**Conclusions**

**Anas’ and Shriram’s Part:**

1. California Summary Statistics:

\_ From our California summary statistics table, we can see that the mean and the median of the age group of “65 and Older” is much higher than those of the “55 and Younger” and “Between 55 and 64” age groups. Also, the oldest age group has a higher variance, standard deviation and SEM than the younger ones because of the higher volatility and higher spread of the values of that group. In truth, this seems logical, as there is usually a higher spread of values that are very different but are around 100,000 than values that are very different but that are around 1,000. However, contrary to our expectations, although it’s the median and the mean of the “Between 55 and 64” group that are higher than those of the “Younger than 55” group, the variance, the standard deviation, the SEM and the total cases of our “Younger than 55” group are higher than those of the “Between 55 and 64” group. Truth has to be said, it’s certainly because of the age range of study difference that we may observe this phenomenon for the state of California (55 years for the group of “Younger than 55” and 9 years for the group of “Between 55 and 64”). More precisely, there are more observations for the “Younger than 55” group than for the “Between 55 and 64” group and that’s why this group has a lower mean but a higher number of total cases than the other one. However, although the “Younger than 55” group has still more observations than the “Older than 65” group, the oldest group has an age range that is not very precise (from 0 to several tens of years), which makes it more likely than the “Between 55 and 64” group to have as much or even more cases than the “Younger than 55” group. At last, the California deaths in total for the “Older than 65” age group are almost 7 times higher than for the “Younger than 55” group and 8 times higher than for the “Between 55 and 64” group.

1. Connecticut Summary Statistics:

\_ About our Connecticut Summary Statistics Table, we can see that it goes exactly as we expected. Unlike for the state of California, the median, the mean, the variance, the standard deviation, the SEM and the total number of cases are all higher for the “Older than 65” group than for the “Between 55 and 64” group and higher for the “Between 55 and 64” group than for the “Younger than 55” group. What we can add to this is that the statistics of California are much bigger than those of the Connecticut all statistics mixed up. Moreover, for the “Younger than 55” group for both Connecticut and California, the minimum number of COVID-19 deaths per day is of 0. For the other groups, the minimum number of COVID-19 deaths per day for both states are very different. For the “Between 55 and 64” group, the minimum number of COVID-19 deaths for California is 8 time higher than the one of Connecticut while for the “65 and Older” group, the minimum number of COVID-19 deaths of California is more than 76 times higher than the one of Connecticut. In addition, for the maximum number of cases, we can see that the highest of the maximum numbers of COVID-19 deaths per day for all age groups concerning COVID-19 deaths is of 5276 for California and 2605 for Connecticut, which tells us how huge is the difference of the impact of the COVID-19 in the mortality of California and Connecticut.

1. Virginia Summary Statistics:

\_ About our Virginia Summary Statistics Table, we can see that it goes exactly as we expected, expect for the minimum statistic part. Indeed, all the statistics were following the same trend of those of the Connecticut and California ones expect for the minimum statistic. For the minimum number of deaths of Virginia, we can see that for all the age groups, it is not bigger than 0. This clearly shows us that for certain time periods, there weren’t any deaths provoked by the COVID-19 in Virginia. This is important, as it leads us to question why there weren’t any COVID-19 deaths in Virginia at this time. This might be because the state of Virginia had restricting policies concerning transportation that were stricter than those of other states, making that state spared from an increase of COVID-19 deaths. At last, for the state of Virginia, all statistics mixed up, it had much lower numbers than the states of Connecticut and California in terms of its statistics. For example, the mean number of cases for the Younger than 55 age group is only of 1, which is incredibly low compared to the states that we previously talked about.

1. Florida Report:

\_ About our Florida Report Table, it is another format of table compared to the previous tables that we analyzed. Instead of giving us statistics like the mean, the median and the variance it gave us the means, the sums and the percentages of the COVID-19 cases, hospitalizations and deaths. The first trend that we can notice for the state of Florida is that there are much more COVID-19 cases in total and less cases in average for the “Younger than 55” group than for the “Between 55 and 64” group. The reason of this is that there are much more samples for the “Younger than 55” than for the “Between 55 and 64”, adding to this the big difference of age range that we talked about in the “California Summary Statistics” part. Another trend, more important this time, is that although the “Younger than 55” group total COVID-19 cases is much higher than for the other two age groups, there are a lot more total hospitalizations and deaths for the “65 and Older” group than for the “Younger than 55” group. This shows that even if young people less than 55 years old are infected by the COVID-19 in a much higher proportion than old retired people, the oldest people infected by the virus die much more than the youngest people. Furthermore, what for me looks very astounding with those statistics is that more than 70% of the Floridian population that is infected by the COVID-19 comprises people younger than 55 but more than 90% of the people dying from the COVID-19 are of 55 and older. This trend is very significant regarding the potential correlation between the mortality from COVID-19 and someone’s age. Clearly, as someone gets older, he has more chances of dying from the COVID-19 than someone that is younger. However, someone that is young has more chances of being infected by the COVID-19 than someone that is older.

1. Georgia Report:

\_ For our Georgia Report, we designed it in the same fashion as the Florida Report. In truth, the state of Georgia follows the same trend than the state of Florida. However, one difference that may be noticeable between the two states are that for Georgia, the “Younger than 55” group has almost the same total number and rate of hospitalization than for the “65 and Older” group. For the state of Florida, there was more hospitalizations in total for the “65 and older” group than for the “Younger than 55” one.

1. ANOVA Test:

\_ For our ANOVA Test, we tried to prove that for the five states that we previously analyzed, the sample of age group are very different from each other. As a result, all of the p-values were very small while the t-stats were abnormally high. This shows that the null hypothesis stating that the mean of those age group populations are equal has to be rejected. Hence, we can conclude that the age groups samples are very different from each other, as we expected.

1. Overall U.S Report (in %):

\_ For our Overall U.S Report, we did it in another fashion. This time, we just put the percentages for each state and the U.S. overall. However, we used other samples that we took from February 2nd to July 4th, 2020. For all the five states and the U.S. overall, a maximum of a little less than 10% of the inhabitants dying from the COVID-19 pertain to the “Younger than 55” group. Also, at least 77% of the inhabitants dying from the COVID-19 pertain to the “65 and older” group. This shows us that even the maximum percentage of COVID-19 cases of the “Younger than 55” age group is largely inferior to the minimum percentage of COVID-19 cases of the “65 and older” group. This totally concords with our expectations that the oldest people have a higher rate of mortality than the youngest people. In fact, if we go state by state and include the overall U.S. statistics, we can see that the “65 and older” group deaths are higher than for the “Between 55 and 64”, which are higher than for the “Younger than 55” group.

1. Overall U.S. Report (in Total):

\_ This report is the same as the previous one but in total numbers. Obviously, the same last trend that we cited above apply to this table. One thing that we can notice is that the COVID-19 deaths for all the age groups in California are higher than the COVID-19 deaths for the age groups of the other states. In truth, it’s not superfluous if the news keep talking about the COVID-19 when it comes to the California total COVID-19 deaths

1. Correlation Coefficients and Equations:

\_ The correlation coefficient of each state values with the U.S values is of very slightly less than 1. This confirms that our trend explained in the last sentence of the “Overall U.S. Report (in %)” part is verified. Also, for the equations depicting these correlations, all of them have positive coefficients. This shows that the correlation between each of the states and the U.S overall is a very strong positive correlation.

1. Independent Two Sample T-test:

\_ For this test, working as the ANOVA test, we tried to prove that for the five states that we previously analyzed, the overall samples are close from the one of the U.S in terms of age groups spread. As a result, all of the p-values were of at least 40% and the t-statistics were all between 2 and -2. In fact, this means that we failed to reject our null hypothesis stating that our population means are close. Clearly, this contributes to show the same conclusions as the one explained in the last sentence of the “Overall U.S. Report (in %)” part and those of the “Correlation Coefficients and Equations” part.

1. Chi-Square Test:

\_ For this test, working as the opposite of the ANOVA Test, we tried to prove that for the five states that we previously analyzed, the overall samples are close from the one of the U.S. In fact, the p-values were of 0 while the t-stats were abnormally high. This shows the same conclusion as the one explained in the last sentence of the “Overall U.S. Report (in %)” part and the one of the “Correlation Coefficients and Equations”, as we failed to reject our null hypothesis.

1. Multiple Bat Plots for All States:

\_ For our multiple bar plots part, one of the first things that may attract our attention is the size of the green bars. For this part, we tried to showcase the difference of size between the age groups for each state. Obviously, for each state, the “65 and Older” group depicted in green is much bigger than the two other groups. Other than that, the “Between 55 and 64” group is slightly bigger than the “Younger than 55” group in a lesser proportion than for the “65 and Older” group. Thanks to this graph, we can more easily see how much bigger is the “65 and Older” group sample than the “Between 55 and 64” group one and how much bigger is the “Between 55 and 64” group sample than the “Younger than 55” group one for each state.

1. U.S Bar Plot:

\_ Sincerely, for our U.S bar plot part, we could have added it in a form of one plot of the “Multiple Bat Plots for All States” part. However, since its size is too big, it would have changed the shape of our previous graph bars (multiple bar plot bars) and have consequently made them less meaningful. Therefore, we chose to plot the overall U.S age groups sizes in a separated bar plot. As a result, we notice that the U.S bar plot follows the same trend as the one explained in the “Multiple Bat Plots for All States” part.

1. Multiple Line Plots:

\_ For our multiple line plots part, one of the things that may be interesting to see is that the state of California has a single neat blue line on top of all of the others, showing that no matter the age groups, the statistics of California are of bigger than those of the other states, as we said for the “Overall U.S. Report (in Total)” part. For the other states, when it comes to the “Younger than 55” and “Between 55 and 64” groups, the states other than California have statistics that are very close. For the “65 and Older” group, it’s the state of Connecticut that comes after California when it comes to the number of COVID-19 deaths, followed by Florida, Georgia and then Virginia.

1. U.S. Line Plot:

\_ For our U.S line plot part, it works the same as the “U.S Bar Plot” part, with the only difference being that it is a line plot this time. Moreover, given the shape of the line plot depicting the evolution of the COVID-19 total deaths in the U.S., we can say that it looks like the one of the other states. Thus, we can conclude that the U.S follows the same trend as the one of the other five states, which we previously explained in the “Multiple Bat Plots for All States” part.

1. Stacked Bar Plots for All States:

\_ For our stacked bar plot part, we can see that although for all the age groups, it is the state of California that has the biggest number of COVID-19 deaths, it’s not very obvious to distinguish which state follows it for the “Younger than 55” and “Between 55 and 64” groups. However, it’s for the “65 and Older” group that we can clearly see the difference in size between the states other than California and come with the same conclusions that we had for the “Multiple Line Plots” part. In truth, this graph is a way to make us avoid going through our “Multiple Bat Plots for All States” and our “Multiple Line Plots” part by directly showing us the similarity of trend that all the states that we previously analyzed have.

1. California Pie Chart:

\_ For the California pie chart, we can see that the “65 and Older” group comprises at least three fourth of the state of California COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little more than one tenth of the state of California COVID-19 deaths. At last, the ”Younger than 55” group has a little less than one tenth of the state of California COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the first multiple bar plot into a pie chart.

1. Connecticut Pie Chart:

\_ For the Connecticut pie chart, we can see that the “65 and Older” group comprises a little less than nine tenth of the state of Connecticut COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little less than one tenth of the state of Connecticut COVID-19 deaths. At last, the ”Younger than 55” group has a little more than 2% of the state of Connecticut COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the second multiple bar plot into a pie chart.

1. Virginia Pie Chart:

\_ For the Virginia pie chart, we can see that the “65 and Older” group comprises at least eight tenth of the state of Virginia COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little more than one tenth of the state of Virginia COVID-19 deaths. At last, the ”Younger than 55” group has a little more than 5% of the state of Virginia COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the third multiple bar plot into a pie chart.

1. Florida Pie Chart:

\_ For the Florida pie chart, we can see that the “65 and Older” group comprises at least eight tenth of the state of Florida COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little less than one tenth of the state of Florida COVID-19 deaths. At last, the ”Younger than 55” group has a little more than 5% of the state of Florida COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the fourth multiple bar plot into a pie chart.

1. Georgia Pie Chart:

\_ For the Georgia pie chart, we can see that the “65 and Older” group comprise a little less of eight tenth of the state of Georgia COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little more than one tenth of the state of Georgia COVID-19 deaths. At last, the ”Younger than 55” group has a little less than one tenth of the state of Georgia COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the fifth multiple bar plot into a pie chart.

1. Overall U.S. Pie Chart:

\_ For the overall U.S. pie chart, we can see that the “65 and Older” group comprises a little more of eight tenth of the overall U.S COVID-19 deaths. Also, the “Between 55 and 64” group constitutes a little more than one tenth of the overall U.S COVID-19 deaths. At last, the ”Younger than 55” group has a little less than one tenth of the overall U.S COVID-19 deaths. About the conclusions that can be drawn from this pie chart, we can say that they are the same of those of the “Multiple Bat Plots for All States” part as that pie chart is just a transformation of the sixth multiple bar plot into a pie chart.

**Anas’ and Shriram’s Final Conclusions:**

\_ Three conclusion that we can draw from Anas’ and Ram’s part are that firstly, to directly answer our hypothesis, we can say that as someone ages, he or she has more chances of dying of the COVID-19 and this is how the correlation between the COVID-19 mortality and age is shown. Secondly, another conclusion that we can draw is that people younger than 55 are much more affected by the COVID-19 than the older than 55 people although very few of them die from it. On the other hand, very fewer older than 55 people than younger than 55 people are affected by the COVID-19 but most of those people die from it. About that, we were very surprised to discover this, as it was not something that initially knew. Thirdly, the state of California, all age groups mixed up, has the highest number of COVID-19 deaths of the five states that we analyzed while the state of Virginia, all age groups mixed up, has the lowest number of COVID-19 deaths of the five states that we analyzed.