Project 1

Leyla Hatfield, Claudia Cabrera, Jimmy Horvath, Lauren Fowler

1/31/2024

# Introduction

For this project, we are analyzing COVID-19 vaccination and case trends by age group across the United States from December 14th, 2020, to October 11th, 2022. The study collected data consisting of five columns.

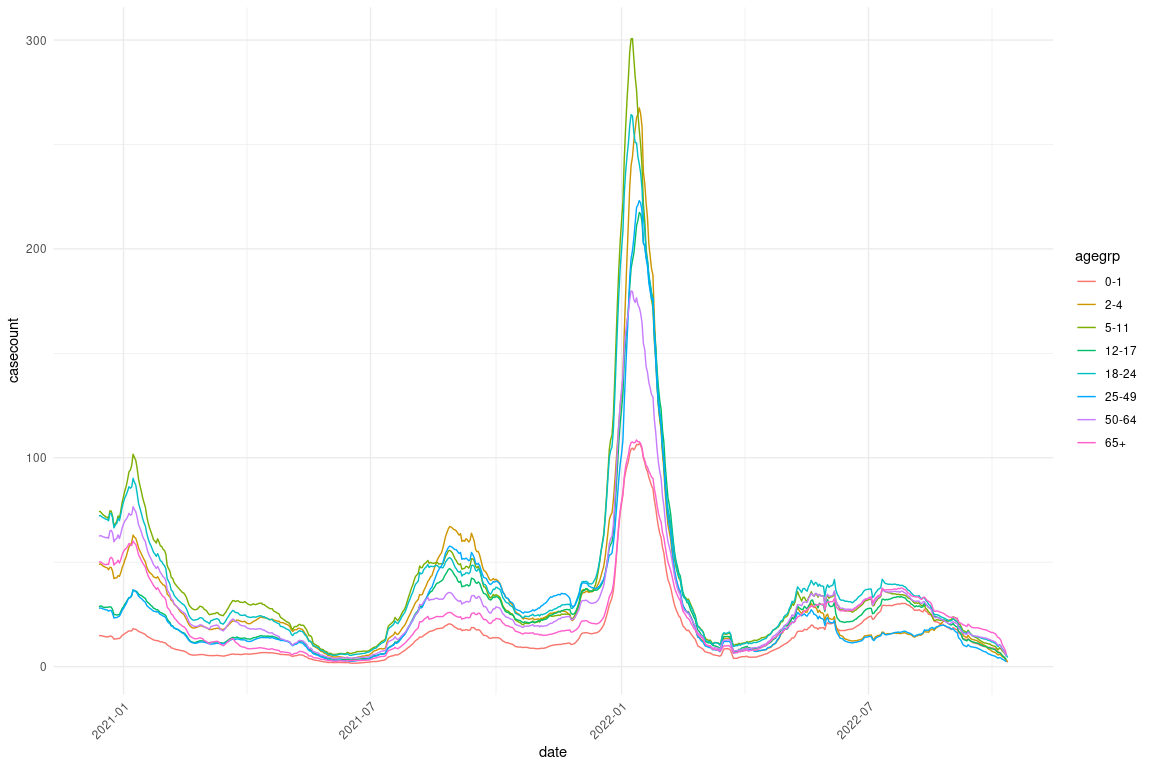
1. Date Administered: This is nominal, categorical data of month, day, and year of vaccine administration.
2. Age group specifics for vaccination data: Age can be argued as both categorical and numerical data. However, for this dataset, our basis of comparison lies between 18-24 year-olds, 25-49 year-olds, and those aged 65 years or more. Thus, the data is ordinal, categorical data illustrating a younger and older population.
3. Seven-day average of the daily cases reported: This is continuous, ratio data consisting of a seven-day average in COVID-19 cases reported per 100,000 population.
4. Percent of population with at least one dose in that specific age group: This is continuous, ratio data of the population percentage with at least one dose in their specific age group.
5. Percent of people who completed a primary series (have a second dose of a two-dose vaccine or one dose of a single-dose vaccine) in an age group: This is continuous, ratio data of the percentage of people within a specific age group who completed their series of COVID-19 vaccinations.

As of May 10, 2023, the CDC Covid Data Tracker has observed that 94.4% of the population of people 65 or older in the United States have completed a primary series of COVID-19 vaccination, while only 79.1% of the population of people who are less than 65, but 18 or older have (1). This percentage continues to drop as age decreases. Several studies have shown that age does have an effect on vaccination hesitancy amongst the population, with findings saying that adults ranging from 18 to 65 are less likely to be fully vaccinated in comparison with those who are 65+. For example, a study published in December 2020 wanted to determine the impact of the COVID-19 pandemic on influenza vaccine uptake by looking at patients’ perceptions of vaccinations, vaccination intentions, and health behaviors through a questionnaire. This study found that factors such as being between the age of 18-49, having an income of less than $20,000 or less, and knowing someone affected by the COVID-19 pandemic significantly affected respondents’ vaccine acceptance and negatively impacted health behavior (2). Thus, we see here that factors such as age do affect vaccine acceptance. Another study looking at factors associated with COVID-19 vaccine hesitancy by race and ethnicity showed that participants ≥ 45 years of age were less likely to be hesitant towards the vaccine than those under 45 (3). From these repeated results, we see that age may play a role in vaccination percentage among the population.

The data we have chosen to look at divides age into eight groups: 0-1, 2-4, 5-11, 12-17, 18-24, 25-49, 50-64, and 65+. Based on past research, we hypothesize that the age groups of 18-24 and/or 25-49 will have a lower percentage of people who have completed a primary series of COVID-19 vaccinations in comparison to the 65+ age group. Due to this, we also hypothesize that these age groups will have a higher average daily case report in comparison to the 65+ age group. With all this in mind, our null hypothesis would be that there would be no significant observed differences between the age groups.

# Results

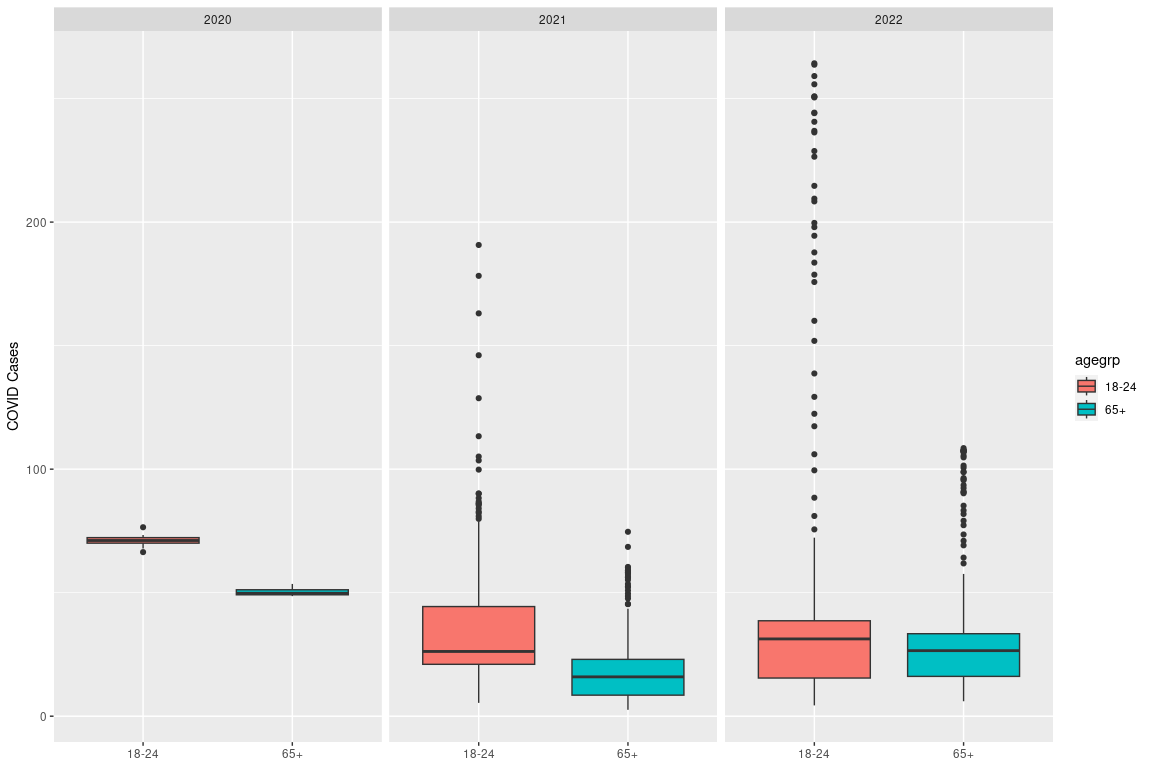
**Figure 1**: Case Count vs. Date



In Figure 1, it is interesting to note that there is a spike in case count at the beginning of 2022. At this time, we also can observe higher case counts in the younger age groups, such as 5-11, 18-24, and 2-4, while the lowest case counts are in the age group of 65+. This generally seems to be the consistent trend throughout the timeline.

**Figure 2**: Case Count in 18-24 year-olds vs 65+

## Joining with `by = join\_by(date, agegrp, casecount, onedosevac, completevac,  
## year, month, onedosevacpercent, completevacpercent)`



In this figure, we see that the average case count is higher in the 18-24 year olds for all years in comparison to 65+. An interesting thing to note, is that the averages appear to be more similar by 2022.

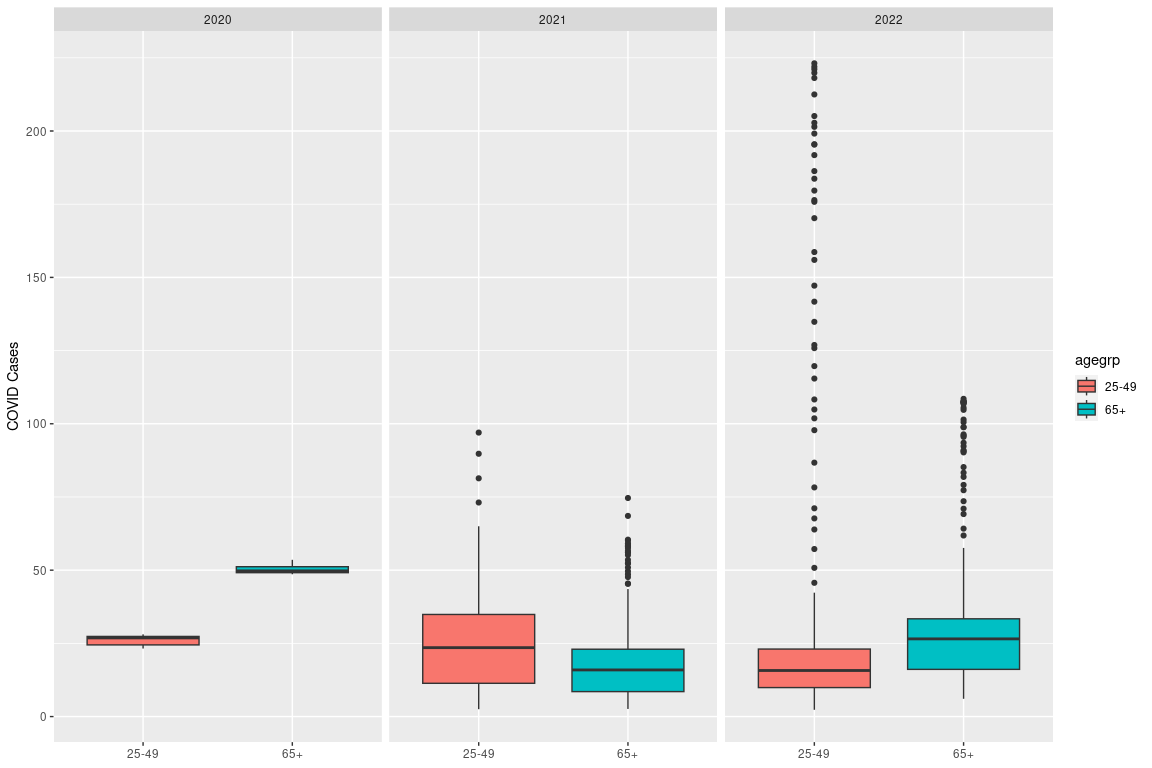
**Table 1**: t-test on Case Count between 18-24 and 65+ in the Year 2022.

##   
## Welch Two Sample t-test  
##   
## data: X7.day\_avg\_group\_cases\_per\_100k by AgeGroupVacc  
## t = 3.7116, df = 363.18, p-value = 0.0002383  
## alternative hypothesis: true difference in means between group 18 - 24 Years and group 65+ Years is not equal to 0  
## 95 percent confidence interval:  
## 7.118206 23.161519  
## sample estimates:  
## mean in group 18 - 24 Years mean in group 65+ Years   
## 46.08190 30.94204

After running the t-test we get a p-value of 0.00024, meaning we are more than 99% confident that the difference in the average case count in the year 2022 between 18-24 years and 65+ years is not due to random chance.

**Figure 3**: Case Count in 25-49 year-olds vs 65+

## Joining with `by = join\_by(date, agegrp, casecount, onedosevac, completevac,  
## year, month, onedosevacpercent, completevacpercent)`



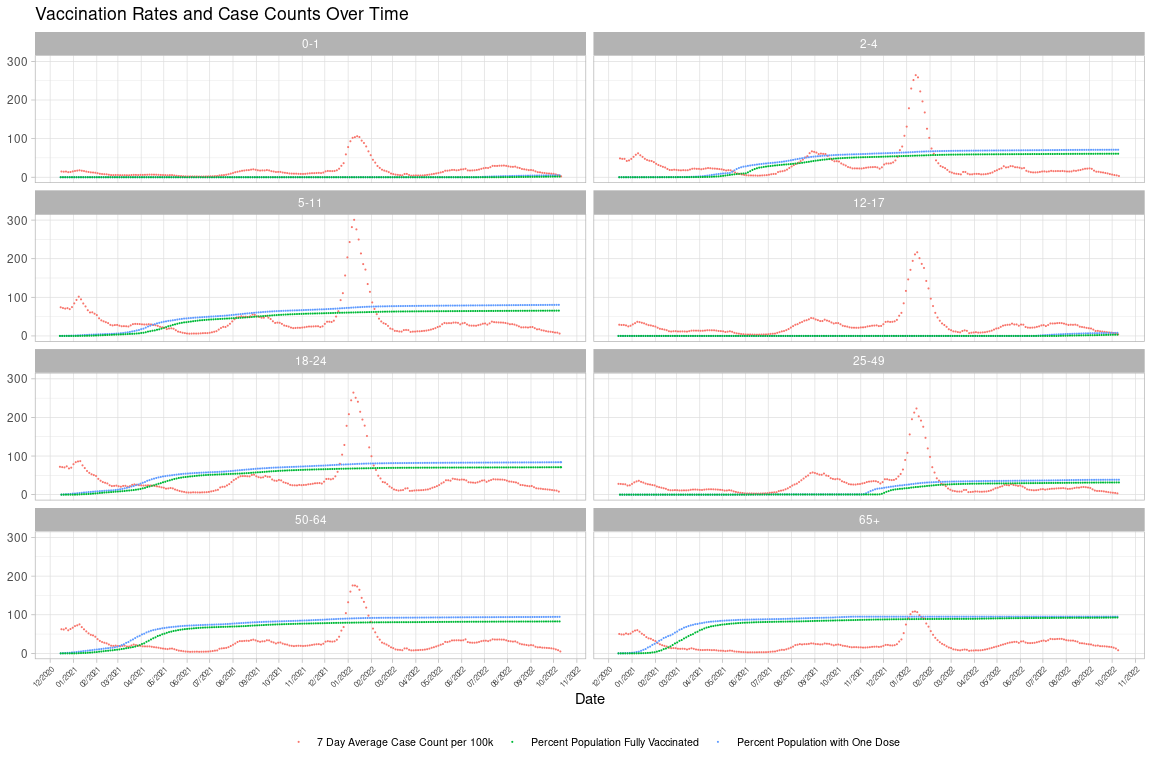
In this figure, we see that the average case count is only higher in the 25-49 year olds compared to 65+ in the year 2021.

**Table 2**: t-test on Case Count between 25-49 and 65+ in the Year 2022.

##   
## Welch Two Sample t-test  
##   
## data: X7.day\_avg\_group\_cases\_per\_100k by AgeGroupVacc  
## t = 4.069, df = 378.49, p-value = 5.748e-05  
## alternative hypothesis: true difference in means between group 25 - 49 Years and group 65+ Years is not equal to 0  
## 95 percent confidence interval:  
## 7.979552 22.902438  
## sample estimates:  
## mean in group 25 - 49 Years mean in group 65+ Years   
## 46.38304 30.94204

After running the t-test we get a p-value of 5.748e-5, meaning we are again morw than 99% confident that the difference in the average case count in the year 2022 between 25-49 years and 65+ years is not due to random chance.

**Figure 4**: Vaccination Rates and Case Counts Over Time



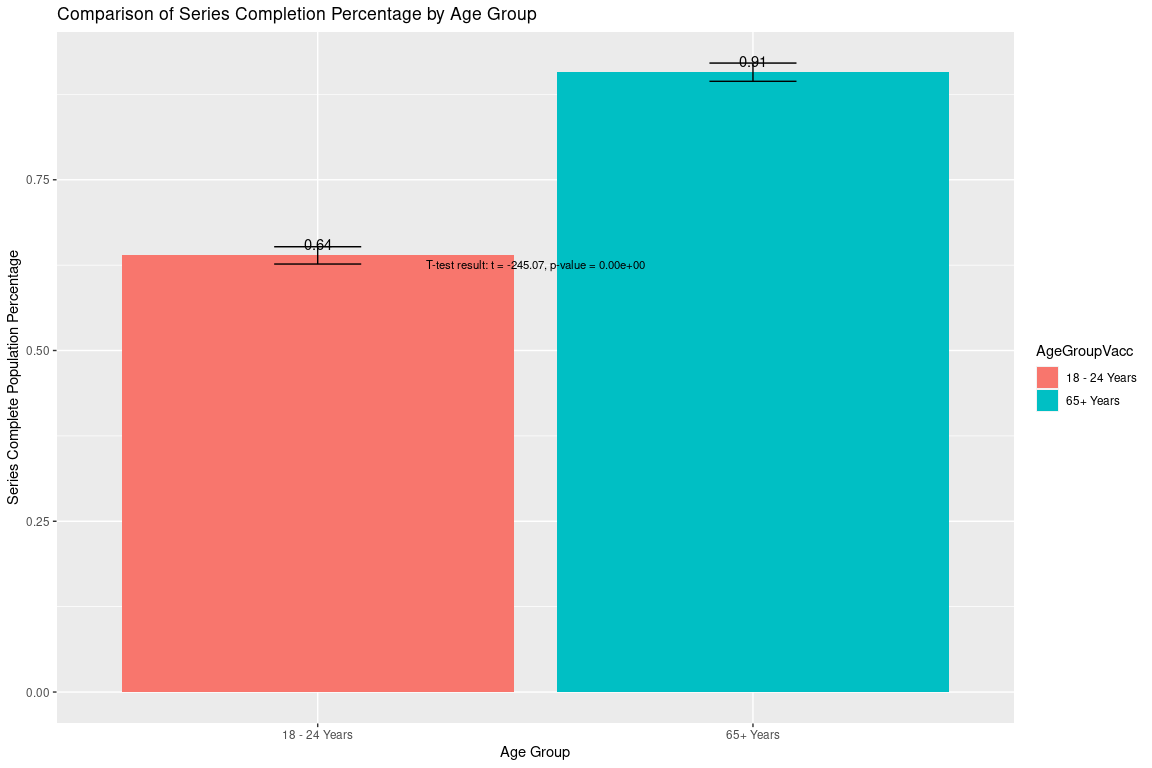
In Figure 4, we can observe how the 65+ age group has high one dose and fully vaccinated percentages as well as appearing to have the lowest average case count in comparison to other age groups. We can also note again the particularly low percentages of vaccinations across the 25-49 age group.

**Table 3**: t-test on Series Complete Percentage between 18-24 and 65+ in 2022

##   
## Welch Two Sample t-test  
##   
## data: Series\_Complete\_Pop\_pct\_agegroup by AgeGroupVacc  
## t = -245.07, df = 562.43, p-value < 2.2e-16  
## alternative hypothesis: true difference in means between group 18 - 24 Years and group 65+ Years is not equal to 0  
## 95 percent confidence interval:  
## -0.2703902 -0.2660904  
## sample estimates:  
## mean in group 18 - 24 Years mean in group 65+ Years   
## 0.6392827 0.9075230

After running the t-test we get a p-value of < 2.2e-16, meaning we are more than 99% confident the difference in the average series complete population percentage between 18-24 years and 65+ years is not due to random chance.

**Figure 5**: Visualization of t-test Comparing Series Complete Population Percentage in 18-24 year olds vs. 65+ in 2022



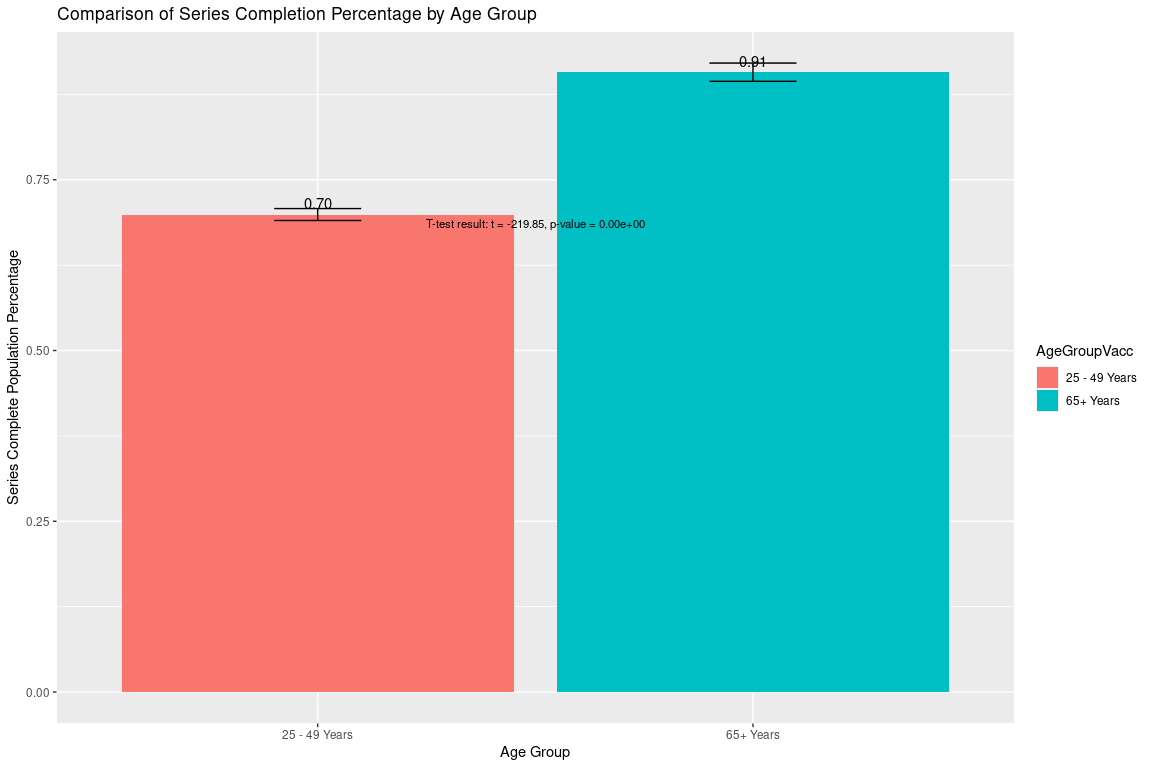
From this visualization, we can see the significance of our result.

**Table 4**: t-test on Series Complete Percentage between 25-49 and 65+ in 2022.

##   
## Welch Two Sample t-test  
##   
## data: Series\_Complete\_Pop\_pct\_agegroup by AgeGroupVacc  
## t = -219.85, df = 485.39, p-value < 2.2e-16  
## alternative hypothesis: true difference in means between group 25 - 49 Years and group 65+ Years is not equal to 0  
## 95 percent confidence interval:  
## -0.2102907 -0.2065651  
## sample estimates:  
## mean in group 25 - 49 Years mean in group 65+ Years   
## 0.6990951 0.9075230

After running the t-test we get a p-value of < 2.2e-16, meaning the difference in the average series complete population percentage between 25-49 years and 65+ years is not due to random chance.

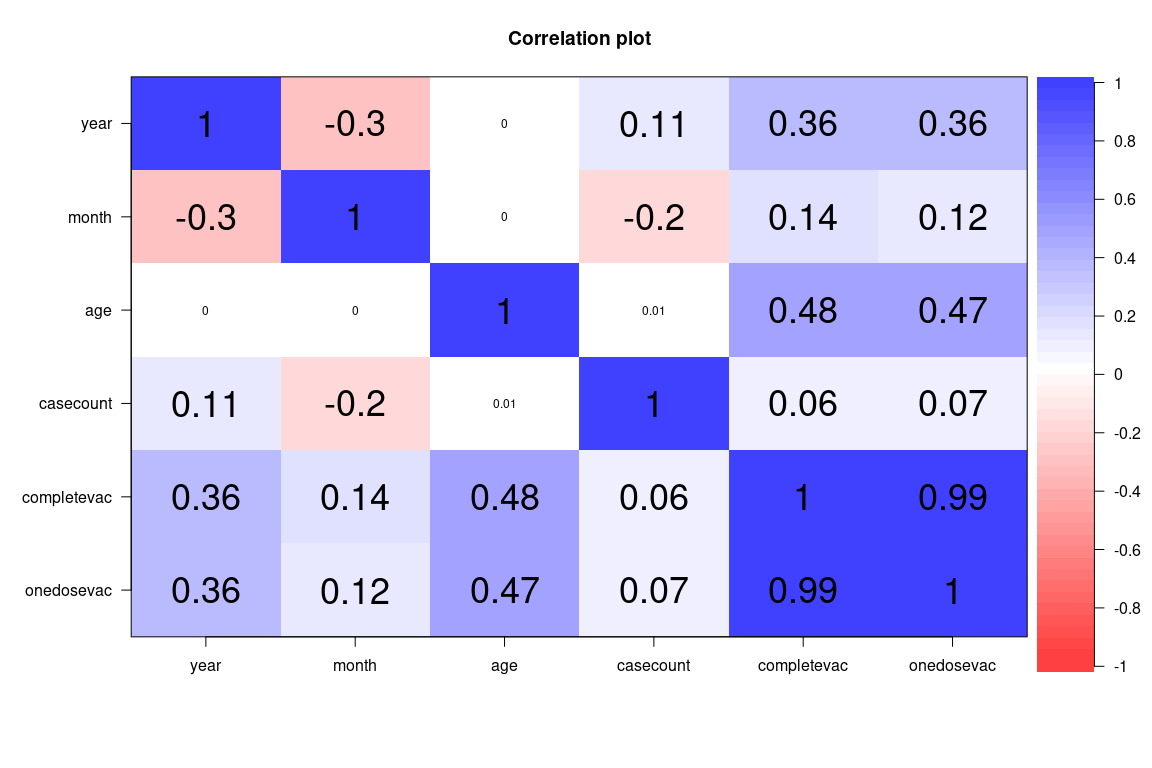
**Figure 6**: Visualization of t-test Comparing Series Complete Population Percentage in 25-49 year olds vs. 65+ in 2022



Again, we see a similar significant result comparing these two age groups.

Looking at correlation.

**Figure 7**: Correlation Plot of Data



When looking at this plot we can see that age does have the highest correlation to complete vaccination and one dose vaccination compared to other variables. There does not seem to be a high correlation between case count and complete or one dose vaccination.

# Methodology

For this study, we first used the CDC website to find data we were interested in analyzing. We then examined the data set, looking at what variables it consisted of when the study had taken place, viewed the data, etc.

After we had chosen the data set we wanted, we began our statistical analysis with data cleaning. We checked for any missing data that may have been in the data set and cleaned the data in a way to make it easier to use, such as changing the format of the dates in the data set. These operations included, but were not limited to adding “year” and “month” to the data set, converting “agegrp” to a factor, and morphing “onedosevac” and “completevac” columns to numeric format. Additionally, we renamed certain variables we deemed necessary for clarification. For example, lines 54-58 renamed “Series\_Complete\_Pop\_pct\_agegroup” to “completevac” indicating the group that was completely vaccinated. Similarly, “Administered\_Dose1\_pct\_agegroup” was renamed to “onedosevac” to depict the group that recieved only one dose of the vaccine. More generally, we also reshaped the data format from wide to long using the pivot\_longer function in Line 94.

Once our data had been cleaned, we used some exploratory visualizations to get a sense of what questions we may want to ask about our data. Once we landed on our research questions and hypothesis, we began our statistical analysis.

We created various line graphs, boxplot, and bar charts that we believed represented the data well as well as used R to run statistical analysis tests like t-tests and correlation tests. After running our analyses, we then interpreted the data and drew some conclusions.

# Discussion

Our first hypothesis was that the age groups of 18-24 and/or 25-49 will have a lower percentage of people who have completed a primary series of COVID-19 vaccinations in comparison to the 65+ age group, which is observed in Figure 4. In Figure 4, it appears as though the percentage of full vaccination and one dose vaccination for the 65+ age group almost reaches 100%, while this is not observed in the 18-24 age group and especially not the 25-49 age group. The 25-49 age group appears to have significantly lower percentages, which is very interesting to note. When looking at Figure 5, we see that ages 18-24 in the year 2022 have a significantly lower series completion percentage in comparison to the 65+ group, with our p-value being extremely low. Therefore, we can reject our null hypothesis, meaning these differences are not random. In Figure 6, we see a similar result when comparing 25-49-year-olds to 65+, and again we can reject our null hypothesis. When looking at Figure 7, we even observed a bit of a correlation, the highest out of all other variables, between age and complete vaccination.

With our other hypothesis, we also thought if these age groups had lower percentages of vaccination completion, then they would report a higher daily average case count. In Figure 1, there does appear to be some evidence for this, especially with the high spike in early 2022. Figure 2 also shows more clear evidence of this with the average case counts for 18-24 year olds being higher than 65+ year-olds every year. When computing the t-test for this, we get a low p-value, which again allows us to reject our null hypothesis, therefore differences in case of counts in 2022 between these age groups are significant (table 1) However, when looking at Figure 3, we observe that the average case counts for 25-49-year-olds in only higher than 65+ in 2021, and it is actually slightly lower in 2022. When running the t-test for 2022, we still get a low p-value, so we can reject our null hypothesis (table 2). This may be because there are a lot of outliers for the 25-49 age group in the year 2022, and the outliers are in the upper quartile range. This may have caused a significant difference in the averages and may be the reasoning behind our result.

When looking at Figure 7 again there is a very small correlation between case count and complete vaccination, but nothing that seems to be very significant.

# Conclusion

While the CDC’s study indicates individuals aged 65 and above experienced fewer COVID-19 cases compared to those aged 18-24, several confounding variables could influence the data. Firstly, younger individuals often engage in more social interactions, including attending social events, parties, and gatherings. This increased social activity can lead to a higher risk of exposure to the virus. Throughout the pandemic, many of these younger individuals were more likely to be students or work in jobs that require frequent interactions with peers, colleagues, or customers. Despite having lower vaccination rates, this age group inherently tended to engage in more interactions, making them more likely to expose themselves to potential risks.

Additionally, young adults aged 18-24 often share housing situations, including dormitories and apartments, potentially increasing virus transmission. In these situations and more generally, those younger may be less consistent with older populations in following preventive measures like wearing a mask and social distancing. Furthermore, younger individuals are and were more likely to be asymptomatic carriers of the virus, contributing to the spread of COVID-19 without obvious symptoms. This can make it challenging to identify and isolate cases promptly.

Moreover, during earlier periods of the pandemic, older individuals were prioritized for vaccination due to their higher risk of severe outcomes. As a result, vaccination coverage among the 65+ age group may be higher than in the 18-24 age group. Given the data has only been updated as of October 2022, perhaps there are uncaptured younger individuals outside the dataset who received vaccinations after October 2022. This would increase the percentage of 18–24-year-olds receiving the vaccine. In October 2023, Travis Kelce, a well-known NFL football player, took a prominent role in a Pfizer advertisement promoting COVID-19 vaccination, including booster shots. Perhaps with the less strict restrictions, there may be a growing number of younger individuals opting to get vaccinated not out of mandate but out of their own choice.

All in all, our data analysis is consistent with prior research that the percentage of those who are fully vaccinated between the ages of 18-24 and 25-49 is significantly lower than those who are 65+. While we may not know the full reasoning behind this, it is an important observation to make when encouraging people to get vaccinated in the future for COVID-19 or other diseases. As for case count, we did find significant evidence of cases being higher in these age groups as well. Despite not finding a significant correlation between complete vaccination percentage and case count, it is important to note this information as we can see that these age groups are at a higher risk for getting the disease based on the data.

With this information, vaccination campaigns may be created to target this demographic in the hope of increasing their full vaccination percentage and lowering average case counts. In future studies, we may want to identify more of the reasoning behind why these individuals are not completing their vaccines as well as try to understand what is causing the case counts to be high in these age demographics to ensure that we can get as many people vaccinated as possible and healthy as possible.

# References

Centers for Disease Control and Prevention. (n.d.-a). *Archive: Covid-19 vaccination and case trends by age group, United States*. Centers for Disease Control and Prevention. <https://data.cdc.gov/Vaccinations/Archive-COVID-19-Vaccination-and-Case-Trends-by-Ag/gxj9-t96f>

Centers for Disease Control and Prevention. (n.d.). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. <https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-people-booster-percent-pop5>

Colvin KM, Camara KS, Adams LS, Sarpong AP, Fuller DG, Peck SE, Ramos AS, Acevedo AL, Badume MA, Briggs SA, Chukwurah TN, Davila-Gutierrez Z, Ewing JA, Frempong JO, Garrett AA, Grampp SJ, Gillespie JW, Herrera EJ, Horsford SME, Maddox EJ, Pelaez JC, Quartey OL, Rodriguez F, Vasquez LA, Piper BJ, Gowtham S. Profiles of COVID-19 vaccine hesitancy by race and ethnicity in eastern Pennsylvania. PLoS One. 2023 Feb 6;18(2):e0280245. doi: 10.1371/journal.pone.0280245. PMID: 36745588; PMCID: PMC9901750.

Mercadante AR, Law AV. Will they, or Won’t they? Examining patients’ vaccine intention for flu and COVID-19 using the Health Belief Model. Res Social Adm Pharm. 2021 Sep;17(9):1596-1605.doi: 10.1016/j.sapharm.2020.12.012. Epub 2020 Dec 30. PMID: 33431259; PMCID: PMC7833824.