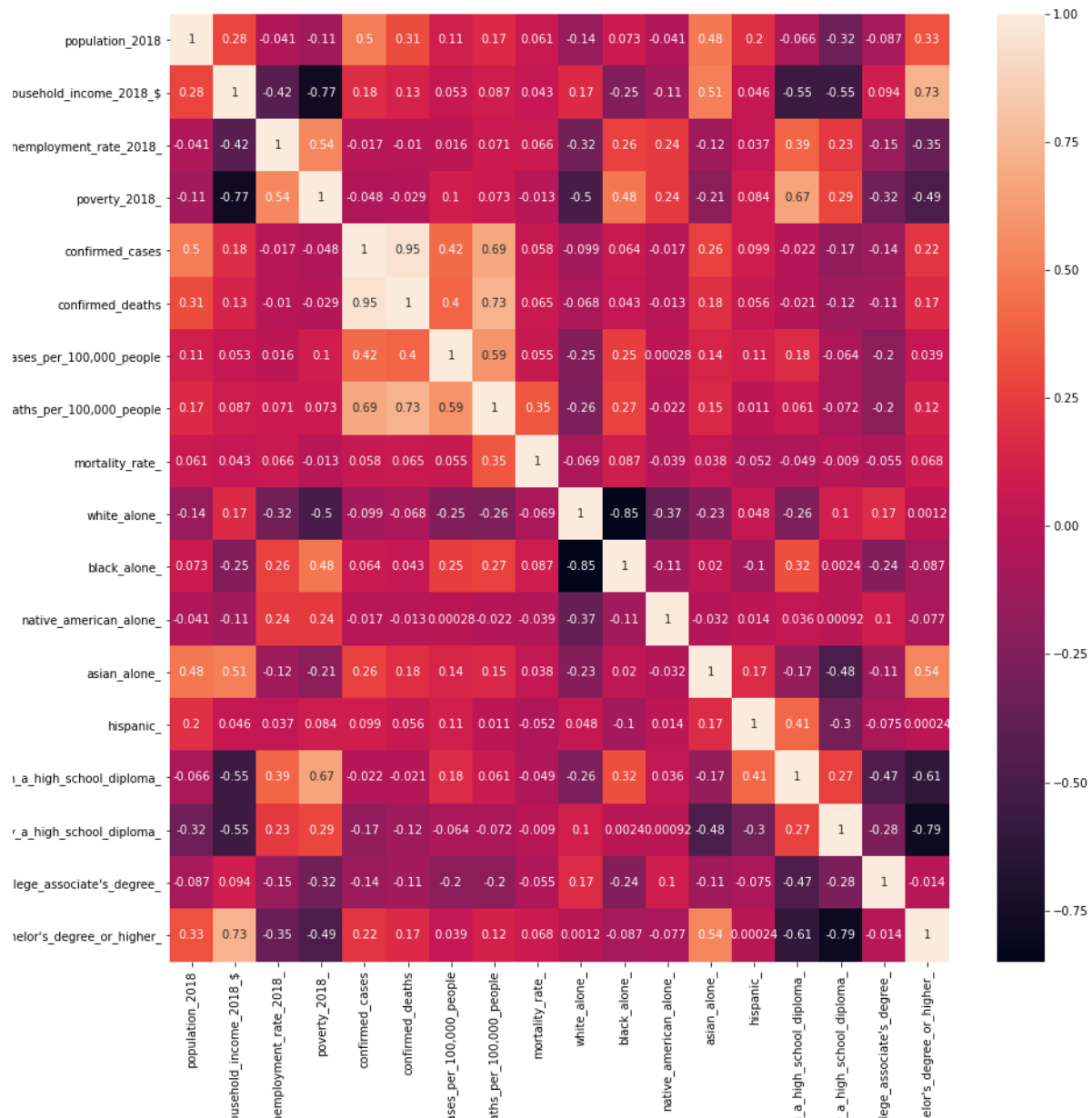
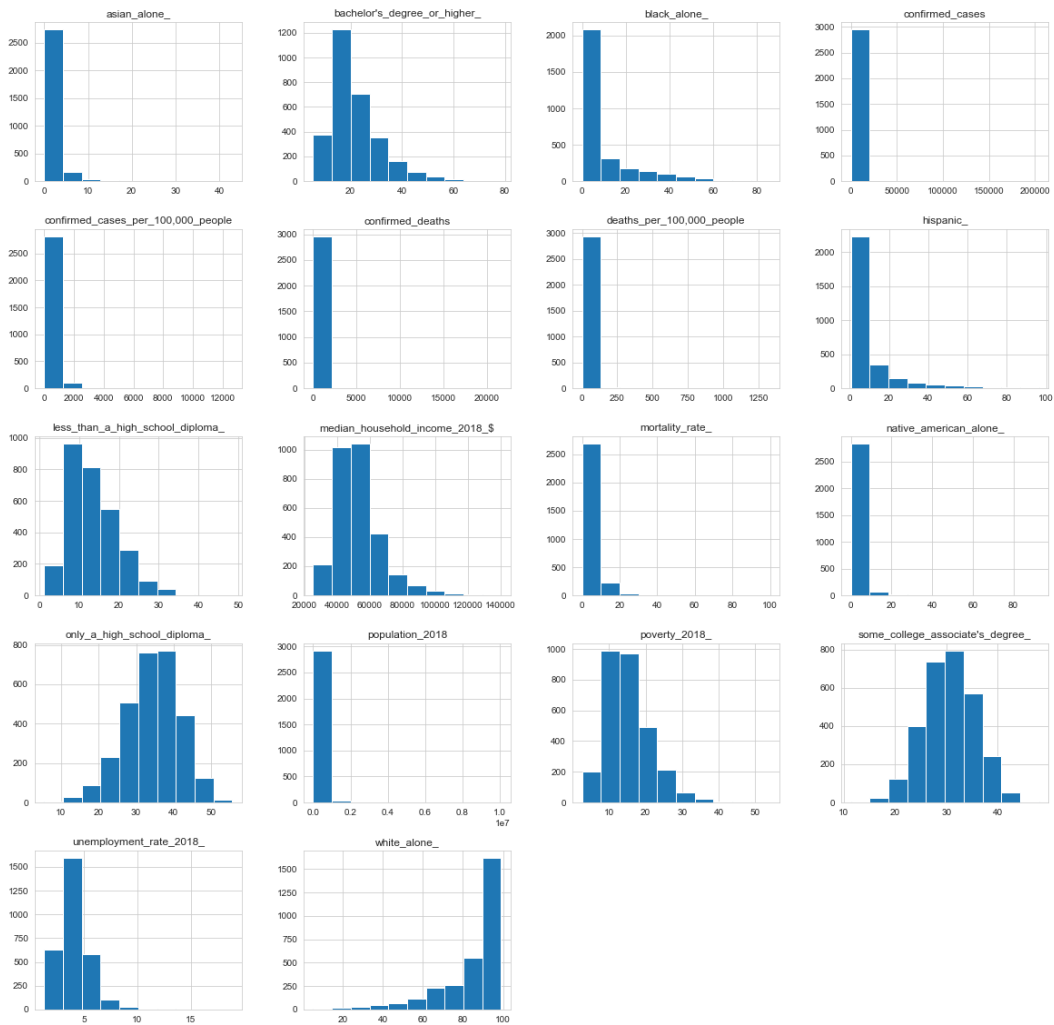


The following correlation table shows the correlation between all the features.



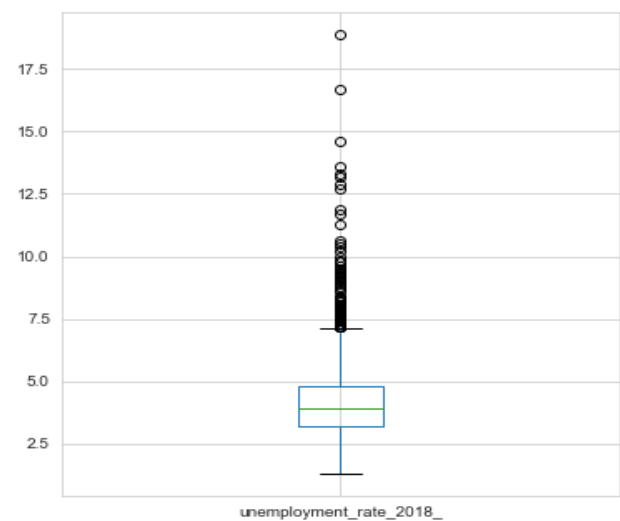
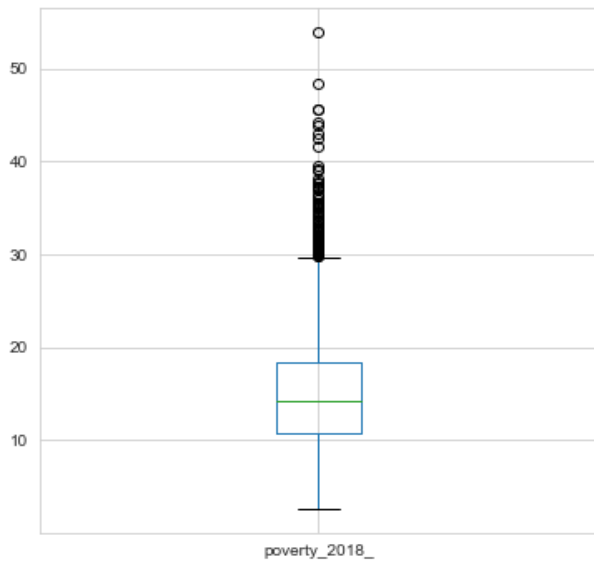
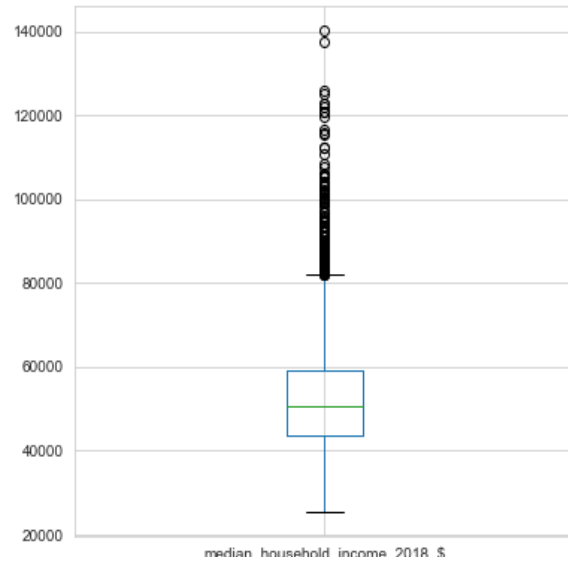
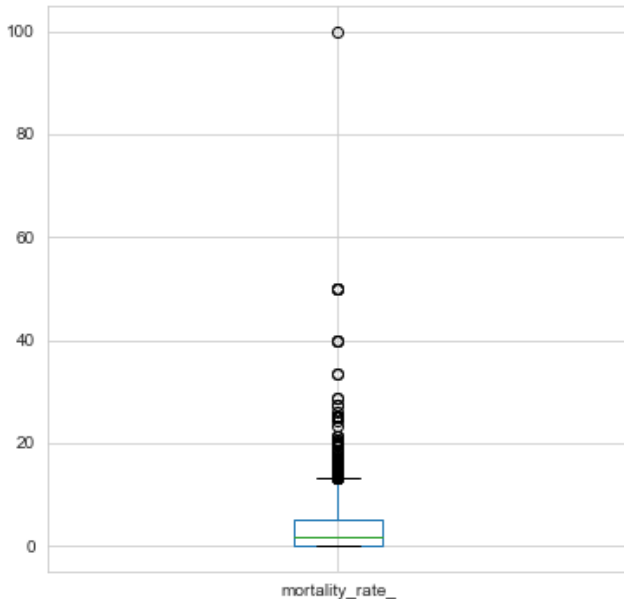
Correlation summarizes the strength and direction of the linear (straight-line) association between two quantitative variables. Denoted by r , it takes values between -1 and +1. A positive value for r indicates a positive association, and a negative value for r indicates a negative association. The closer r is to 1 the closer the data points fall to a straight line; thus, the linear association is stronger. The closer r is to 0, making the linear association weaker. For example, Poverty of the county is negatively correlated to the median household income of the county i.e., Higher the median household income, lower the poverty in that county. Counties with higher "white only" population has lower "black only" population and vice-versa. Counties with higher "white only" population is negatively correlated to poverty, which is the complete opposite in the case of counties with "black only" population. Unemployment is Positively correlated to the poverty i.e., higher the unemployment higher the poverty.

The following chart shows the histograms of all the features.



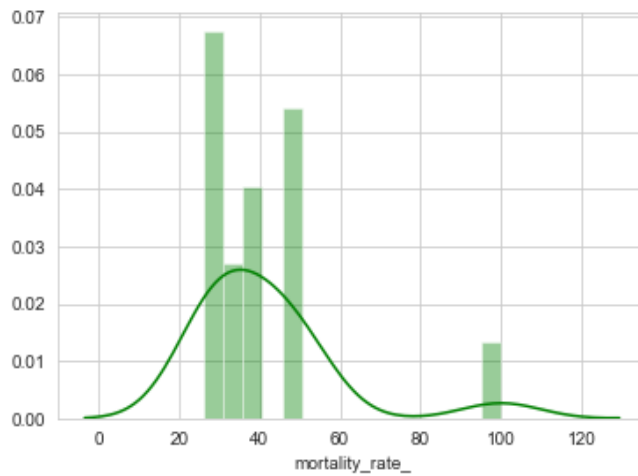
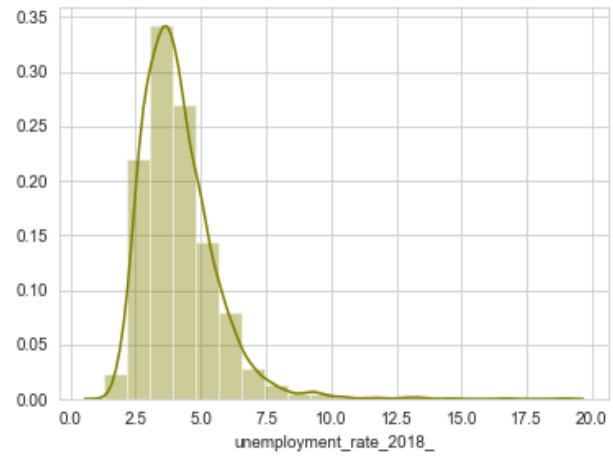
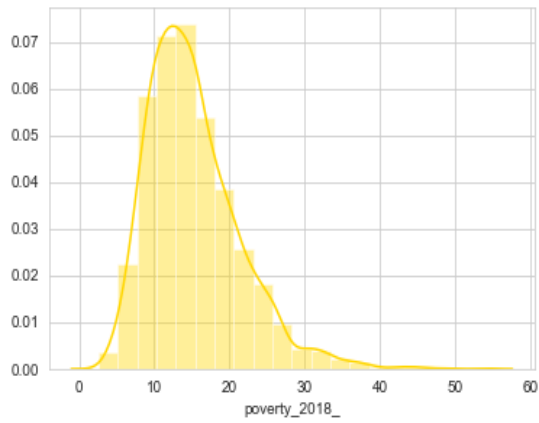
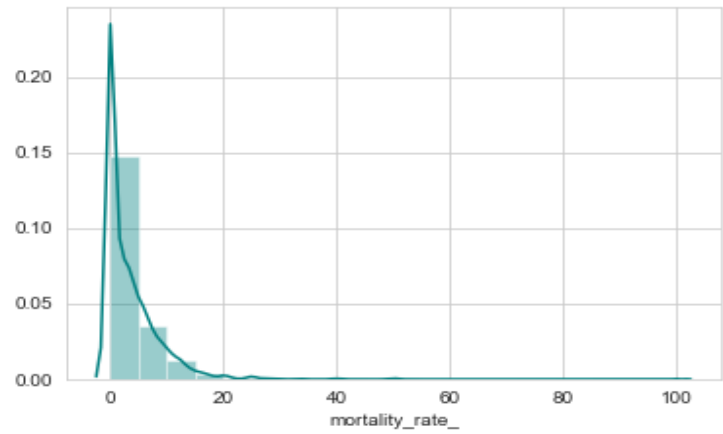
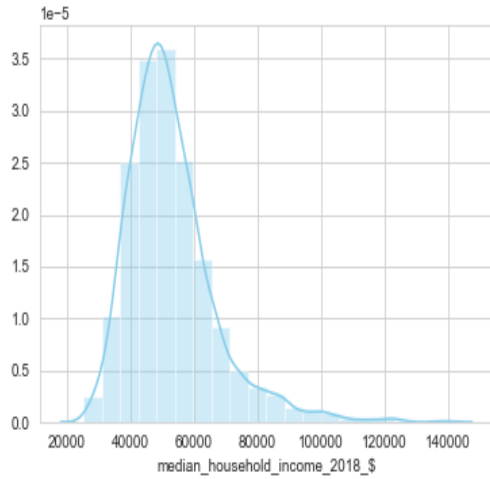
A histogram is basically used to represent data provided in a form of some groups. It is accurate method for the graphical representation of numerical data distribution. It is a type of bar plot where X-axis represents the bin ranges while Y-axis gives information about frequency. As we can see from the above histograms, most of the features in our dataset were not equally distributed it.

Using Boxplots to find outliers

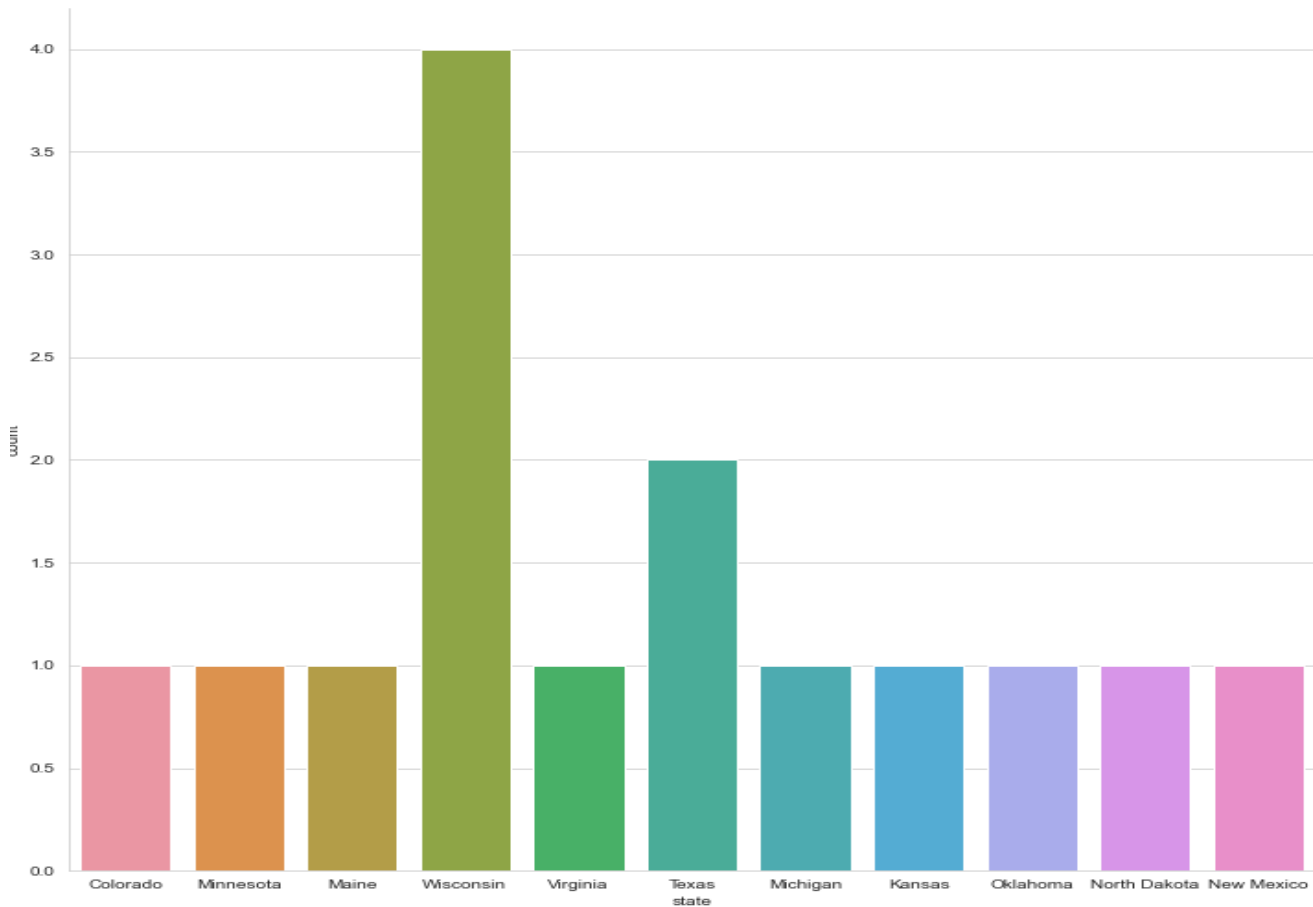


Boxplots divide the data set into three quartiles. This graph represents the minimum, maximum, median, first quartile and third quartile in the data set. It is also useful in comparing the distribution of data across data sets by drawing boxplots for each of them. An Outlier is a Data Point or value that differs considerably from all or most other data in a dataset. These won't make the model fail. There are outliers across all the columns which are plotted.

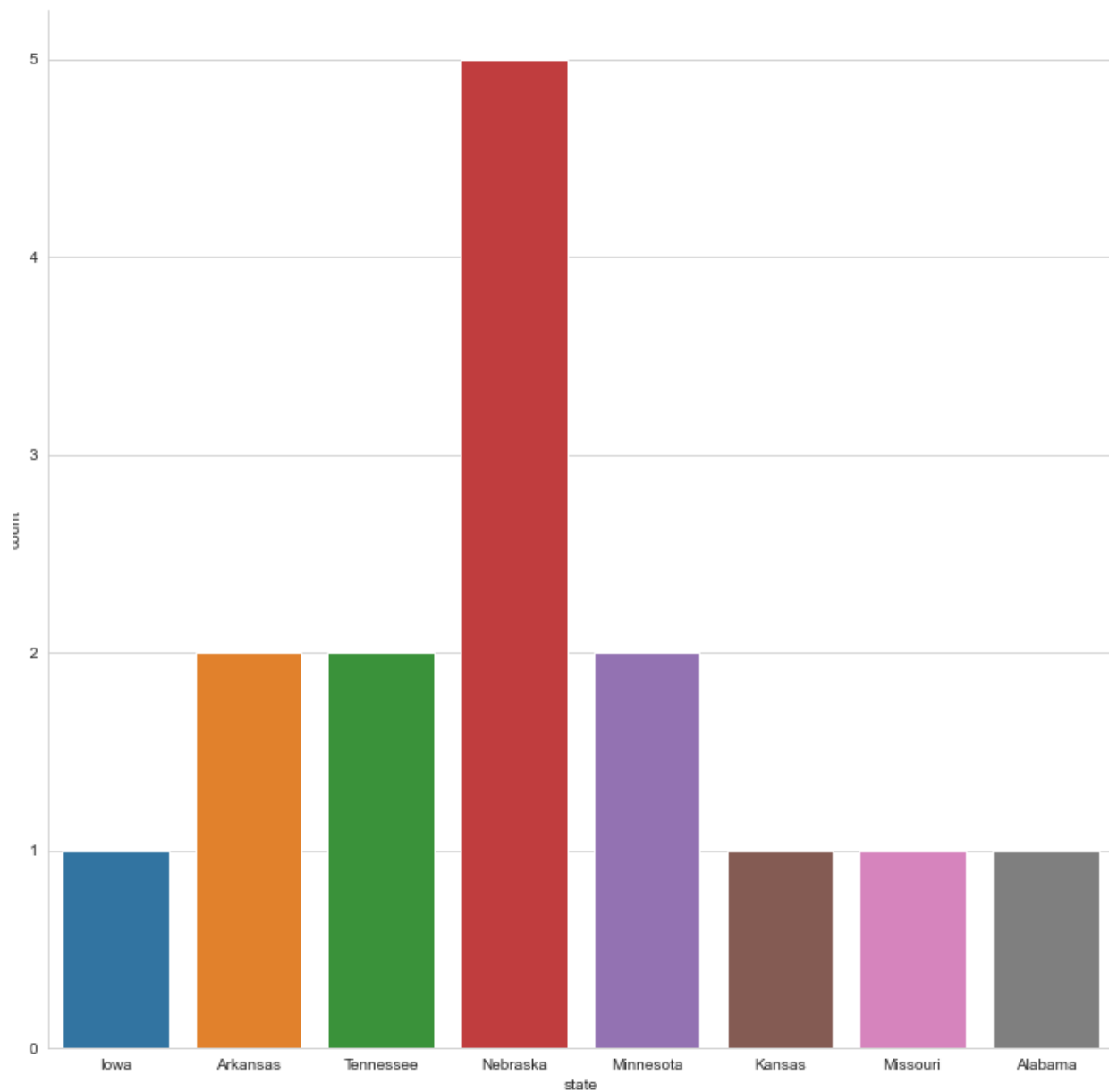
Here are some more histograms we used



Below we can see the Distribution of Mortality rate for Top Countries.



A bar plot or bar chart is a graph that represents the category of data with rectangular bars with lengths and heights that is proportional to the values which they represent. The bar plots can be plotted horizontally or vertically. A bar chart describes the comparisons between the discrete categories. One of the axes of the plot represents the specific categories being compared, while the other axis represents the measured values corresponding to those categories. From the above graph we can see that Wisconsin and Texas have 4, 2 counties respectively in the top 15 counties with high mortality rate



From the Below graph we can see that Nebraska has 5 counties with the lowest mortality rate in the USA.