex for Anne Arundel and Baltimore Counties. Percentage indicates the percentage of sex in each county with a particular degree.

Like with everything else, Baltimore and Anne Arundel counties are extremely similar when it comes to education. There are two things to note between these two counties, however. First, there is a higher percentage of those without a degree in Baltimore than in Anne Arundel for both sexes. Second, a higher percentage of Anne Arundel's female population has a bachelor's degree or above than Baltimore's female population. Baltimore's female population has slightly more people with a master's or above degree than their male population. The opposite is true for those with bachelor's degree. Overall, Baltimore is roughly equal in education level when it comes to sexes while Anne Arundel has a slightly better-educated female population. To preserve the picture, Cecil and Charles are listed on the next page.

Figure 6.2. Distribution of Education by Sex in Charles and Cecil Counties

Distribution Of Education By Sex In Cecil County
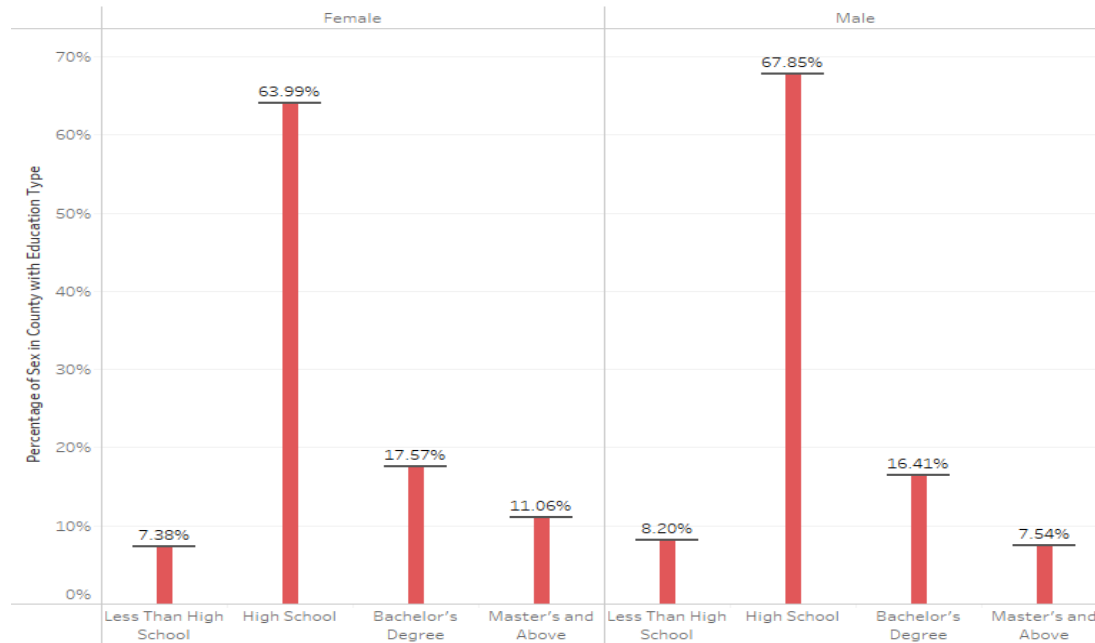
Distribution Of Education By Sex In Charles County

Education distribution bar charts separated by sex for Cecil and Charles Counties. Percentage indicates the percentage of sex in each county with a particular degree.
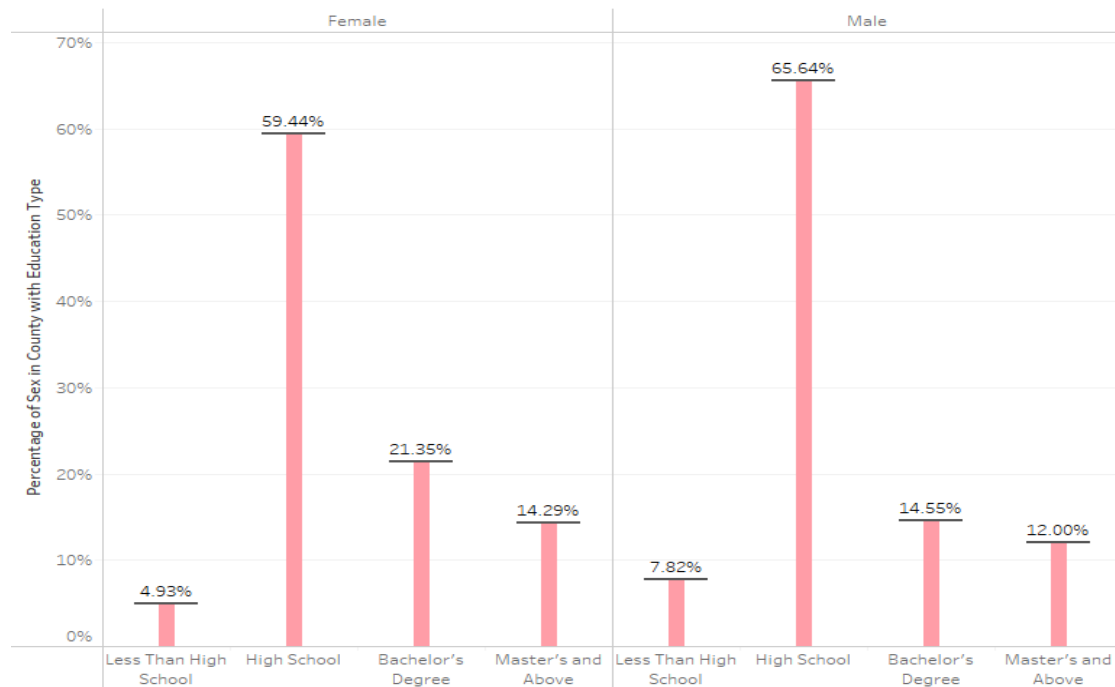
The education distribution across sexes is very similar in Cecil County. Females are slightly more educated with a higher percentage of the female population having bachelor's/above degrees than the male population percentage in the same categories. A similar case can be made with Charles County. However, Charles County has bigger differences in education by sex percentages. The percentage of their male population without a high school diploma is 7.82%, only 0.38% ahead of Cecil's male population. By contrast, only 4.93% of Charles County's female population is without a degree, which is a 2.45% difference from Cecil County and a 2.89% difference from the male population percentage without a degree. Focusing solely on Charles County, 21.35% of their female population has a bachelor's degree compared to only 14.55% of their male population. Compared to Anne Arundel and Baltimore counties, the percentage of Cecil/Charles County populations, regardless of sex, with a master's degree or above is small.

Figure 6.3. Distribution of Education by Sex in Montgomery County
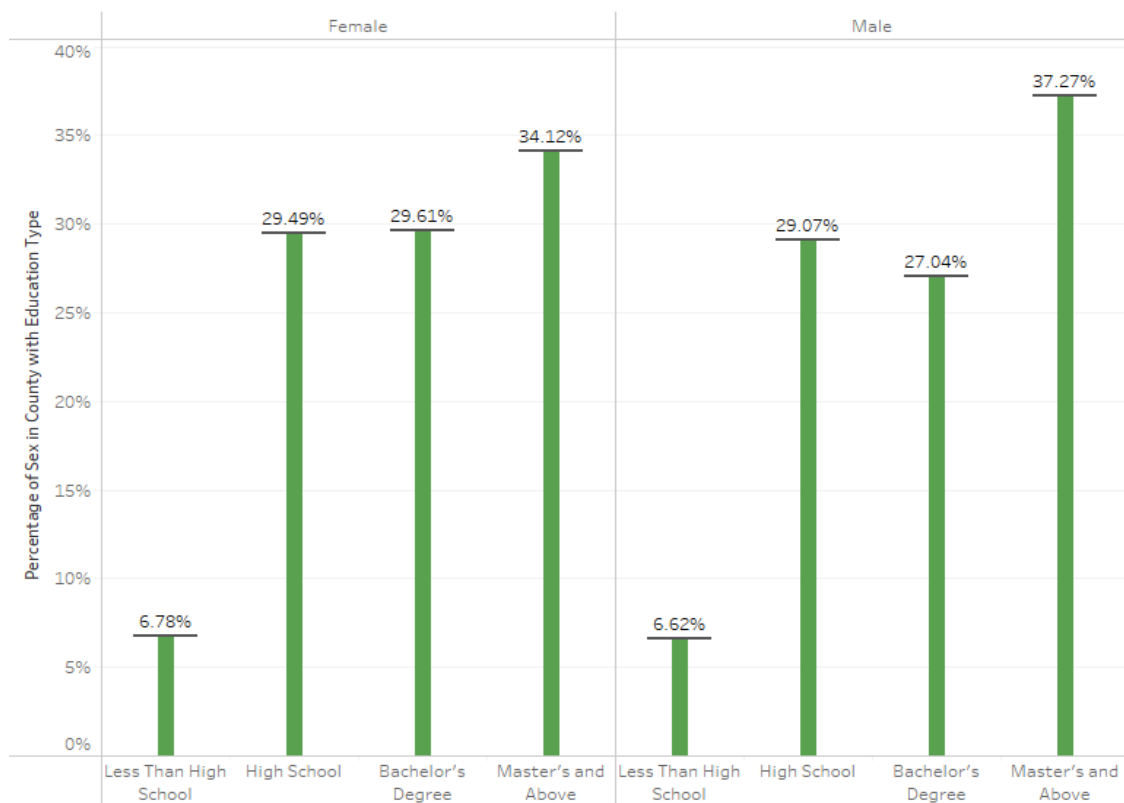
Education distribution bar charts separated by sex for Montgomery County. Percentage indicates the percentage of sex in Montgomery County with a particular degree.

Putting aside the massive jump in college education between counties, the education distribution between sexes in Montgomery County is roughly equivalent. Both sexes have a similar percentage of their populations with high school diplomas and without any degree at all. In terms of college degrees, a slightly higher percentage of the female population has a bachelor's degree than the male population. A slightly higher percentage of the male population has a master's degree or above than the female population.

The female populations in Anne Arundel, Charles, and Cecil counties are more educated than their male populations. In this, the education difference between females and males is most prevalent in Charles County, with each education category being better than the respective male education category. By contrast, both sexes in Baltimore and Montgomery counties are educated at roughly the same level in each.

Conclusion

This report found the differences between Maryland counties, specifically Anne Arundel, Baltimore, Cecil, Charles, and Montgomery, through the use of direct comparisons and subsample analyses.

When constructing a prediction model, it was found that race and education are the most factors by a random forest. When it came to races, Anne Arundel and Baltimore were very similar, but Baltimore had a slightly higher black population. Cecil County was predominantly white, more-so than the other counties. Charles had a massive black population whereas Montgomery had a large other population.

When it comes to income, Cecil County has the lowest median income. They are also the worst when it comes to gender income disparity. Charles County is better when it comes to overall income. Charles is also the best when it comes to gender income disparity. Anne Arundel and Baltimore are very similar when it comes to income, but Anne Arundel is slightly better. Montgomery is the best when it comes to income, even when comparing different races across counties. Each county experiences disparity in income between races. In most counties, the 99% confidence interval cutoffs (median, 75th percentile, maximum) are smaller for the Black groups than it is for the Other groups, and smaller with Other than it is for the White groups.

Overall education follows a similar pattern when comparing overall education. The education disparity by race in these counties is also prevalent, with the Black and Other groups being less educated than the White group. The education distribution by sex, however, is slightly in favor of females, with Charles County having the biggest education gap with females being more educated than males. Furthermore, Charles has an extremely low no-degree population, one whose percentage surpasses that of the extremely well-educated Montgomery County.

The median ages between the counties by race is very similar in each county. Anne Arundel and Cecil counties are slightly younger than the rest while Charles County is slightly older. These are the differences in Maryland counties.

References

1. *The History of Baltimore* (n.d.).
   https://www.baltimorecity.gov/sites/default/files/5_History.pdf

2. Most Violent Cities In America 2021. (n.d.). Worldpopulationreview.com.
   https://worldpopulationreview.com/us-city-rankings/most-violent-cities-in-america

3. Kristin. (2018, May 18). *History of Annapolis.* Annapolis.com.
   https://www.annapolis.com/history/

4. *History.* (n.d.). Historic Annapolis. Retrieved May 5, 2024, from
   https://www.annapolis.org/about-us/history/

5. Washington, D.C. - Capital, Founding, Monumental | Britannica. (n.d.).
   Www.britannica.com. https://www.britannica.com/place/Washington-
   DC/History#ref282698

6. Montgomery | county, Maryland, United States. (n.d.). Encyclopedia Britannica.
   https://www.britannica.com/place/Montgomery-county-Maryland

7. *Charles County, Maryland - Historical Chronology.* (n.d.). Msa.maryland.gov.
   https://msa.maryland.gov/msa/mdmanual/36loc/ch/chron/html/chchron.html

8. *Cecil County, Maryland - Historical Chronology.* (n.d.). Msa.maryland.gov. Retrieved
   May 5, 2024, from
   https://msa.maryland.gov/msa/mdmanual/36loc/ce/chron/html/cechron.html

9. Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace
   Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. IPUMS USA:
   Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024.
   https://doi.org/10.18128/D010.V15.0

10. Maryland Population By County in 2022. (n.d.). Beautifydata.com. Retrieved May 8,
    2024, from https://beautifydata.com/united-states-population/by-
    county/maryland/estimate/2022

11. ai2html. (n.d.). Media.nbcnewyork.com. Retrieved May 8, 2024, from
    https://media.nbcnewyork.com/assets/editorial/national/2023/DC/black-migration/demo-
    lines/demo-lines-ai2html/index.html

12. Brown, D. J. (2024, March 13). *Women earn less than men throughout Md.*; lawmakers
    say wage transparency may help. Maryland Matters.
    https://www.marylandmatters.org/2024/03/13/women-earn-less-than-men-throughout-
    md-lawmakers-say-wage-transparency-may-help/

Important Tables (And Other Figures)

Some tables (and other figures) are listed below. They are here if the reader wants to look at them.

Table 1. County Population Breakdown of Race and Sex

County Name	SEX	Black		Other		White	
Anne Arundel	Female	■	7.48%	■	8.00%	■	36.58%
	Male	■	6.52%	■	7.74%	■	33.68%
Baltimore	Female	■	11.89%	■	7.72%	■	34.26%
	Male	■	8.70%	■	6.24%	■	31.19%
Cecil	Female	■	2.85%	■	5.26%	■	42.43%
	Male	■	4.28%	■	4.39%	■	40.79%
Charles	Female	■	22.95%	■	6.90%	■	22.69%
	Male	■	17.69%	■	6.13%	■	23.64%
Montgomery	Female	■	7.91%	■	17.19%	■	28.02%
	Male	■	6.07%	■	14.85%	■	25.96%

County Name: Anne Arundel, Baltimore, Cecil, Charles, Montgomery

Percentages indicate how much percentage of each county's population belong to a particular race and sex. For example, 7.48% of Anne Arundel's population is black and female.

Table 2. Income Summary Statistics

County Name	Mean Income	Income Standard Deviation	Min. Income	Income 25th Percentile	Median Income	Income 75th Percentile	Max. Income
Anne Arundel	74,790	88,818	0	18,550	53,000	101,760	1,022,688
Baltimore	62,996	82,288	0	13,780	42,400	84,800	1,022,900
Cecil	50,308	58,582	0	12,720	36,040	72,080	616,920
Charles	64,259	64,887	0	18,020	47,276	94,234	654,020
Montgomery	95,961	119,486	0	19,928	61,480	127,200	1,022,900

County Name: Anne Arundel, Baltimore, Cecil, Charles, Montgomery

Income summary statistics for each county.

Figure EX.1. Conditional Inference Tree for Cecil (15) and Charles(17) counties
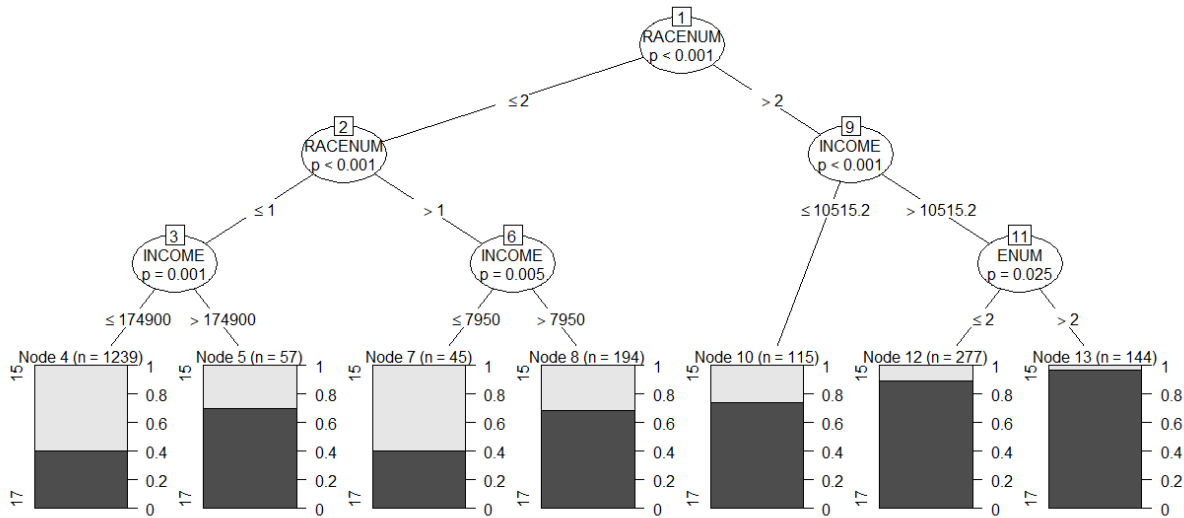
Conditional inference tree with numeric race (RACENUM) and education type (ENUM) to find how Cecil (15) and Charles (17) counties differ. Races White, Other, Black, are coded as 1,2, and 3 respectively. Education categories Less Than High School, High School, Bachelor's Degree, Master's And Above are coded as 1, 2, 3, and 4 respectively. The sexes are encoded in SEXNUM as 1 for females and 2 for males. The summary of this tree is that if a person is white or other with a low income, chances are they live in Cecil County. If a person is black, chances are high they live in Charles. If the person is black and has a college degree this already-high chance jumps very close to 100%.

Figure EX.2. Conditional Inference Tree for Anne Arundel (3) and Baltimore (5) counties
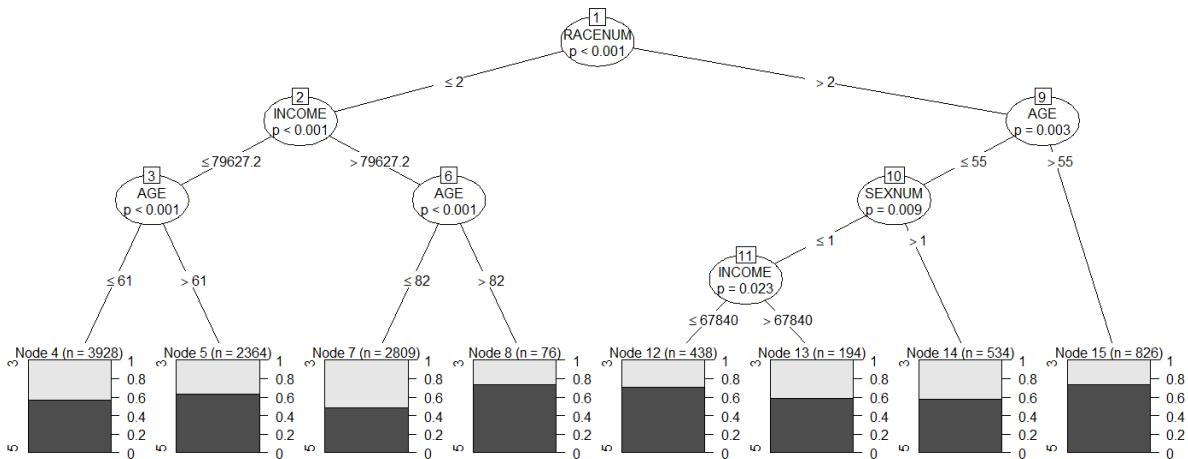
Same variable encoding as the previous inference tree. As can be seen, there are really no good indicators between these two counties. One of the best indicators is if a person is white or other, making more than $80,000

a year, and over 82 then there is around a 75% chance they live in Baltimore. Another indicator is that if a person is black and over 55, there is around a 75% chance they live in Baltimore.

Pre-Analysis Code

Below is the pre-analysis code. I tried using a conditional inference tree, but needed numeric values to get good results. When I converted race/education/sex back to numeric, the way the tree represented the categorical numbers (such as ≤ 2) seemed a little off to me report-wise so I just used the ctrees in the other figures section above so that if the reader wants to see it they can.

```
finalData <- read.csv("C:/Users/natek/Downloads/clean (7).csv")
head(finalData)

#Data for random forest
forestData <- finalData[c('COUNTYFIP', 'AGE', 'INCOME','RACE','SEX','EDUCATION')]

library(randomForest)
modelRF <- randomForest(as.factor(COUNTYFIP)~., data = forestData, importance=T, ntree=)
importance(modelRF)
varImpPlot(modelRF, main = "Importance in Determining County")



blackLoc <- finalData$RACE=='Black'
otherLoc <- finalData$RACE=='Other'
whiteLoc<- finalData$RACE=='White'

finalData$RACENUM[whiteLoc] <- 1
finalData$RACENUM[blackLoc] <- 3
finalData$RACENUM[otherLoc] <- 2
```

```
lessHSLOC<-finalData$EDUCATION=='Less Than High School'
HSLOC <- finalData$EDUCATION=='High School'
BALOC <- finalData$EDUCATION=='Bachelor\'s Degree'
MALOC <- finalData$EDUCATION=="Master\'s and Above"


finalData$ENUM[lessHSLOC]<-1
finalData$ENUM[HSLOC]<-2
finalData$ENUM[BALOC]<-3
finalData$ENUM[MALOC]<-4

femaleLoc <- finalData$SEX =="Female"
maleLoc <- finalData$SEX =="Male"

finalData$SEXNUM[femaleLoc] = 1
finalData$SEXNUM[maleLoc] = 2

library(party)
forestData <- finalData[c('COUNTYFIP', 'SEXNUM','ENUM', 'AGE',
'INCOME','RACE','SEX','EDUCATION', 'RACENUM')]
forestDataAB <- forestData[forestData$COUNTYFIP== 15| forestData$COUNTYFIP ==17,]


model.ctree <- ctree(as.factor(COUNTYFIP)~., data=forestDataAB[,-
match(c('RACE','EDUCATION', 'SEX'), names(forestData))])
print(model.ctree)
plot(model.ctree)


forestDataAC <- forestData[forestData$COUNTYFIP== 3| forestData$COUNTYFIP ==5,]


model.ctree <- ctree(as.factor(COUNTYFIP)~., data=forestDataAC[,-
match(c('RACE','EDUCATION', 'SEX'), names(forestData))])
```

```
print(model.ctree)
plot(model.ctree)
```

Data Cleaning Code

```
DATA WORK.IMPORT;
  set '/home/u63783283/Final Project/usa_00009.sas7bdat';
RUN;

/*Limit to Maryland*/
data MARYLAND;
set WORK.IMPORT;
if STATEFIP=24;
run;

/*Limit to listed counties*/
proc freq data=MARYLAND;
table COUNTYFIP;
run;
data MARYLANDCOUNTIES;
set MARYLAND;
if COUNTYFIP = 3 or COUNTYFIP = 5 or COUNTYFIP = 17 or COUNTYFIP = 31 or
COUNTYFIP = 15;
run;

proc freq data=MARYLANDCOUNTIES;
table COUNTYFIP;
run;

/*Clean the Data*/
```

```
data MARYLANDCOUNTIES;
set MARYLANDCOUNTIES;
/*Clean income*/
if INCTOT < 999999;
if INCTOT >=0;
INCTOT = 1.06*INCTOT;

/*Rename for simplicity*/
RENAME INCTOT = INCOME;

/*Clean age*/
if AGE >=18;
if AGE >=60 then
AGEDEC =  6;
else if AGE >=50 then
AGEDEC = 5;
else if AGE >=40 then
AGEDEC = 4;
else if AGE >=30 then
AGEDEC=3;
else if AGE >=20 then
AGEDEC=2;
else if AGE >=10 then
AGEDEC=1;
else
AGEDEC=0;

/*Rename for simplicity*/
RENAME AGEDEC=AGE_DECADE;

/*Simplify Table*/
drop cluster;
drop strata;
```

```
drop GQ;
drop PERNUM;
DROP PERWT;

/*Educational Attainment*/
Length EDUCATION $21.;
if EDUCD < 62 then
EDUCATION = "Less Than High School";
else if EDUCD < 101 then
EDUCATION = "High School";
else if EDUCD < 114 then
EDUCATION = "Bachelor's Degree";
else if EDUCD <= 116 then
EDUCATION = "Master's and Above";
else
EDUCATION = '';

/*Get name of County*/
if COUNTYFIP=3 then
COUNTY_NAME  = "Anne Arundel";
else if COUNTYFIP = 5 then
COUNTY_NAME = "Baltimore";
else if COUNTYFIP = 15 then
COUNTY_NAME = "Cecil";
else if COUNTYFIP = 17 then
COUNTY_NAME = "Charles";
else if COUNTYFIP = 31 then
COUNTY_NAME = "Montgomery";
else
COUNTY_NAME ='';

/*Also for future reference, snag the total FIPSCODE.*/
if COUNTYFIP=3 then
FIPSCODE=24003;
```

```
else if COUNTYFIP = 5 then
FIPSCODE=24005;
else if COUNTYFIP = 15 then
FIPSCODE=24015;
else if COUNTYFIP = 17 then
FIPSCODE=24017;
else if COUNTYFIP = 31 then
FIPSCODE=24031;
else
FIPSCODE=0;


/*Race*/
length Race2 $12.;
if RACE=1 then
Race2="White";
else if RACE=2 then
Race2="Black";
else if RACE=3 | RACE=4 | RACE=5 | RACE=6 | RACE=7 | RACE=8 | RACE=9 then
Race2 = "Other";
else
Race2='';

/*Sex*/
length Sex2 $6.;
if SEX=1 then
Sex2 = "Male";
else if SEX=2 then
Sex2="Female";
else
Sex2='';

DROP RACE;
DROP SEX;
```

```
Rename Sex2=SEX;
rename Race2=RACE;
run;


/*Basic Summary Statistics*/
proc means data=MARYLANDCOUNTIES N Mean Stddev Min P25 Median P75 Max;
var INCOME;
run;

proc means data=MARYLANDCOUNTIES;
var AGE;
run;

proc freq data=MARYLANDCOUNTIES;
table AGE_DECADE;
run;

proc freq data=MARYLANDCOUNTIES;
table EDUCATION;
run;

proc freq data=MARYLANDCOUNTIES;
table RACE;
run;

proc freq data=MARYLANDCOUNTIES;
table FIPSCODE;
run;


proc freq data=MARYLANDCOUNTIES;
table COUNTY_NAME;
```

```
run;


proc freq DATA = MARYLANDCOUNTIES;
table SEX;
run;



/*Export as csv for Tableau*/

proc export data=MARYLANDCOUNTIES
    outfile="/home/u63783283/Final Project/clean.csv"
    dbms=csv
    replace;
run;
```