CAPSTONE PROJECT REPORT

MIS 6940 Capstone Project

Prof. Venu Gopal Balijepally

Priyanka Sabnani

Abstract

Why is the mortality rate higher in developed countries? It is clear that the Coronavirus pandemic will remain on the world agenda for a few more months. When we look at the information shared by different institutions, organizations, or experts, it can be seen that we have not yet reached satisfactory information about the pandemic and its spread. However, it is accepted that all kinds of contact reduction and isolation are effective in terms of protection. In this project we analyzed government data, healthcare data and data shared by Dr. Anupam Sule from St Joseph Mercy Hospital, to understand the spread of virus in all counties of USA. Not only this report has trends, pandemic variables calculated like waves, severity, and magnitude of pandemic but it also allows to draw a meaningful insight for the government and health care system of United States America to formulate policies and measure to control the spread of the virus. The counties are distinguished on Rural Urban Continuum Code and we analyzed four counties where first county is most urban ones followed by second, third and fourth. My analysis is focused on County 1and County 4.

COVID-19-Mobility Data Analysis

Organization: St Joseph's Mercy Hospital

Foundation: In 1911, four Sisters of Mercy arrived in Ann Arbor from Dubuque, Iowa. They came at the invitation of local medical and religious leaders who dreamed of founding a community hospital to serve area residents. That dream became a reality on November 21, 1911, when the Sisters opened St. Joseph's Sanatorium, a small hospital located in a former student rooming house at the corner of State and Kingsley streets. Saint Joseph Mercy Health System (SJMHS) is one of the largest health care networks based in southeast Michigan, United States. It consists of five prime hospitals, nine Urgent Care Centers, and five Health Centers spread around metro Detroit, providing health care in six counties that include Livingston, Macomb, Oakland, St. Clair, Washtenaw, and Wayne.

Problem Description:

In this project we have dived deep into 'What does data say about Covid-19 situation in United States of America?'. And with available data we came up with some observations and conclusions. This analysis mainly focuses on:

What is the virus situation in all parts of United States of America?
A detailed county wise comparison (total county type is 4)
Probable reasons: government norms/ demographics that play a role in the spread of the virus.
Impact of regulation imposed to control the spread.

Organization Sponsors:

Dr Venugopal Balijepally
The University of Texas at Arlington (Ph.D.)
Management Development Institute, India (M.B.A)
IIT Mumbai (M.S)

Dr Anupam Sule
M.D. Information and Outcomes at St Joseph Mercy Hospital Oakland
William Beaumont School of Medicine at Oakland University

System Capabilities:

The deliverables are correlation matrixes between the demographic variables, three metrics and Pandemic variables that helps understand and effectiveness of these factors. It gives insights on possible ways to control this virus. Another is a dashboard that can help to analyze the data at granular levels. It can help to understand the wave and its surge with timelines, trends, county waves and duration on day basis. As discussed earlier there are four types of counties from urban to rural and this project draws a clear distinction between the different types of counties. The calculation of pandemic variables like MOP, SOP, and wave calculations in all parts of United States of America that provides information for counties and states.

Business Benefits:

Insights to government and health care section to formulate policies for the betterment of society and to monitor and control the virus in four different types of counties. Detailed analysis of the variables like MOP, SOP and all the trends of waves: Wave1, Wave2, Wave3. Segregates all counties based on demographic variables like Social Economic Factors, Vote Count, Household Composition and Disability, Housing Type and Transportation and Total Population Ethnicity and Area for each county. Calculation of waves for each county, insights like duration of waves, trends severity of wave and its magnitude. These variables are further calculated at different timelines of the Pandemic like at day 15, 30, 60 and 120 days, respectively. The two important variables of this research are MOP (Magnitude of Pandemic) No of Cases in each County and SOP (Severity of Pandemic) No of Deaths in each Counties. We have drawn trends between SOP and MOP Cases Trends vs Death Cases on day-by-day count, Total cases and deaths for each county, Duration of waves, Top affected Sates, States with highest magnitude and deaths. Lastly, we correlated the MOP and SOP with our three important metrics reduction in visits, reduction in encounters, reduction in POI: to understand these measures taken by government and degree of their effectiveness.

PHASE 1

Demographic Variables

Demographic Variable Sheet: This data was collected using government sites and consolidated sheet shared by Dr Anupam. We calculated Area of the counties, Republican vote counts, percentage of population and medical facilities like patients count for Ambulatory care and Diabetes each county. The vote count data was collected from government website election lab, Covid19.census and Social Vulnerability index. We focussed on Republican data and the vote count percentage is of republican votes only. Also the data is from 2018.

Demographic Variables 2018 data

□□Socioeconomic Status
o Below Poverty EP-POV
o Unemployed EP_UNEMP
o Income EP_PCI
o No High School Diploma EP_NOHSDP

• Household Composition & Disability
o Aged 65 or Older EP_AGE65
o Aged 17 or Younger EP_AGE17
o Civilian with a Disability EP_DISABL
o Single-Parent Households EP_SNGPNT

• Minority Status & Language
o Minority EP_MINRTY
o Speaks English "Less than Well" EP_LIMENG

• Housing Type & Transportation
o Multi-Unit Structures EP_MUNT
o Mobile Homes EP_MOBILE
o Crowding EP_CROWD
o No Vehicle EP_NOVEH
o Group Quarters EP_GROUPQ

Socioeconomic - RPL_THEME1
• Household Composition & Disability - RPL_THEME2
• Minority Status & Language - RPL_THEME3
• Housing Type & Transportation - RPL_THEME4
Overall RPL_THEMES

PHASE 2

Calculation of Waves and Pandemic

Phase 2 started with understanding the data and calculation of waves. We used moving averages to calculate waves; there were three waves derived for the duration of data which is from February 24, 2020 to January 14, 2021. In many of the county's wave three is still going on.

Algorithm of Start and End wave, to reach this variable we first calculated ***start of pandemic*** (Cumulative cases above 10 per 100K) it is number of cumulative cases more than 10 over 100,000 population. We created a Column Start of Pandemic where cases count was more than 10 over population of 100k (see sheet URBANCOUNTY1_final_variables and URBANCOUNTY4_final_variables) The column has value 0 and 1 which implies if the pandemic started 1 else 0. ***Moving average,*** was calculated on average for 7 days, moving average for 14 days and moving average for 28 days with a condition that MA7>MA14>MA28 was consecutive for 7 days and created a column in the excel file. So the algorithm is MA7>MA14>MA28 for 7 consecutive days and the pandemic is 1/pandemic has started.

Start of Wave1: when the Pandemic has started, and the value is 1 and the 7-day moving average is more than 14 day moving average which is more than 28 day moving average for 7 days.

End of wave1: when the pandemic is 1 and the 7-day moving average is less than 14 day moving average which is less than 28 day moving average for 7 days.
Start of Wave 2 when pandemic is 1 the 7-day moving average is more than 14 day moving average which is more than 28 day moving average for 7 days *and it is less than End Wave1*.
End of Wave 2 when the pandemic is 1 and the 7-day moving average is less than 14 day moving average which is less than 28 day moving average for 7 days and *it is less than Start wave 1*.

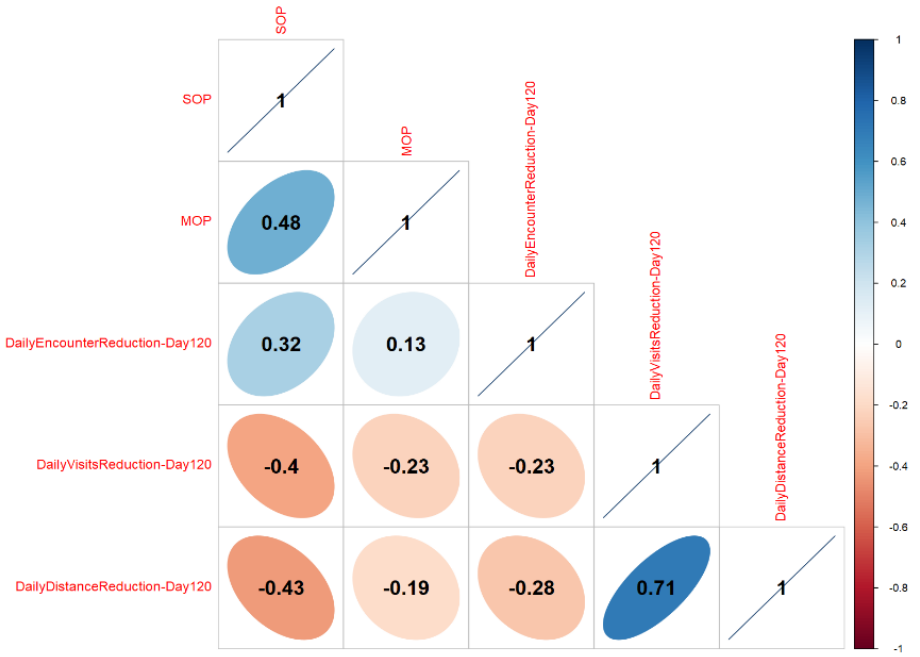
Calculation of Pandemic Variables

We calculated **MOP** (Magnitude of Pandemic) which is Cumulative cases per 100,000 population at day 120 after crossing threshold point of start of pandemic.
Another important variable is **SOP** (Severity of Pandemic) which is Cumulative deaths per 100,000 population at day 120 after crossing threshold point of start of pandemic.
Magnitude of Wave **MOW** that is cumulative cases per 100k population at various days like at day 15, day 30, day 60 and day 120 after crossing the threshold of Start of Wave.
Severity of Waves **SOW** which is deaths per 100 covid cases at day 120 after crossing the threshold point of start of wave.
So, the variables are **MOW15, MOW30, MOW60, MOW120 and SOW15, SOW30, SOW60, SOW120. And SOP AND MOP for WAVE1 WAVE2 and WAVE3 and peak date where the number of cases were maximum for each county.**

PHASE 3

Correlation Matrix

Correlation Matrix between SOP, MOP and three metrics from Uncast Mobility Data



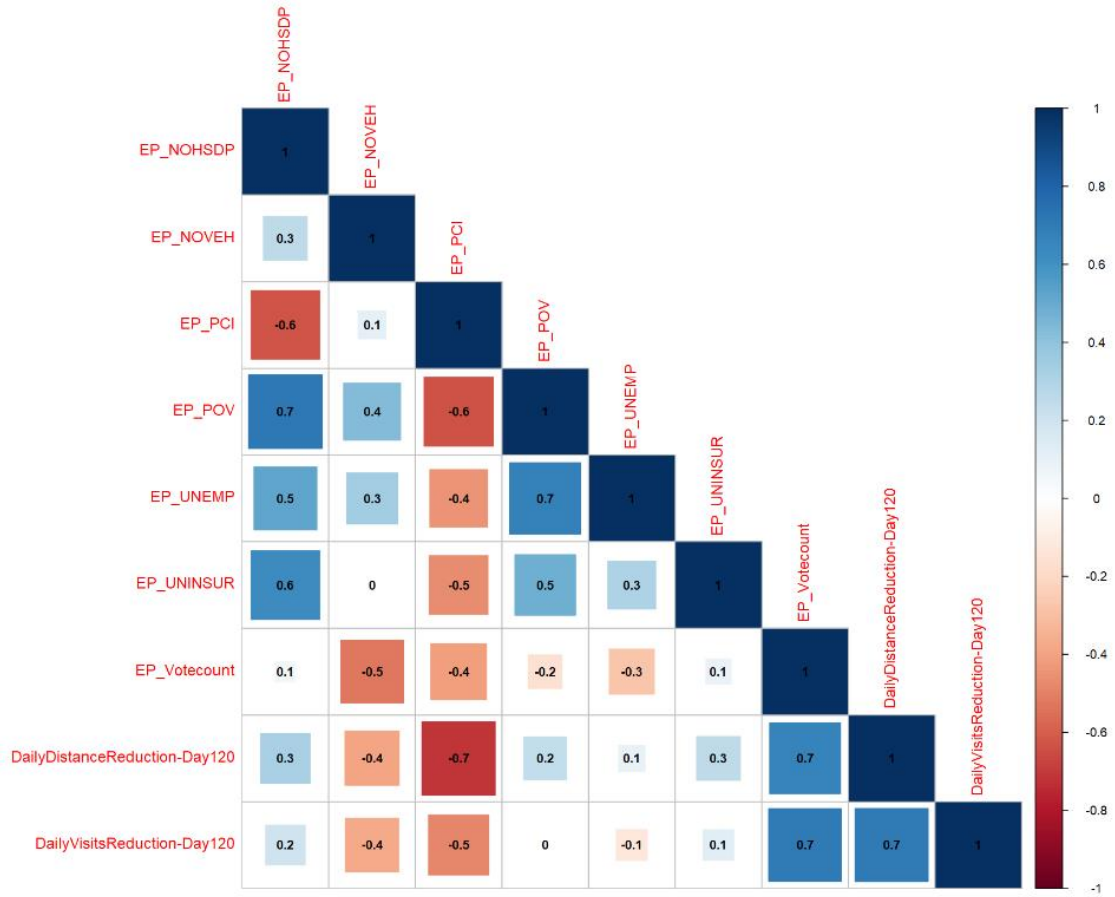
Analysis:

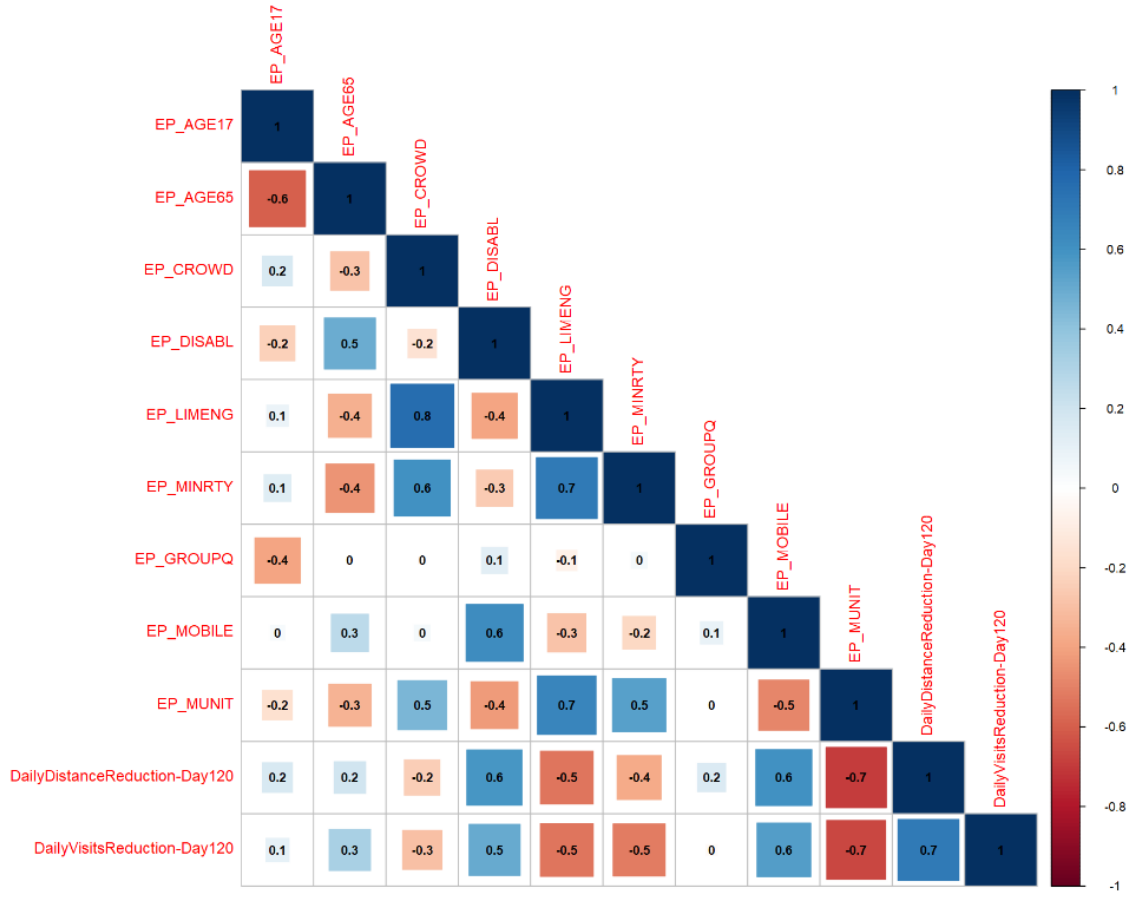
Correlation between MOP and Daily Distance Reduction is 19% that means if daily reduction is reduced government can control the count of cases by 19%. The sigma value was 0.000 that means this is accurately there in the data and did not appeared by any chance.
Correlation between SOP and Daily Distance Reduction; it is a very powerful correlation and suggests that if the daily reduction in distance that is not driving and moving out of the house is imposed government can control the death rate by 43%.
Correlation between MOP and Daily Visits Reduction is 23% suggests if daily visits (Pont if Interest) are reduced that is going out for shopping, malls, meeting friends and different places of fund and leisure, it can help to control the spread of the pandemic by 23%
Correlation between SOP and Daily Visits Reduction is 40% again a very powerful number or correlation signifies that by implementing the policy of lock down and closing the business and stores and other places of interests can reduce the death by 40%

Correlation Matrix between Demographic Variables and Uncast Metrics

Another important **Correlation Matrix is between three metrics from Uncast Data and all demographics variables**. **County 1**

Part A (County 1)


EP_NOHSDP
EP_NOVEH
EP_PCI
EP_POV
EP_UNEMP
EP_UNINSUR
EP_Votecount
DailyDistanceReduction-Day120
DailyVisitsReduction-Day120
1 0.3 -0.6 0.7 0.5 0.6 0.1 0.3 0.2
0.3 1 0.1 0.4 0.3 0 -0.5 -0.4 -0.4
-0.6 0.1 1 -0.6 -0.4 -0.5 -0.4 -0.7 -0.5
0.7 0.4 -0.6 1 0.7 0.5 -0.2 0.2 0
0.5 0.3 -0.4 0.7 1 0.3 -0.3 0.1 -0.1
0.6 0 -0.5 0.5 0.3 1 0.1 0.3 0.1
0.1 -0.5 -0.4 -0.2 -0.3 0.1 1 0.7 0.7
0.3 -0.4 -0.7 0.2 0.1 0.3 0.7 1 0.7
0.2 -0.4 -0.5 0 -0.1 0.1 0.7 0.7 1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


EP_AGE17
EP_AGE65
EP_CROWD
EP_DISABL
EP_LIMENG
EP_MINRTY
EP_GROUPQ
EP_MOBILE
EP_MUNIT
DailyDistanceReduction-Day120
DailyVisitsReduction-Day120
1 -0.6 0.2 -0.2 0.1 0.1 -0.4 0 -0.2 0.2 0.1
-0.6 1 -0.3 0.5 -0.4 -0.4 0 0.3 -0.3 0.2 0.3
0.2 -0.3 1 -0.2 0.8 0.6 0 0 0.5 -0.2 -0.3
-0.2 0.5 -0.2 1 -0.4 -0.3 0.1 0.6 -0.4 0.6 0.5
0.1 -0.4 0.8 -0.4 1 0.7 -0.1 -0.3 0.7 -0.5 -0.5
0.1 -0.4 0.6 -0.3 0.7 1 0 -0.2 0.5 -0.4 -0.5
-0.4 0 0 0.1 -0.1 0 1 0.1 0 0.2 0
0 0.3 0 0.6 -0.3 -0.2 0.1 1 -0.5 0.6 0.6
-0.2 -0.3 0.5 -0.4 0.7 0.5 0 -0.5 1 -0.7 -0.7
0.2 0.2 -0.2 0.6 -0.5 -0.4 0.2 0.6 -0.7 1 0.7
0.1 0.3 -0.3 0.5 -0.5 -0.5 0 0.6 -0.7 0.7 1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Analysis:

There is a positive correlation between no vehicle and no high school diploma by 30%. Those who do not have high school diploma have 30% probability of not having a vehicle. Income is negatively correlated with No high school diploma by 60% which suggests that if no high school diploma increases, they will have less jobs and the income probability to drop is significantly high that is 60%. Poverty and No diploma are positively correlated poor households will have no vehicle by 40%. Uninsured households are positively correlated with households without high school diploma. It signifies no diploma will increase unemployment and that will reduce income, therefore resultant will be these households will not buy insurance. If number of republicans is increased the households with vehicles will increase by 50%. Also vote count is negatively correlated with per capita income and increase in republican may decrease the per capita by 40%. Vote count is also positively correlated with Reduction in Daily Distance travelled indicates they follow the guidelines of not travelling much by 70%.

Daily distance reduction is also negatively correlated with per capita income by 70% if individuals moving out is reduced, they will not be able to get jobs (positive correlation 10%) and in result poverty will increase by 20% and per capita income will by 70% (Daily distance is positively correlated with Per capita income by 70%)

Daily visit reduction: This factor is positively correlated with republican vote means they follow guidelines and do not go to various point of interest. And negatively correlated with per capita income and household with no vehicles, suggests if increase in reducing moving out to POI places will decrease the per capita income by 50% (economy will take a hit) and increase the households without vehicles by 40% (clearly, we saw that automobiles sales were very down during the pandemic, it faced worst decline since1980)

Part B (County 4)

Correlation within variables County Type 4
Correlation between three metrics and Demographic Variables

The correlation between daily distance reduction and republican vote count is positive 35% like County1. If travel is reduced the poverty might shoot by 13%, the data also shows a positive relation in unemployment which is a obvious factor. In these conditions the per capita of the households will reduce up to 39%. There is no relation with MOP magnitude of the Pandemic and Daily distance reduction. There is no correlation between MOP and SOP with Reduction in daily distance travelled or Daily Visit Reduction unlikely to County 1.

Similarities between County 1 and County 4

County 1 and County 4 republican people follow the norms and will not travel much.

County 1 and County 4 if distance reduced increases the poverty of the households and as a result there is also increase in unemployment.

Daily distance reduction decreases the Per capita Income of the county 1 and County 4.

Dissimilarities between County 1 and County 4

County 1- Daily distance reduction is correlated with MOP and SOP and County 4 these variables do not show any correlation.

County 1- Daily visit reduction is correlated with MOP and SOP and County 4 these variables do not show any correlation.

PHASE 4

POWER BI DASHBOARD ANALYSIS

Sheet 1: Sheet1 has overall data, total number of the cases and total deaths that took place during the time and the dates in the data are February 24, 2020 to January 14, 2021. It also gives the count of cases in the state, cases in the counties, deaths in the state and deaths in each county. Number of days for each wave or individual county. Start of pandemic for each state and county. It also shows the waves- Start and End and duration of the days a county stayed in each wave. There are few counties in wave 1 and wave 2 and not in Wave 3 that means the wave 2 is still going on and wave 3 has not started in that county.

Slide 2 Sheet two shows trend of cases vs trends of deaths on county and state level per day and consolidated. The trend shows overall spread of cases and deaths across USA. Maximum cases and maximum deaths trends can be understood by analyzing these trends. For Example, New York had the first hit whereas California stayed calm till election time. New Jersey had 1800 deaths in June 27. The overall trend of peak of virus was seen in summer, the month of June have been peaking for most of the States. Ohio had more deaths as compared to cases registered. South Carolina have multiple highs and lows after summer in death trend.

Slide 3 Geographical representation of the data based on states: It gives the count of cases in each state and the top 5 states that have maximum cases:

California with 2.3 million cases
Texas with 1.3 million cases
Florida with 1 million cases

New York with 1 million cases
Illinois with 750K cases
New Jersey with 500k cases

Slide 4

Another geographical graph represents deaths states wise.

Top 5 states with highest cases of deaths are:
New York with approx. 36000 deaths
California with 25000 deaths
New Jersey with 18000 deaths
Florida with 14000 deaths
And interestingly New Jersey is on third with 18000 deaths but it was 6th position in terms of cases

Analysis: The New York took the earliest hit. On April 3, it had 13000 cases and on April 14, 2400 deaths; the severity and magnitude at this time was high because the NY is port of entry for many international flights. So, flights from Europe and China landed in New York before the restrictions and led to this massive spread. Especially lot of Chinese returned during this time as they celebrated their New Year Eve in February or so, unknowingly they were the ones carrying virus and New York became the center.

California on the other hand remained a little controlled till July 6, there was a peak 13500 cases were registered but there was a massive surge in the cases on December 27. There were 44 thousand cases reported and this led to surge in severity of the deaths a result January 7, 833 deaths were reported.

Lessons Learned:

It is important to check the credibility of data before starting the project.
The agile format of working with every week changes and findings.
Always try different ways to do the same thing to check accuracy of the output.
The hit and trial needs patience and lot of attention.
Sometimes it is very important to step and look at the problem from a holistic approach.

Limitations:

Excel is not a very powerful tool handling the huge data.
The data can be improvised with updated data *2020 election* for republican vote count (the data used if from 2018)
Cumulative Encounter data can be revised.

Future Work

Prediction Model: It can be used to mine data and predict the next waves and peaks, the pandemic magnitude and severity. It will also be helpful to understand the reaction of counties and impact of virus and policies implemented to control the surge. Trends could be analyzed to understand why some states have higher SOP than others for Example New Jersey; It was at 6 position in maximum cases but second highest in deaths.

Appendix:

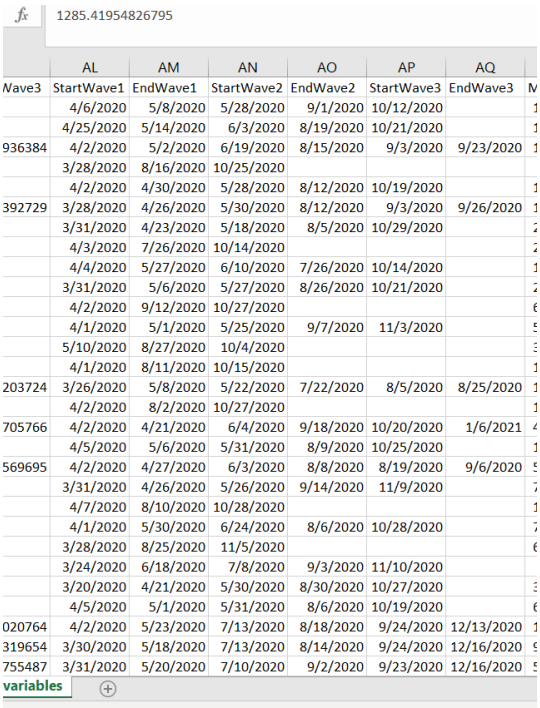
PHASE 1

Demographic Variable Sheets

	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1	Demographic details														
2	EP_NOHSDP	EP_AGE65	EP_AGE17	EP_DISABL	EP_SNGPNT	EP_MINRTY	EP_LIMENG	EP_MUNIT	EP_MOBILE	EP_CROWD	EP_NOVEH	EP_GROUPQ	EP_UNINSUR	RPL_THEME1	RPL_THEI
3	11.3	14.6	24.2	19.3	7.5	25	80.00%	3.8	18.4	1.4	5.6	1	7.1	0.3631	
4	9.7	19.5	21.9	14	5.7	17	50.00%	18.3	11.7	1.3	3.4	1.4	10.2	0.2232	
5	27	18	21.1	22.2	12.8	53.9	160.00%	1.3	29.2	3.4	9.2	11.2	11.2	0.978	0
6	16.8	16.3	20.7	16.7	6	25.4	30.00%	2.4	28.5	0.8	6	9.3	7.9	0.7694	0
7	19.8	17.8	23.4	14.2	7	12.9	170.00%	0.9	25.2	1.6	4.2	0.9	11	0.6143	0
8	24.8	15.6	20.8	14.4	19.5	78.5	40.00%	1.7	33.4	0	11.7	4.2	10.8	0.9771	0
9	15.4	19	22.8	17.7	10.5	48.1	50.00%	1.3	26.2	1.8	7.8	1.6	10.2	0.8455	0
10	15.9	16.8	21.9	20.8	10.4	27.5	100.00%	3.7	14.7	1.7	5.8	2.7	9.4	0.7866	0
11	18.6	18.9	20.7	16.7	9.7	44.2	10.00%	4	14	3	7.3	1.5	10.8	0.6901	0
12	19.8	21.7	20.4	20	4.7	8.2	40.00%	2	29.2	1.5	4.1	1.1	8.3	0.6879	0
13	17.8	15.8	24.2	17.7	10.3	19.9	190.00%	0.7	32.3	2.7	4.9	0.9	14.2	0.7685	0
14	18.6	22.1	20.3	28.4	8.7	43.6	40.00%	1.1	31.6	1.4	10.1	1.4	11.4	0.9057	0
15	18.8	19	22.1	17.1	7.9	47.1	10.00%	1.8	26.5	1.8	10.2	1.5	10.8	0.9436	0
16	24.5	19.8	21.1	15.3	9.7	19.7	40.00%	0.4	20.9	0.9	4.2	1.9	10.2	0.8162	0
17	22.6	18.8	23	22.6	5.5	7.4	70.00%	0.4	30.3	0.9	4	1.1	12.8	0.8694	0
18	13.8	16.3	23.7	17.7	9.7	29.4	140.00%	0.9	13.6	1.4	6	1.2	10.1	0.5373	0
19	15.6	19.3	21.3	18.7	8.2	21.3	30.00%	2.2	9.3	0.4	6.1	0.8	8.1	0.6519	0
20	19.2	21.6	21.2	22.8	10.8	49.9	0.00%	0.5	28.7	1.1	6.8	0.4	14.5	0.9529	0
21	20.4	21.8	17	22.6	5	34.7	40.00%	0.2	30.7	0.9	4.4	3.6	7.8	0.693	0
22	17.1	20.5	21.9	20.9	9	16.6	40.00%	1.3	21.6	2.1	5.2	1.8	11.2	0.8232	
23	21.6	18.3	22.5	21.1	8.9	28.8	120.00%	0.4	27.3	0.9	6.6	1.2	10	0.7271	0
24	18.2	18	22.6	17.6	6.7	7.9	60.00%	2.5	22.9	1.8	4.2	1.5	11.8	0.6895	0
25	14.2	16.1	23.4	21.3	8.2	31	100.00%	1.2	18.2	1.7	5.5	2.6	11.4	0.779	0
26	19.3	16.8	24.4	17.8	13.8	72.5	10.00%	1.5	20.5	3.1	13.5	1.5	12	0.9669	0
27	25.9	16.3	24.6	15.5	9.4	19.3	350.00%	1.1	23.9	2.9	5.3	1.3	14.5	0.8035	0
28	13.6	14.5	22.5	17.9	9.2	26.8	70.00%	2.6	15.1	1.4	4.7	5.4	8.2	0.4911	0
29	18.5	17.3	22.3	18.8	8.5	39.8	50.00%	1	17.5	2	4.9	7.6	15.5	0.9557	0
30	15.5	18.3	21.6	19.6	8.9	21.9	120.00%	3.1	11	1.2	6.3	1.3	10.4	0.7169	0

PHASE 2

CALCULATION OF WAVES: START WAVE 1, END WAVE 1 , START WAVE 2, END WAVE 2



fx 1285.41954826795

	AL	AM	AN	AO	AP	AQ	
Wave3	StartWave1	EndWave1	StartWave2	EndWave2	StartWave3	EndWave3	M
	4/6/2020	5/8/2020	5/28/2020	9/1/2020	10/12/2020		1
	4/25/2020	5/14/2020	6/3/2020	8/19/2020	10/21/2020		1
936384	4/2/2020	5/2/2020	6/19/2020	8/15/2020	9/3/2020	9/23/2020	1
	3/28/2020	8/16/2020	10/25/2020				
	4/2/2020	4/30/2020	5/28/2020	8/12/2020	10/19/2020		1
392729	3/28/2020	4/26/2020	5/30/2020	8/12/2020	9/3/2020	9/26/2020	1
	3/31/2020	4/23/2020	5/18/2020	8/5/2020	10/29/2020		2
	4/3/2020	7/26/2020	10/14/2020				2
	4/4/2020	5/27/2020	6/10/2020	7/26/2020	10/14/2020		1
	3/31/2020	5/6/2020	5/27/2020	8/26/2020	10/21/2020		2
	4/2/2020	9/12/2020	10/27/2020				6
	4/1/2020	5/1/2020	5/25/2020	9/7/2020	11/3/2020		5
	5/10/2020	8/27/2020	10/4/2020				3
	4/1/2020	8/11/2020	10/15/2020				1
203724	3/26/2020	5/8/2020	5/22/2020	7/22/2020	8/5/2020	8/25/2020	1
	4/2/2020	8/2/2020	10/27/2020				1
705766	4/2/2020	4/21/2020	6/4/2020	9/18/2020	10/20/2020	1/6/2021	4
	4/5/2020	5/6/2020	5/31/2020	8/9/2020	10/25/2020		1
569695	4/2/2020	4/27/2020	6/3/2020	8/8/2020	8/19/2020	9/6/2020	5
	3/31/2020	4/26/2020	5/26/2020	9/14/2020	11/9/2020		7
	4/7/2020	8/10/2020	10/28/2020				1
	4/1/2020	5/30/2020	6/24/2020	8/6/2020	10/28/2020		7
	3/28/2020	8/25/2020	11/5/2020				6
	3/24/2020	6/18/2020	7/8/2020	9/3/2020	11/10/2020		
	3/20/2020	4/21/2020	5/30/2020	8/30/2020	10/27/2020		3
	4/5/2020	5/1/2020	5/31/2020	8/6/2020	10/19/2020		6
020764	4/2/2020	5/23/2020	7/13/2020	8/18/2020	9/24/2020	12/13/2020	1
319654	3/30/2020	5/18/2020	7/13/2020	8/14/2020	9/24/2020	12/16/2020	9
755487	3/31/2020	5/20/2020	7/10/2020	9/2/2020	9/23/2020	12/16/2020	5

variables

Calculation of Start of Pandemic MOP SOP, MOW'S SOW'S at day 15, 30, 60, 120, Maximum cases (Peak Date)

MOP	MOW1-Day15	MOW1-Day30	MOW1-Day60	MOW1-Day120	MOW2-Day15	MOW2-Day30	MOW2-Day60	MOW2-Day120	MOW3-Day15	MOW3-Day30	MOW3-Day60	MOW3-Day120
1509.332857	142.8954184	192.0157185	339.3766187	1737.07243								
1326.392972	76.09033999	84.73696953	290.5267527	1900.529174								
1339.245521	83.2808139	121.5449716	231.8357792	1573.332133	553.7048708	4004.231566	553.7048708	553.7048708	3292.968398	6761.501756	3292.968398	3292.968398
1367.04663	90.65054292	127.3966591	254.034101	1579.931154								
1135.043346	58.09276968	81.55331129	142.9975869	1369.648762								
1082.213301	97.84016683	146.071235	215.8914479	1267.788077	292.6018135	2762.032503	292.6018135	292.6018135	2559.462017	7292.996849	2559.462017	2559.462017
2065.458667	130.6654492	149.5568395	541.5531871	2249.649722								
2422.897864	55.53556483	102.2202187	234.8947054	2659.620717								
1633.789913	53.37205508	101.1259991	203.116323	1777.699989								
2177.040976	254.4051715	389.9489104	765.3008028	2350.119904								
618.0710082	63.48241429	102.0146243	207.619206	665.9969402								
574.4994044	54.70184461	78.62848345	125.7015447	656.8555889								
330.3205198	40.44741059	57.55977661	151.4185114	522.7049984								
1680.996128	108.1171861	241.2067129	547.8176495	1851.479419								
1238.669995	59.1130721	86.15826849	161.1120985	1285.419548	220.6115305	2531.816742	220.6115305	220.6115305	2260.978418	3049.539073	2260.978418	2260.978418
1069.152802	47.26528895	83.00553076	203.8610797	1139.783077								
402.9332537	32.6363383	39.91675223	54.47758009	470.9674666								
1453.524848	115.2376843	180.2840344	336.0795549	1543.424004								
556.4225048	57.72980133	71.3246541	92.07130146	632.7083137	124.9953288	1466.11789	124.9953288	124.9953288	1198.859836	3426.740496	1198.859836	1198.859836
786.5240097	65.2783085	81.19984715	138.5173863	956.8844733								
1453.521308	72.38249885	117.5183536	277.0534176	1570.443354								
781.7681296	62.51628808	111.1633661	224.094083	858.7527297								
601.7816366	98.91679305	161.5338455	273.1555478	672.9064408								
565.242971	80.48809442	126.0153958	239.1161703	623.0326401								
326.6848285	68.31437268	99.64457853	128.7961939	378.6597726								
655.7823129	64.39909297	78.00453515	97.95918367	717.4603175								
1069.728519	166.2089478	334.3505579	636.0391248	1151.286863								
943.8157754	151.9974413	363.8495865	647.8929012	1019.052986								

SOP	SOW1	SOW2	SOW3	PeakDate
8.930963651	13.39644548	44.65481825		12/20/2020
5.187977726	10.37595545	25.93988863		12/20/2020
11.25416404	13.50499685	69.77581705	121.5449716	1/11/2021
28.39472617	30.97606492			12/20/2020
8.937349182	10.05451783	41.33523997		12/20/2020
12.86161818	13.32096168	22.96717531	47.77172465	12/20/2020
86.58553864	97.60551629	125.9426017		12/14/2020
40.28613635	46.57318143			1/3/2021
26.36190575	30.46744845	46.24137566		1/3/2021
31.27932437	35.44990095	95.92326139		1/14/2021
10.65020711	10.88953761			12/25/2020
8.929144207	10.05612357	17.51152553		12/20/2020
0.518556546	1.037113092			12/14/2020
42.83249496	45.34267839			1/2/2021
16.22711783	20.86343721	43.27231422	50.22679329	8/5/2020
17.75990871	19.5233039			12/25/2020
3.514682587	3.765731343	12.30138905		12/10/2020
27.200465	29.83146236	48.81512022		12/22/2020
6.958502839	8.569267386	27.44742787	46.841033	12/27/2020
3.184307731	6.368615463	17.51369252		1/12/2021
18.48551777	22.33857854			12/22/2020
15.69647099	16.7149443	25.94111427		1/13/2021
5.898707843	6.3524546			1/13/2021
14.87138211	14.87138211	21.00256597		1/13/2021
8.610619487	9.181202706	16.18381494		1/13/2021
17.68707483	19.04761905	24.94331066		12/24/2020
32.27545848	33.04852335	52.56841141		11/16/2020
53.61031999	54.82873635	60.00700589		11/19/2020
45.41261619	45.41261619	61.02320301		11/22/2020

Pandemic Variable were calculated using MS Access.

SQL Queries for Pandemic Variables used in MS Access

MOP -

```
SELECT COVID_MOBILITY_URBAN4_7.[countyFIPS], COVID_MOBILITY_URBAN4_7.[COVIDCases_Date], COVID_MOBILITY_URBAN4_7.[CumCasesPer100k]
FROM COVID_MOBILITY_URBAN4_7, COUNT4DatesDay120
WHERE COVID_MOBILITY_URBAN4_7.[countyFIPS] = COUNT4DatesDay120.[countyFIPS]
and     COVID_MOBILITY_URBAN4_7.[COVIDCases_Date] = COUNT4DatesDay120.[Day120];
```

SOP-

```
SELECT COVID_MOBILITY_URBAN4.[countyFIPS], COVID_MOBILITY_URBAN4.[COVIDCases_Date],
((COVID_MOBILITY_URBAN4.[Cdeaths]*100)/(COVID_MOBILITY_URBAN4.[population])) AS CDeathsper100
FROM COVID_MOBILITY_URBAN4, COUNT4DatesDay120
WHERE COVID_MOBILITY_URBAN4.[countyFIPS] = COUNT4DatesDay120.[countyFIPS]
and     COVID_MOBILITY_URBAN4.[COVIDCases_Date] = COUNT4DatesDay120.[Day120];
```

START WAVE 1-

```
SELECT COVID_MOBILITY_URBAN4_7.countyFIPS, Min(COVID_MOBILITY_URBAN4_7.COVIDCases_Date) AS StartWave1
FROM COVID_MOBILITY_URBAN4_7
WHERE (((COVID_MOBILITY_URBAN4_7.StartWave)=1))
GROUP BY COVID_MOBILITY_URBAN4_7.countyFIPS;
```

END WAVE 1-

```
SELECT COVID_MOBILITY_URBAN4_7.countyFIPS, Min(COVID_MOBILITY_URBAN4_7.COVIDCases_Date) AS EndWave1
FROM COVID_MOBILITY_URBAN4_7 RIGHT JOIN [COUNTY4-StartWave1Dates] ON COVID_MOBILITY_URBAN4_7.countyFIPS = [COUNTY4-StartWave1Dates].countyFIPS
WHERE (((COVID_MOBILITY_URBAN4_7.EndWave)=1) And ((COVID_MOBILITY_URBAN4_7.COVIDCases_Date)>[COUNTY4-StartWave1Dates].StartWave1))
GROUP BY COVID_MOBILITY_URBAN4_7.countyFIPS;
```

START WAVE 2-

```
SELECT COVID_MOBILITY_URBAN4_7.countyFIPS, Min(COVID_MOBILITY_URBAN4_7.COVIDCases_Date) AS StartWave2
FROM COVID_MOBILITY_URBAN4_7 LEFT JOIN [COUNTY4-EndWave1Dates] ON COVID_MOBILITY_URBAN4_7.countyFIPS = [COUNTY4-EndWave1Dates].countyFIPS
WHERE (((COVID_MOBILITY_URBAN4_7.StartWave)=1) And ((COVID_MOBILITY_URBAN4_7.COVIDCases_Date)>[COUNTY4-EndWave1Dates].EndWave1))
GROUP BY COVID_MOBILITY_URBAN4_7.countyFIPS;
```

END WAVE 2-

```
SELECT COVID_MOBILITY_URBAN4_7.countyFIPS, Max(COVID_MOBILITY_URBAN4_7.COVIDCases_Date) AS EndWave2
FROM COVID_MOBILITY_URBAN4_7 RIGHT JOIN [COUNTY4-StartWave2Dates] ON COVID_MOBILITY_URBAN4_7.countyFIPS = [COUNTY4-StartWave2Dates].countyFIPS
WHERE (((COVID_MOBILITY_URBAN4_7.EndWave)=1) And ((COVID_MOBILITY_URBAN4_7.COVIDCases_Date)>[COUNTY4-StartWave2Dates].StartWave2))
GROUP BY COVID_MOBILITY_URBAN4_7.countyFIPS;
```

SEVERITY OF WAVE 1-

```
SELECT COVID_MOBILITY_URBAN4_7.[countyFIPS], COVID_MOBILITY_URBAN4_7.[COVIDCases_Date],
((COVID_MOBILITY_URBAN4_7.[Cdeaths]*100)/(COVID_MOBILITY_URBAN4_7.[population])) AS CDeathsper100
FROM COVID_MOBILITY_URBAN4_7, COUNT4Datesthreshholdwavedate
WHERE COVID_MOBILITY_URBAN4_7.[countyFIPS] = COUNT4Datesthreshholdwavedate.[countyFIPS]
and     COVID_MOBILITY_URBAN4_7.[COVIDCases_Date] = COUNT4Datesthreshholdwavedate.[Day120];
```

MAGNITUDE OF WAVE-

```
SELECT COVID_MOBILITY_URBAN4_7.[countyFIPS], COVID_MOBILITY_URBAN4_7.[COVIDCases_Date], COVID_MOBILITY_URBAN4_7.[CumCasesPer100k] AS MOW120
FROM COVID_MOBILITY_URBAN4_7, COUNT4Datesthreshholdwavedate
WHERE COVID_MOBILITY_URBAN4_7.[countyFIPS] = COUNT4Datesthreshholdwavedate.[countyFIPS]
and     COVID_MOBILITY_URBAN4_7.[COVIDCases_Date] = COUNT4Datesthreshholdwavedate.[Day120];
```

PEAK DATE OF WAVE-

```
SELECT COVID_MOBILITY_URBAN4_7.countyFIPS, Max(COVID_MOBILITY_URBAN4_7.Conf) AS MaxOfCases,
first(COVID_MOBILITY_URBAN4_7.COVIDCases_Date) AS MaxCaseDate
FROM COVID_MOBILITY_URBAN4_7, COUNT4Dates
WHERE (((COVID_MOBILITY_URBAN4_7.countyFIPS)=COUNT4Dates.countyFIPS) And
((COVID_MOBILITY_URBAN4_7.COVIDCases_Date)>=COUNT4Dates.StartWave1 And
(COVID_MOBILITY_URBAN4_7.COVIDCases_Date)<=COUNT4Dates.EndWave1))
GROUP BY COVID_MOBILITY_URBAN4_7.countyFIPS;
```

PHASE 3

Correlation Matrix between SOP, MOP and three metrics from Uncast Mobility Data

		countyFIPS	DailyDistanceReduction-Day120	DailyVisitsReduction-Day120	DailyEncounterReduction-Day120	DailyDistanceReduction-StartWave1	DailyVisitsReduction-StartWave1
countyFIPS	Pearson Correlation	1	0.038	.104*	0.031	-.132**	-0.088
DailyDistanceReduction-Day120	Pearson Correlation	0.038	1	.709**	-.278**	.170**	.140**
DailyVisitsReduction-Day120	Pearson Correlation	.104*	.709**	1	-.225**	-0.063	.305**
DailyEncounterReduction-Day120	Pearson Correlation	0.031	-.278**	-.225**	1	0.028	0.011
DailyVisitsReduction-EndWave3	Pearson Correlation	0.083	.651**	.959**	-.611**	-0.128	0.123
DailyEncounterReduction-EndWave3	Pearson Correlation	0.054	-.446**	-.594**	.993**	0.096	-0.071
MOP	Pearson Correlation	0.026	-.188**	-.227**	.126**	.208**	.168**
	Sig. (2-tailed)	0.585	0.000	0.000	0.009	0.000	0.000
	N	432	432	432	432	432	432
SOP	Pearson Correlation	-0.038	-.430**	-.396**	.324**	.187**	.143**
	Sig. (2-tailed)	0.432	0.000	0.000	0.000	0.000	0.003
	N	432	432	432	432	432	432
SOW1	Pearson Correlation	-0.032	-.422**	-.394**	.323**	.162**	.121*
SOW2	Pearson Correlation	-0.047	-.316**	-.314**	.271**	.188**	.184**
SOW3	Pearson Correlation	-0.166	-0.226	-0.243	.428**	.250*	0.212

Correlation Matrix between Demographic Variables and Uncast Metrics

Another important **Correlation Matrix is between three metrics from Uncast Data and all demographics variables. County 1**

Part A (County 1)

		population	POVALL_2018	Percent of adults with high school diploma or less	Civilian_labor_force_2018	Employed_2018	Unemployed_2018	Unemployment_rate_2018	Med_HH_Income_Percent_of_State_Total_2018	PovRate	SAHDate	SchClosDate	DailyDistanceReduction-Day1
	N	432	432	432	432	432	432	432	432	432	431	432	
SAHDate	Pearson Correlation	-.171**	-.136**	.177**	-.172**	-.172**	-.168**	-.167**	-0.059	.134**	1	.132**	.3
	Sig. (2-tailed)	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.220	0.005		0.006	0
	N	431	431	431	431	431	431	431	431	431	431	431	
SchClosDate	Pearson Correlation	0.052	0.048	0.012	0.049	0.049	0.034	-.119*	0.061	-0.026	.132**	1	
	Sig. (2-tailed)	0.284	0.317	0.797	0.311	0.305	0.477	0.013	0.207	0.591	0.006		0
	N	432	432	432	432	432	432	432	432	432	431	432	
DailyDistanceReduction-Day120	Pearson Correlation	-.434**	-.357**	.650**	-.451**	-.453**	-.392**	.182**	-.344**	.227**	.316**	.124*	
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	
	N	432	432	432	432	432	432	432	432	432	431	432	
DailyVisitsReduction-Day120	Pearson Correlation	-.462**	-.413**	.563**	-.468**	-.470**	-.423**	0.074	-.169**	-0.014	.243**	0.058	.7
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.124	0.000	0.777	0.000	0.228	0
	N	432	432	432	432	432	432	432	432	432	431	432	
DailyEncounterReduction-Day120	Pearson Correlation	.170**	.210**	-.102*	.177**	.178**	.167**	0.015	-0.001	.131**	-0.078	-0.006	-.2
	Sig. (2-tailed)	0.000	0.000	0.035	0.000	0.000	0.000	0.759	0.982	0.007	0.107	0.907	0
	N	432	432	432	432	432	432	432	432	432	431	432	
DailyDistanceReduction-StartWave1	Pearson Correlation	0.070	0.083	.241**	0.067	0.067	0.073	.251**	-.212**	.333**	0.018	-0.038	.1
	Sig. (2-tailed)	0.147	0.085	0.000	0.164	0.166	0.132	0.000	0.000	0.000	0.714	0.426	0
	N	432	432	432	432	432	432	432	432	432	431	432	
DailyVisitsReduction-StartWave1	Pearson Correlation	-0.028	-0.015	.342**	-0.028	-0.028	-0.017	.264**	-.234**	.252**	-0.023	-.096*	.1
	Sig. (2-tailed)	0.563	0.755	0.000	0.565	0.559	0.721	0.000	0.000	0.000	0.638	0.045	0

Part B (County 4)

Correlation within variables County Type 4

			StartWave1	EndWave1	StartWave2	EndWave2	StartWave3	EndWave3	MOP	SOP	MOW15	MOW30	MOW60	MOW120	SOW	Cum_Dist120
30		Sig. (2-tailed)	0.438	0.075	0.495	0.388	0.106		0.000	0.000	0.000		0.000	0.000	0.001	0.284
31		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
32	MOW60	Pearson Correlation	0.045	0.010	0.103	0.065	-0.059	.	.703**	.449**	.691**	.867**	1	.692**	.460**	-0.056
33		Sig. (2-tailed)	0.509	0.885	0.139	0.399	0.627		0.000	0.000	0.000	0.000		0.000	0.000	0.418
34		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
35	MOW120	Pearson Correlation	.139*	.173*	.168*	-0.076	0.069	.	.911**	.551**	.353**	.458**	.692**	1	.628**	0.127
36		Sig. (2-tailed)	0.042	0.012	0.015	0.328	0.568		0.000	0.000	0.000	0.000	0.000		0.000	0.065
37		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
38	SOW	Pearson Correlation	-0.072	0.050	0.121	-0.113	0.065	.	.637**	.909**	.177**	.219**	.460**	.628**	1	-0.010
39		Sig. (2-tailed)	0.295	0.474	0.083	0.142	0.591		0.000	0.000	0.010	0.001	0.000	0.000		0.889
40		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
41	Cum_Dist120	Pearson Correlation	0.050	0.093	-0.013	0.009	0.128	.	0.095	-0.028	-0.094	-0.074	-0.056	0.127	-0.010	1
42		Sig. (2-tailed)	0.467	0.181	0.854	0.910	0.287		0.166	0.679	0.169	0.284	0.418	0.065	0.889	
43		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
44	Cum_visit120	Pearson Correlation	-0.070	-0.035	-0.054	-0.076	-0.149	.	0.003	0.068	-0.007	-0.006	-0.030	-0.021	0.056	0.019
45		Sig. (2-tailed)	0.307	0.612	0.439	0.326	0.215		0.961	0.321	0.913	0.933	0.659	0.760	0.418	0.785
46		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
47	Cum_encounters120	Pearson Correlation	-0.084	0.046	0.000	-0.039	-0.019	.	0.107	0.018	-0.017	-0.019	-0.020	0.085	0.004	0.070
48		Sig. (2-tailed)	0.222	0.510	0.999	0.613	0.875		0.119	0.792	0.810	0.783	0.776	0.215	0.956	0.309
49		N	214	210	207	169	71	0	214	214	214	214	214	214	214	214
50	Fullrepublicanmerged_ID	Pearson Correlation	0.092	.173*	.188**	-.190*	0.110	.	.387**	.262**	0.006	0.012	.148*	.399**	.302**	.152*
51		Sig. (2-tailed)	0.181	0.012	0.007	0.014	0.362		0.000	0.000	0.933	0.859	0.031	0.000	0.000	0.026

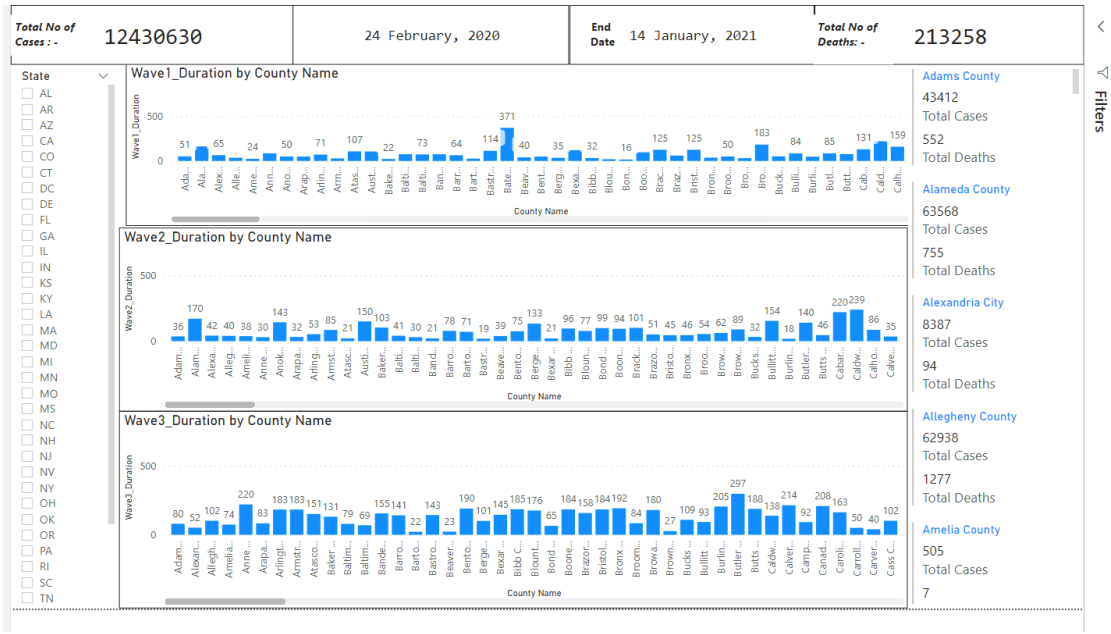
Correlation between three metrics and Demographic Variables

	candidatevotes	totalvotes	EP_Votecount	EP_POV	EP_UNEMP	EP_PCI	EP_NOHSDP	EP_AGE65	EP_AGE17	EP_DISABL	EP_SNGPNT	EP_MINRTY	EP_LIMENG	EP_MUNIT	EP_MOBILE	EP_CR
	0.221	0.245	0.872	0.872	0.851	0.982	0.402	0.271	0.186	0.293	0.483	0.710	0.113	0.977	0.108	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
MOW60	-.171*	-.150*	-0.102	0.025	0.017	-.139*	.255**	-0.133	.187**	-0.095	.184**	.222**	.340**	0.001	-0.009	
	0.012	0.028	0.137	0.711	0.804	0.042	0.000	0.053	0.006	0.164	0.007	0.001	0.000	0.992	0.895	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
MOW120	-.282**	-.260**	-.152*	0.058	0.032	-.353**	.536**	-.269**	.377**	-0.058	.381**	.569**	.503**	-0.019	.209**	
	0.000	0.000	0.026	0.398	0.641	0.000	0.000	0.000	0.000	0.401	0.000	0.000	0.000	0.787	0.002	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
SOW	-.187**	-0.107	-.257**	0.018	-0.002	-.295**	.565**	-0.124	.377**	0.060	.459**	.595**	.511**	-.208**	.179**	
	0.006	0.118	0.000	0.793	0.972	0.000	0.000	0.071	0.000	0.382	0.000	0.000	0.000	0.002	0.009	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
Cum_Dist120	-0.062	-.241**	.355**	.136*	0.112	-.393**	0.122	-.184**	0.049	0.134	-0.010	0.048	-0.085	0.075	.219**	
	0.369	0.000	0.000	0.048	0.101	0.000	0.074	0.007	0.477	0.051	0.879	0.486	0.216	0.277	0.001	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
Cum_visit120	0.107	0.035	.163*	0.022	0.038	.195**	-0.021	.276**	0.004	0.111	-0.014	-0.098	-0.042	-.135*	0.005	
	0.119	0.614	0.017	0.751	0.584	0.004	0.758	0.000	0.950	0.106	0.839	0.154	0.542	0.049	0.943	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
Cum_encounters120	0.056	-0.025	0.074	0.040	0.034	-0.075	0.109	0.009	0.016	0.113	.149*	0.050	-0.078	.141*	-0.063	
	0.417	0.716	0.282	0.561	0.619	0.274	0.111	0.893	0.812	0.099	0.030	0.463	0.255	0.040	0.359	
	214	214	214	214	214	214	214	214	214	214	214	214	214	214	214	
Fullrepublicanmerged_ID	-.310**	-.206**	-.290**	.160*	0.124	-.338**	.413**	-0.133	0.075	0.097	.296**	.554**	.323**	.159*	.321**	
	0.000	0.002	0.000	0.020	0.070	0.000	0.000	0.052	0.273	0.158	0.000	0.000	0.000	0.020	0.000	

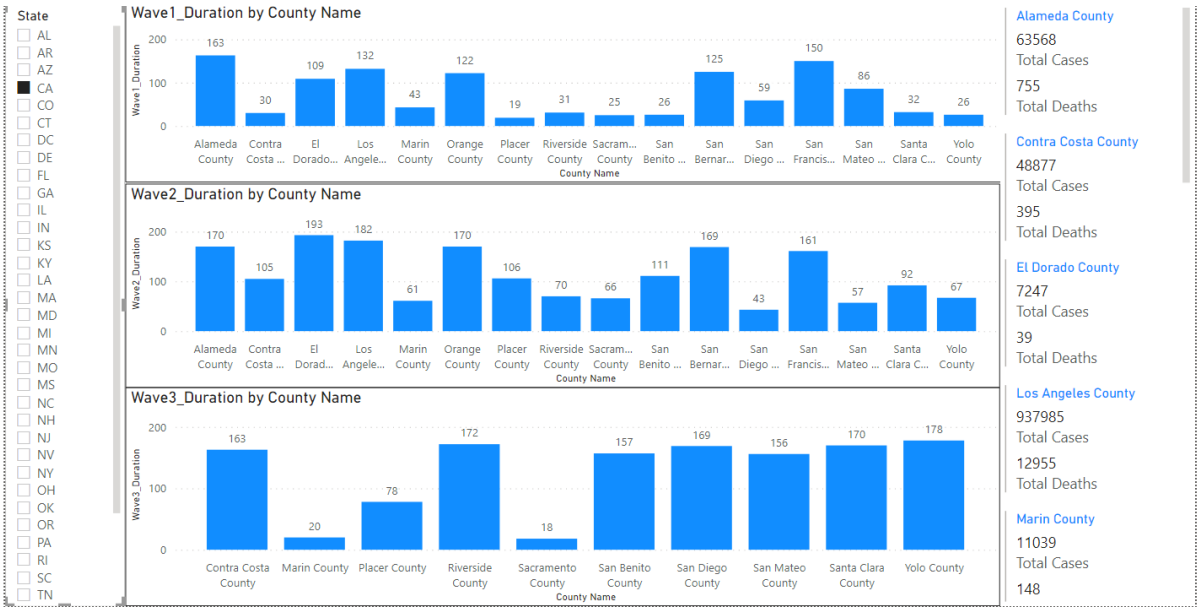
PHASE 4

POWER BI DASHBOARD ANALYSIS

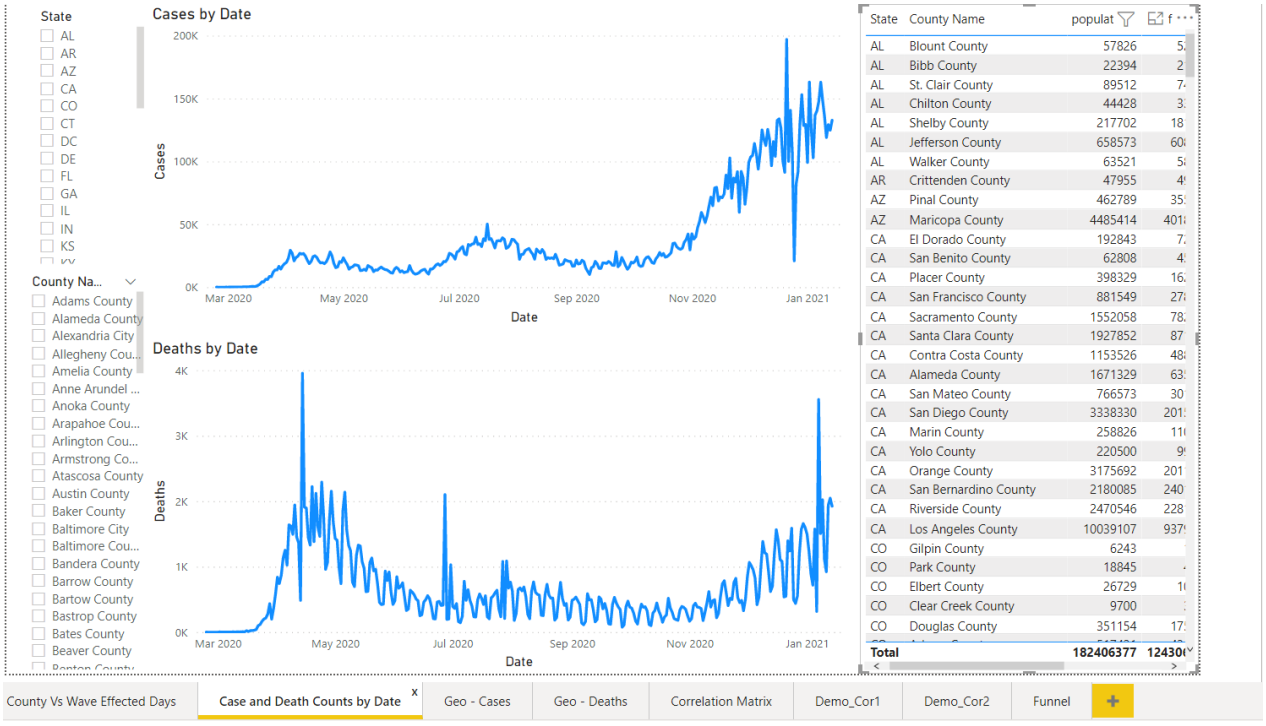
Sheet 1 Over view of Waves for all Counties and States



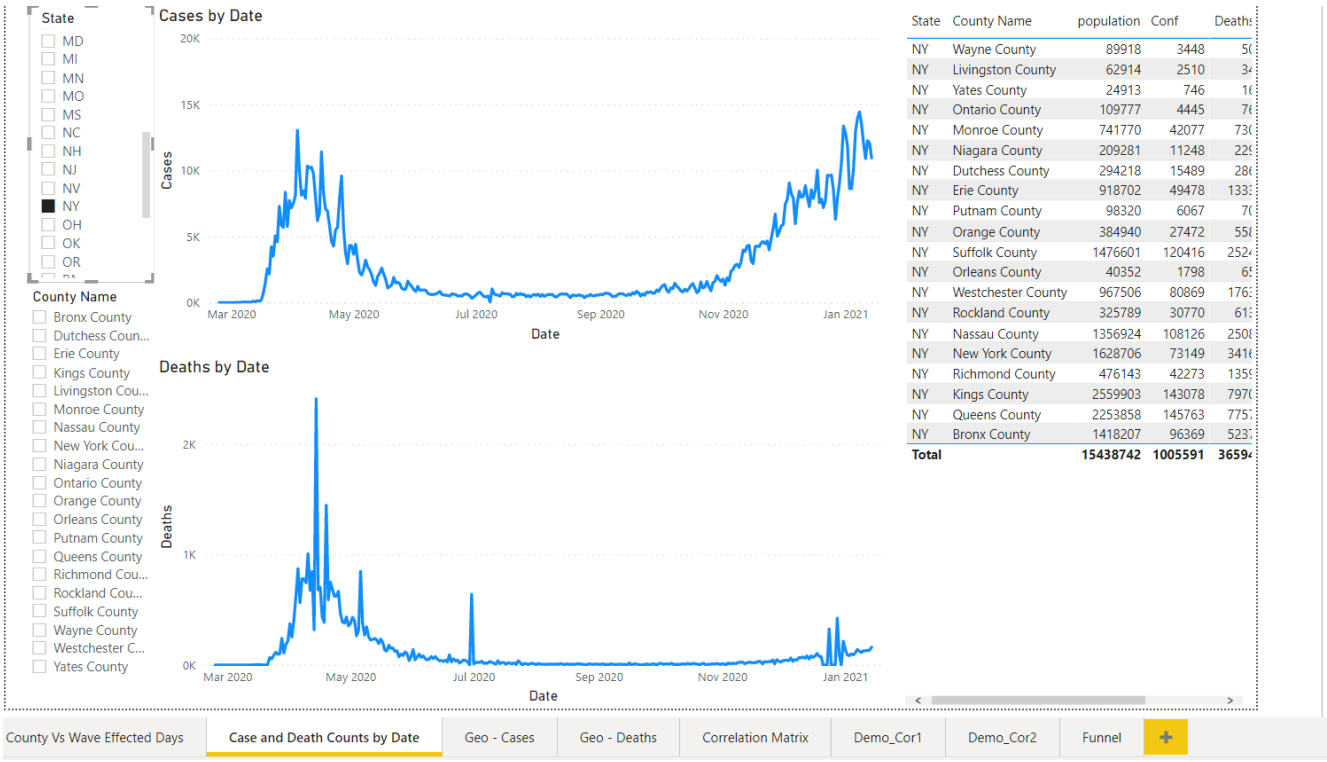
Slide 1Example for California State



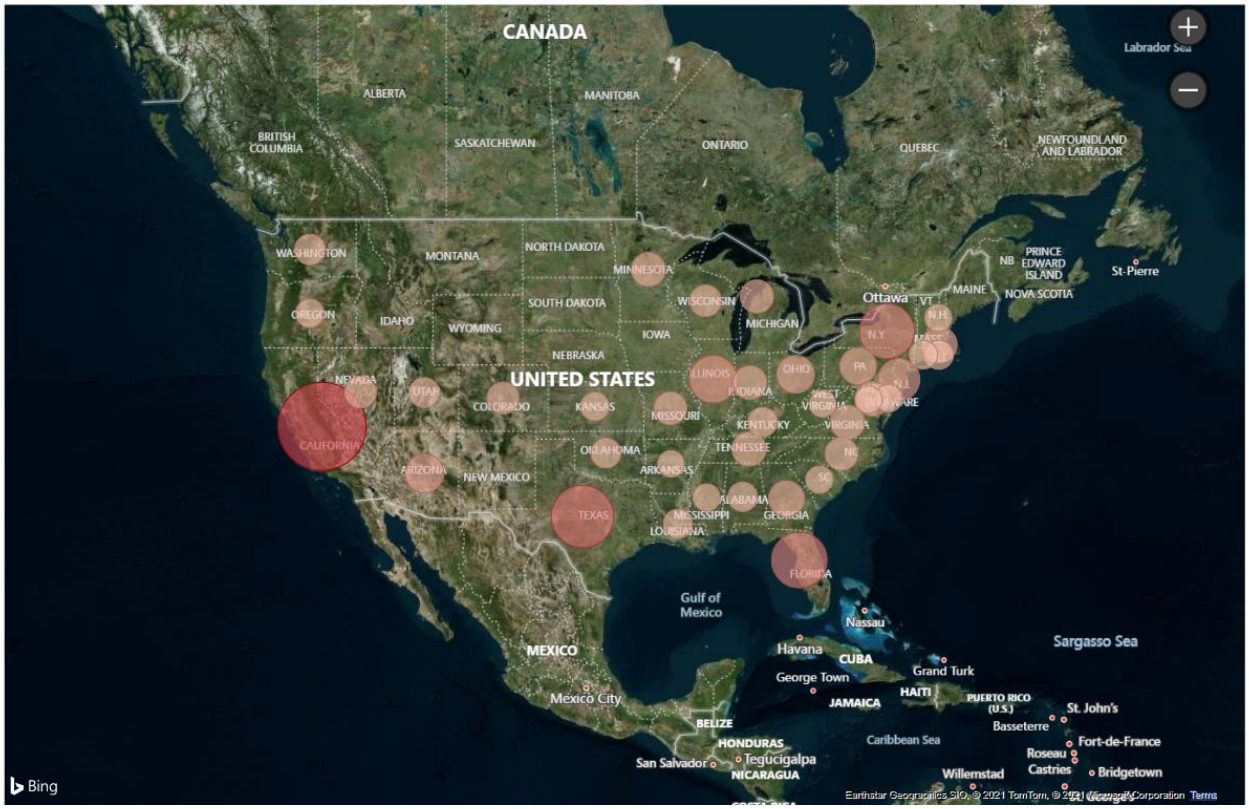
Slide 2 Cases vs Deaths Trends, Duration of Waves



For State New York

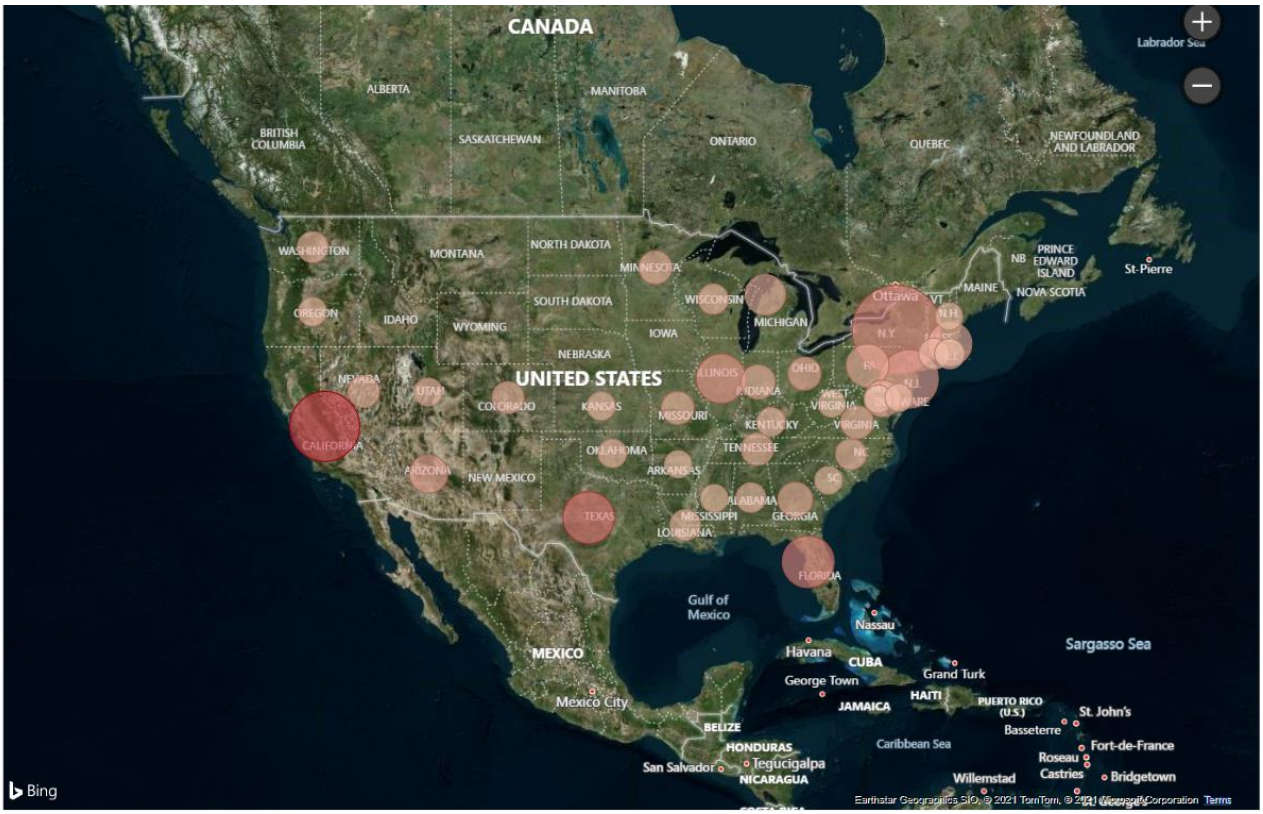


Slide 3 Geographical Representation of States with Maximum Number of Cases during the data time frame



State	Conf
CA	2194136
TX	1272288
FL	1036201
NY	1005591
IL	752588
NJ	546543
AZ	437405
OH	406758
PA	364711
GA	338603
MA	329802
MD	280553
TN	276914
VA	275452
MN	265054
IN	260870
MI	232191
NC	225187
MO	225149
NV	196134
CO	193085
WI	187261
WA	143492
UT	125488
OK	117521
KY	103478
AL	102935
LA	92746
RI	91678
CT	69881
KS	69771
OR	61681
DE	38955

Slide 4 Geographical Representation of States with Maximum Number of Deaths during the data time frame



State	Deaths
NY	36594
CA	25183
NJ	18420
FL	14606
TX	14512
IL	13101
MA	9856
PA	9371
MI	7771
AZ	6688
MD	5640
GA	4477
IN	4169
OH	3580
MN	3522
VA	3434
MO	3274
TN	3107
CO	2816
NV	2762
CT	2393
WA	2181
NC	2138
LA	2106
RI	1913
WI	1863
AL	1489
KY	885
OK	845
DC	842
OR	808
KS	795
UT	630