

Project 1

Henry Lambson, Jonas Moros, Blake Miller



February 26, 2023

Table of Contents

[Abstract 2](#_Toc128331801)

[1. Business Understanding 2](#_Toc128331802)

[2. Data Understanding 3](#_Toc128331803)

[2.1 Data Description 3](#_Toc128331804)

[2.2 Data Quality 4](#_Toc128331805)

[2.3 Visualization of Attributes 6](#_Toc128331806)

[2.4 Exploring Relationships among Various Data Types 14](#_Toc128331807)

[3. Data Preparation 22](#_Toc128331808)

[4. Exceptional Work 24](#_Toc128331809)

# Abstract

In this paper the census and COVID-19 statistic data will be analyzed to explore similarities and differences in how Florida and California responded to the pandemic. The main focus of this paper will be comparing how the states’ differing populations were affected by the pandemic. Race, gender, age, and their relationship to the number of cases and deaths in each state will be explored to try to determine which state faired better or worse. This will also allow a look into which race, gender, and age group were at the most risk in each state.

# 1. Business Understanding

The COVID-19 outbreak began in late 2019 and early 2020. It is a contagious disease caused by a virus SARS-CoV-2. The first case of COVID-19 was documented in Wuhan, China, and due to the virus’ highly contagious nature, the disease spread worldwide in the ensuing months. This spread led to the COVID-19 Pandemic.

Once the United States Federal Government saw how widespread the disease had become, it began taking steps to “flatten the curve,” or lessen the spread. This took the form of “social distancing,” where citizens were expected to limit their exposure to people outside their immediate household, and when that could not be done, stay at least six feet apart. Due to these regulations, places of public gathering, such as churches, gyms, and restaurants, began closing.

In this report, the data on the spread of the virus and its affects will be analyzed to answer questions like: How does the virus effects people differently in different parts of the country? How effective is social distancing? Are there regions that do better or worse against the virus? Can we predict the virus’s impact in one region given the data in another region? Answering these questions through observing data is important because it allows people to make decisions with evidence instead of guessing or using their own “common sense”.

This report is intended for policy makers, doctors, educators, or any concerned citizens wondering about the virus. With the information in this report, policy makers could make decisions about whether to put social distancing or other spread prevention policies in place; doctors could learn more about how the virus spreads and those who are at highest risk; educators could make decisions about in-person vs. distance learning; and concerned citizens can learn more about how the virus might affect their lives.

# 2. Data Understanding

## 2.1 Data Description

The COVID-19\_cases\_plus\_census dataset contains data from the U.S. Census as well as data relating to the COVID outbreak, such as confirmed cases and deaths. The dataset contains 259 features and 3142 attributes. A truncated description of the dataset is provided in Table 1.

Table 1 The Description of the Data

|  |  |
| --- | --- |
| $ county\_fips\_code | Factor w/ 3142 levels "01001","01003",.. |
| $ county\_name | Factor w/ 1878 levels "Abbeville County",.. |
| $ state | Factor w/ 51 levels "AK","AL","AR",.. |
| $ state\_fips\_code | Factor w/ 51 levels "01","02","04",.. |
| $ date | Date, format |
| $ confirmed\_cases | num |
| $ deaths | num |
| $ geo\_id | Factor w/ 3142 levels "01001","01003",.. |
| $ nonfamily\_households | num |
| $ family\_households | num |
| $ median\_year\_structure\_built | num |
| $ rent\_burden\_not\_computed | num |
| $ rent\_over\_50\_percent | num |
| $ rent\_40\_to\_50\_percent | num |
| $ rent\_35\_to\_40\_percent | num |
| $ rent\_30\_to\_35\_percent | num |
| $ rent\_25\_to\_30\_percent | num |
| $ rent\_20\_to\_25\_percent | num |
| $ rent\_15\_to\_20\_percent | num |
| $ rent\_10\_to\_15\_percent | num |
| $ rent\_under\_10\_percent | num |
| $ total\_pop | num |
| $ male\_pop | num |
| $ female\_pop | num |
| $ median\_age | num |
| $ white\_pop | num |
| $ black\_pop | num |
| $ asian\_pop | num |
| $ hispanic\_pop | num |
| $ amerindian\_pop | num |
| $ other\_race\_pop | num |
| $ two\_or\_more\_races\_pop | num |

Due to the large size of the dataset, this report will only focus on a few important features, rather than taking a broad look at many features. The focus of this report will be to visualize the correlation between confirmed cases, deaths, and population by state, sex, and race. Table 2 shows a description of the updated dataset with only the features that will be analyzed in more detail.

Table 2 Description of Smaller Dataset

|  |  |
| --- | --- |
| $ state | Factor w/ 51 levels "AK","AL","AR",.. |
| $ confirmed\_cases | num |
| $ deaths | num |
| $ total\_pop | num |
| $ male\_pop | num |
| $ female\_pop | num |
| $ white\_pop | num |
| $ black\_pop | num |
| $ asian\_pop | num |
| $ hispanic\_pop | num |
| $ amerindian\_pop | num |

## 2.2 Data Quality

To continue the preprocessing of the data, it is necessary to verify the quality of the data. Duplicates and missing data must be accounted for in order to see quality results once the analysis begins.

The small dataset was verified to find any missing or duplicated values, and there were none of either. The data is clean and ready to be worked with. Just to be safe, the full dataset will also be checked for duplicates and missing data so that it can be worked with if need be.

The full dataset also had no duplicated values, but it did contain missing values. The rows with these missing values have been dropped from the dataset, so now the full dataset has been cleaned and can also be worked with. Now that both datasets have been cleaned, the small dataset will be split into the two states that will be analyzed further: Florida and California. Table 3 contains a summary of the California data, and Table 4 contains a summary of the Florida data.

Table 3 Summary of California Dataset

|  |  |  |
| --- | --- | --- |
| state  CA: 58  AK: 0  AL: 0  AR: 0  AZ: 0  CO: 0  (Other): 0 | confirmed\_cases  Min. : 40  1st Qu.: 2673  Median : 11114  Mean : 51668  3rd Qu.: 39784  Max. : 1002614 | deaths  Min. : 0.0  1st Qu.: 27.5  Median : 119.0  Mean : 581.7  3rd Qu.: 440.5  Max. : 13936.0 |
| total\_pop  Min. : 1203  1st Qu.: 47268  Median : 182486  Mean : 672118  3rd Qu.: 677036  Max. : 10105722 | male\_pop  Min. : 664  1st Qu.: 23726  Median : 92134  Mean : 333907  3rd Qu.: 336391  Max. : 4979641 | female\_pop  Min. : 539  1st Qu.: 23541  Median : 91938  Mean : 338212  3rd Qu.: 340644  Max. : 5126081 |
| white\_pop  Min. : 777  1st Qu.: 30575  Median : 101253  Mean : 254786  3rd Qu.: 289446  Max. : 2676982 | black\_pop  Min. : 4.0  1st Qu.: 545.8  Median : 3695.5  Mean : 37266.5  3rd Qu.: 16494.5  Max. : 799579.0 | asian\_pop  Min. : 0  1st Qu.: 732  Median : 9925  Mean : 93585  3rd Qu.: 54587  Max. : 1442577 |
| hispanic\_pop  Min. : 117  1st Qu.: 9684  Median : 49339  Mean : 260446  3rd Qu.: 248389  Max. : 4893579 | amerindian\_pop  Min. : 28.0  1st Qu.: 550.2  Median : 1218.5  Mean : 2376.1  3rd Qu.: 2651.5  Max. : 19915.0 |  |

Table 4 Summary of Florida Dataset

|  |  |  |
| --- | --- | --- |
| state  FL: 67  AK: 0  AL: 0  AR: 0  AZ: 0  CO: 0  (Other): 0 | confirmed\_cases  Min. : 845  1st Qu.: 2466  Median : 6886  Mean : 23668  3rd Qu.: 23409  Max. : 347965 | deaths  Min. : 4.0  1st Qu.: 34.0  Median : 156.0  Mean : 364.7  3rd Qu.: 414.5  Max. : 4622.0 |
| total\_pop  Min. : 8347  1st Qu.: 27432  Median : 116754  Mean : 302663  3rd Qu.: 334997  Max. : 2702602 | male\_pop  Min. : 5104  1st Qu.: 14735  Median : 58883  Mean : 147976  3rd Qu.: 162600  Max. : 1311997 | female\_pop  Min. : 3131  1st Qu.: 12862  Median : 57871  Mean : 154688  3rd Qu.: 173556  Max. : 1390605 |
| white\_pop  Min. : 5933  1st Qu.: 17984  Median : 98930  Mean : 166044  3rd Qu.: 229463  Max. : 799018 | black\_pop  Min. : 998  1st Qu.: 4116  Median : 9616  Mean : 46708  3rd Qu.: 37357  Max. : 518277 | asian\_pop  Min. : 0  1st Qu.: 139  Median : 1785  Mean : 7963  3rd Qu.: 7414  Max. : 66304 |
| hispanic\_pop  Min. : 406  1st Qu.: 1784  Median : 11801  Mean : 74851  3rd Qu.: 50581  Max. : 1823038 | amerindian\_pop  Min. : 11.0  1st Qu.: 101.0  Median : 347.0  Mean : 622.5  3rd Qu.: 903.0  Max. : 3156.0 |  |

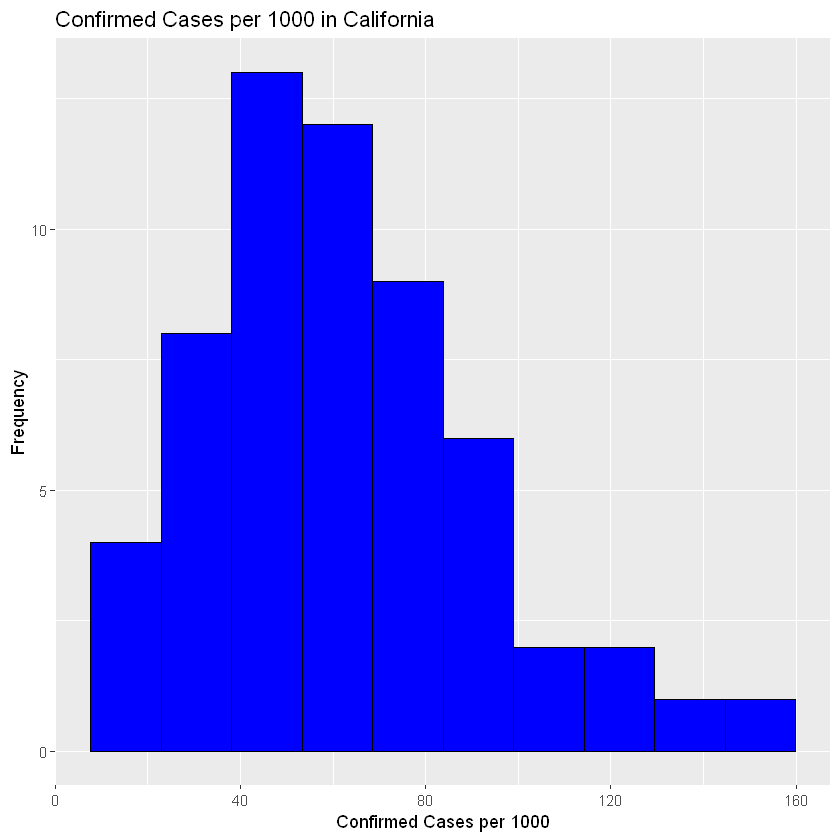
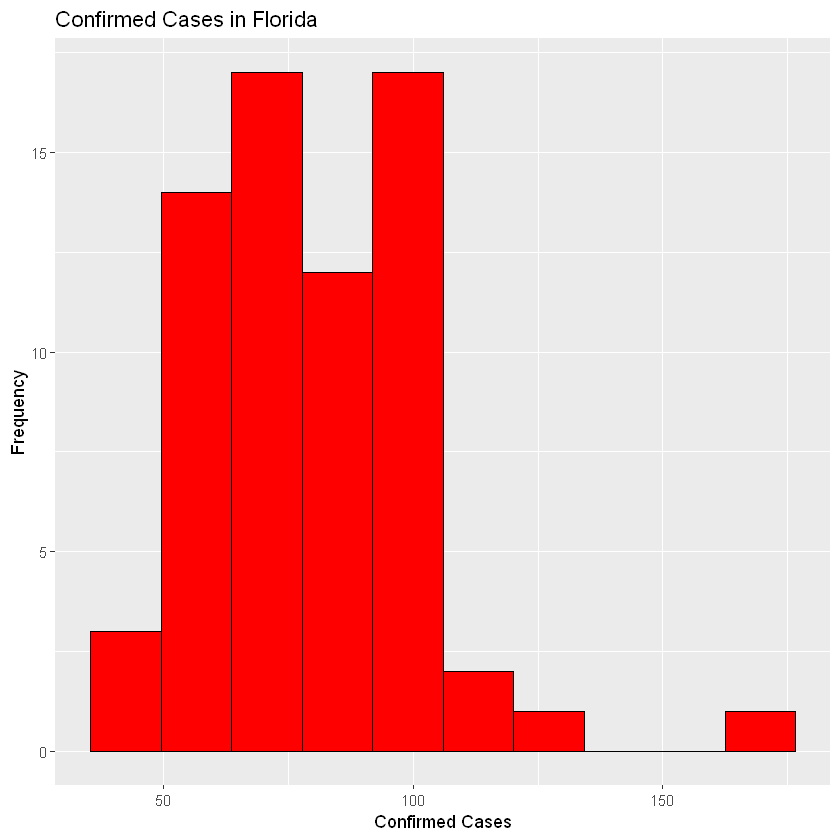
## 2.3 Visualization of Attributes

In this section, the individual attributes of both the California and Florida datasets will be analyzed and compared. In order to get a good comparison between the states, without their difference in population affecting the graphs, the data was normalized. This was done by dividing the attribute by the total population and then multiplying by 1,000 to get an average per 1,000 individuals.

*A.* *Confirmed Cases (Ratio)*

The data for confirmed cases is a ratio, so a histogram will be used for analysis. The data for California and Florida are represented in the histograms in Figures 1 and 2.  The distribution for California looks similar to a uniform distribution with a skew to the right. The majority of the California data is in the 40-80 range, with some outlier counties reaching up to 180 cases per 1,000 individuals.

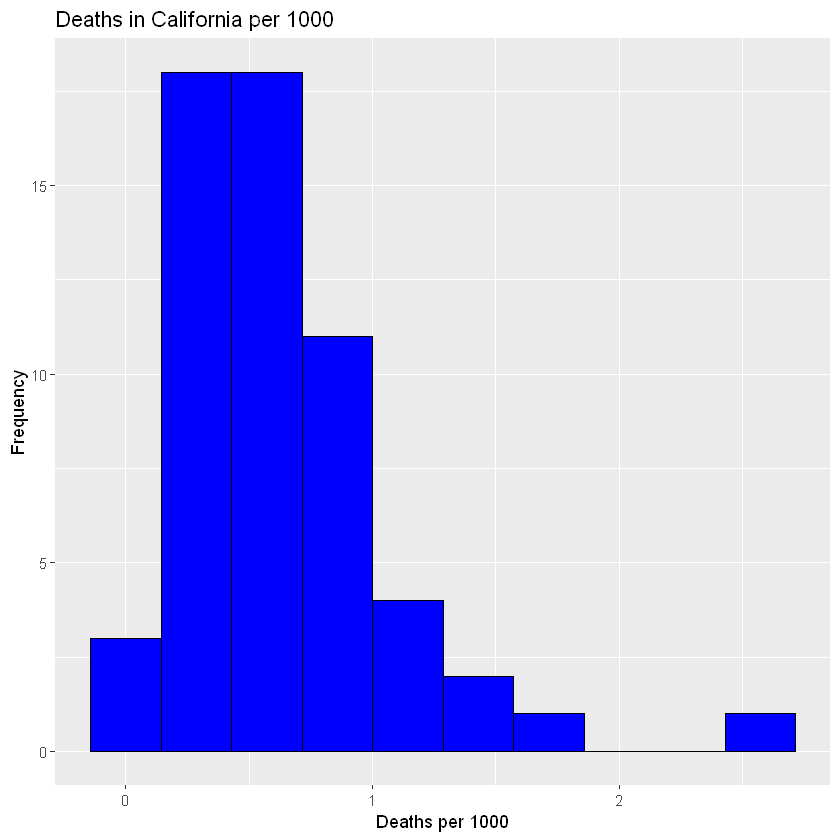
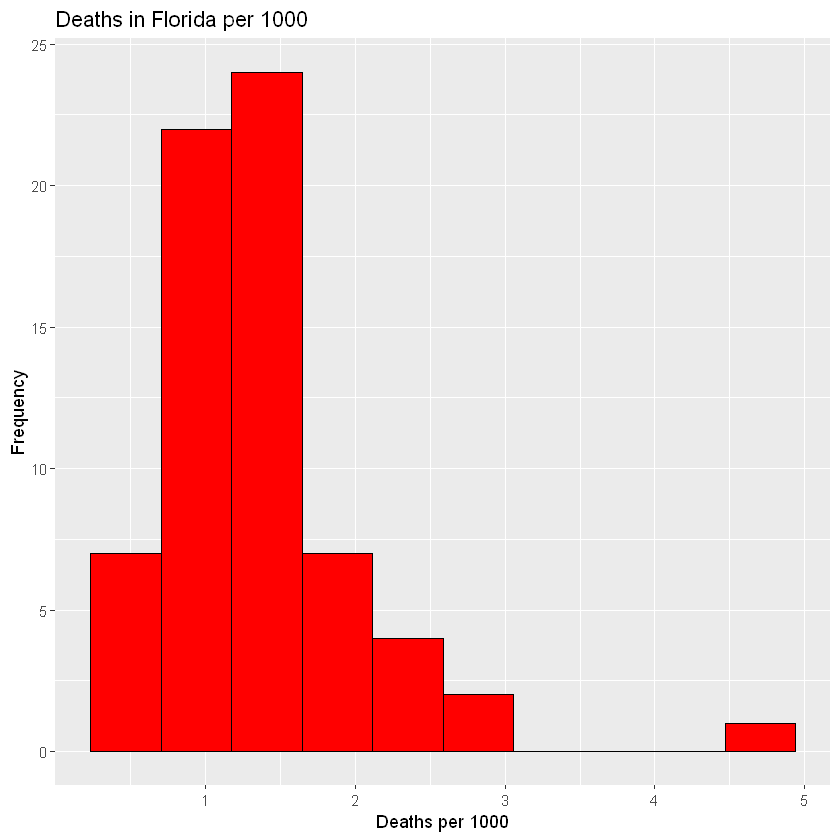
The Florida data is less similar to a normal distribution than the California dataset, but still shares a comparable structure. Most of the data is in the 50-100 range, with a single outlier around the 175 mark, causing a skew to the right. It is difficult to judge which state fared better in terms of confirmed cases just based off of these graphs alone, because they are quite similar. California had more counties in the 120-180 range, but had more counties on the lower end of the distribution than Florida did.

       Figure 1 Confirmed Cases CA          Figure 2 Confirmed Cases FL     

*B.* *Deaths (Ratio)*

The data for deaths in each state are shown in Figures 3 and 4. Since this attribute is a ratio, a histogram will again be used. These histograms share a similar structure, with the majority of the data being on the lower end with an outlier causing a significant skew to the right. It is important to point out the difference in the scale of the axes for these two graphs. While the structure of the graphs is very similar, the Florida graph has a maximum of 5 deaths per 1000, while the California graph has a maximum of 3 deaths per 1000. Because of this, the axes are different, leading to the majority of the data for California to be in the 0-1 range, while the majority of the data for Florida is in the 0.5-2 range. It is clear based on these graphs that Florida had more deaths per 1000 people than California.

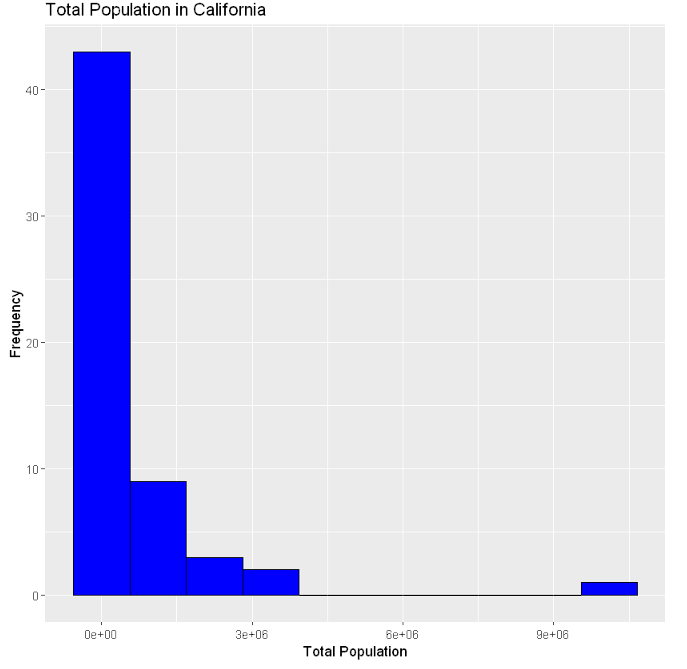
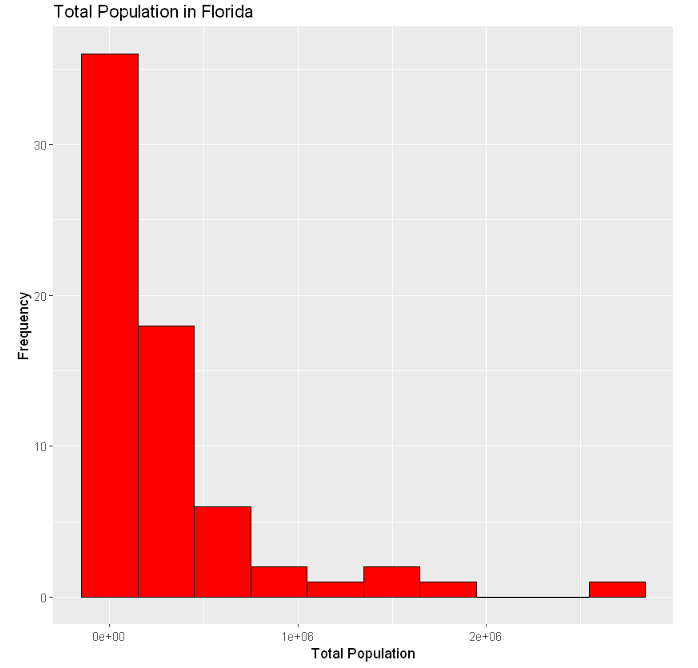
Figure 3 Deaths per 1000 FL Figure 4 Deaths per 1000 CA



*C.* *Total Population (Ratio)*

The data for total population by county is shown below in Figures 5 and 6. The population is a ratio, as shown in the histogram presenting the data. This data was not normalized, as this was the attribute that the rest of the data was normalized off of. As shown in the histograms, California has a larger range of population by county than Florida. The two histograms are also right-skewed, meaning there are a large number of low population counties, and a small number of high population counties.

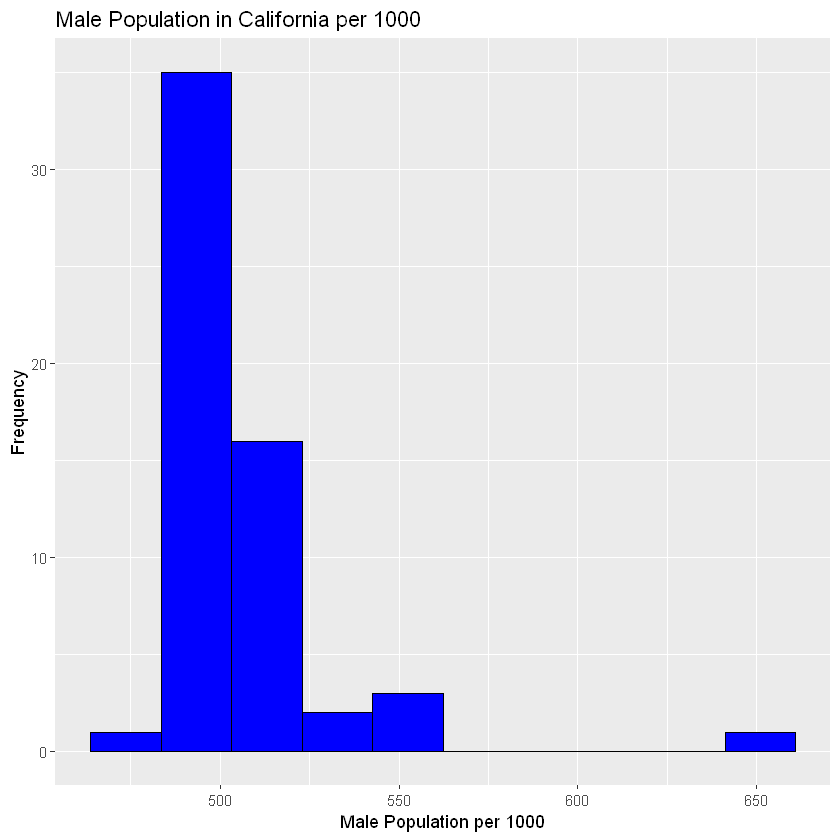
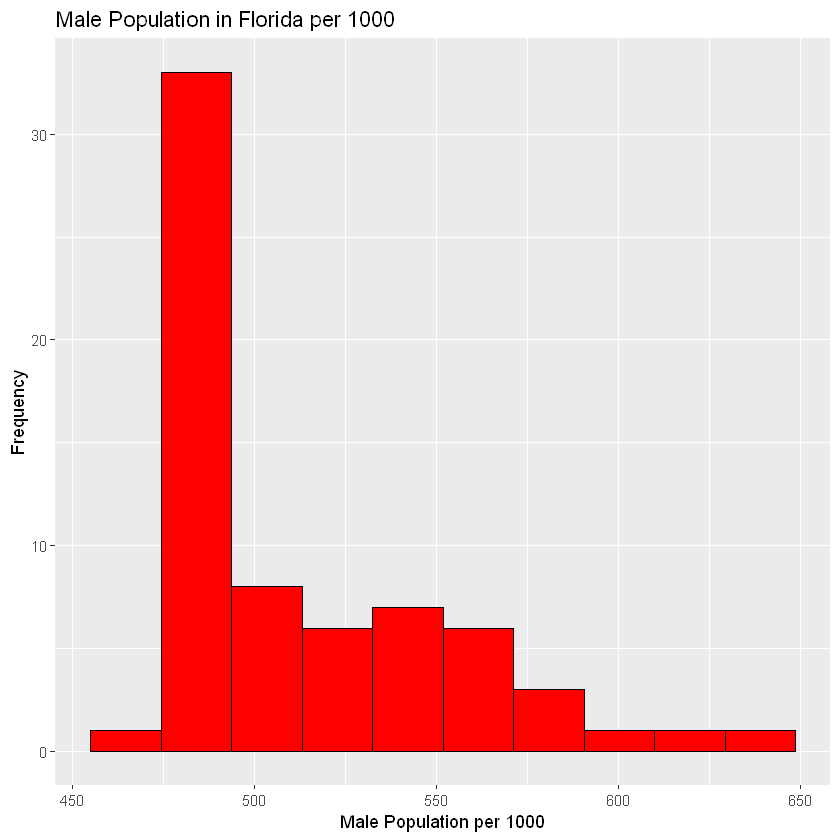
Figure 5 Total Pop FL Figure 6 Total Pop CA

**

*D.* *Male Population (Ratio)*

The data for male population by county is shown below in Figures 7 and 8. The male population is a ratio, so shown in the histogram presenting the data. As shown in the histograms, the distribution of the male populations follow the same distributions as the total population. Analyzing the graphs it can be shown that the overall male population in Florida is more distributed per county compared to the distribution of males in California counties. It can also be seen that the majority of counties have a male population of under 500 per 1000, but there are a few outlier counties with majority men.

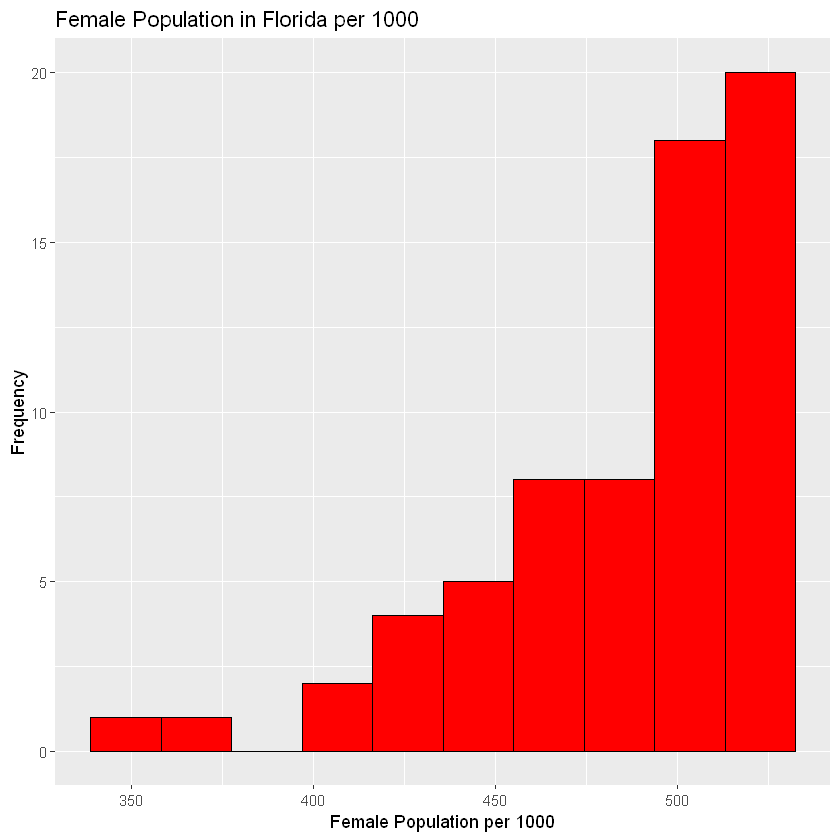
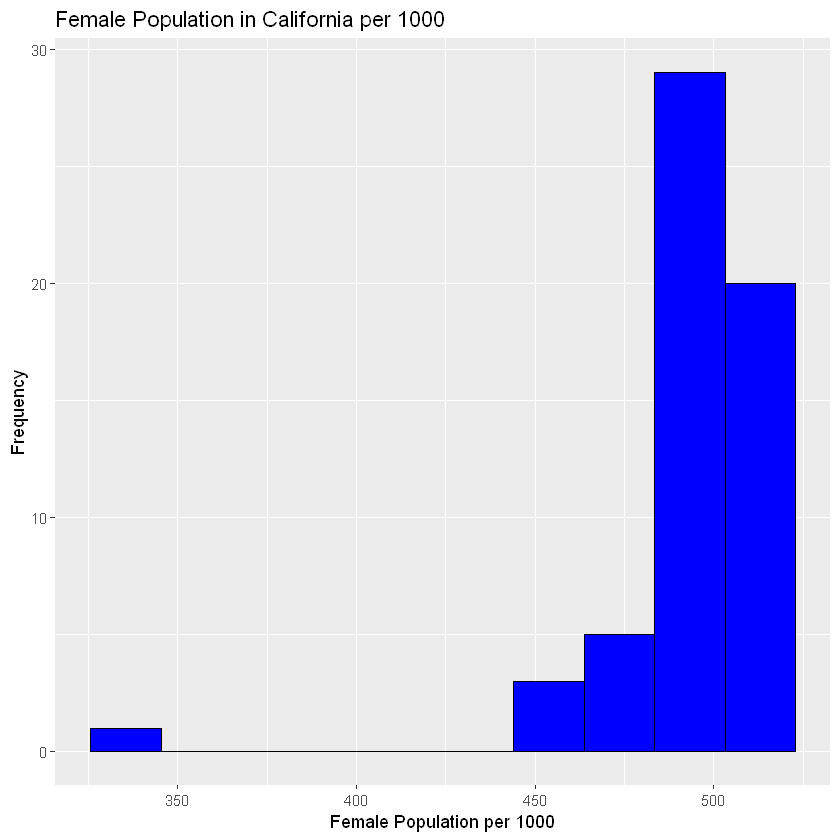
Figure 7 Male Pop FL Figure 8 Male Pop CA

**

*E.* *Female Population (Ratio)*

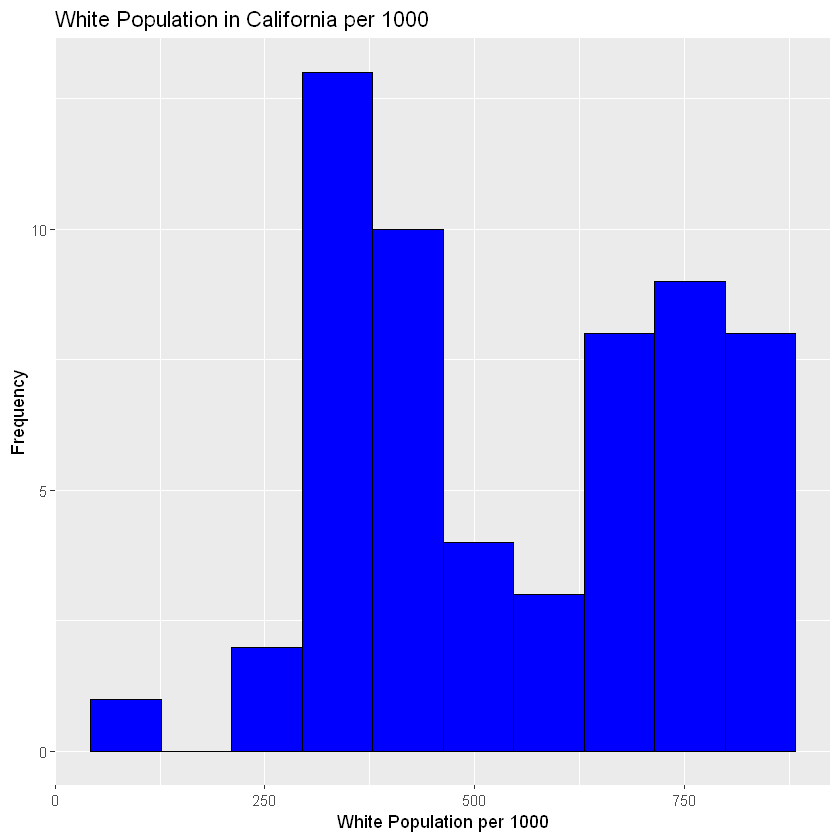
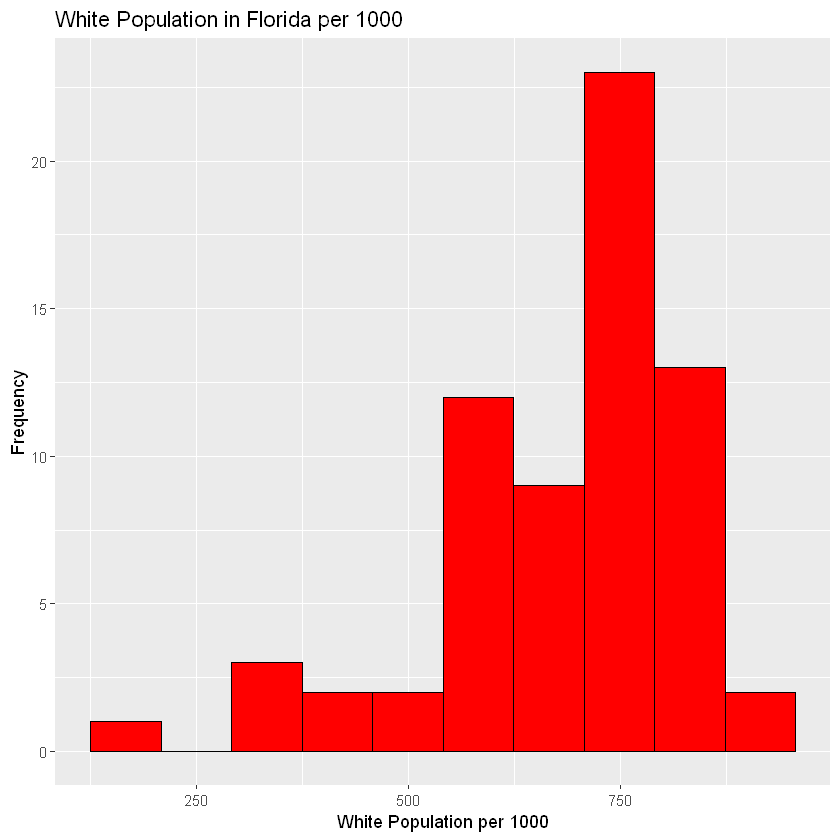
The data for female population by county is shown below in Figures 9 and 10. The female population is a ratio, as shown in the histogram presenting the data. As shown in the histograms, the distribution of the female population is very different from the distribution of the male population. The graphs are skewed left instead of being skewed right like the male population. The majority of counties have a female population of over 500 per 1000, but there are a few outlier counties where women are the minority. That being said, similar to the male graphs, the majority of the data for both states is around 500, which is expected. It is interesting to see that both states follow the same general distribution for both male and female populations even when the male and female graphs are drastically different from each other.

Figure 9 Female Pop FL Figure 10 Female Pop CA



*F.* *White Population (Ratio)*

The data for white population by county is shown below in Figures 11 and 12. Again, since the population is a ratio, it can be presented using a histogram. The graphs are no longer following the same structure as each other, so there are some differences to analyze.

 Figure 11 White Pop CA Figure 12 White Pop FL

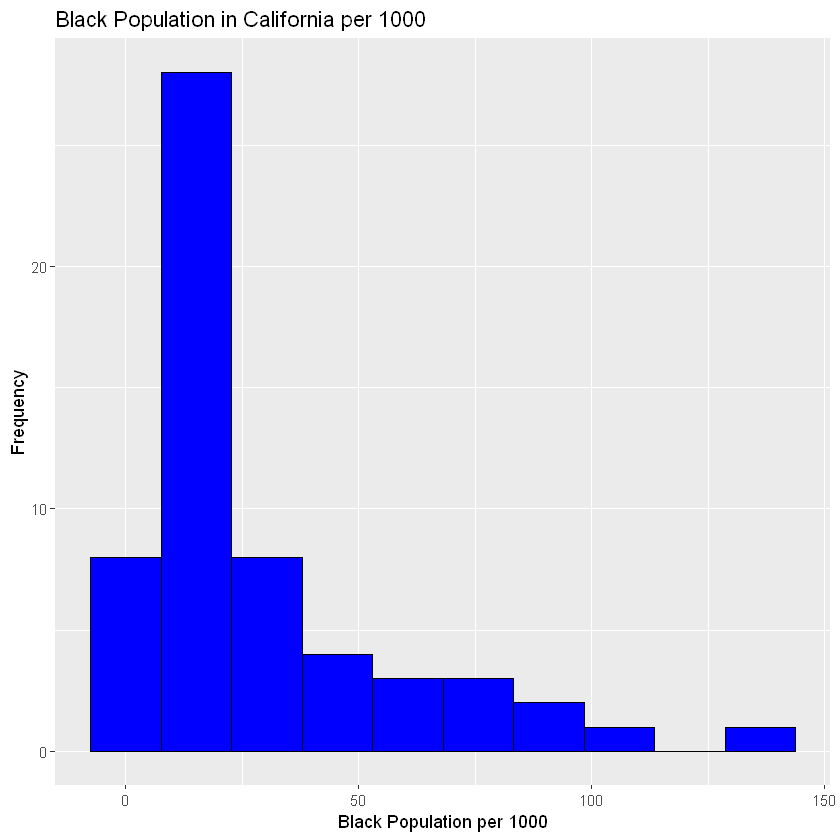
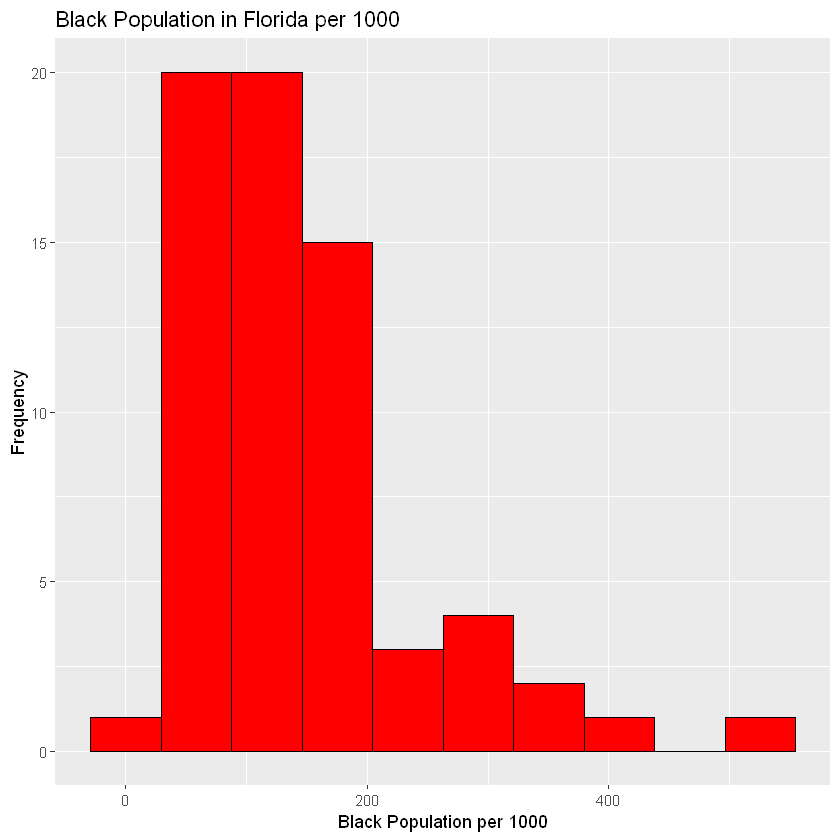
The Florida distribution follows a somewhat uniform distribution with a left skew. The majority of the data is around the 750 per 1000 mark. The skew is due to an outlier on the low end of the distribution, as well as a few counties under the 500 mark.

The California distribution does not follow any distinct distribution, as it contains two sections where most of the data falls under. One of these is in the approximately 275-475 range, and the other is in the approximately 600-900 range. There are not many counties in between these two ranges, and only a few below 275 per 1000.

*H.* *Black Population (Ratio)*

The data for black population by county is shown below in Figures 13 and 14. Analyzing the histograms shows that Florida has many more black people per 1000 than California does. Looking at the scale of the x-axis of the graphs, the California distribution only goes up to 150, while Florida has a county with close to 500 per 1000. The majority of the Florida distribution is in the 0-200 range, while the majority of the California distribution is below 50.

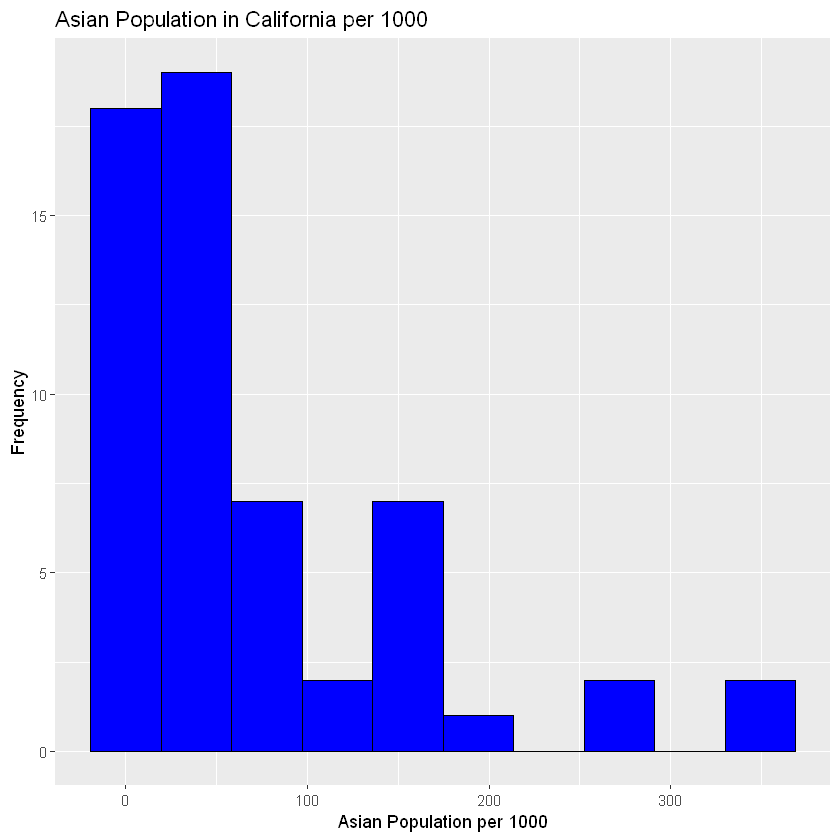
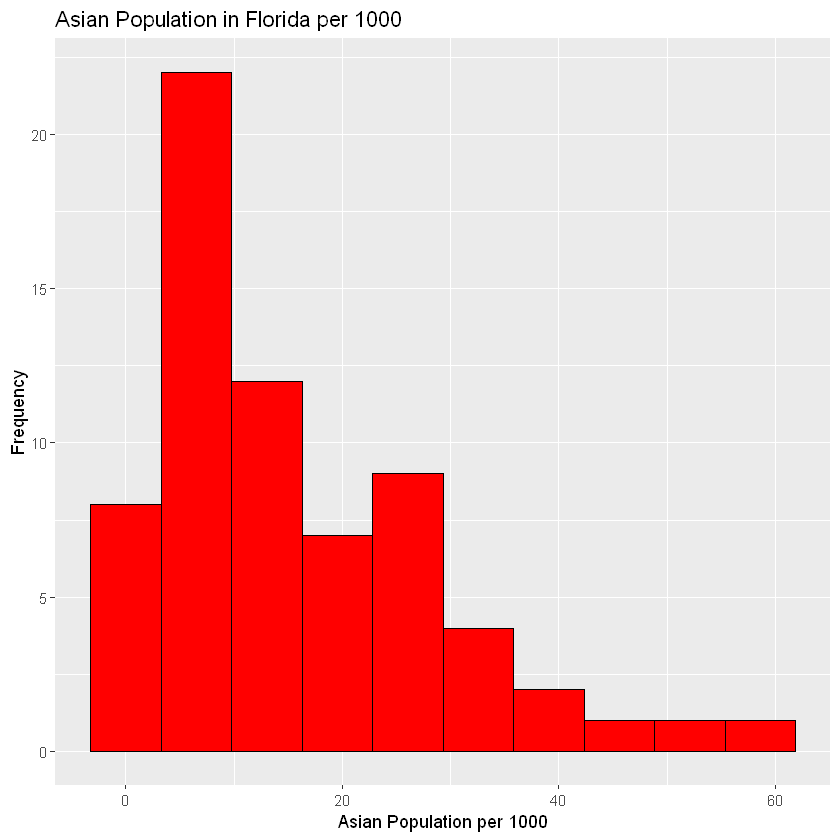
Figure 13 Black Pop FL Figure 14 Black Pop CA

**

*G.* *Asian Population (Ratio)*

The data for Asian population by county is shown below in Figures 15 and 16. Again, since the population is a ratio, it can be presented using a histogram. From the histograms, it can be seen that the Asian population in Florida is sparse. When analyzing the California graph we see that not only does California have a higher population of Asians but some counties have Asian populations as high as around 350 per 1000 people. All of the data in the Florida distribution falls under 100, with the majority being under 20 per 1000 people.

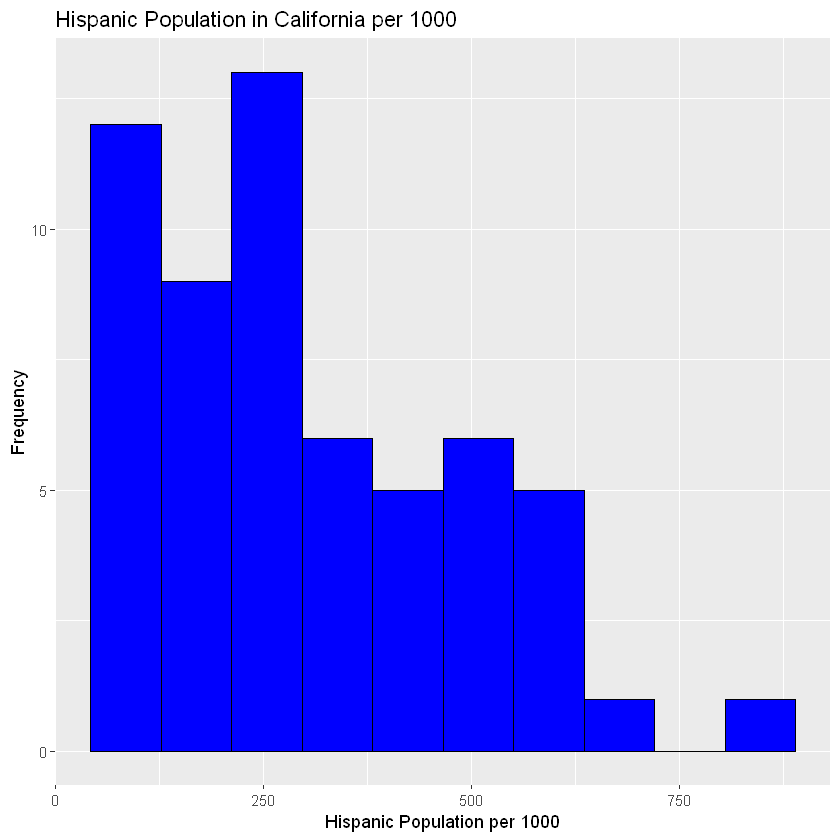
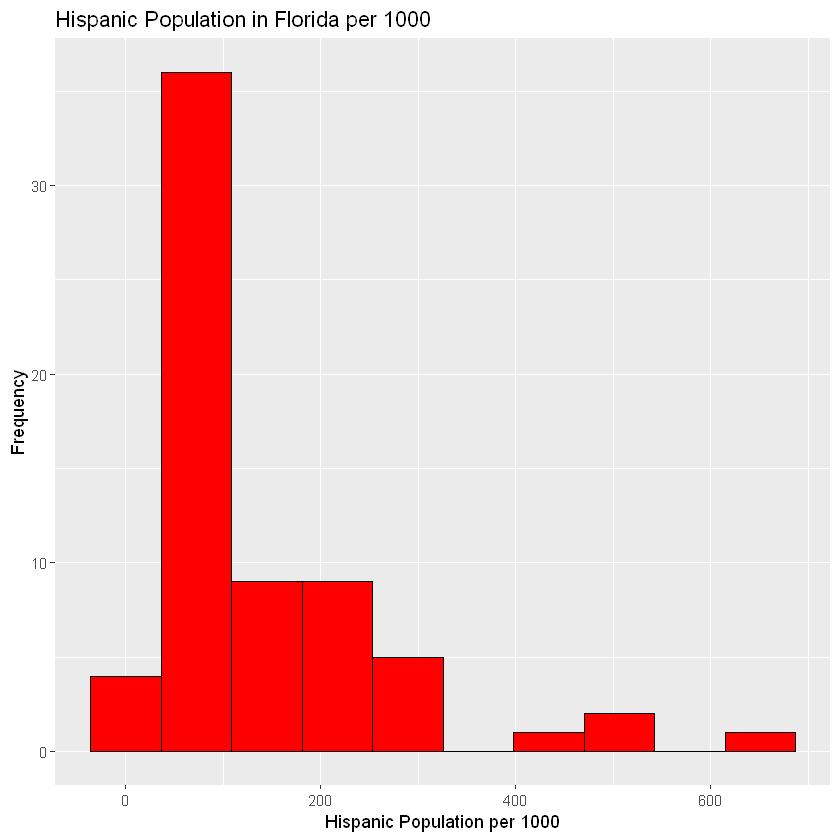
Figure 15 Asian Pop FL Figure 16 Asian Pop CA

**

*H.* *Hispanic Population (Ratio)*

The data for Hispanic population by county is shown below in Figures 17 and 18. Again, since the population is a ratio, it can be presented using a histogram. Analyzing the data we see that California far outnumbers Florida in Hispanic population. The Hispanic population in Florida is around 50-200 per thousand compared to California having around 250-500 per thousand. On top of this the graphs show a larger distribution of Hispanics in counties in California where the Florida graph suggests there are more concentrated populations of Hispanics.

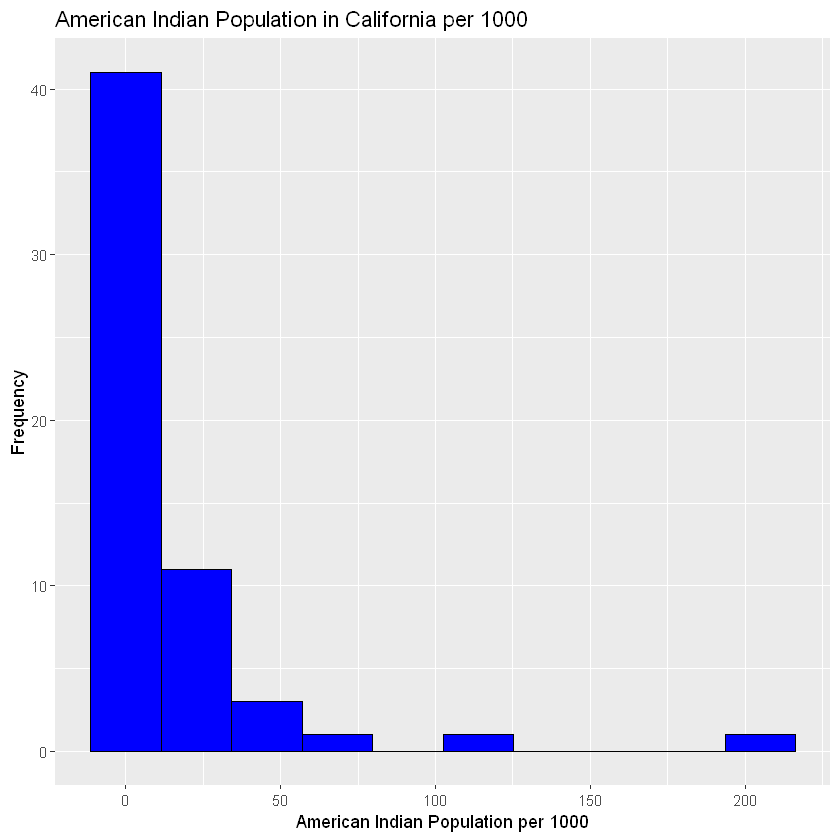
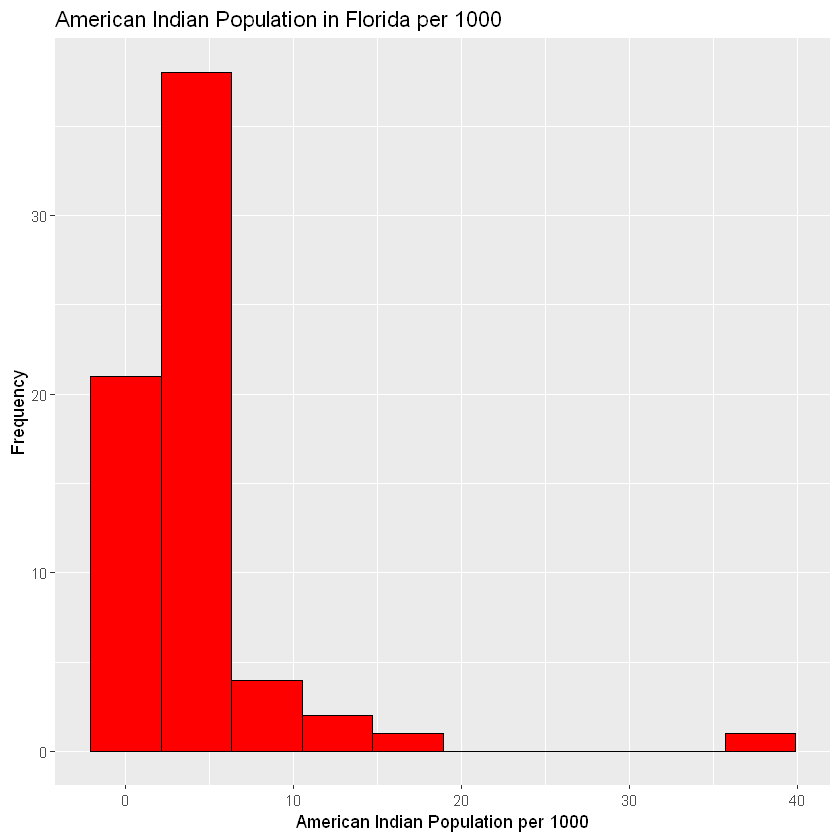
Figure 17 Hispanic Pop FL Figure 18 Hispanic Pop CA



*I.* *American Indian Population (Ratio)*

The data for the American Indian population is shown below in figures 19 and 20. Analyzing the histograms, it can be seen Florida has a much lower American Indian population. Also, the American Indian population is very sparsely distributed throughout Florida, with the most population at 35 per 1000. In California, the American Indian population is much larger and more concentrated in specific counties.

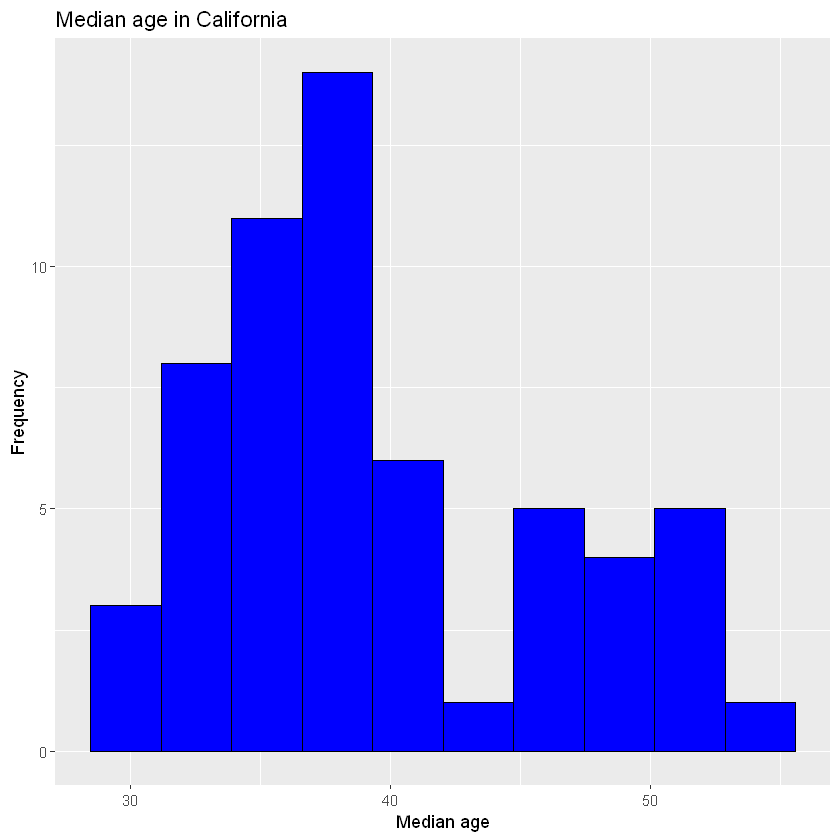
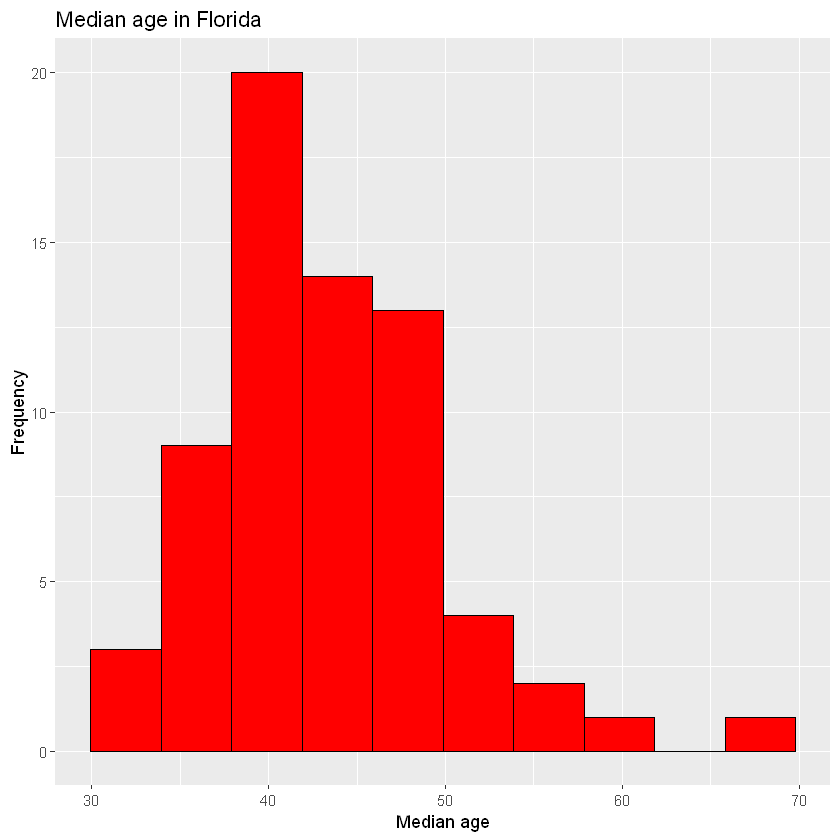
Figure 19 A. Indian Pop FL Figure 20 A. Indian Pop CA



*J.* *Median Age (Ratio)*

The median age attribute is a ratio, so it will be represented by a histogram in Figures 20 and 21. The two graphs share a similar shape, with the majority of the data on the low end with a skew to the right. The main difference between the two is the scale of the x-axis. Florida has a county with a median age of 70 which really stretches out the axis of its histogram. The majority of Florida’s counties fall between median age of 40-50, while the majority of California’s data falls below 40. There are a few counties in California near the 50 mark, but not nearly as many as Florida.Overall, the median age is much higher in Florida than in California.

Figure 20 Median Age FL Figure 21 Median Age CA



## 2.4 Exploring Relationships among Various Data Types

The relationship between attributes will now be explored in order to find which attributes are correlated, and which are not. To do this, Figure 22 and 23 will show a heat map of the correlations between the attributes for California and Florida respectively.

Figure 22 California Correlation Heat Map

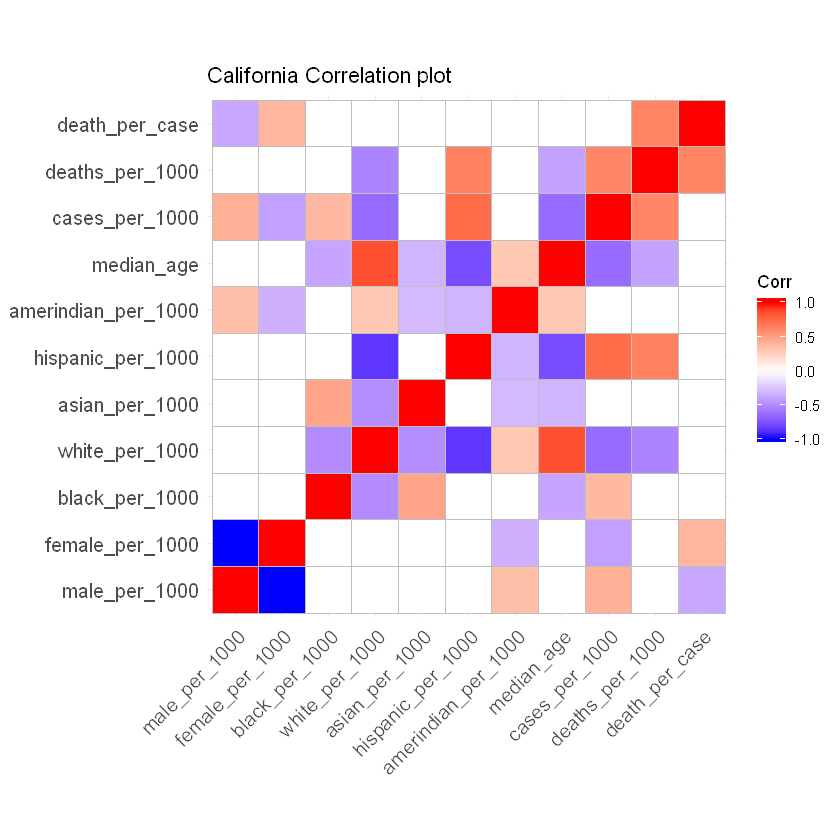
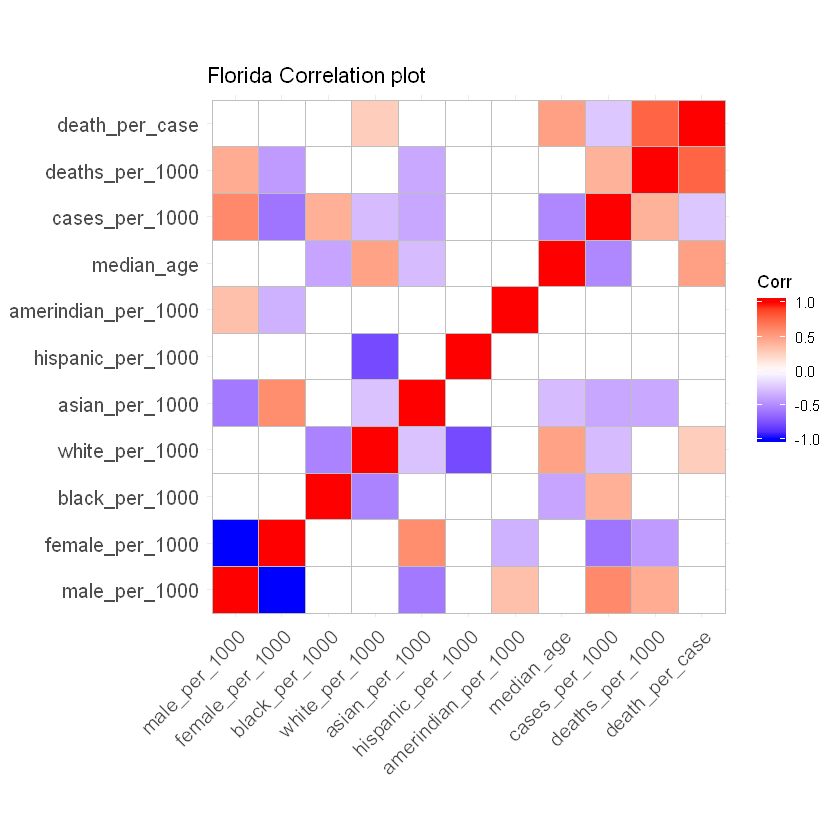


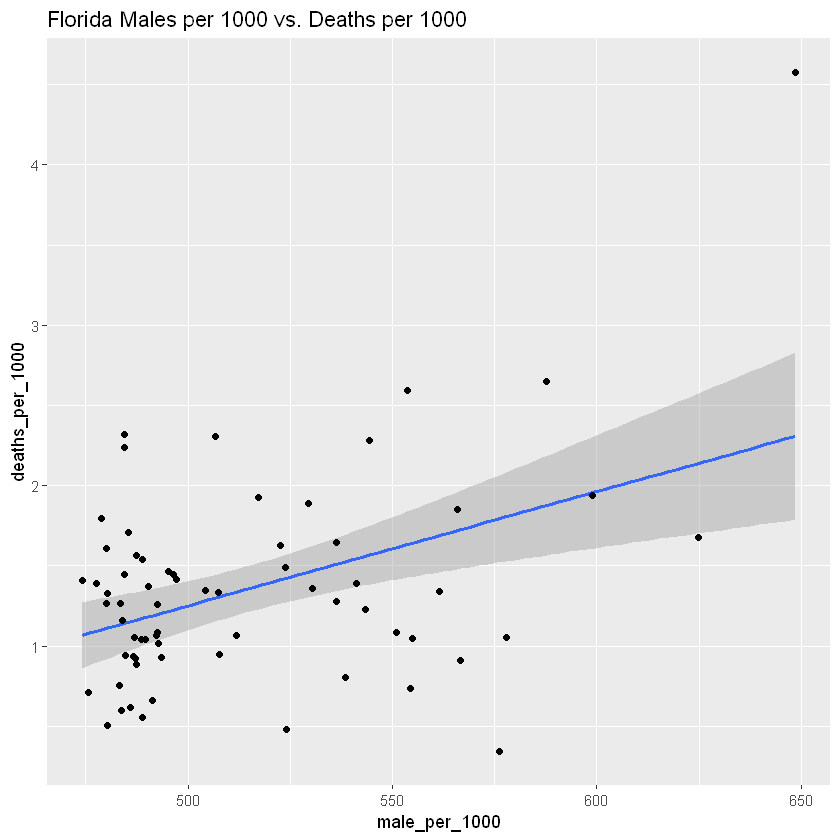
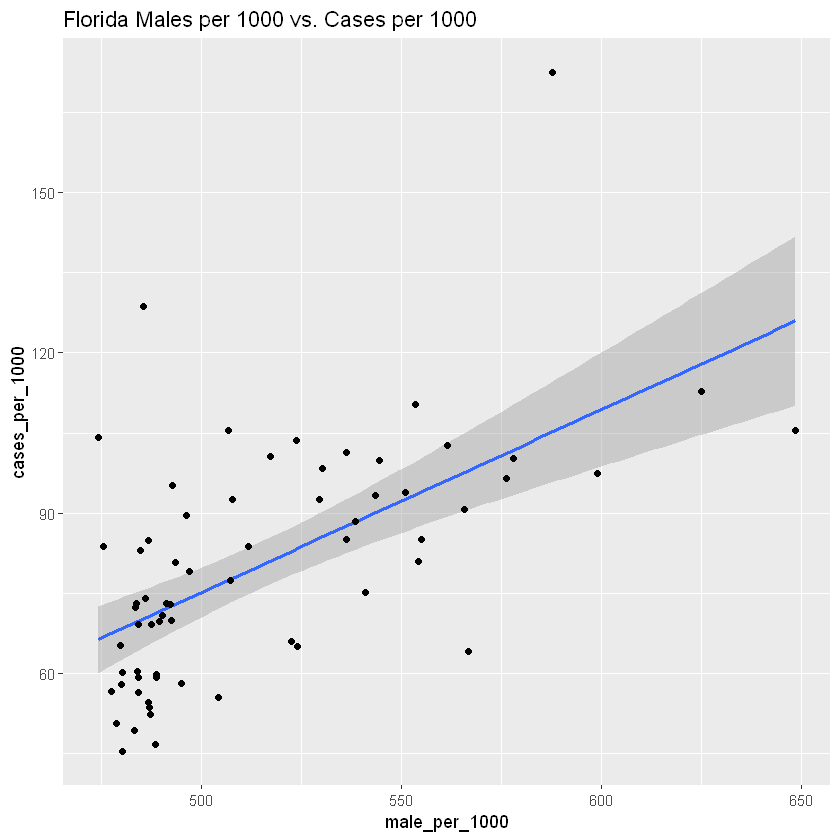
Figure 23 Florida Correlation Heat Map



*Female Vs Male for Cases and Deaths in FL and CA*

From the Florida heat map, it can be seen that males are positively correlated with confirmed cases per 1000 and deaths per 1000, while females are negatively correlated. Figures 24 and 25 show in more detail the correlation between males and cases, and males and deaths. Figures 26 and 27 will show detail for females.

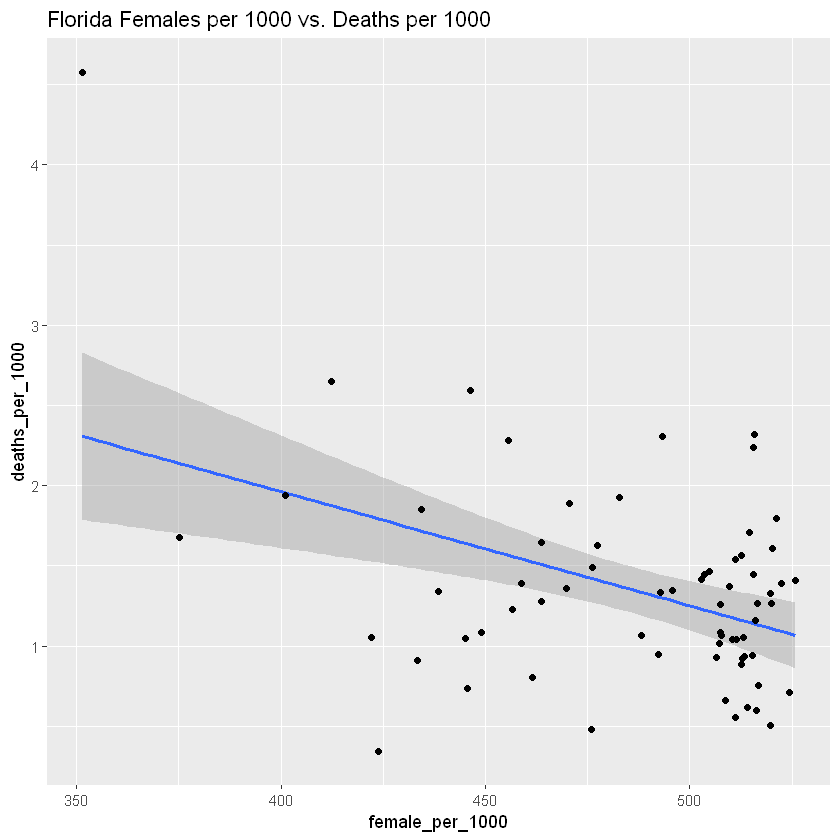
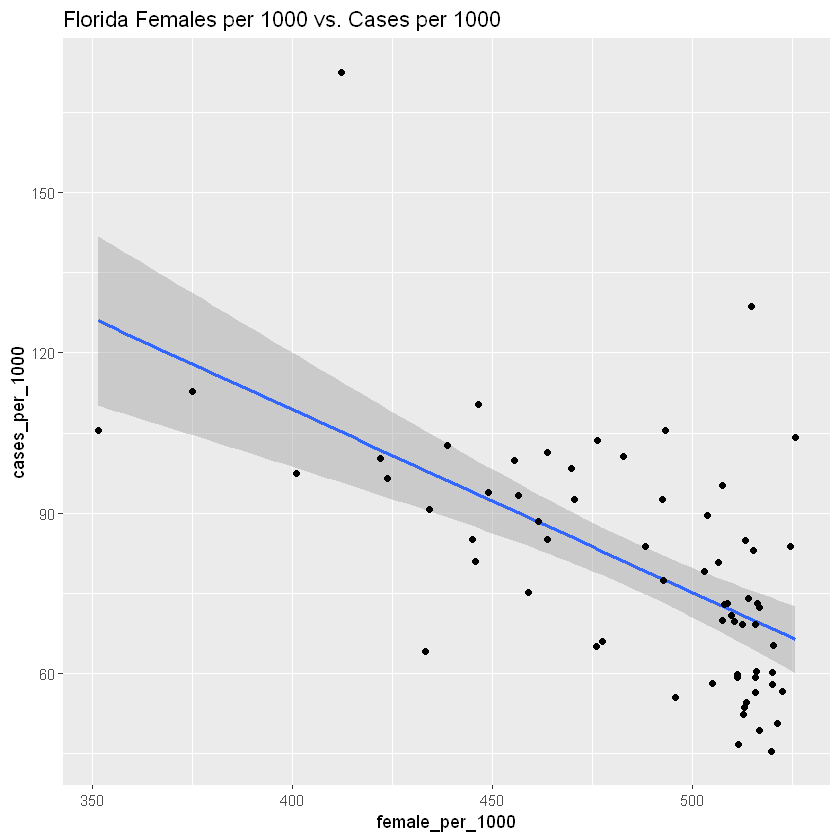
Figure 24 Males vs Cases per 1000 FL Figure 25 Males vs Deaths per 100 FL



Further observation of deaths and cases in Florida males shows a strong positive trend in the line of best fit. When observing the counties themselves, the trend in the regression plot keeps climbing. This shows that the slope of the line is determined by a multitude of data points, rather than a single outlier influencing the line. Along with a strong trend in cases, there is also a slight positive trend in deaths. Although marginal, it needs to be mentioned when comparing this data to the female data below.

Females on the other hand have the exact opposite trend in Florida. In cases and deaths, they negatively trend down. In Florida, the cases and deaths graphs seem to be quite correlated, as they have very similar slopes.

Figure 26 Females vs Cases per 1000 FL Figure 27 Females vs Deaths per 1000 FL



Now looking at the graphs for males in California, shown in figures 28 and 29, there is a positive correlation in cases, however this is mainly due to the influence of one outlier county. Without this county, the line of best fit would be much closer to a slope of 0, showing that there is not much of a correlation between males and cases, and males and deaths in California. The same can be said with the female graphs, shown in figures 30 and 31. The negative slope of the line of best fit is due to the single outlier county exerting influence over the rest of the graph. Again, without this county, this slope of the line for cases would be close to 0. It is interesting to see that the male and female graphs for deaths per 1000 do not share the slope of the cases graph, showing that while the outlier county had a large number of cases, it did not have a large amount of deaths per 1000 individuals.

 Figure 28 Males vs Cases per 1000 CA Figure 29 Males vs Deaths per 1000 CA

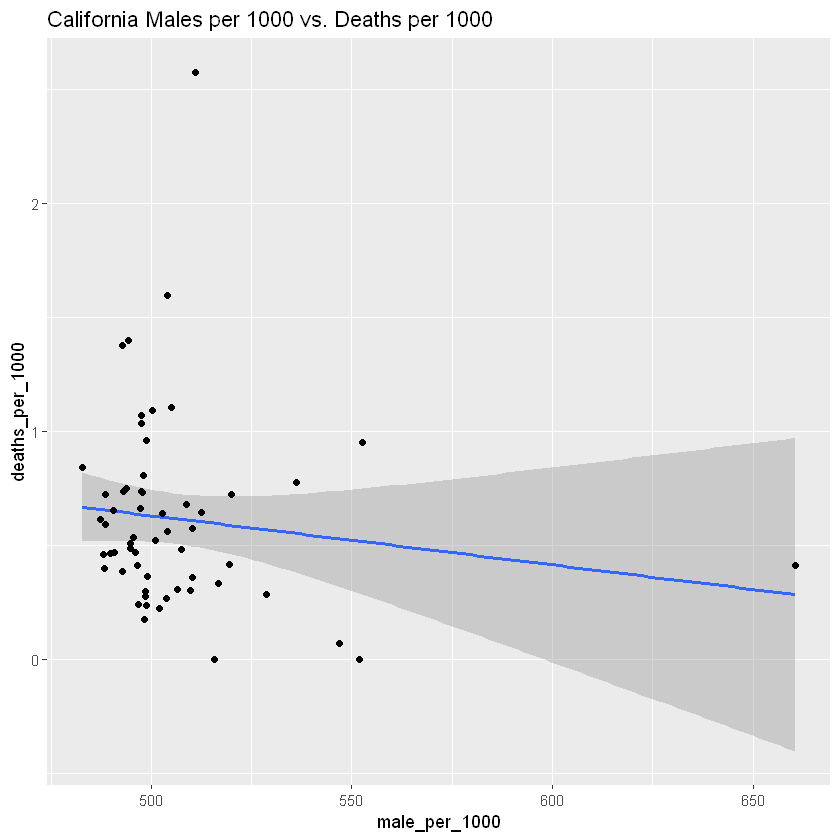
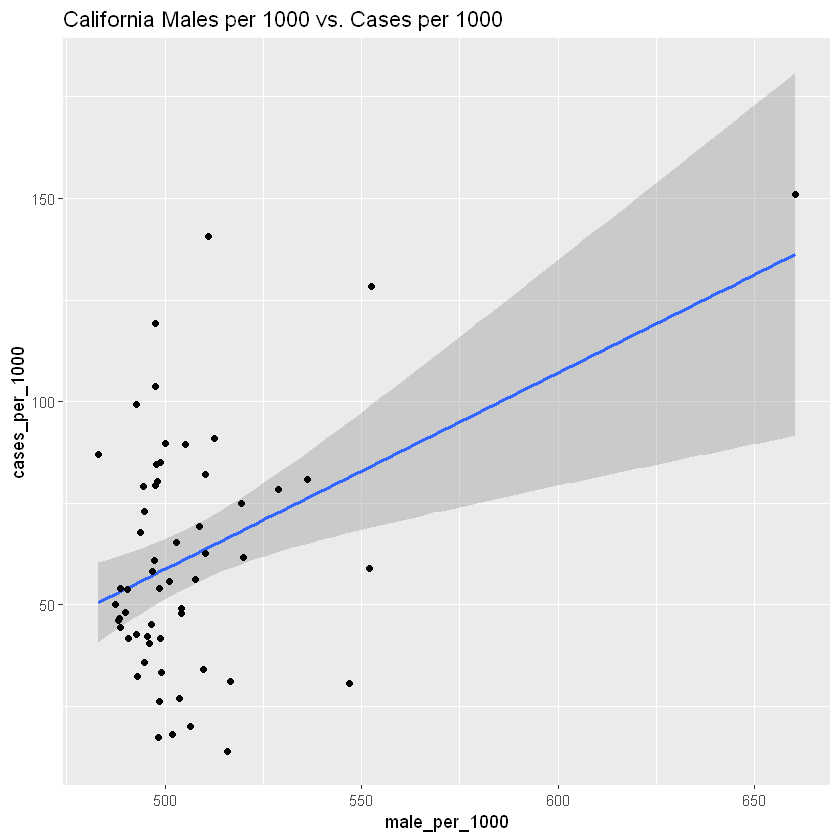
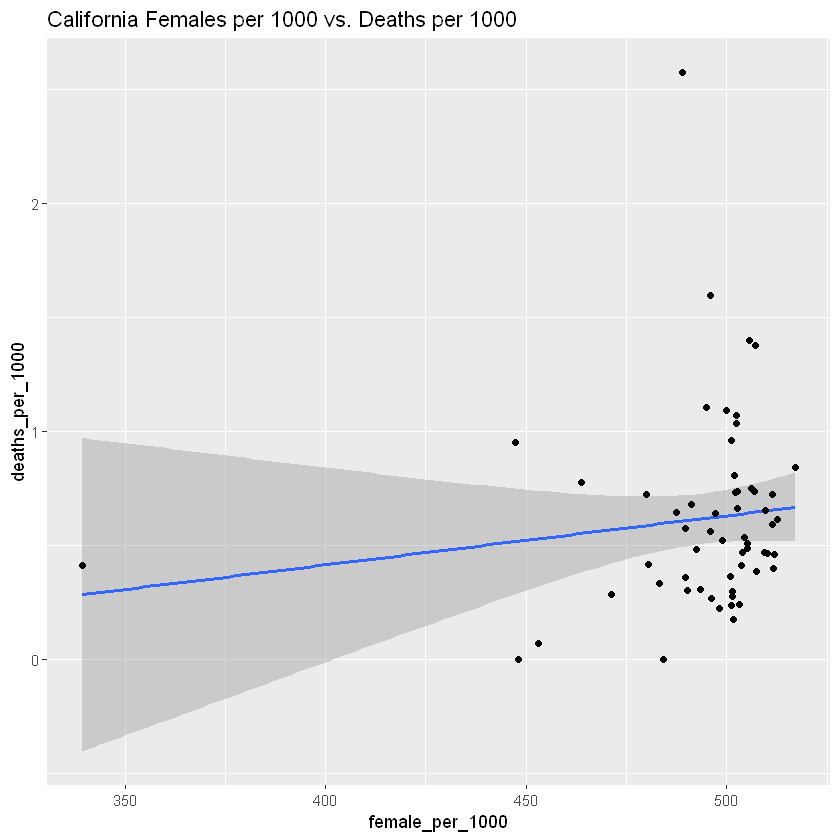
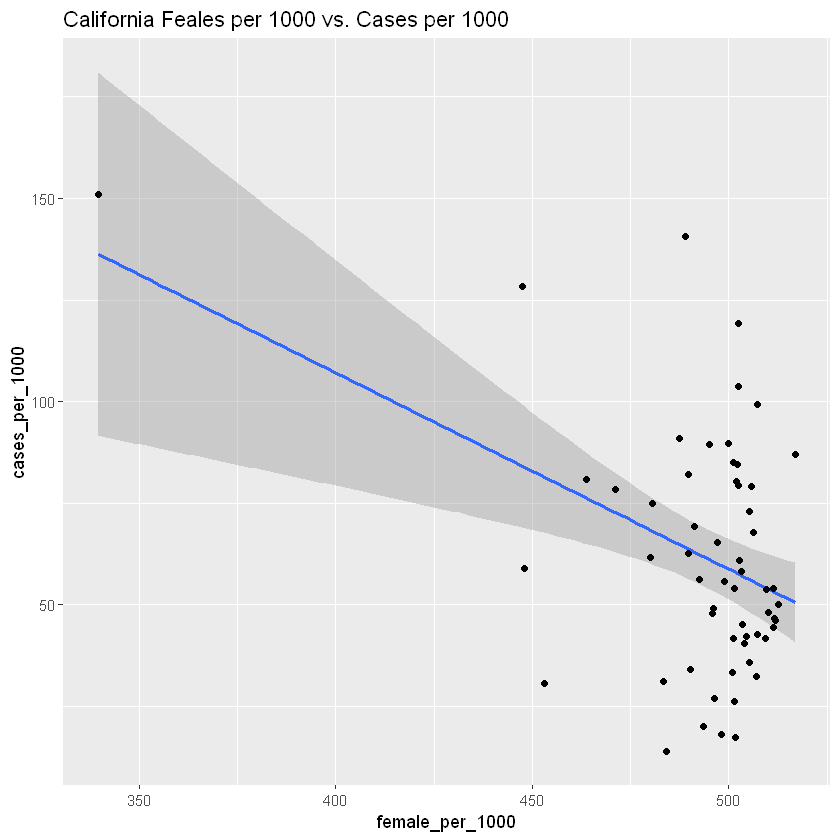


 Figure 30 Females vs Cases per 1000 CA Figure 31 Females vs Deaths per 1000 CA



*Cases and Deaths vs Median Age*

The next attributes to be investigated in further detail areMedian Age compared to cases and to deaths. The graphs for Florida cases and deaths will be shown in Figures 32 and 33, and the graphs for California cases and deaths will be shown in Figures 34 and 35.

Beginning with Florida, a negative correlation can be seen between median age and cases per 1000. This suggests that the more young people in a county, the more cases will be confirmed. Older people tended to stay isolated and distanced more than young people because they were at more of a risk from COVID, causing the younger population to contract more cases. Surprisingly, there is almost no correlation between median age and deaths in Florida. It was expected to see a positive slope, and while there is a slight positive trend, it is not significant enough to draw any conclusions from it. In fact, the county with the highest number of deaths had a median age of just under 40, which is on the lower end for Florida as discussed above in the analysis of the attribute.

Moving on to the California graphs, again the negative correlation between median age and deaths can be seen. Despite the differences in how the states were run during this time, the two share similar data when it comes to confirmed cases. The deaths in California have a slight negative correlation with median age, which again is surprising given how older people were at higher risk of death from COVID. Again the county with the most deaths is on the low end of median age, being close to 32.

Figure 32 Median Age vs Cases per 1000 FL Figure 33 Median Age vs Deaths per 1000 FL

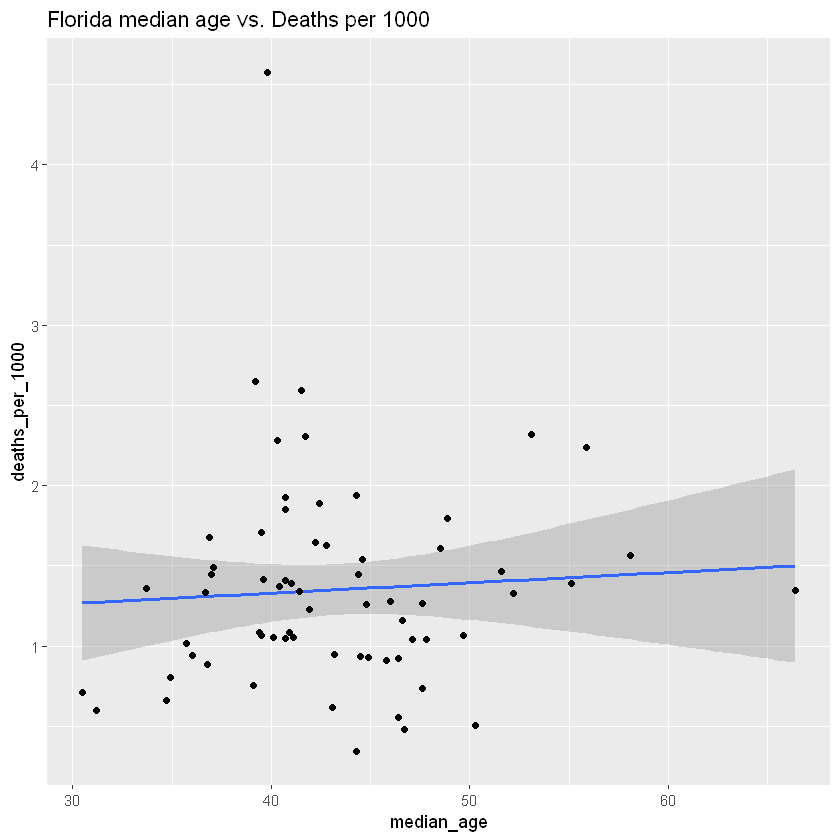
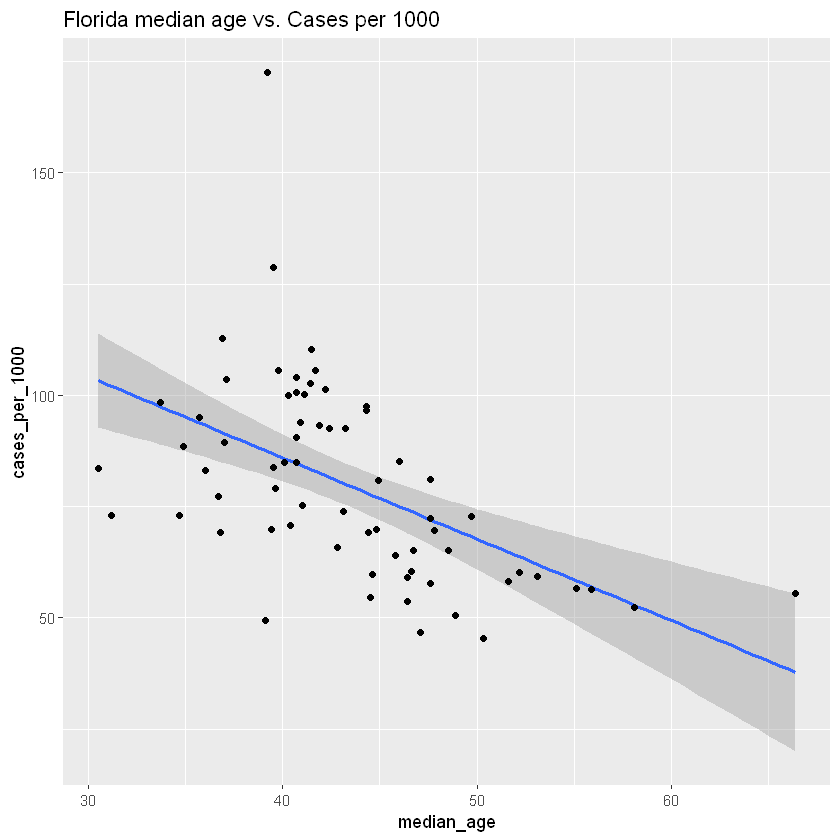
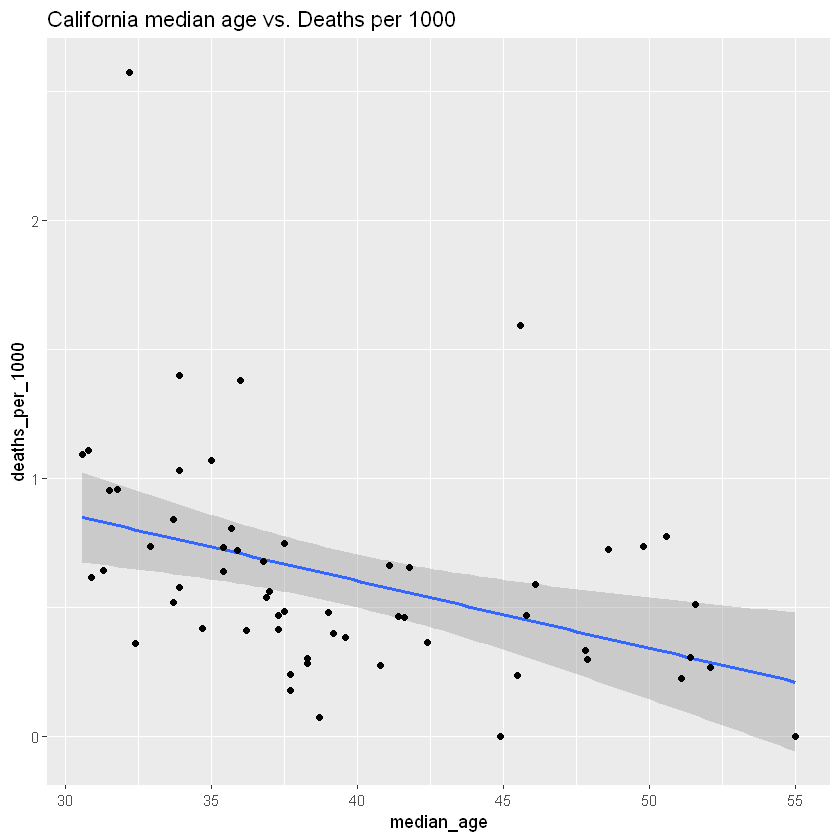
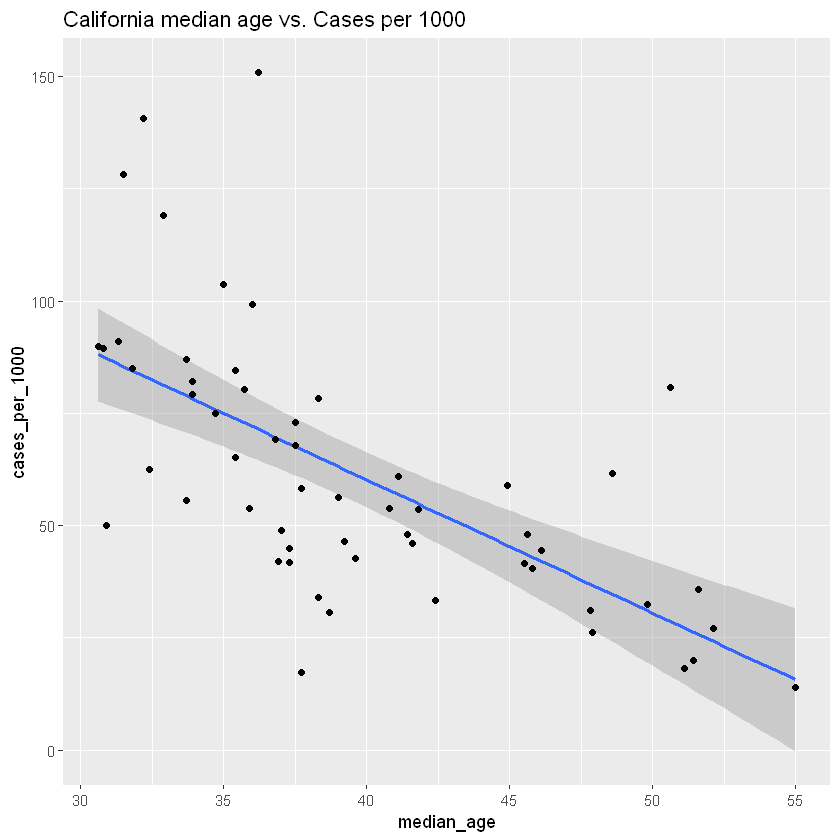


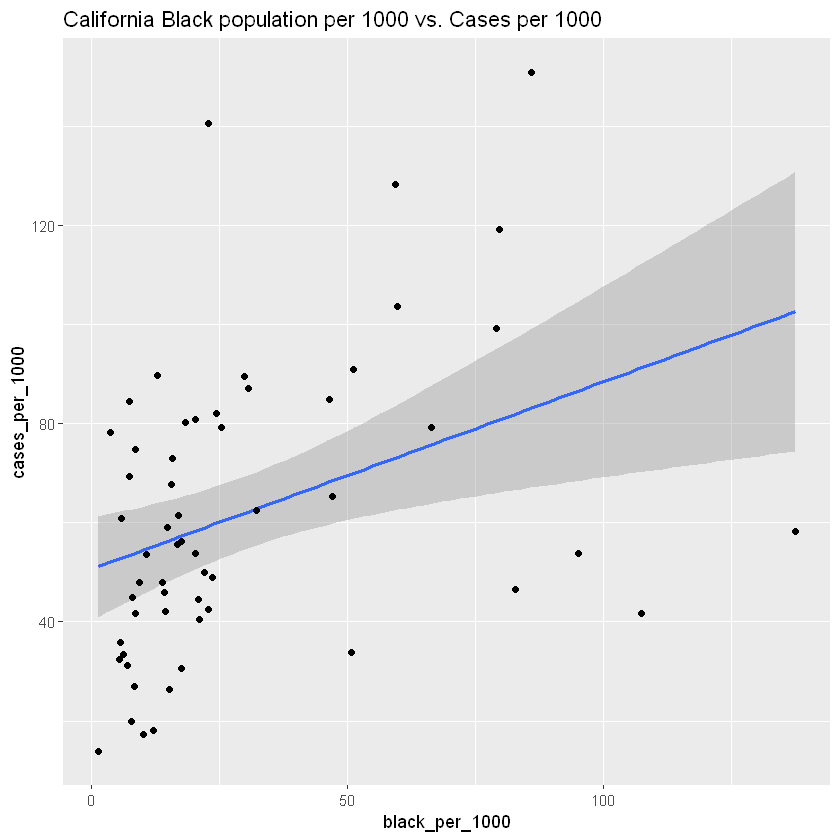
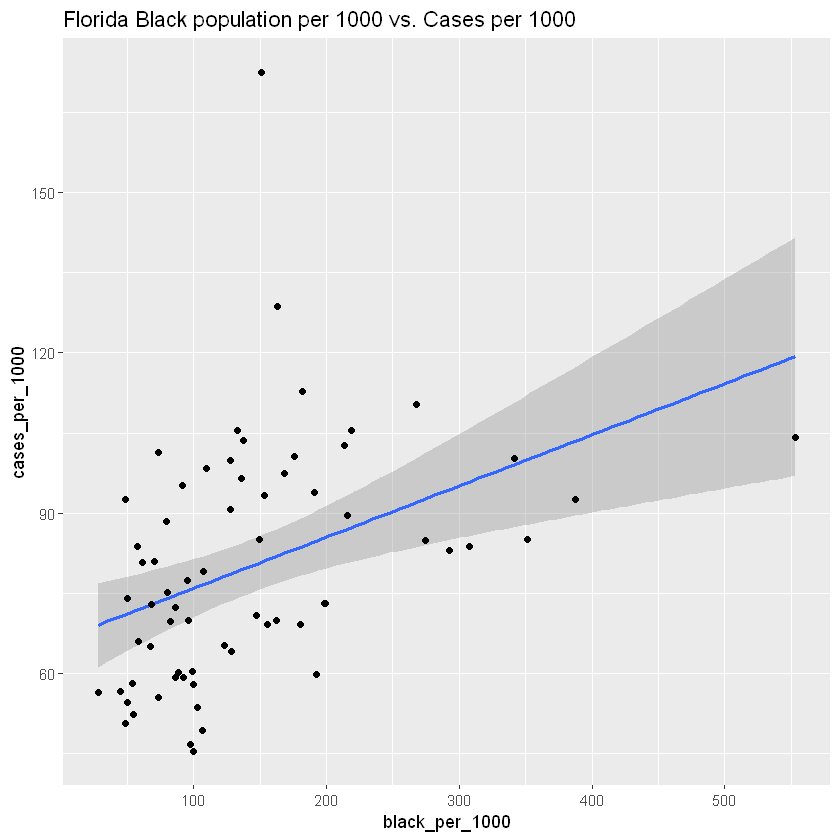
Figure 34 Median Age vs Cases per 1000 CA Figure 35 Median Age vs Deaths per 1000 CA



*C. Racial Demographics vs. Cases*

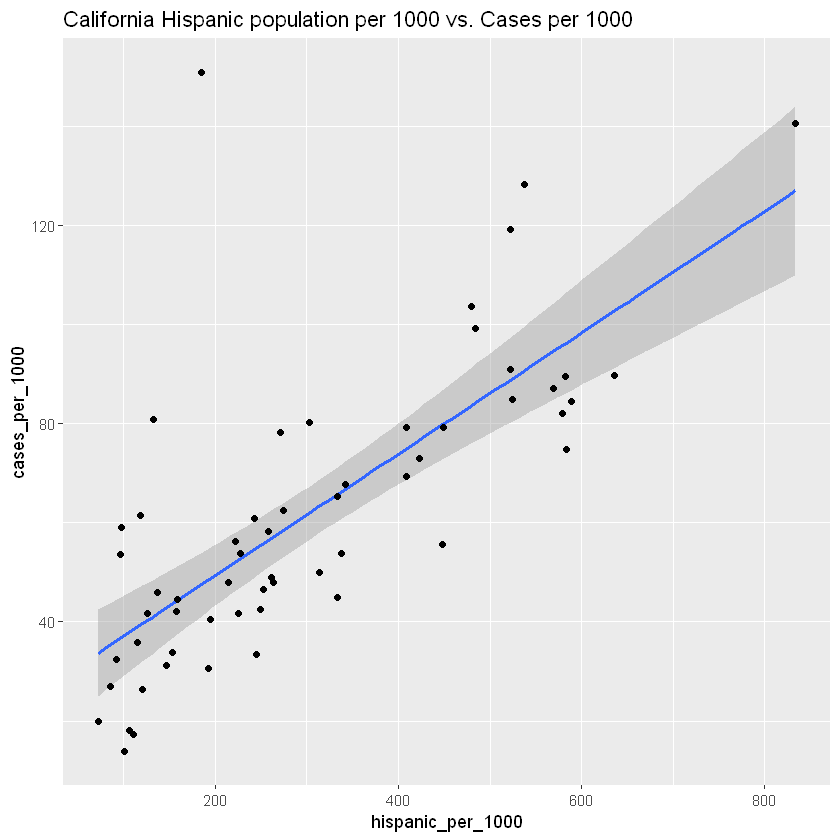
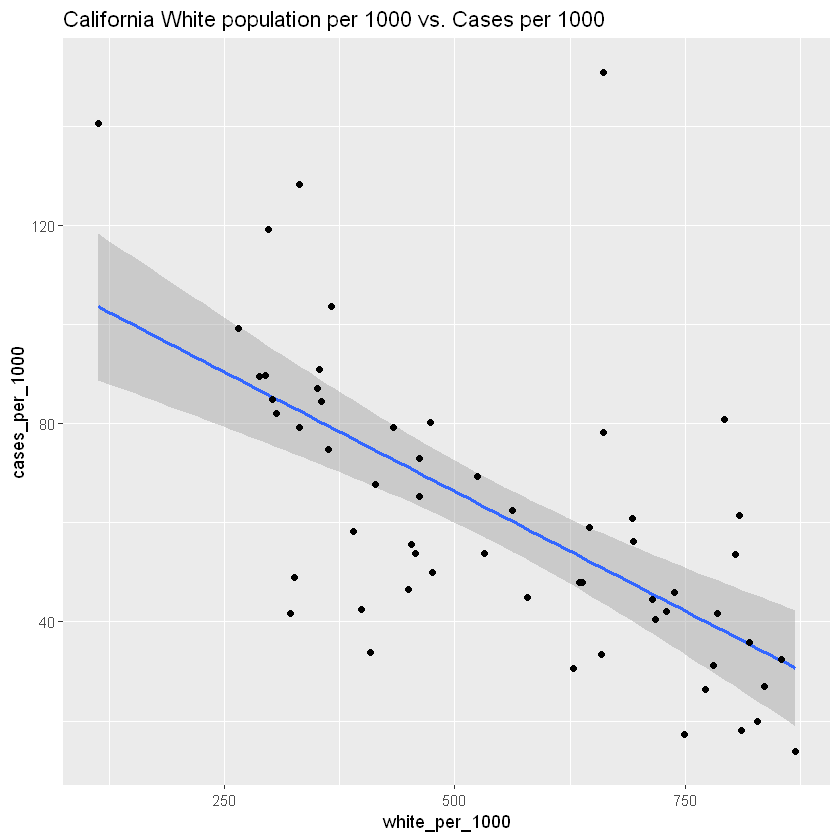
The relationships between racial demographics and confirmed cases will now be explored in more detail. Figures 36 and 37 show the Florida and California black population per 1000 individuals vs cases per 1000 individuals. As can be seen, the cases per 1000 slightly trends up as the black population per 1000 increases in Florida. This same trend is not seen in the California graph, as the data is much more scattered.

Figure 36 Black vs Cases per 1000 FL Figure 37 Black vs Cases per 1000 CA



Figures 38 and 39 show the Hispanic population vs cases and white population vs cases in California. As can be seen in the graphs, the cases strongly trend upward as the Hispanic population increases. The cases strongly trend downwards as the white population increases.

Figure 38 Hispanic vs Cases per 1000 CA Figure 39 White vs Cases per 1000 CA

**. 

These graphs of the racial demographics vs cases show that not all groups of people were affected the same by COVID-19. From the data, we could conclude that counties with larger minority populations had more COVID-19 cases than those with larger white populations. The trends for cases in California vs Hispanic population and cases in California vs white population show that it is possible that lockdowns could have had a larger negative effect on the hispanic population than the white population.

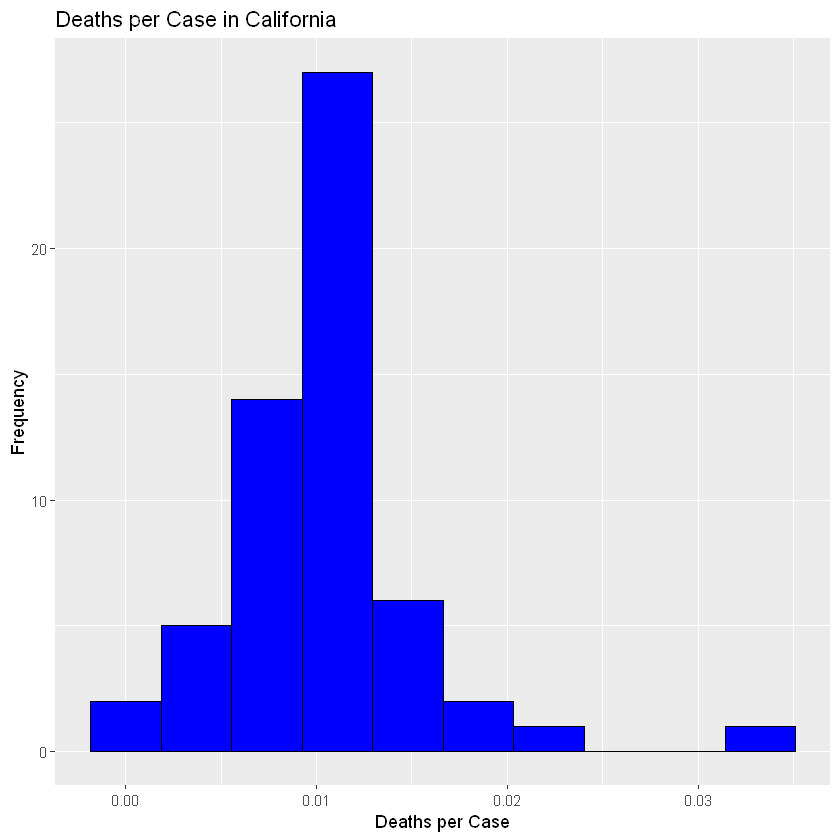
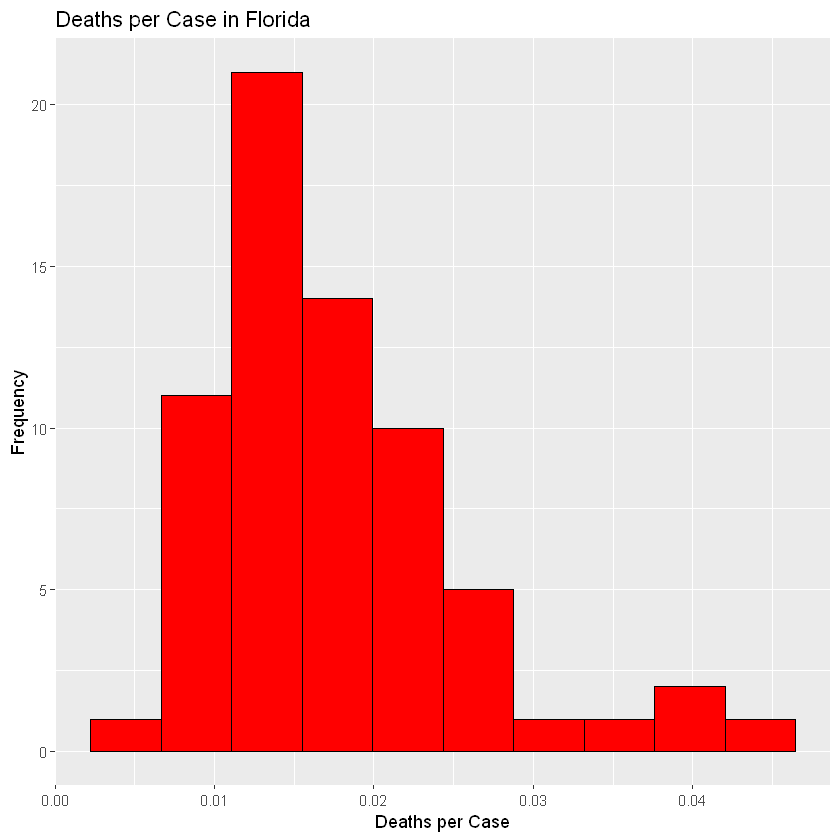
# 3. Data Preparation

In the previous section, the dataset was modified in order to normalize the data so that a fair comparison between California and Florida could be made without the gap in population between the two states affecting the results. In this section, the dataset will be modified slightly further in order to produce more attributes for the small dataset based on the analysis of the previous section.

Figures 40 and 41 show the visualization of deaths per case in Florida and California.

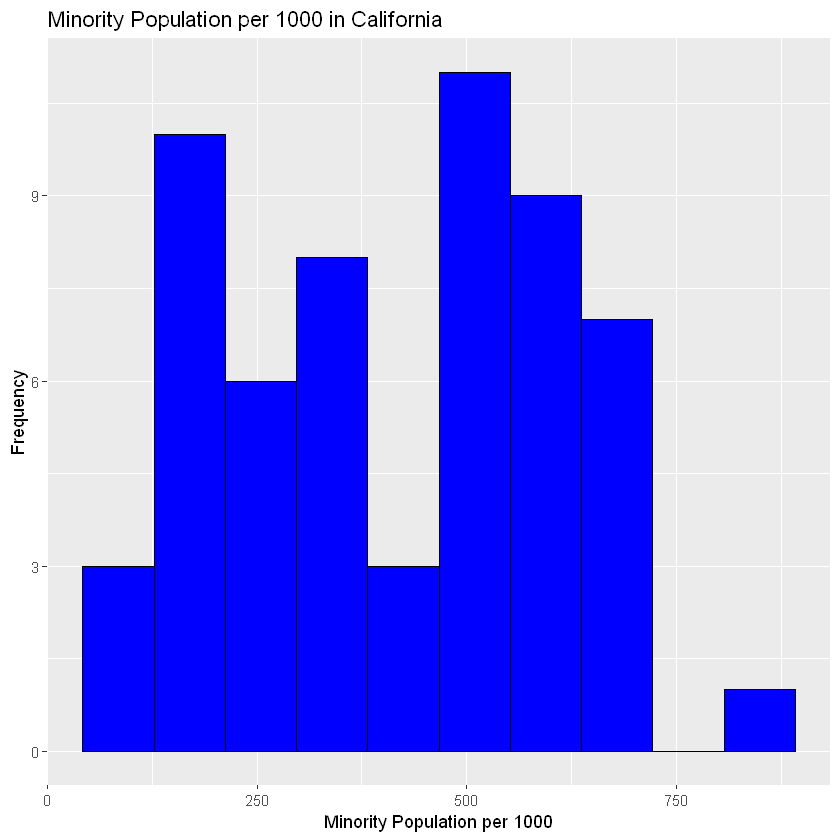
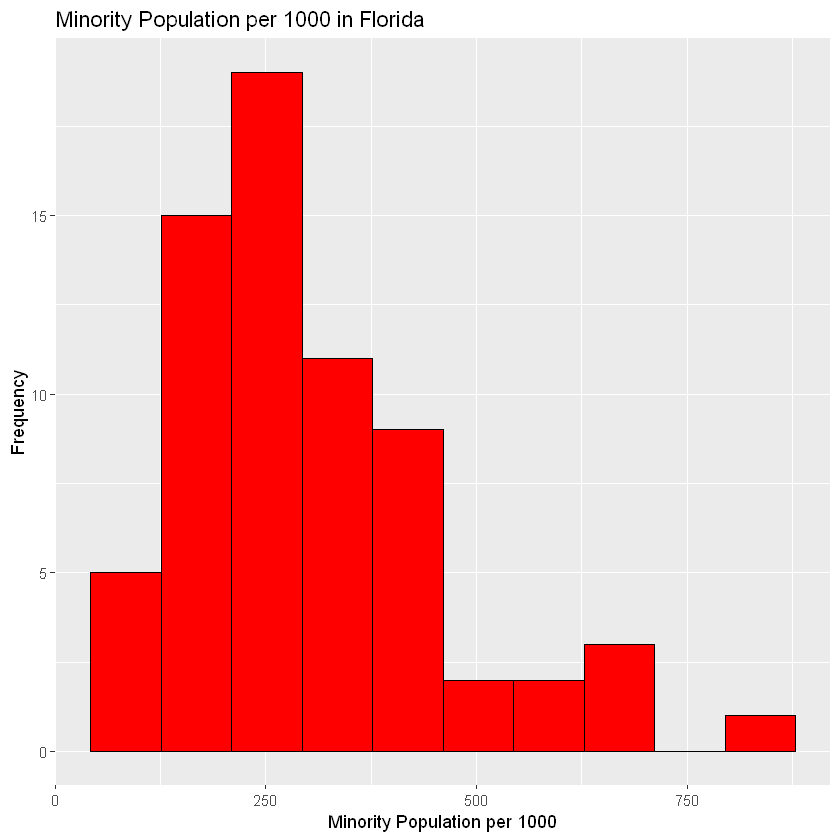
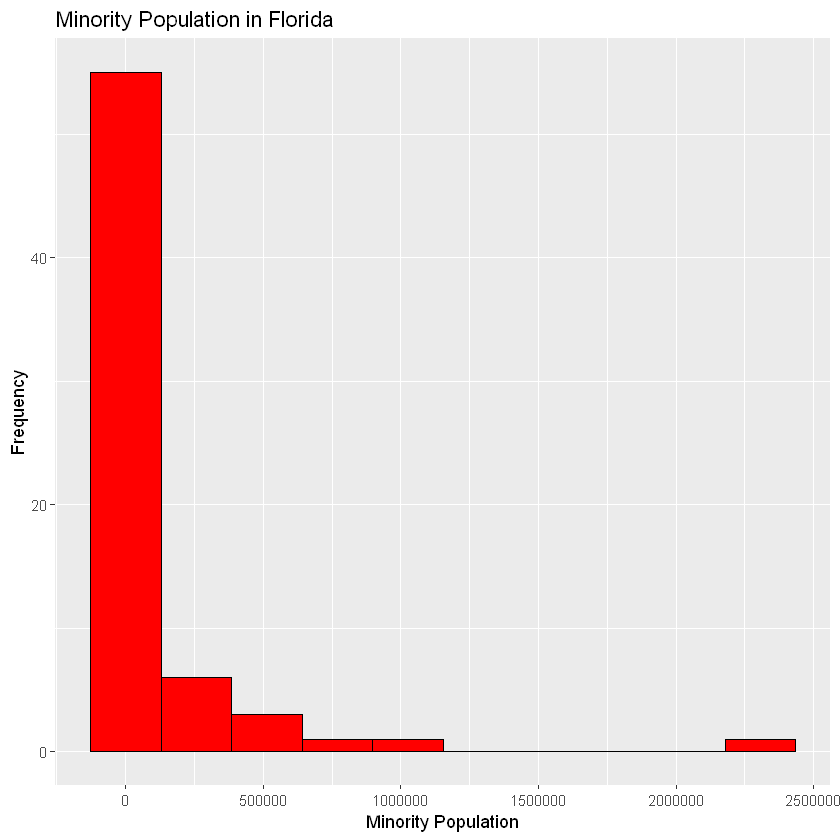
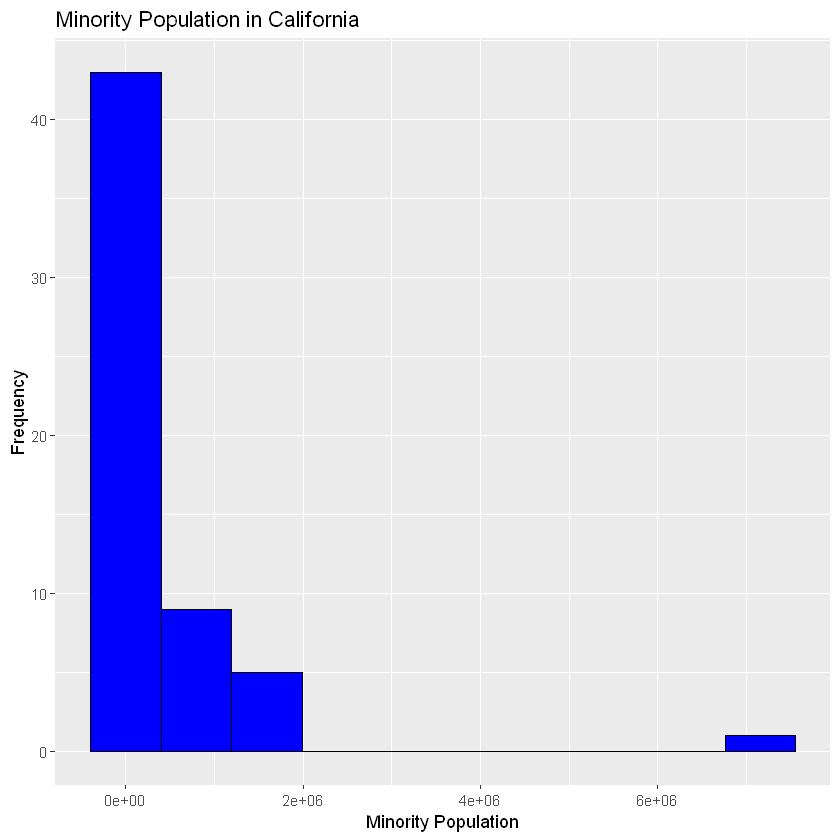
Analyzing the graphs, we see that the average deaths per case in Florida is slightly higher than the average deaths per case in California. Firstly, there are many factors of why this could be one being that Florida had a more relaxed approach to covid where California enforced strict lockdowns. Another also could be that California's hospital beds per capita is larger than Florida which could have allowed California to better manage the COVID-19 cases and provide more effective treatment to a larger number of patients. Also another reason could be that Florida's average age is higher and is home to a large senior citizen population. The exact demographic that got hit the hardest by covid which were people above the age of 50.

Figure 40 Deaths per Case FL Figure 41 Deaths per Case CA



In the previous section, the racial demographics of the states were explored. For a more general representation of this, all non-white racial demographics have been combined into one attribute. Figures 42 and 43 show the non-normalized and normalized data for Florida, and Figures 44 and 45 show the non-normalized and normalized data for California. Looking at the non-normalized graphs, they are showing just the raw population data for each state. Because they have not been normalized, the two graphs are quite similar, which highlights the need for the normalization. The vast majority of the counties are on the low end of the graph, with one county in each state causing a significant skew to the right. To get a clearer picture of how the minority population differs between the two states, the normalized data is necessary. In the normalized graphs, it is clear that there are distinct differences in the population of the two states. Florida has the majority of its counties below the 500 mark, while California has around half of its counties above the 500 mark. California’s population is clearly more diverse than Florida’s, which is easy to see due to the normalization.

Figure 42 Non-Normalized Minority Pop FL Figure 43 Normalized Minority Pop FL

  
  
  
Figure 44 Non-Normalized Minority Pop CA Figure 45 Normalized Minority Pop CA  
  


# 4. Exceptional Work

  For further work, the question “What aspects of everyday life were affected the most by the pandemic?” will be explored.

Figure 46 shows Florida’s change from baseline travel to parks, retail/recreation, and workplace environments. As is shown in the graph, the pandemic caused all categories to dip to around -30% to -40% in April 2020. As the summer started, everything spiked 10-20% and then evened out to about -20% by January 2021.

Figure 46 Change from Baseline FL

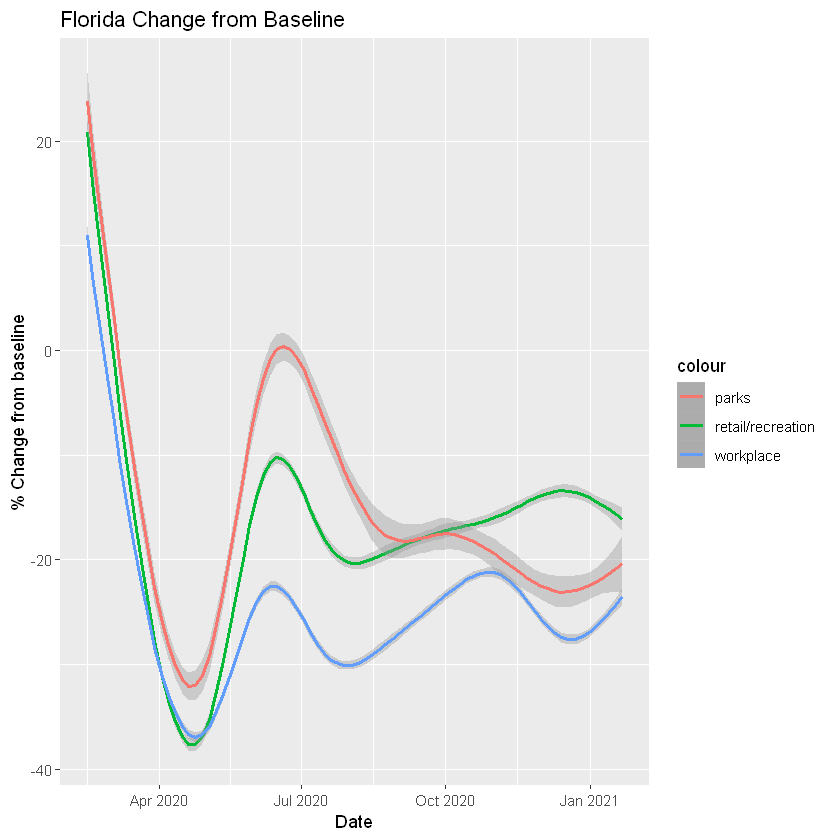
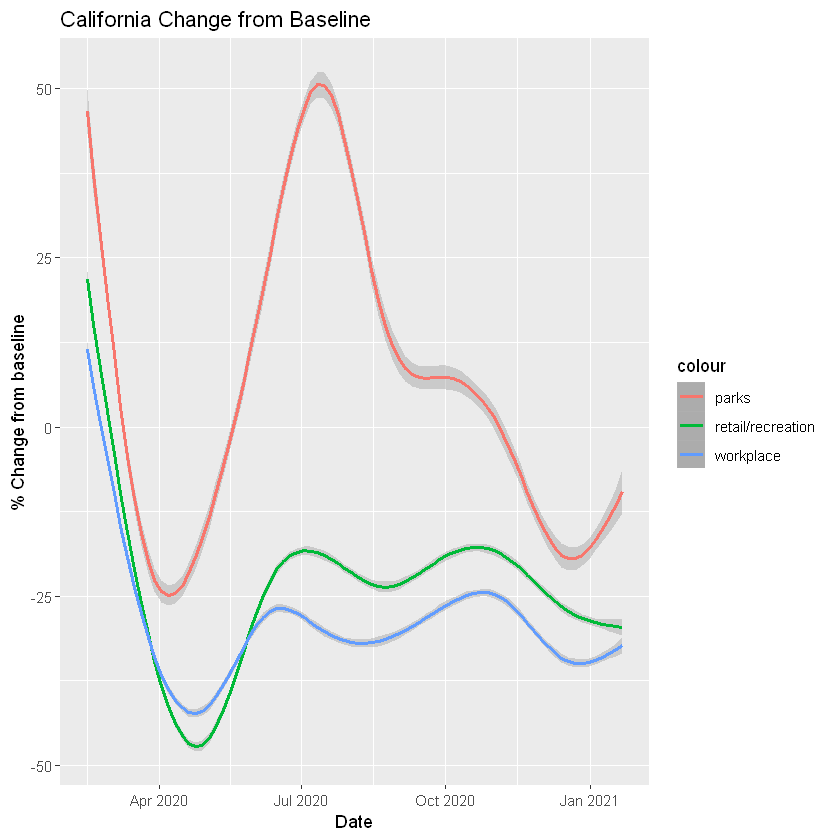


Figure 47 shows California’s change from baseline travel to parks, retail/recreation, and workplace environments. As is shown in the graph, the pandemic caused parks to dip to around -25% and work and retail to around -45% in April 2020. As the summer started, work and retail slightly spiked up to -25%, but the parks shot all the way up to +50%. By January 2021, the parks were about even, and retail and workplace stayed around -25%.

Figure 47 Change from Baseline CA



From these graphs, it could be concluded that more lockdowns had a negative impact on business, but not by much. During the pandemic, the work and retail change from baseline dipped almost down to -50% in California and evened out around -25%, while Florida only dipped to almost -40% and evened out around -20%. But it can clearly be seen that the lockdowns did not have much impact on people traveling to parks in California during the summer. In July 2020, California went up to +50% from baseline in the parks category, and ended up at about even. This is likely because outdoor activities became popular as COVID-19 is mostly spread through the air indoors.