Problem Set- 4

Sonika Mahat

2025-03-19

Report on COVID-19, Mask use, and Vaccination

Introduction

The COVID-19 pandemic has affected U.S. counties differently, with varying mortality rates influenced by key public health measures such as mask-wearing and vaccination rates. This report analyzes how these factors shaped COVID-19 outcomes in 2022, drawing insights from three publicly available data sources: the New York Times COVID-19 data set, the CDC Vaccination data set, and a mask usage survey from July 2020.

The purpose of this analysis is to evaluate the effectiveness of mask-wearing and vaccinations in reducing COVID-19 deaths at the county level and provide data-driven insights for public health decision-making. By employing statistical methods, we will assess whether counties with higher percentages of residents who "always" wore masks and had higher vaccination rates experienced fewer deaths per 100,000 residents. This analysis will also explore disparities across counties by examining differences in social vulnerability and population size.

Table 1: Summary Statistics of COVID-19 Data

fips	county	state	deaths	always.mask	population	vax.complete	svi.index	deaths.scaled
Min. : 1001	Length:3142	Length:3142	Min. : 0.00	Min. :0.1150	Min. : 169	Min. :11.30	Length:3142	Min. : 0.00
1st Qu.:18182	Class :character	Class :character	1st Qu.: 12.00	1st Qu.:0.3930	1st Qu.: 11519	1st Qu.:44.30	Class :character	1st Qu.: 74.05
Median :29180	Mode :character	Mode :character	Median : 29.00	Median :0.4970	Median : 26296	Median :52.10	Mode :character	Median :103.60
Mean :30482	NA	NA	Mean : 84.14	Mean :0.5077	Mean : 103730	Mean :53.43	NA	Mean :113.16
3rd Qu.:46005	NA	NA	3rd Qu.: 70.00	3rd Qu.:0.6130	3rd Qu.: 69685	3rd Qu.:61.00	NA	3rd Qu.:142.16
Max. :78030	NA	NA	Max. :7034.00	Max. :0.8890	Max. :10039107	Max. :95.00	NA	Max. :755.03
NA	NA	NA	NA	NA's :9	NA's :93	NA's :93	NA	NA's :93

From the table above, 3,142 U.S. counties reveals notable trends. Counties reported a median of 29 total COVID-19 deaths, but with a highly skewed distribution, where some urban counties experienced up to 7,034 deaths. Adjusting for population, the median number of deaths per 100,000 residents was 103.6, with some counties experiencing as high as 755 deaths per 100,000. Mask usage varied widely, with the proportion

of people who "always" wore masks ranging from 11.5% to 88.9% (median: 49.7%), while vaccination rates ranged from 11.3% to 95% (median: 52.1%).

Statistical modeling confirms that higher vaccination rates were significantly associated with lower COVID-19 death rate, while higher mask usage also played a role in reducing deaths. Additionally, counties with higher social vulnerability (SVI) had lower vaccination rates and higher death rates, highlighting inequities in public health outcomes. These findings emphasize the importance of vaccination efforts, mask usage, and targeted interventions in vulnerable communities to mitigate pandemic-related death.

Data: Description and Summary of Key Variables

This report utilizes publicly available county-level data from different datasets to analyze the relationship between COVID-19 deaths, mask usage, and vaccination rates in 2022. The dataset is constructed by merging three primary sources:

- COVID-19 Deaths Data (New York Times COVID-19 dataset): Contains daily reported COVID-19 deaths at the county level in the United States. The dataset is sourced from the New York Times, which compiles data from state and local health agencies.

- Mask Usage Data (New York Times Mask Survey - July 2020): Provides estimates of mask-wearing behavior based on a survey conducted by the New York Times, measuring the proportion of people in each county who reported "always" wearing a mask in public.

- Vaccination Data (CDC COVID-19 Vaccination Data): Contains county-level records of COVID-19 vaccine administration, including the percentage of the population that is fully vaccinated. The data is sourced from the Centers for Disease Control and Prevention (CDC) and includes information on vaccine coverage and social vulnerability levels.
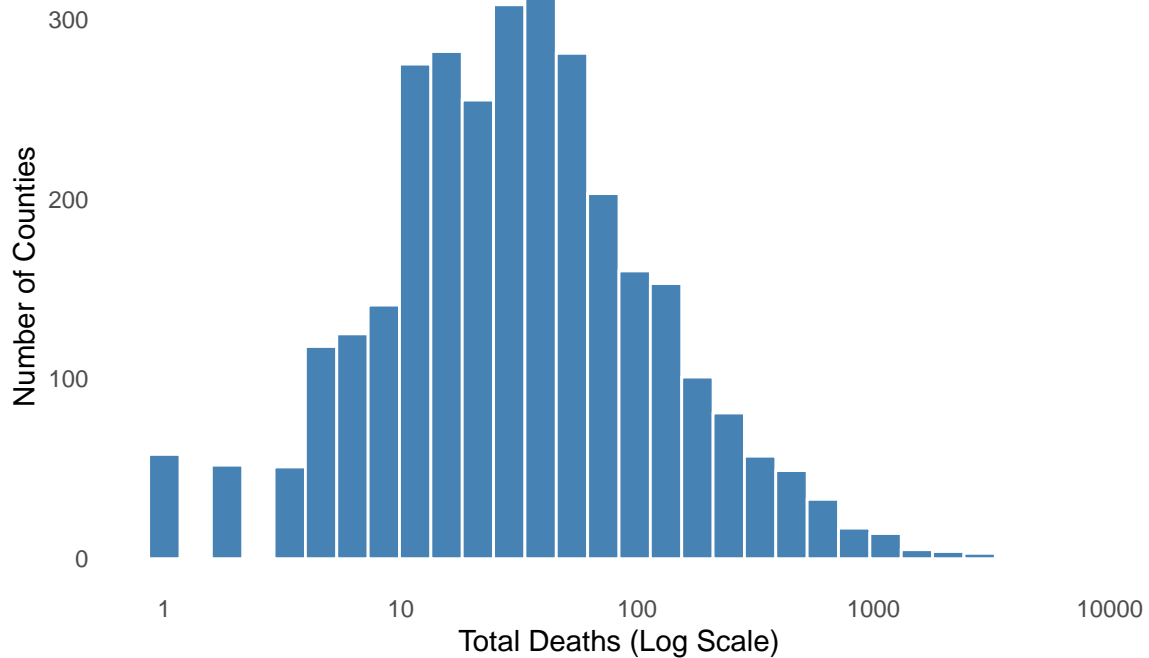
COVID-19 Deaths

The number of COVID-19 deaths across U.S. counties in 2022 varied a lot. While some counties reported no deaths at all, others mostly large urban areas had extremely high death counts. For example, the highest reported death count in a single county was over 7,000, while the median number of deaths was just 29. The average (or mean) was much higher at 84, which suggests that a few counties with very high deaths pushed the overall average up. Most counties had death counts somewhere between 12 and 70.

To better understand the pattern, we visualized in a histogram using a log scale as shown below. The distribution is highly skewed, meaning, a lot of counties had lower deaths, while a smaller number had extremely high deaths. In the plot, we can see that most counties fall between 10 to 100 deaths, but a few extend far out to the right, crossing into the thousands.

Overall, this shows is that COVID-19 did not affect all counties equally. Some regions were hit much harder than others, and these differences are important as we explore whether factors like mask usage and vaccination rates played a role in shaping these outcomes.

Distribution of COVID−19 Deaths per County


```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   12.00   29.00   84.14   70.00 7034.00
```

Mask Usage

Mask usage behavior also varied across counties, but the distribution was more balanced compared to COVID-19 deaths. The percentage of people who said they "always" wore a mask in public were from 11.5% to 88.9%, but most counties fell somewhere in the middle. In fact, the median was just under 50%, meaning half the counties had less than half their population consistently wearing masks, and half had more.

To present the data visually, we have used histogram as shown below, which shows a pretty even, bell-shaped curve, with the majority of counties clustered between 40% and 60% mask adherence. Very few counties were on either extreme (less than 20% or more than 80%), suggesting that while mask behavior differed, most counties hovered around a similar range.

Distribution of Mask Adherence Across Counties
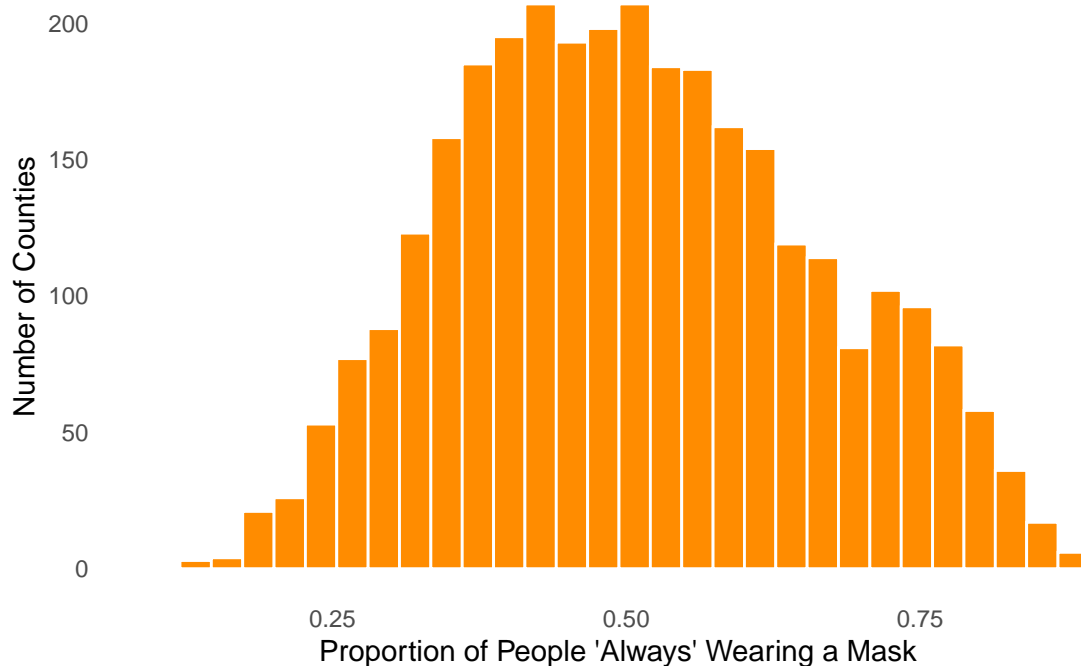

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.1150  0.3930  0.4970  0.5077  0.6130  0.8890       9
```

Highest/Lowest Mask-Wearing Counties

Looking at the range in mask usage across the country, in different counties, Inyo County, California had the highest reported mask usage, with 88.9% of people saying they "always" wore a mask in public. On the other end, Valley County, Montana had the lowest at just 11.5%. This wide gap shows just how differently communities responded to mask usage during the pandemic, likely influenced by state policies, public messaging, and local attitudes toward COVID-19. These county-level differences are important context for understanding broader patterns in pandemic outcomes.

Table 2: Counties with Highest and Lowest Mask Adherence

always.mask	state	county
0.889	California	Inyo
0.115	Montana	Valley

Vaccination Rates

COVID-19 vaccination rates also varied widely across U.S. counties. According to summary data, the percentage of fully vaccinated residents ranged from as low as 11.3% to as high as 95%. The median vaccination rate was 52.1%, which means half of the counties had more than half of their population vaccinated, while the other half didn't. The average (mean) was similar, at 53.4%, suggesting that the distribution of vaccination coverage was fairly balanced across counties. Most counties fell between 44.3% and 61%, which gives us a useful range to understand where the majority stood.

To represent data visually, we used histogram. It shows a bell-shaped curve, with most counties concentrated between 45% and 60% vaccination coverage. Only a few counties had rates above 80% or below 30%, which were more of an exception than the norm. Not every county reached high levels of vaccination, many made decent progress. That said, the counties with low vaccination rates still represent gaps in public health outreach, and understanding where and why these gaps occurred is key for designing more equitable health responses in the future.

Distribution of COVID−19 Vaccination Rates
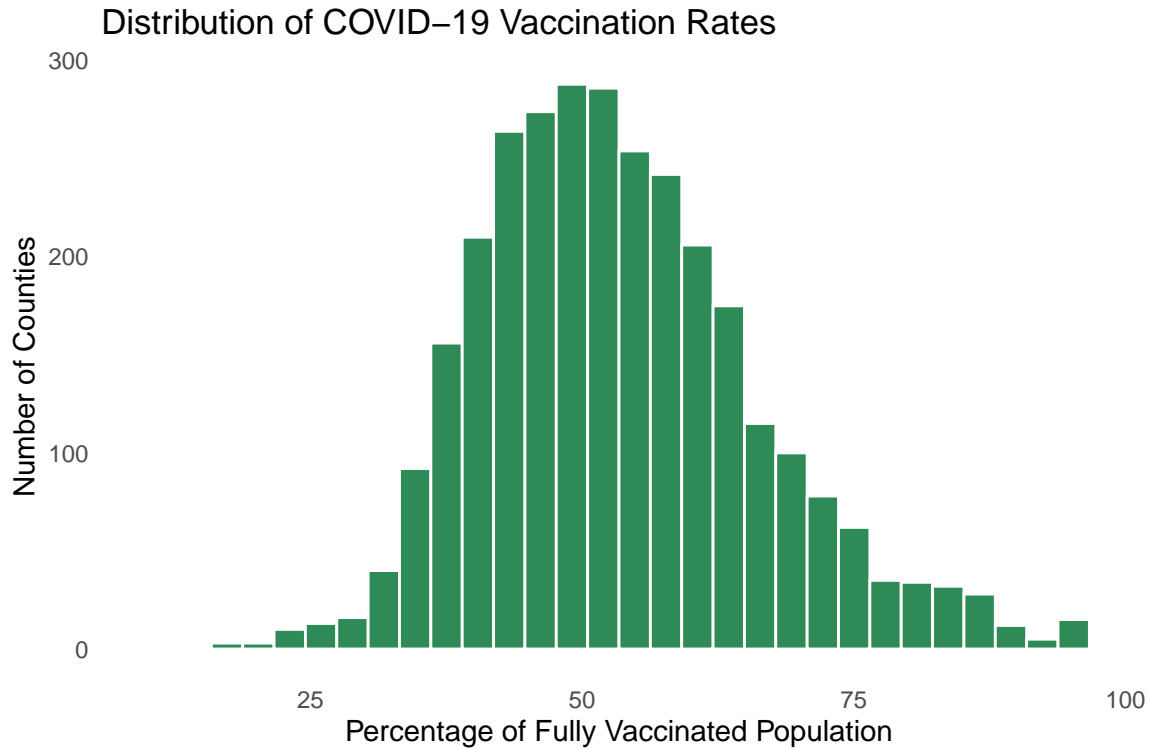

```
##    Min. 1st Qu.  Median   Mean 3rd Qu.   Max.   NA's
## 11.30   44.30   52.10  53.43   61.00  95.00     93
```

Vaccination Rates by Social Vulnerability Category

The social vulnerability categeory is an important dimension to understand the variablity of vaccination. The Social Vulnerability Index (SVI) ranks counties based on factors like income, education, and access to resources, with A being the most vulnerable and **D the least.

From the boxplot, we can see that counties in the higher vulnerability categories (A and B) tend to have slightly lower median vaccination rates compared to those in the lower vulnerability categories. For example, counties in group A, the most socially vulnerable have a visibly lower median vaccination rate than those in group D. The distribution of rates is fairly similar across groups, but the trend suggests that social vulnerability may have played a role in shaping vaccine access or uptake.

This matters because it points to potential inequities in the public health response. If more vulnerable counties consistently lagged in vaccination, it could have left them more exposed to severe outcomes, not because of personal choice alone, but due to structural barriers like healthcare access, information availability, or trust in institutions.

Vaccination Rates by Social Vulnerability Index
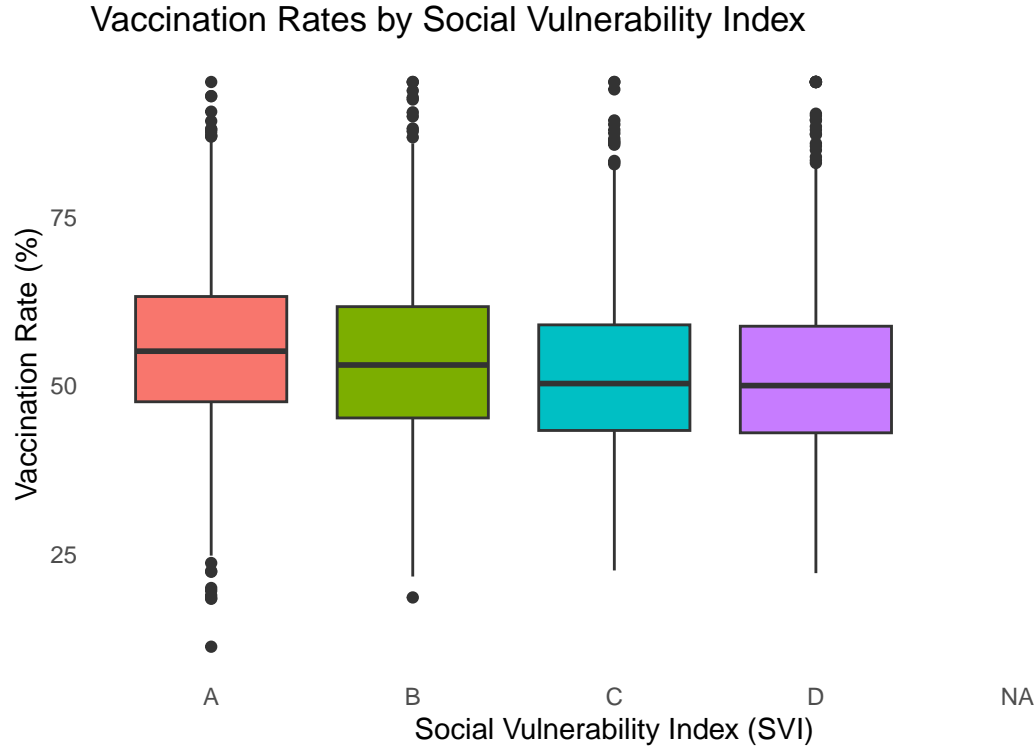
Highest/Lowest Vaccination Rates in Counties

To better understand the upper end of vaccine coverage, we looked at counties that reported the highest vaccination rates, specifically, those that reached 95% full vaccination, the maximum in the dataset. It's interesting to know how geographically diverse these counties are. Places like Apache County in Arizona, Imperial County in California, and Chattahoochee County in Georgia all hit this milestone.

This suggests that high vaccination rates were possible across a range of locations, not just in large urban centers or high-income regions. While national trends often focus on averages, these counties show that local-level efforts, such as strong public health messaging, mobile clinics, or community partnerships may have made a real difference. It also challenges assumptions that rural or socially vulnerable areas couldn't achieve high coverage. These examples make us understand that public health outcomes can be shaped and improved through local action, even in the middle of a pandemic.

Table 3: Counties with Highest and Lowest Vaccination Rates

vax.complete	state	county
95.0	Arizona	Apache
95.0	Arizona	Santa Cruz
95.0	California	Imperial
95.0	Colorado	San Juan
95.0	Georgia	Chattahoochee
95.0	Kansas	Geary
95.0	New Mexico	McKinley
11.3	North Dakota	Slope
95.0	Texas	Brooks
95.0	Texas	Irion

	m1	m2	m3
always.mask	-112.887***		-83.296***
	(10.212)		(10.834)
population	-0.000***	-0.000***	
	(0.000)	(0.000)	
svi.indexB	11.415***	9.968***	9.581***
	(2.884)	(2.878)	(2.859)
svi.indexC	15.142***	14.399***	13.128***
	(3.111)	(3.103)	(3.080)
svi.indexD	19.811***	18.353***	17.491***
	(3.401)	(3.393)	(3.367)
vax.complete		-1.123***	-0.923***
		(0.093)	(0.099)
Num.Obs.	3049	3049	3049

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

95.0	Texas	Maverick
95.0	Texas	Presidio
95.0	Texas	Starr
95.0	Texas	Webb
95.0	Wyoming	Teton

Impact on 2022 COVID Deaths

To understand what actually reduced COVID-19 deaths in 2022, we looked at how factors like mask usage, vaccination rates, population size, and social vulnerability were related to county-level death rates. The regression models give us a clearer picture of which factors made the biggest difference.

The results show that mask adherence had a strong and consistent impact. In the models where mask use was included, higher mask usage was significantly associated with lower death rates. For instance, in model 1, a full one-unit increase in mask adherence (from 0 to 100%) would correspond to around 113 fewer deaths per 100,000 people. Even partial increases would still translate to meaningful reductions in mortality.

Population size had a small but significant negative relationship with death rates, which might reflect the fact that larger counties had more resources or better health infrastructure in place.

Social vulnerability also played a role. Counties in SVI categories B and C (moderate to high vulnerability) experienced higher death rates than less vulnerable counties, suggesting that inequities in access, infrastructure, or baseline health may have made these communities more exposed to the worst outcomes.

Overall, these results reinforce what the earlier data suggested that public health behaviors like masking and structural factors like social vulnerability both mattered significantly in shaping how deadly the pandemic was in different places.