COVID-19 CASES AND DEATHS

Section 1: COVID-19 Cases and Deaths (as of 3/1/2020 and 3/31/2020)

Looking at the CDC datasets on COVID-19 cases and deaths, I initially cleaned the data by removing observations for each state that were not allocated to a specific county (listed as "Statewide Unallocated" in the data). This removed 52 observations and left a total of 3,142 observations. Using a frequency table of case counts by county, we found that as of 3/1/2020, 3102 counties (98.73%) had zero cases reported. The maximum case count was 130 cases, found in one county (Los Angeles Co., CA). This represents 0.03% of all counties observed. In terms of deaths, as of 3/31/2020, there were 3,139 counties with zero deaths (99.9%). The maximum number of deaths (2) occurred in Santa Clara Co., CA. This represented 0.03% of all observations.

Overall, California had the highest occurrence of cases and deaths at this time when compared to the other states. There were 485 cases overall in California (96.23% of all recorded cases at the time). Of the 4 deaths reported overall at this time, 2 were in California. Of cases in California, 64% of them were in and around the urban centers of Los Angeles and San Francisco.
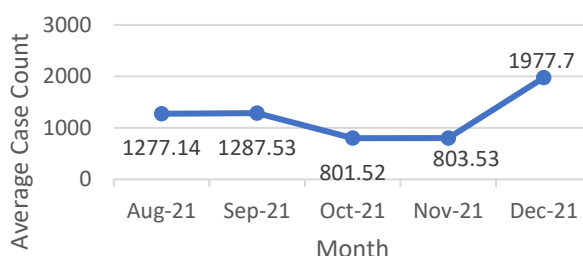
At the start of March 2020, the average number of cases and deaths among the counties were 0.16 cases and 0.001 deaths. However, by the end of March, the situation had changed radically. As of March 31, 2021, the average case and death counts were 61.04 cases and 1.16 deaths. The largest hotspot had also moved from California to the New York City Metro Area (see table).

March 1, 2020		March 31, 2020	
COUNTIES WITH THE MOST CASES			
Los Angeles (CA)	130	Queens (NY)	13,869
Santa Clara (CA)	73	Kings (NY)	11,160
Orange (CA)	59	Westchester (NY)	9,967
COUNTIES WITH THE MOST DEATHS			
Santa Clara (CA)	2	Queens (NY)	376
King (WA)	1	Bronx (NY)	262
La Crosse (WI)	1	Kings (NY)	261

Section 2: COVID-19 Cases (8/1/2021 - 12/31/2021)

For the months of August through December of 2021, the average number of cases dropped around 37% between the months of August and November (approx. 1277 cases in August to 803 cases in November). However, averages spiked back up again in December, more than doubling to an average of 1977 cases.

Average COVID Cases Per County

As we investigate the counties with the top three case counts over this period, we find several recurring areas. Los Angeles County, which started out having the most cases in the nation in March of 2020, made this list every month. Overall, these top areas are all large metropolitan areas: Los Angeles, Houston, Miami, Phoenix, and Chicago. It should be noted however, that these areas all fall into the top ten most populous counties in the U.S., and in particular, Los Angeles, Cook, and Harris Counties are the top three. Given this, the prevalence in cases could very likely be due to the relative size of the populations. Additional analysis would be recommended to investigate if there may be other factors than population that are contributing to this trend.

COUNTIES WITH THE MOST COVID CASES (AUG 2021 – DEC 2021)

AUGUST 2021		SEPTEMBER 2021		OCTOBER 2021	
Los Angeles (CA)	83,756	Harris (TX)	58,494	Maricopa (AZ)	44,474
Miami-Dade (FL)	78,018	Maricopa (AZ)	52,592	Los Angeles (CA)	34,309
Harris (TX)	72,560	Los Angeles (CA)	46,574	Cook (IL)	23,324

NOVEMBER 2021		DECEMBER 2021	
Maricopa (AZ)	63,593	Los Angeles (CA)	21,3877
Cook (IL)	34,447	Miami-Dade (FL)	16,8546
Los Angeles (CA)	32,053	Cook (IL)	15,8009

Section 3: COVID-19 Deaths (8/1/2021 - 12/31/2021)

Between the start of August and end of December 2021, the lives of 18,4142 Americans were lost to COVID-19. The three counties worst affected all had more than 2,000 deaths: Maricopa County, Arizona; Harris County, Texas; and Los Angeles County, California. These counties are also among the 4 most populous in the US.

Aug-Dec Deaths

Obs	county_name	interval_deaths
1	Maricopa County	3229
2	Harris County	2966
3	Los Angeles County	2628

While more fatal cases happen in places with more people, the rate is not uniform: Maricopa County experienced 730 fatalities per million while Los Angeles County had 262 per million.

In the months reviewed, August had the least deaths, at 24,546, while September had the most, with 44,569. There was a reduction in October and November but an increase in December: no clear trend. The states with the most COVID deaths in August through December 2021 were Texas, California, Georgia, Ohio, and Pennsylvania. These states are all in the top 8 for population. Those other states are (in descending population) Florida, New York, and Illinois. Their absence suggests that there may be factors or policies in these states that are more effective at reducing COVID fatalities. Additionally, California had the least deaths per capita of the 5 most-impacted states: 304 deaths per million, compared to the range of 682 – 852 deaths per million for the others.

Most State Deaths

State	state_deaths
TX	22106
CA	11875
GA	9303
OH	8952
PA	8855

Section 4: Conclusions

In conclusion, we can certainly see a trend of increased prevalence and impact to more populated areas. Given this, the overarching recommendation would be to focus resources and efforts into tackling these urban areas. As noted above, additional investigation and analysis into the correlation between case counts and population is recommended to determine if any other factors can be attributed to the trend.

Note: The CDC deaths-by-county data has some anomalies. Cumulative deaths reported decreased across some months for some counties. The biggest discrepancy in the months reviewed was with Arapahoe County, Colorado: at the end of July 2021, they reported 775 cumulative deaths, but reported 765 a month later, suggesting that negative ten people died. This is possibly an accounting error, but it was not alone: that month alone, 15 different counties had decreases in cumulative deaths.