**Conclusions**

1. **Anas and Ram’s Part:**
2. California Summary Statistics:

\_ From the 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 age groups of “55 and Younger” and “Between 55 and 64”. Also, the oldest age group has a higher variance, standard deviation and SEM than the younger ones because of the higher volatility and the higher spread of the values of that group. In truth, this is 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 certain states like 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 it. However, although the “Younger than 55” group has still more observations than the “Older than 65 group, the “Older than 65 group” has an age range that is not precise (from 0 to several tens of years), which make 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 oldest age group are almost seven times higher than the youngest age group and eight times higher than the “Between 55 and 64” age group.

1. Connecticut Summary Statistics:

\_ About the 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 number of California are much bigger than those of the Connecticut all statistics mixed up. Although for both states “Younger than 55” group, the minimum statistic is of 0. For the other groups, the minimums of both states are very different. For the “Between 55 and 64” group, the minimum number of cases of California is 8 time higher than the number of cases of Connecticut while for the “65 and Older” group, the minimum number of cases of California is more than 76 times higher than the one of Connecticut. For the maximum number of cases, we can see that for the highest maximum number of deaths is of 5276 for California and 2605 for Connecticut, which tells how huge is the difference of the impact of the COVID-19 in the mortality of California and Connecticut.

1. Virginia Summary Statistics:

\_ About the Virginia Summary Statistics Table, we can see that it goes exactly as we expected, expect for the minimum. All the statistics were following the same trend of those of the Connecticut expect for the minimum statistic. For the minimum number of case of Virginia, we can see that for all the age groups, it is of 0. This shows that for certain time periods, there weren’t any deaths of 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. At last, for the state of Virginia, all statistics mixed up, it had much lower numbers than the state of Connecticut. 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 the 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 gives 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 in total 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 stated for the previous analysis (55 > 9). Another trend, more important this time, is that the “Younger the 55” group total COVID-19 cases are much higher than for the other two age groups but there are a lot more hospitalizations and deaths for the “65 and Older” group than for the “Younger than 55” group. This shows that although young people less than 55 years old are infected by the COVID-19 in a much higher proportion than old retired people, the old people infected by the virus die much more than the young people. What for me looks very astounding with those statistics is that more than 70 percent of Floridian population is infected by the COVID-19 comprises people younger than 55 but more 90 percent 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. 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 than someone that is older.

1. Georgia Report:

\_ For the 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 the 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 is rejected. Hence, the age groups samples are very different from each other, as we expected.

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

\_ For the 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 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 show us that even the maximum percentage of the “Younger than 55” age group is largely inferior to the minimum percentage of the “65 and older” group. This totally concords with our hypothesis that older people have a higher rate of mortality than the younger 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 trends 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 number of 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 true. Also, for the equations depicting this correlation, all of them have positive coefficients. This shows that the correlation 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. 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., like for the “Independent Two Sample T-test”. In fact, the p-values are of 0 while the t-stats are abnormally high. This shows the same conclusion as the one explained in the last sentence of the “Overall U.S. Report (in %)” part, those of the “Correlation Coefficients and Equations” part and the one of the “Independent Two Sample T-test”, which is a strong correlation between each state and the U.S.