Final Project: COVID-19 And Races

NickyNie

12/08/2021

Introduction

America is a smelting furnace of different races. Racial diversity brings in communication between different cultures but also results in conflicts between them. Racism has been a great problem in America for a long time. With the pandemic, racism violence increased a lot. For example, COVID-19 was called the "Chinese Virus" by some racists, and more people were hostile to all the Asians they saw in the street. I even saw some Asian YouTubers being splashed water and even being beaten when they were doing live streaming. This project aims to study the association between COVID-19 and races.

The dataset I used is a collaboration between the COVID Tracking Project and the Boston University Center for Antiracist Research, which shows the counts of cases/deaths of COVID-19 categorized by races for each state in the United States. The data was collected from April 12, 2020, to March 07, 2021.

Main Question: Is death from COVID-19 associated with races?

Sub Questions:

1. Are the trends of incidence/death of different races the same across the year at the country level?
2. Is the trend in CA similar to that at the country level?
3. Which race with the highest death rate in each state?
4. Is there any spatial finding across the whole country?

Method

Download data

I download data in csv form from The COVID Tracking Project https://covidtracking.com/race

Preprocessing data

Total_objects	NA_total	NA_white	NA_black	NA_latinx	NA_asian	NA_AIAN	NA_NHPI	NA_multi
5320	232	652	733	3248	950	2057	3139	3795

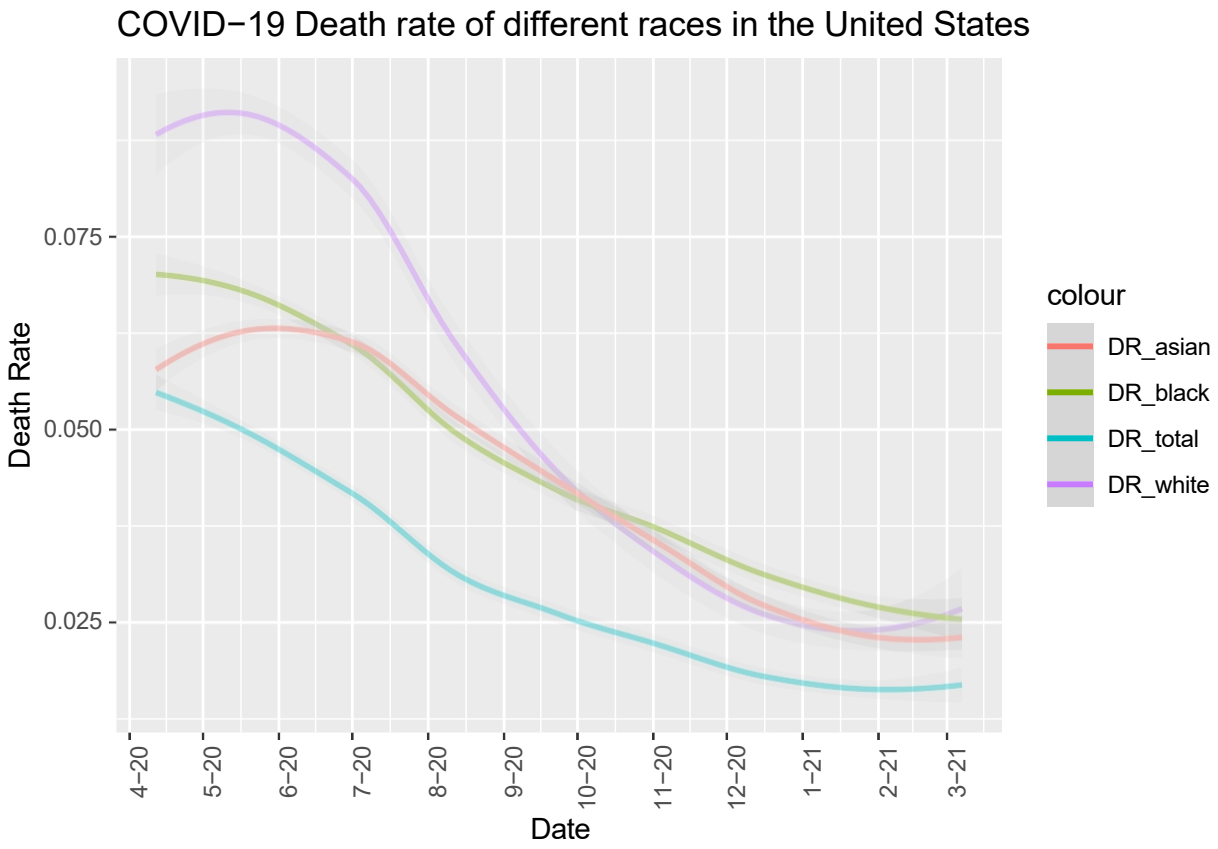
This table shows the total number of objects and the number of missing data in each race. Based on this table, I decided to take a close look at data of Total, White, Black, and Asian since races like Latinx, AIAN,

NHPI and multiracial have more than half of the missing data, which is not suitable for my further analysis. And for the three races categories left, I dropped all the NAs for further analysis.

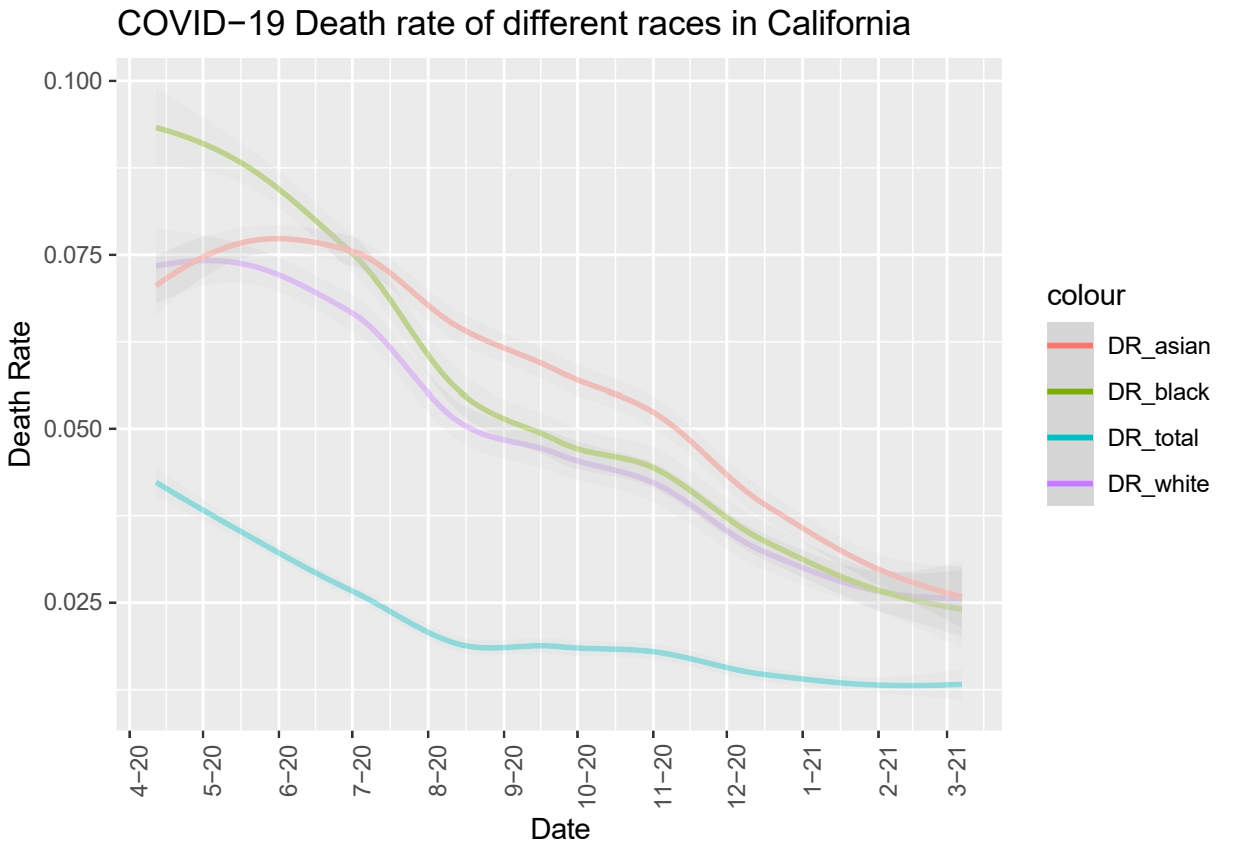
I first created two new datasets, one is based on state(race_state) and another(race_country) is based on country. Because of the different population bases and population distribution in each state, the original counts of deaths and cases maybe not be very meaningful to analyze, so I choose to calculate the death rate instead. I also created a dataset of CA to see if the pattern in CA is similar to the pattern of the whole country. And since the number of cases and deaths are cumulative, I chose the latest data (2021.03.07) and created a new dataset called "latest" to find the race with the highest death rate in each state.

Inference of missing data *I think racial inequality may contribute to the huge number of missing data in minority groups. Some states may think that few people care about data from minority groups and are not willing to devote much attention and labor force to collect data from them. Or the population of those minority groups in those states is too small, hence they ignored it.*

Preliminary Results



From this graph, we can see that the trend of all death rates is gradually decreasing. The total death rate is always the lowest one since there are a lot of cases of other or unknown races which largely increase the denominator. For the relatively higher death rate in white at the beginning of the study, the reason might be the greater old population in white in the United States compared to other races since the old are at higher risk to die from COVID-19. And although the death rate of white is higher than the other two races initially, they approach the same level finally with a little bit of difference.



The general trend is similar to that of country, but it seems that four lines go up slightly at last. But we can still see a trend of the death rate of three races approaching the same level at the final stage of the data.

State	DR_total	DR_white	DR_black	DR_asian
AK	0.0051406	0.0069399	0.0060040	0.0118512
AL	0.0203033	0.0294985	0.0268511	0.0105587
AR	0.0163753	0.0200919	0.0154203	0.0123584
AZ	0.0197567	0.0261499	0.0167992	0.0184548
CA	0.0154578	0.0303423	0.0294305	0.0327237
CO	0.0137369	0.0212970	0.0151143	0.0179519
CT	0.0270003	0.0633329	0.0461045	0.0274925
DC	0.0248678	0.0098058	0.0383356	0.0153173
DE	0.0166716	0.0242453	0.0166431	0.0059718
GA	0.0174951	0.0280037	0.0244460	0.0166592
HI	0.0159550	0.0109709	0.0068966	0.0402109
IA	0.0163629	0.0257178	0.0131501	0.0179570
ID	0.0108656	0.0176252	0.0089286	0.0139459
IL	0.0192050	0.0259433	0.0331343	0.0272217
IN	0.0190885	0.0201121	0.0211149	0.0077890
KS	0.0162644	0.0177400	0.0163992	0.0115837
KY	0.0117334	0.0138986	0.0138005	0.0062500
LA	0.0224720	0.0265559	0.0272827	0.0044953
MA	0.0277616	0.0547782	0.0302473	0.0258121
MD	0.0205386	0.0302754	0.0244214	0.0312394

ME	0.0154169	0.0165973	0.0044405	0.0059761
MI	0.0253905	0.0296323	0.0488960	0.0172747
MN	0.0133670	0.0165998	0.0081703	0.0120633
MO	0.0169793	0.0204635	0.0252862	0.0155200
MS	0.0228778	0.0260481	0.0272974	0.0127208
MT	0.0136849	0.0134534	0.0021739	0.0029240
NC	0.0131877	0.0161873	0.0181097	0.0088106
NE	0.0103716	0.0130129	0.0113348	0.0106800
NH	0.0154044	0.0281319	0.0126829	0.0115830
NJ	0.0290103	0.0481572	0.0535786	0.0387883
NM	0.0203721	0.0339408	0.0207627	0.0159953
NV	0.0170060	0.0251505	0.0207908	0.0244575
OH	0.0180445	0.0259273	0.0211917	0.0102879
OK	0.0105688	0.0133303	0.0117220	0.0072593
OR	0.0145977	0.0200269	0.0121340	0.0119588
PA	0.0256672	0.0400241	0.0348485	0.0234298
RI	0.0198328	0.0305781	0.0133676	0.0157985
SC	0.0166469	0.0198065	0.0224426	0.0104287
SD	0.0167270	0.0176819	0.0027789	0.0108992
TN	0.0147380	0.0178243	0.0195069	0.0096584
TX	0.0165441	0.7144365	0.3355429	0.9624866
UT	0.0052714	0.0055744	0.0035993	0.0083017
VA	0.0163838	0.0255112	0.0228102	0.0168635
VT	0.0129329	0.0145269	0.0032626	0.0067024
WA	0.0146314	0.0369776	0.0147274	0.0305788
WI	0.0114308	0.0127723	0.0117882	0.0081656
WY	0.0124534	0.0215528	0.0047059	0.0213523

Looking back to the table of death rates in each state, I find that Texas has a ridiculous result in the death rate of White (0.7144365) death rate of Black (0.3355429), and death rate of Asian (0.9624866), which is abnormally high. Especially for Asians, which means that if you were an Asian in Texas and got COVID-19, you would have been almost a hundred percent to die. So, I think that there must be some problems in this data and hence I delete the data of Texas when I list the top 5 states with the highest death rate in each race and draw the map of each race. The total death rate of Texas is not problematic, so I include it in the list of top 5 states with the highest total death rate and draw the map of a total death rate.

The top5 states with the highest death rate of total are NJ, MA, CT, PA, MI

The top5 states with the highest death rate of White are CT, MA, NJ, PA, WA

The top5 states with the highest death rate of Black are NJ, MI, CT, DC, PA

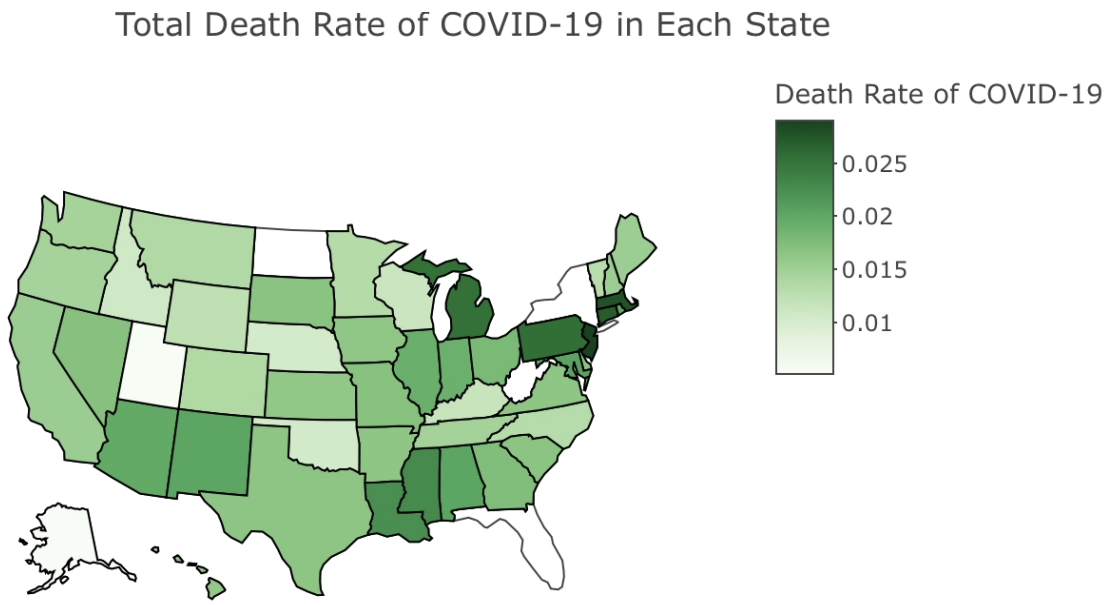
The top5 states with the highest death rate of Asians are HI, NJ, CA, MD, WA

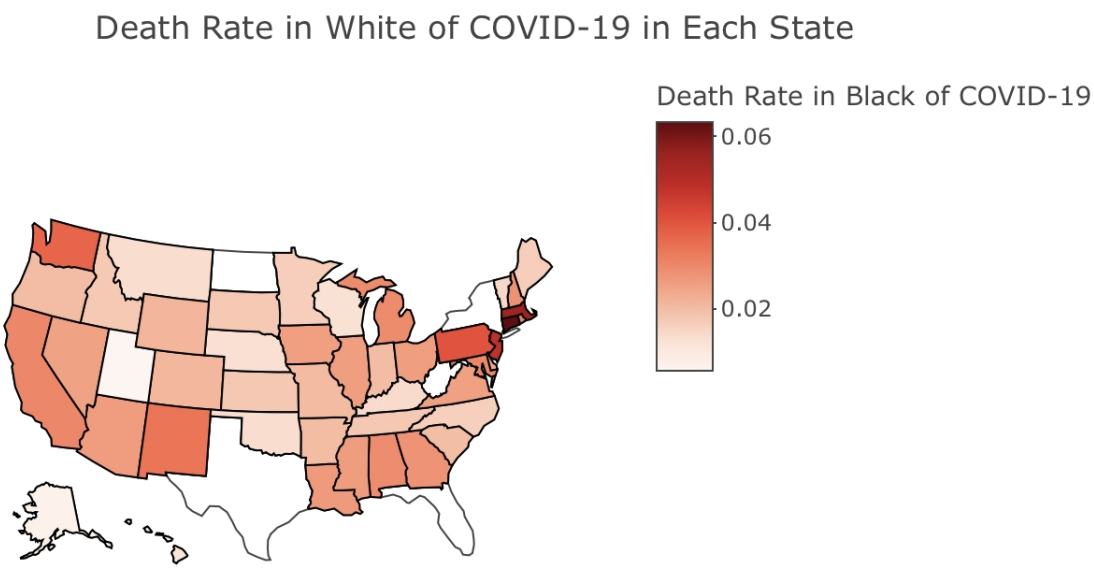
For the top five states with the highest death rate in each race, MA, NJ, CT, PA, MI, and WA appear more than one time. Hence, in my opinion, the death rate is highly dependent on the action of the State rather than race, for example, the burden of the medical system, the policy to restrict COVID-19, etc.

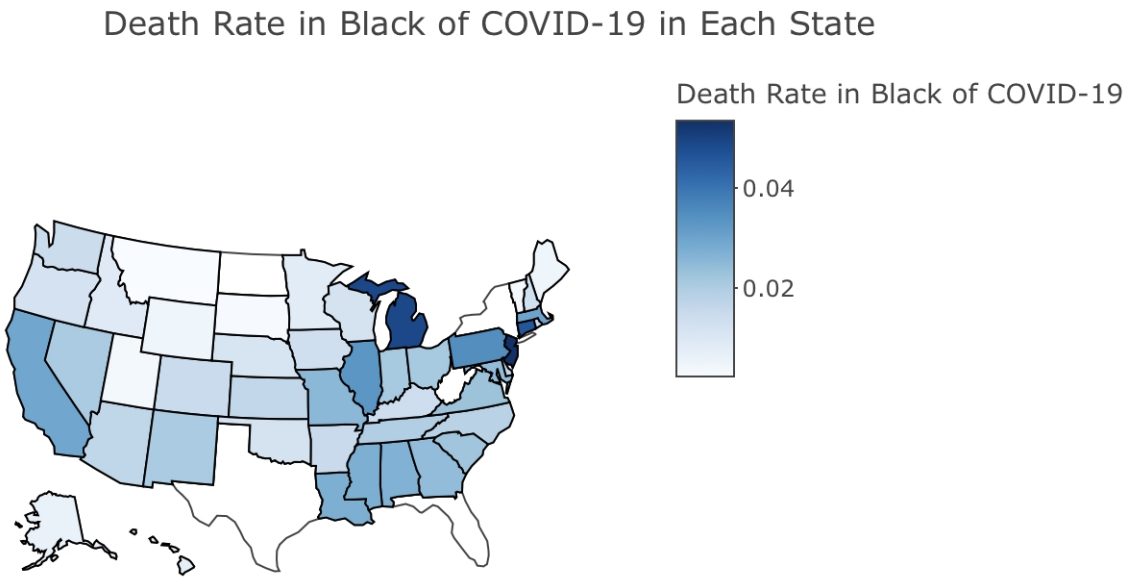
highest_DR	N_states	avg_DR_white	avg_DR_black	avg_DR_asian
Asian	6	0.1330899	0.0676491	0.1811356
White	30	0.0249392	0.0159199	0.0143990
Black	11	0.0236851	0.0304532	0.0152750

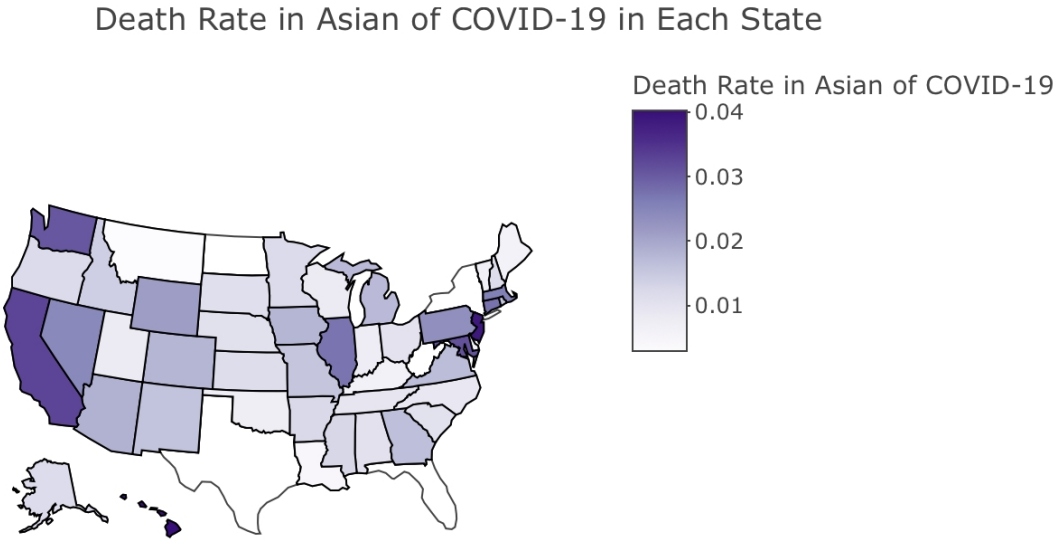
This table shows the number of states with the highest death rate in each race and the average death rate of each race within each race category. Although the number of states with DR_white to be the highest DR among three races is greatest (30), the difference among average DR is not that so great. This probably is caused by samples collection. The majority of the United States population is white, hence the data collected in white might be greater, leading to a relatively higher DR.

Death rate map









Generally, it shows a radial pattern: the middle part has the lowest death rate and gradually becomes higher while approaching the coastline. And total death rate of the eastern and north-eastern areas is the highest. Relatively higher populations and more samples of those states along the coastline may lead to a greater burden on the medical system and finally lead to a higher death rate. Also, there is more international transportation for those states along the coastline, which may bring in more cases and deaths of COVID-19.

The patterns of death rates in those three maps with different races are similar to the pattern of the total death rate. The difference such as the high death rate of black in Michigan, the high death rate of Asians in California and Hawaii is caused by the relatively higher population of such races in those states, which provides more data.

Conclusion

Just from these data, I think there is no clear association between races and the death rate of COVID-19. Both at CA and country level, the death rates of all three races approach the same level finally. But from the huge number of missing data from minority groups, we can still get some insights of unequaled treatment and ignorance of those groups from the government. Compared to races, the action taken by each state's government and the location of each state may contribute more to the death rate of COVID-19.

I saw a lot of data showed a different result from these data, which indicated that the death rate of black is the highest among all races. That might be caused by inequality of medical distribution at the stage of the onset of COVID-19. As the burden of medical resources relieved and the government took action to restrict COVID-19 and gave supports to citizens, I believe that the death rate that approaches the same level is the right trend.