COVID VACCINATION & DEATH ANALYSIS

Big Data Analytics Symposium - Fall 2021

HEMANT PALAVALI AND RAJAN CHATURVEDI

Abstract

- Although a safe and effective vaccine holds the greatest promise for resolving COVID-19 pandemic, hesitancy to accept vaccine remains common

- With our analysis, we try to prove how effective or derogative vaccines are

- We used Hive, Hadoop, and Google Colab to perform this analysis.

Motivation

- This pandemic has haunted us for too long and having finally had a promising breakthrough in containing it, it is important that we analyze how effective it has been until now.

- This analysis can be used by government bodies as well as general public in deciding whether or not vaccinations are effective (spoiler alert: they are)

Goodness

- To corroborate the accuracy of our analysis, we analyzed results against a more trusted source, <u>our world in data</u>

- This source is used by Google, NY Times among other corporate giants for their covid data needs

- Our results closely matched with the analysis on this repository.

Data Sources

Name: COVID-19 Vaccinations in the United States, County
Description: This dataset is provided by Centre for Disease Control (CDC), which is a government body. It provides data on number of vaccinations given out, and is updated every week
Size: 500 MB

Name: United States COVID-19 Cases, Deaths, and Laboratory Testing (NAATs) by State, Territory, and Jurisdiction
Description: Managed by CDC, provides detailed data on number of confirmed COVID deaths per day
Size: 100 MB

DATA SAMPLE 1

Table 1

Date	Recip_State	Series_Complete_Pop_Pct	Series_Complete_Yes	Completeness_pct	Administered_Dose1_Recip	Administered_Dose1_Pop_Pct	Administered_Dose1_Recip_12Plus	Administered_Dose1_Recip_12PlusPop_Pct	Administered_Dose1_Recip_18Plus	Administered_Dose1_Recip_18PlusPop_Pct	Administered_Dose1_Recip_65Plus	Administered_Dose1_Recip_65PlusPop_Pct	SVI_CTGY	Series_Complete_Pop_Pct_SVI	Series_Complete_12PlusPop_Pct_SVI	Series_Complete_18PlusPop_Pct_SVI	Series_Complete_65PlusPop_Pct_SVI	Metro_status	Series_Pop_y
12/08/2021	OH	39.8	7435	98.6	7943	42.5	7897	49.6	7569	53	2926	82.2	B	6	7	8	8	Non-metro	
12/08/2021	OR	51.2	5940	97.8	6624	57.1	6554	68.8	6074	72.1	1558	80.1	D	16	16	16	16	Non-metro	
12/08/2021	AR	52.1	9117	90.4	10558	60.4	10434	70.6	9710	72	2993	84.3	D	16	16	16	16	Non-metro	
12/08/2021	WV	69.9	14248	76	17566	86.1	17534	98.5	16563	99.9	6276	99.9	C	12	12	12	12	Non-metro	
12/08/2021	MN	64.1	13508	94.5	14710	69.8	14326	81.3	13413	84.5	5077	99.9	A	4	4	4	4	Metro	
12/08/2021	MN	48.4	7346	94.5	7915	52.2	7848	61.3	7422	64.6	2473	89.9	A	3	4	4	4	Non-metro	
12/08/2021	IA	43.5	3355	97.4	3596	46.7	3568	54.0	3434	57.9	1350	86.3	B	7	8	8	8	Non-metro	
12/08/2021	TX	35.1	19847	99.1	22497	39.8	22397	46.3	21585	49.5	8321	72.4	C	10	11	11	11	Non-metro	
12/08/2021	CT	74.9	706760	98.3	849318	90	823244	99.9	751867	99.9	173330	99.9	B	8	8	8	8	Metro	
12/08/2021	OK	33.4	3704	93.9	4183	37.7	4163	43.7	4039	46.6	1657	63.2	D	14	14	15	15	Non-metro	
12/08/2021	TX	38.2	33729	99.1	39400	44.7	39080	54.1	36183	56.2	8516	74.7	D	14	15	15	15	Metro	
12/08/2021	ID	34.9	295	97.6	330	39.1			329	49.9	82	50	D	14	15	15	14	Non-metro	
12/08/2021	TX	51.1	26366	99.1	29930	58	29567	67.1	27581	69.3	7518	85.9	C	12	12	12	12	Metro	
12/08/2021	SC	32.1	18367	93.9	21240	37.1	21113	43.3	20170	45.6	6824	69.6	C	10	10	10	11	Non-metro	
12/08/2021	DC	62.5	441120	95.3	566421	80.3	552096	90.5	525289	90.9	90060	99.9	C	12	12	12	12	Metro	
12/08/2021	KS	56.9	69619	93.6	92236	75.4	89386	83.0	83650	83.5	27216	99.9	A	4	4	4	4	Metro	
12/08/2021	IN	0	46129	98.7	59438	0	58173	0.0	56133	0	13830	0							
12/08/2021	MO	31.8	2653	91.2	3138	37.6	3113	44.0	3008	47.5	1331	69.9	C	10	10	11	11	Non-metro	
12/08/2021	KS	45.8	5446	93.6	6171	51.9	6089	59.1	5761	61.4	2258	79.7	A	3	4	4	4	Non-metro	
12/08/2021	OH	39.8	7435	98.6	7943	42.5	7897	49.6	7569	53	2926	82.2	B	6	7	8	8	Non-metro	

DATA SAMPLE 2

United_States_COVID-19_Cases_and_Deaths_by_State_over_Time (2)

submission_date	state	tot_cases	conf_cases	prob_cases	new_case	pnew_case	tot_death	conf_death	prob_death	new_death	pnew_death	created_at	consent_cases	consent_deaths
02/12/2021	UT	359641	359641	0	1060	0	1785	1729	56	11	2	02/13/2021 02:50:08 PM	Agree	Agree
03/01/2021	CO	438745	411869	26876	677	60	5952	5218	734	1	0	03/01/2021 12:00:00 AM	Agree	Agree
08/22/2020	AR	56199			547	0	674			11	0	08/23/2020 02:15:28 PM	Not agree	Not agree
07/12/2020	FL	277072			12161	43	5143			156	0	07/12/2020 12:00:00 AM	Not agree	Not agree
06/25/2021	MP	183	183	0	0	0	2	2	0	0	0	06/27/2021 12:00:00 AM	Agree	Agree
05/22/2021	MA	704796	659246	45550	451	46	17818	17458	360	5	0	05/23/2021 01:37:59 PM	Agree	Agree
10/28/2020	PR	35112	34791	321	619	1	805	624	181	3	0	10/28/2020 12:00:00 AM	Agree	Agree
08/01/2021	GA	1187107	937515	249592	3829	1144	21690	18725	2965	7	0	08/02/2021 02:03:57 PM	Agree	Agree
05/27/2021	NYC	948436	782257	166179	394	95	33203	28130	5073	6	0	05/25/2021 12:00:00 AM	Agree	Agree
09/14/2021	AS	0			0	0	0			0	0	09/15/2021 02:35:57 PM		
09/23/2021	NYC	1075197	881626	193571	1755	168	34153	28965	5188	20	-7	09/24/2021 01:56:18 PM	Agree	Agree
07/20/2020	GA	147804	143697	4107	2766	317	3176	3176	0	3	0	07/21/2020 10:50:17 AM	Agree	Agree
05/29/2020	CO	25968	23464	2504	346	29	1436	1129	307	15	1	05/29/2020 12:00:00 AM	Agree	Agree
03/19/2020	GA	493	493	0	115	0	13			7		03/26/2020 12:22:39 PM	Agree	Agree
04/13/2021	MA	662699	622662	40037	1598	197	17427	17077	350	8	0	04/15/2021 12:00:00 AM	Agree	Agree
03/07/2021	MP	146	146	0	2	0	2	2	0	0	0	03/09/2021 12:00:00 AM	Agree	Agree
10/23/2021	GA	1625399	1257310	368089	0	0	28519	24357	4162	0	0	10/25/2021 03:25:56 PM	Agree	Agree
04/16/2020	AK	300			7		9			0		04/16/2020 12:00:00 AM	N/A	N/A
08/01/2021	AS	0			0	0	0			0	0	08/02/2021 02:03:57 PM		
06/07/2021	PR	140304	123121	17183	9	4	2511	2157	354	4	0	06/07/2021 12:00:00 AM	Agree	Agree
06/17/2021	HI	35862			31	0	504			0	0	06/18/2021 01:48:26 PM	Not agree	Not agree
11/18/2021	AK	142203			517	0	812			0	0	11/19/2021 03:04:49 PM	N/A	N/A

Design Diagram
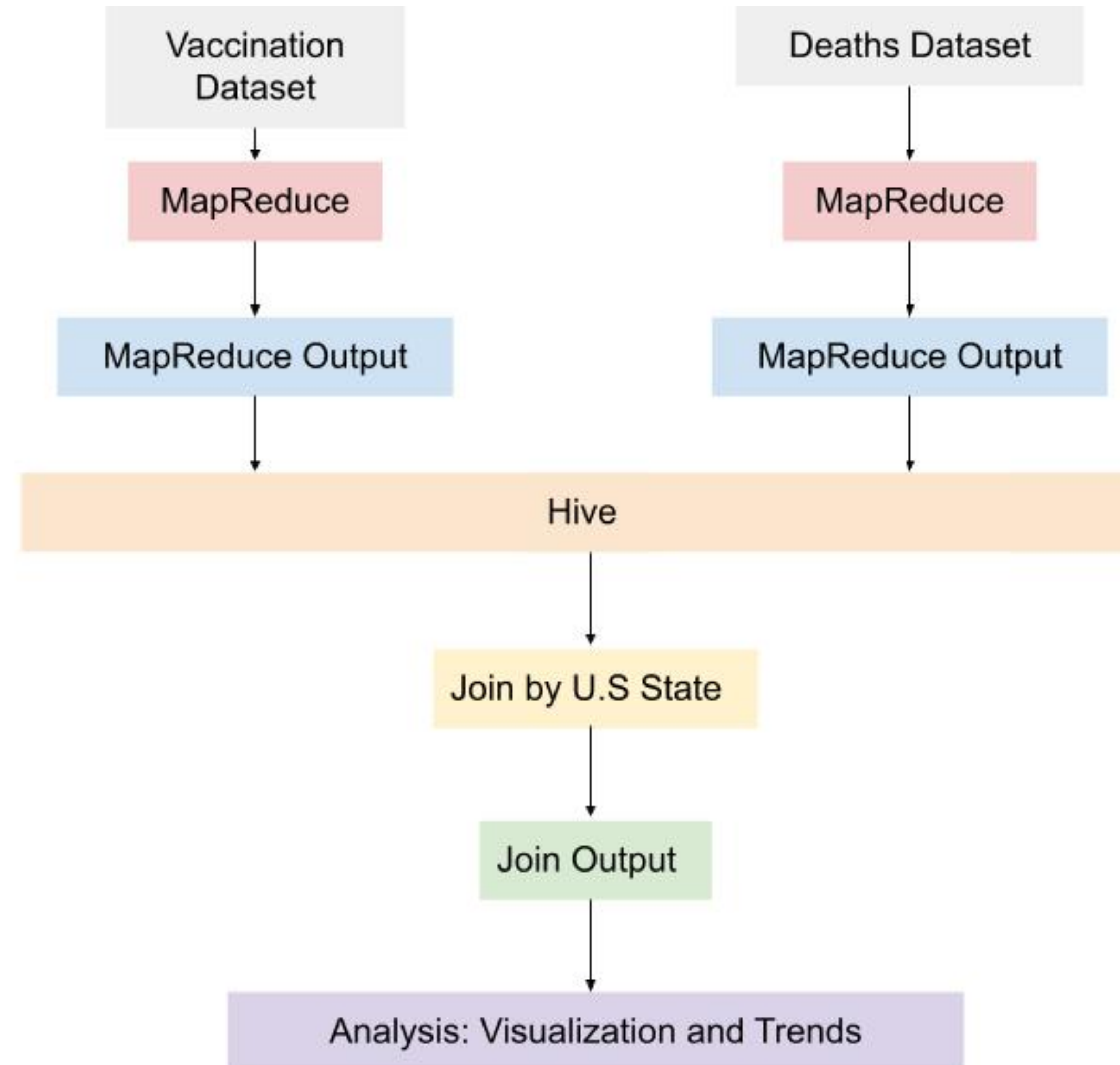
Code Challenge 1

```sql
SELECT * , SUM(people_fully_vaccinated)
OVER
    (ORDER BY vdate_c rows BETWEEN 50 preceeding AND current row)
FROM vcf_f
GROUP BY location
```

Code Challenge 2

- Visualizing the data turned out to be a bigger challenge than anticipated. Mapping state-wise data on the shape-file of the USA required the data consistent format across multiple files.

- This meant running multiple hive queries to get the data consistent across the board and helped us visualize the trends better.

Obstacles

- Cleaning the datasets

- Merging the results of our dataset

RESULTS

VACCINATION TREND - 2021
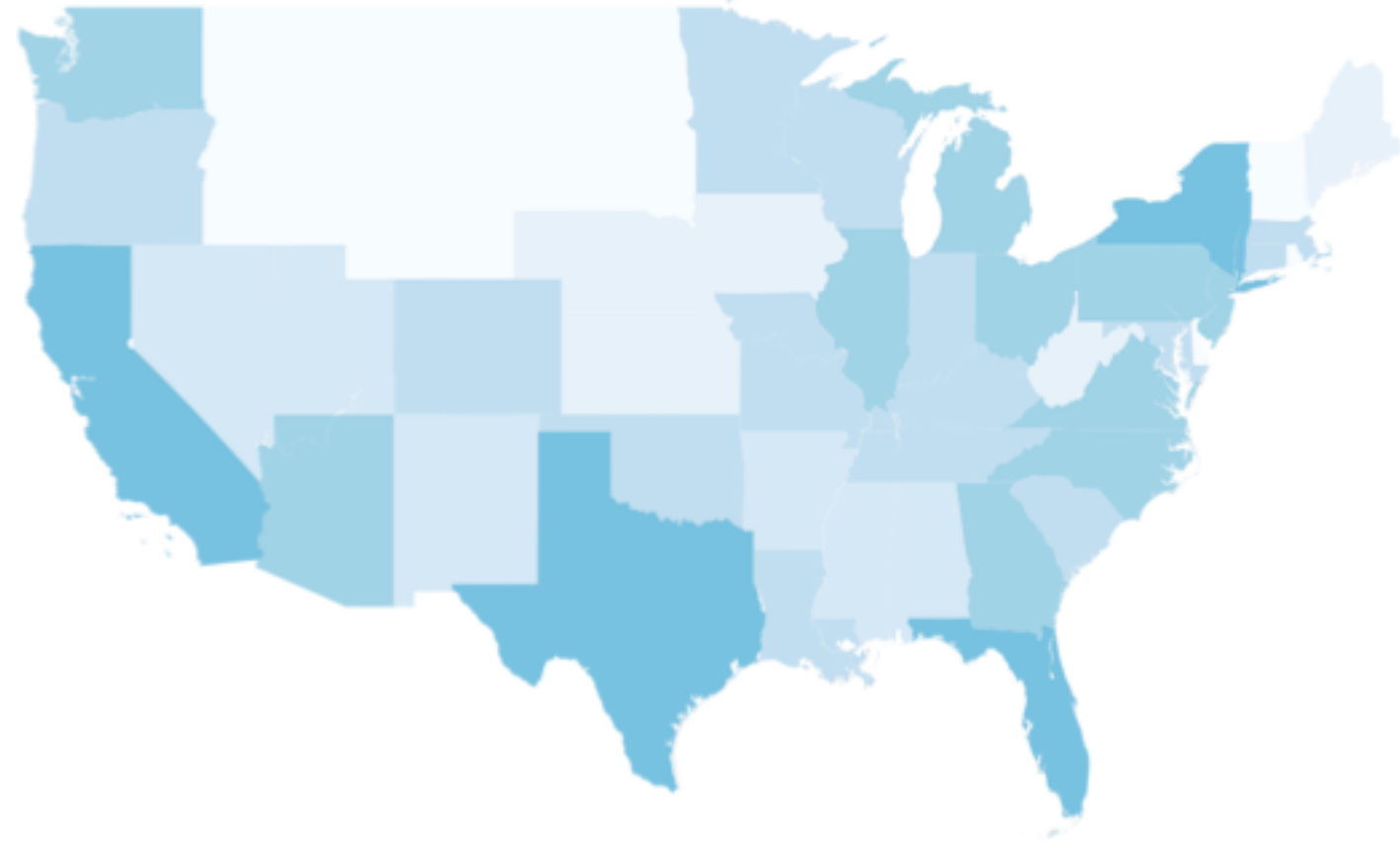
JANUARY ~2.7M
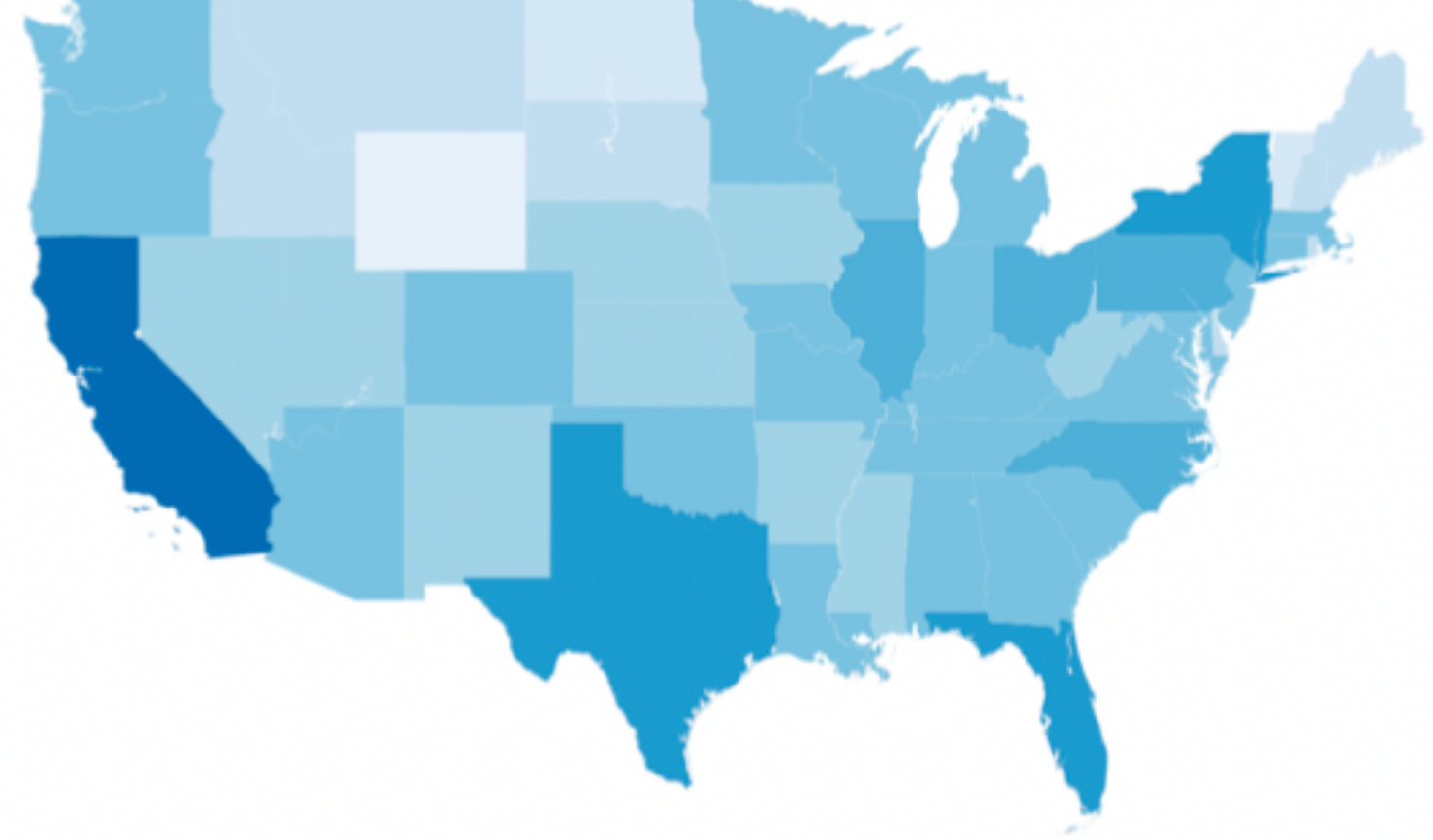
MARCH
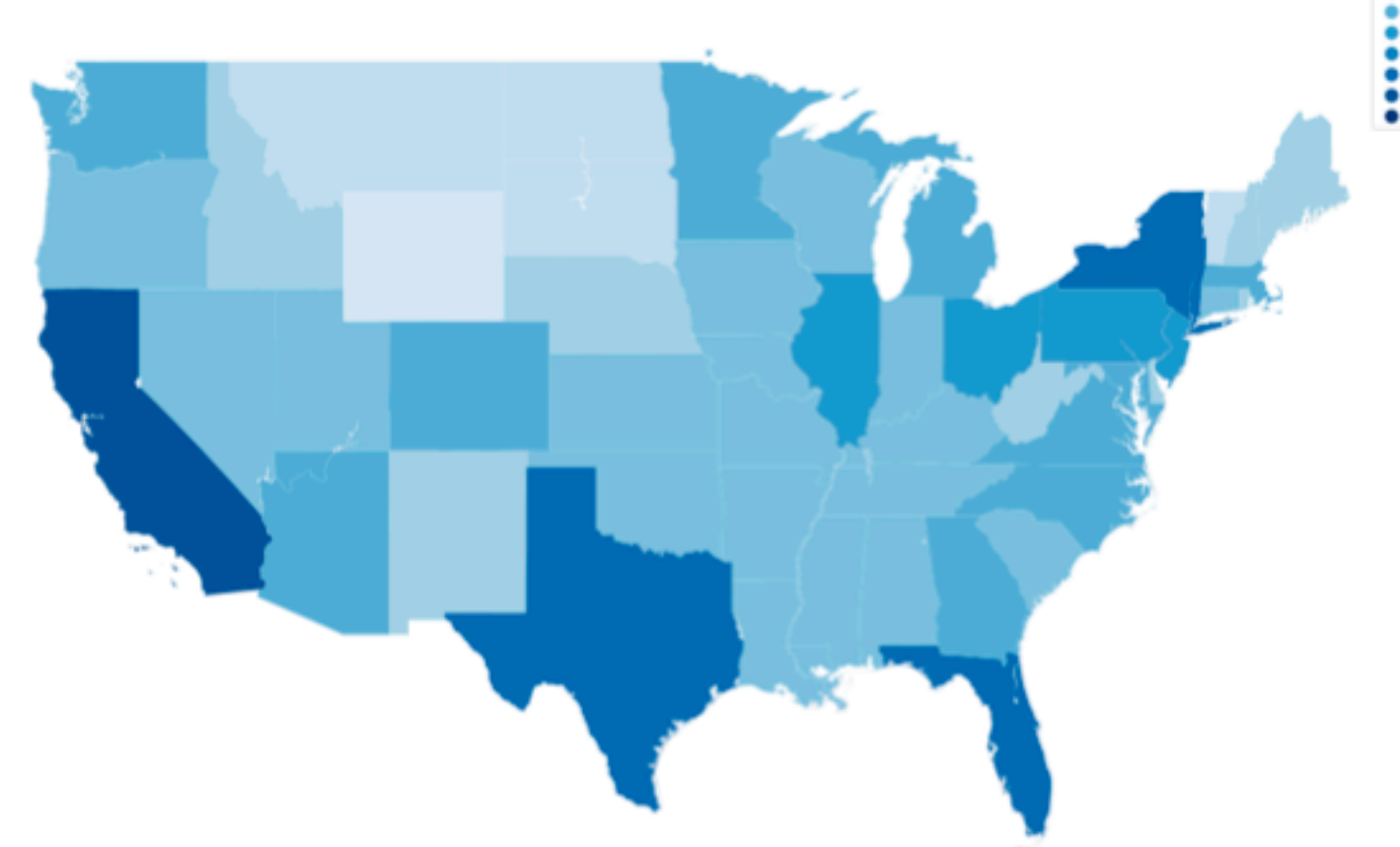
MAY
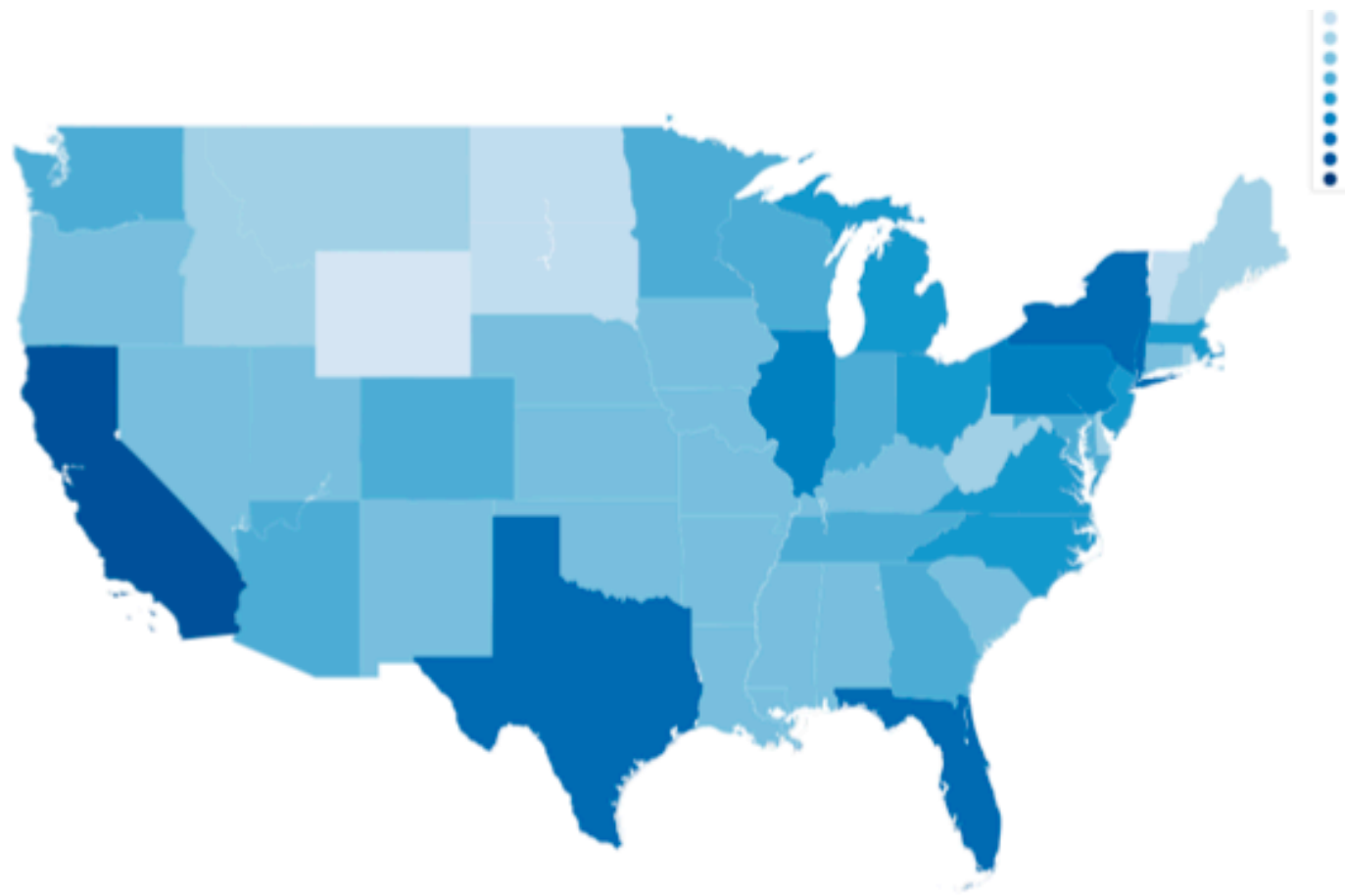
JULY
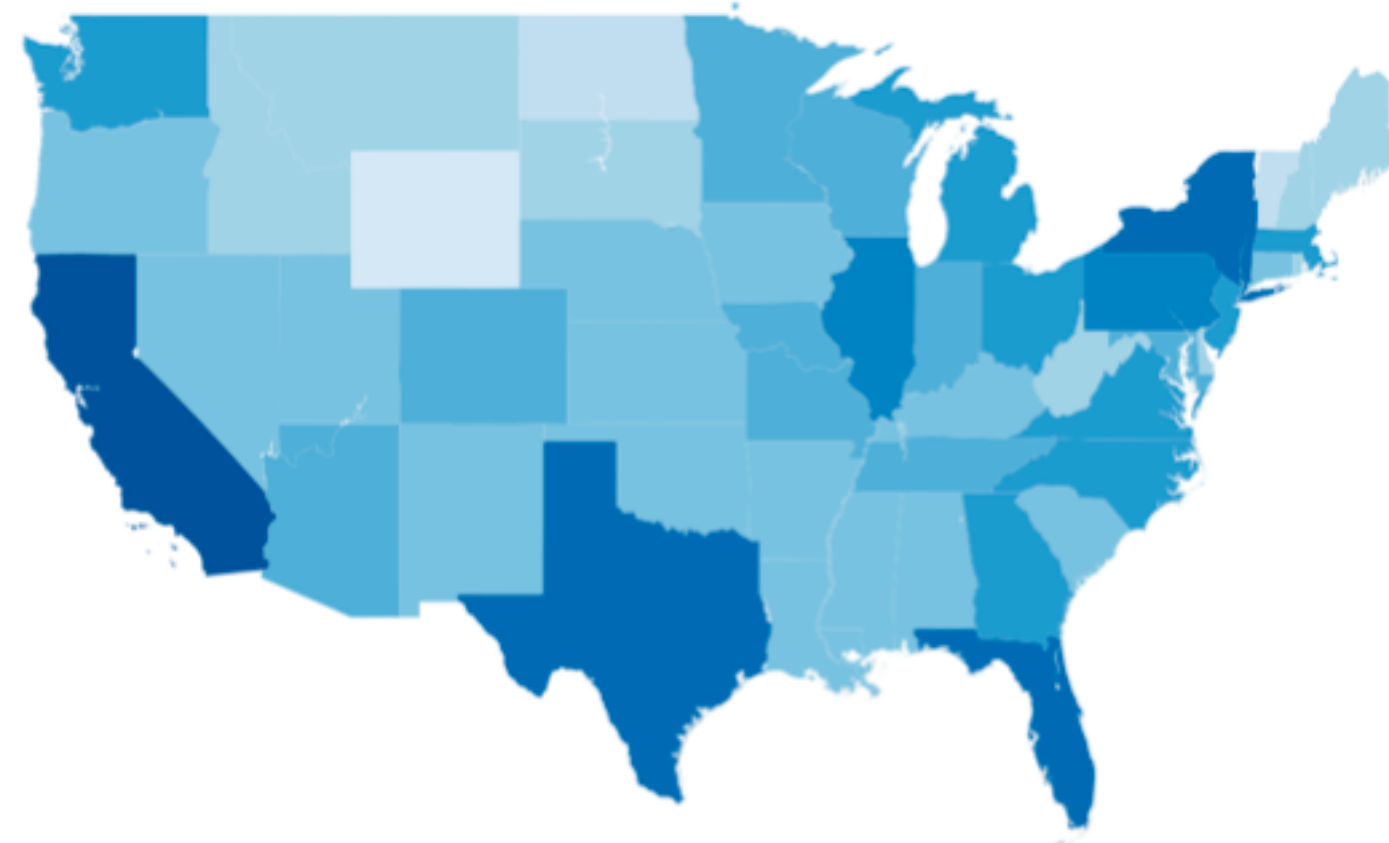
SEPTEMBER
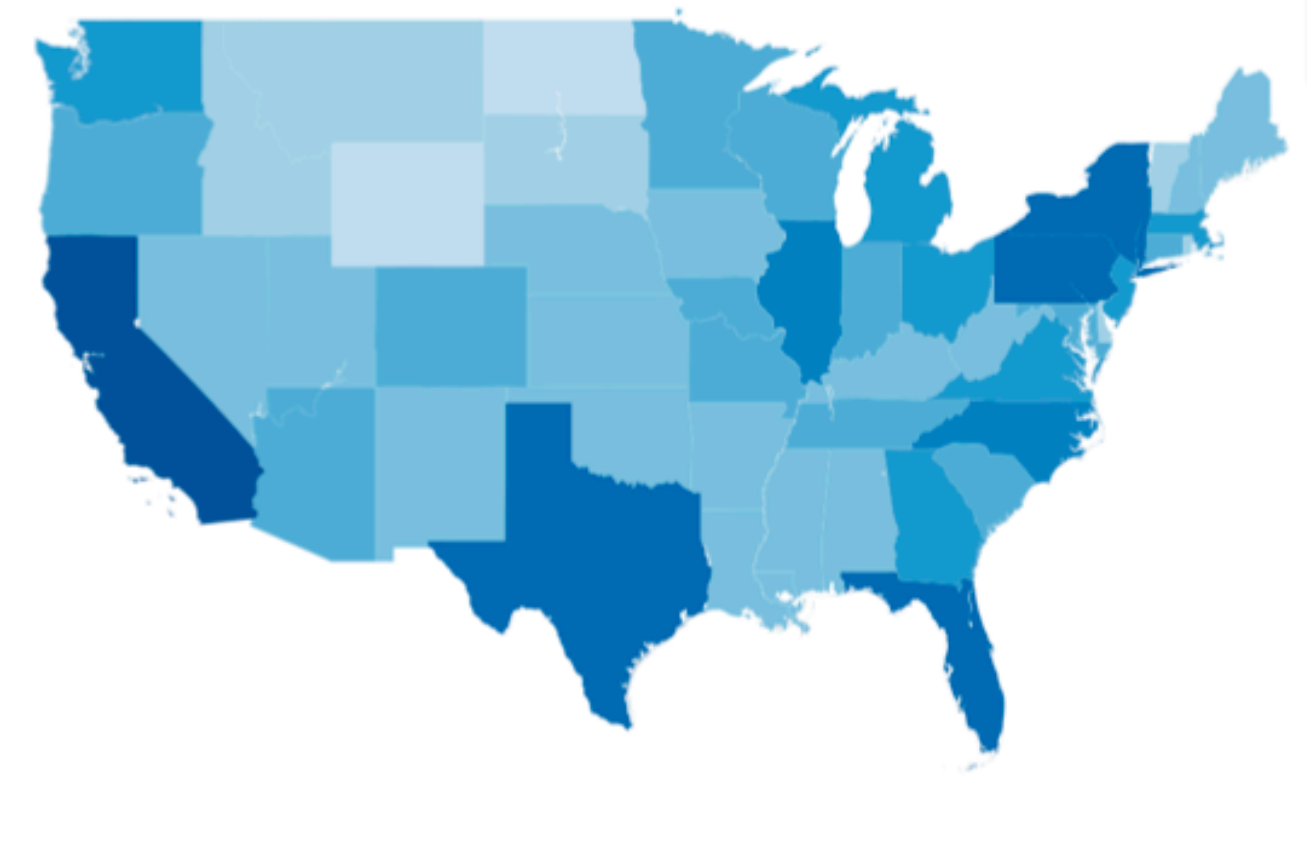
DECEMBER ~290M

COVID DAILY-DEATH TREND - 2020/21
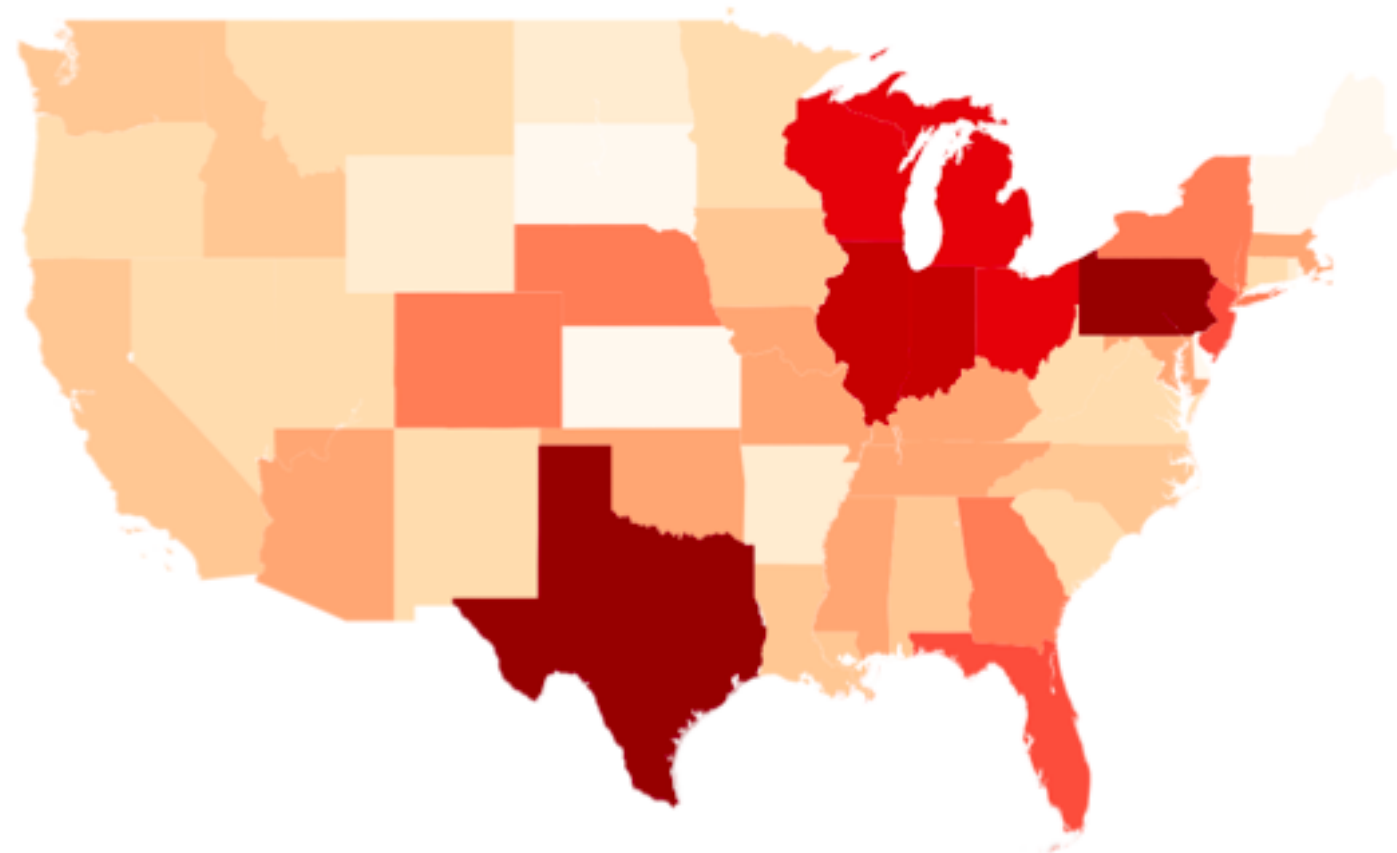
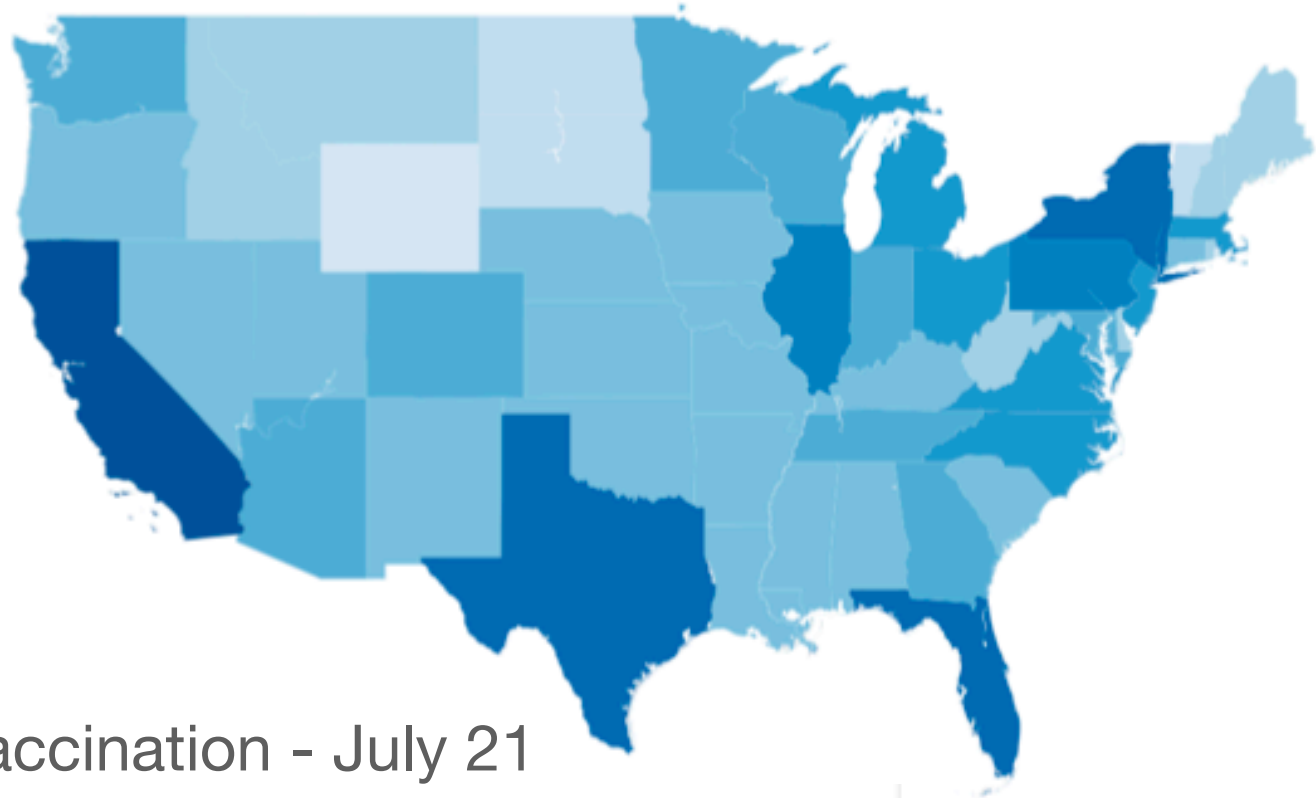DECEMBER - 20 ~3749
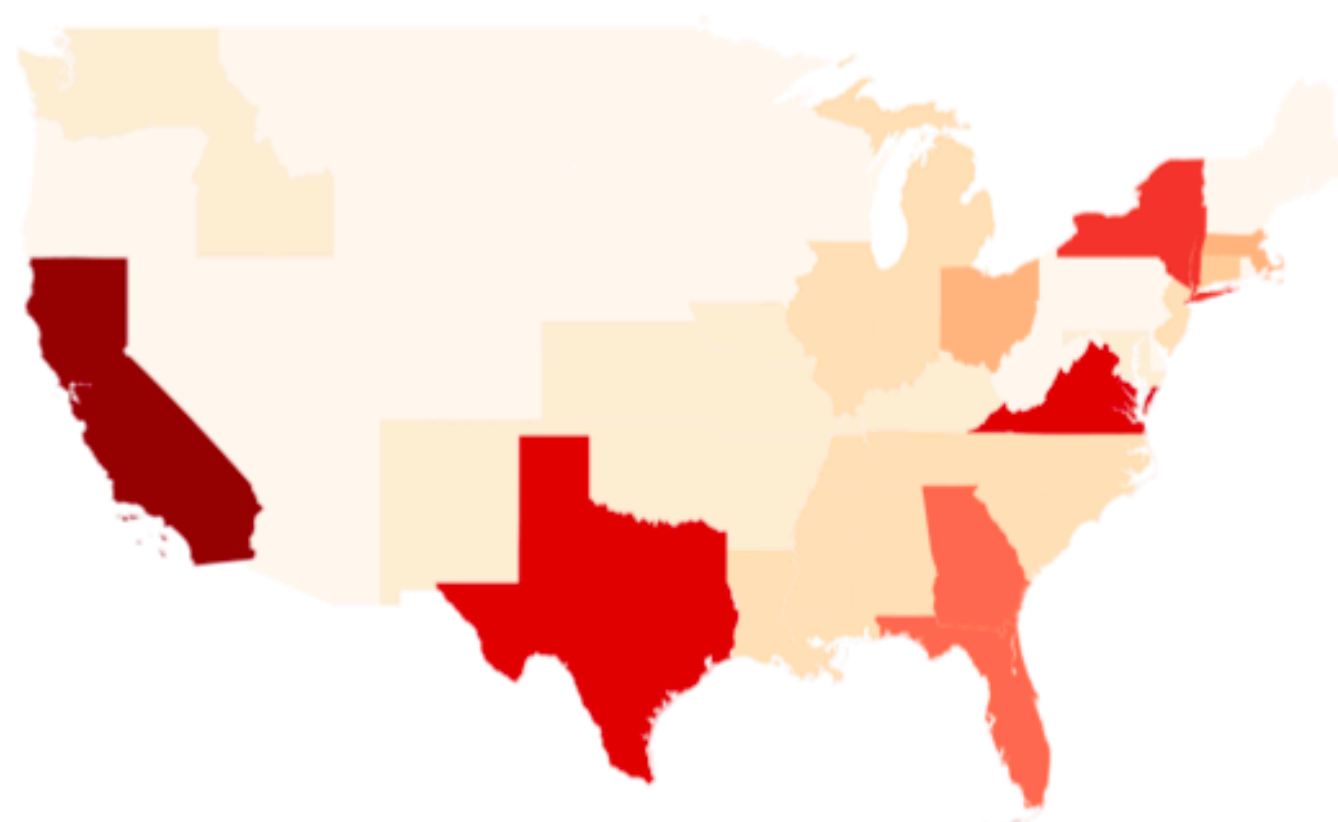
MARCH
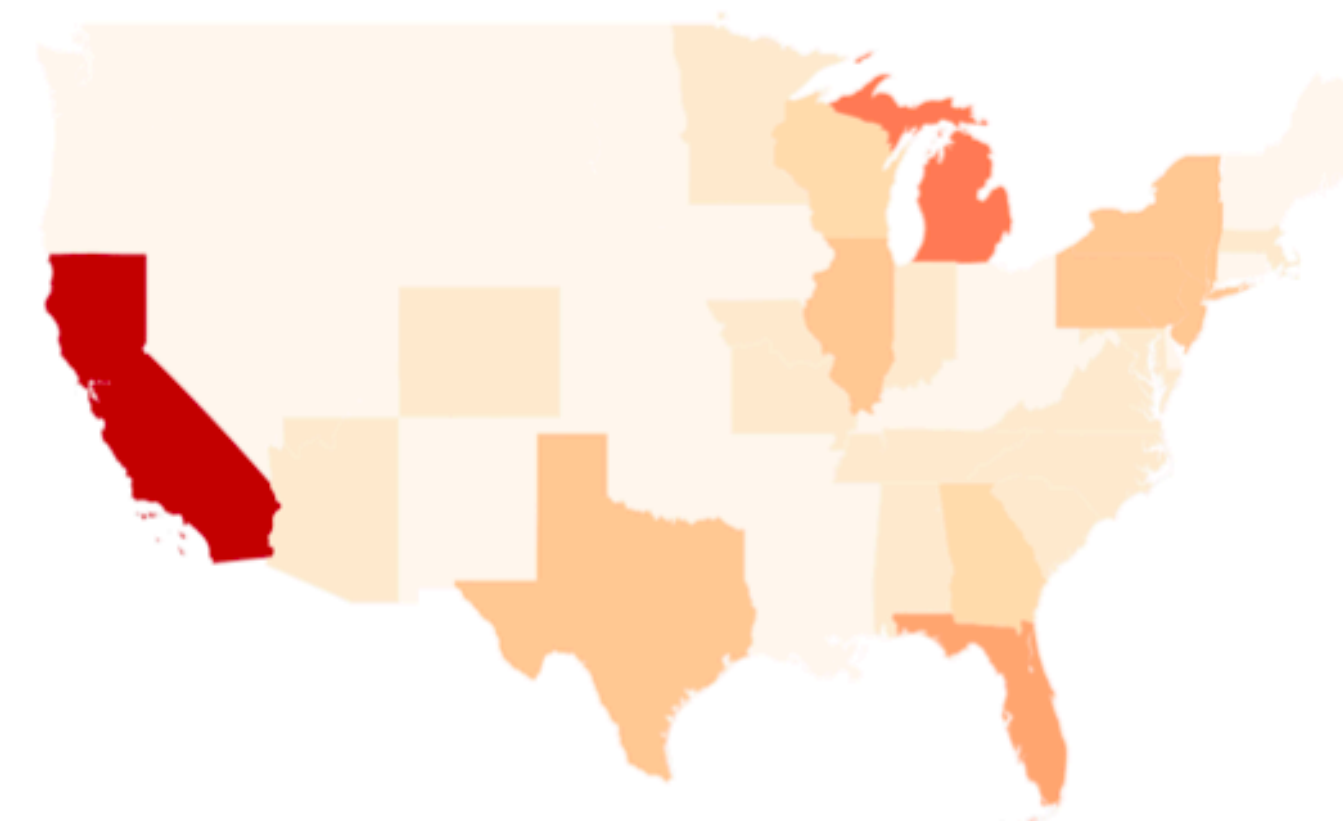
MAY

JULY ~332
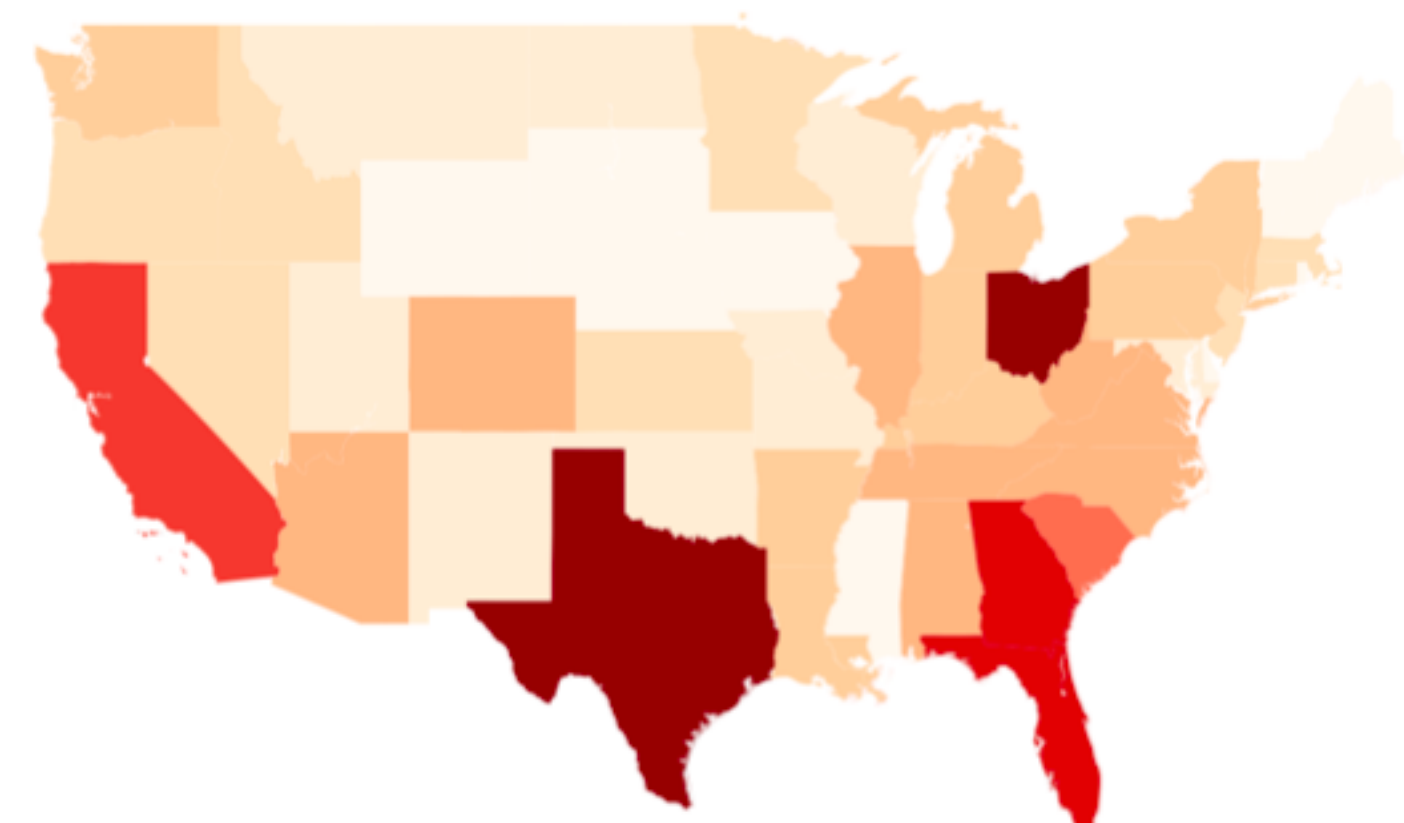
OCTOBER
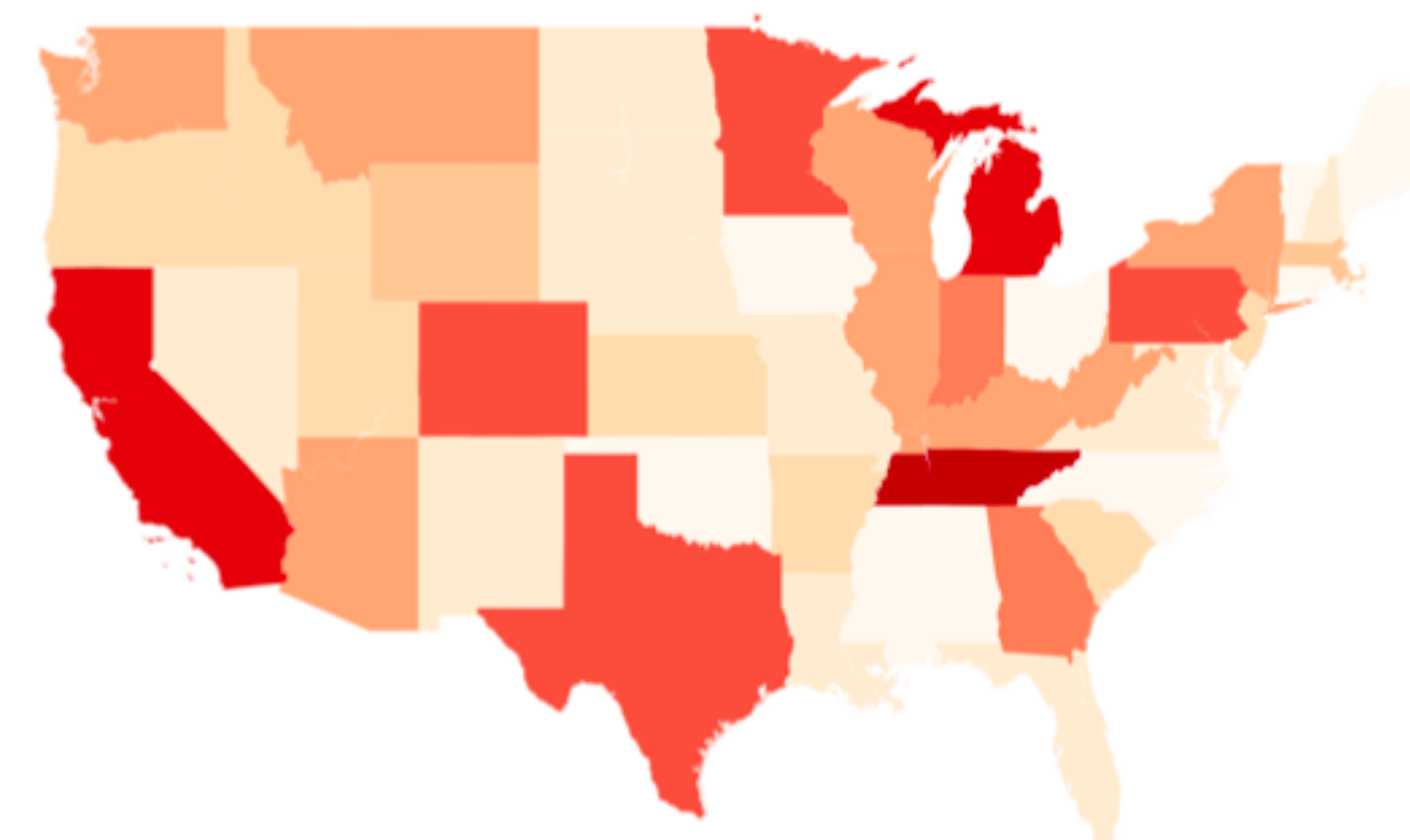
DECEMBER ~1851
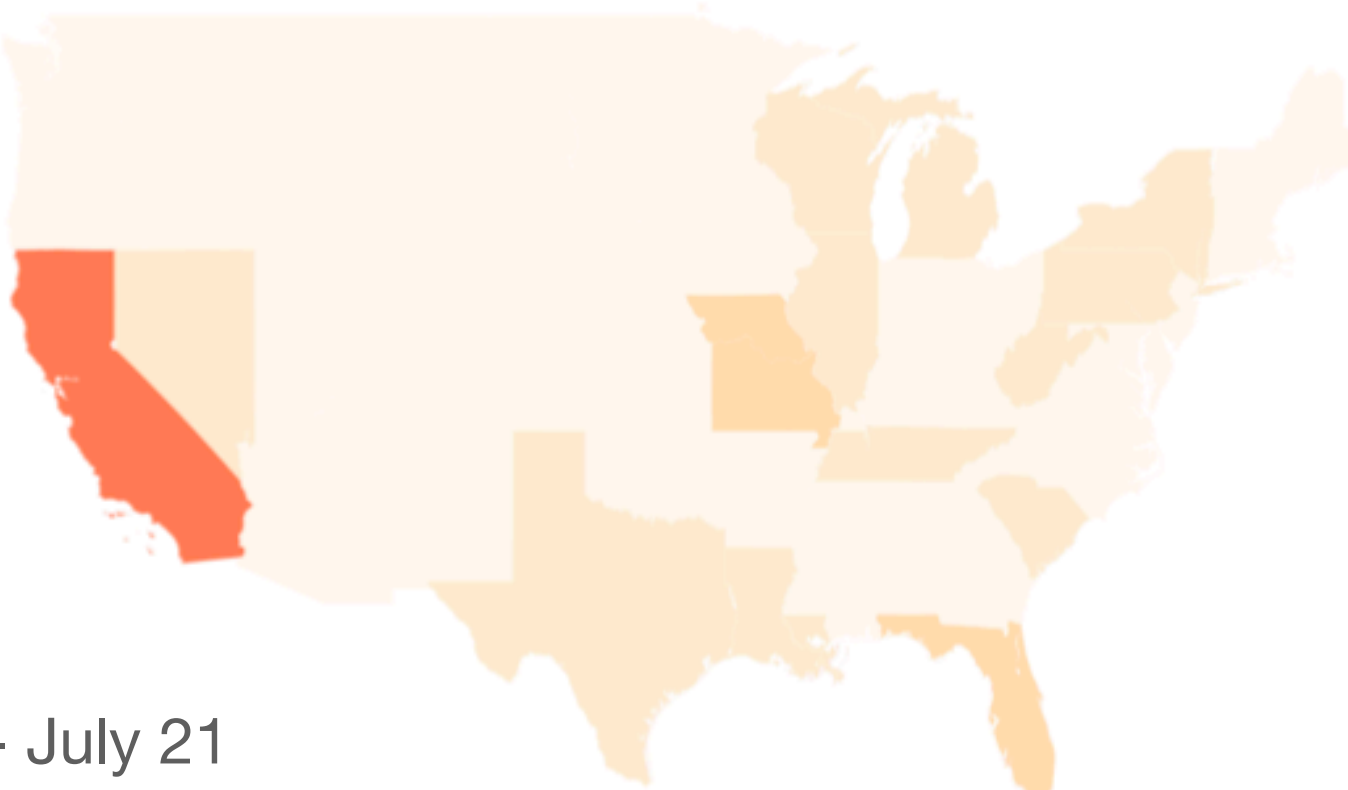
Vaccination / Deaths Relation - US States
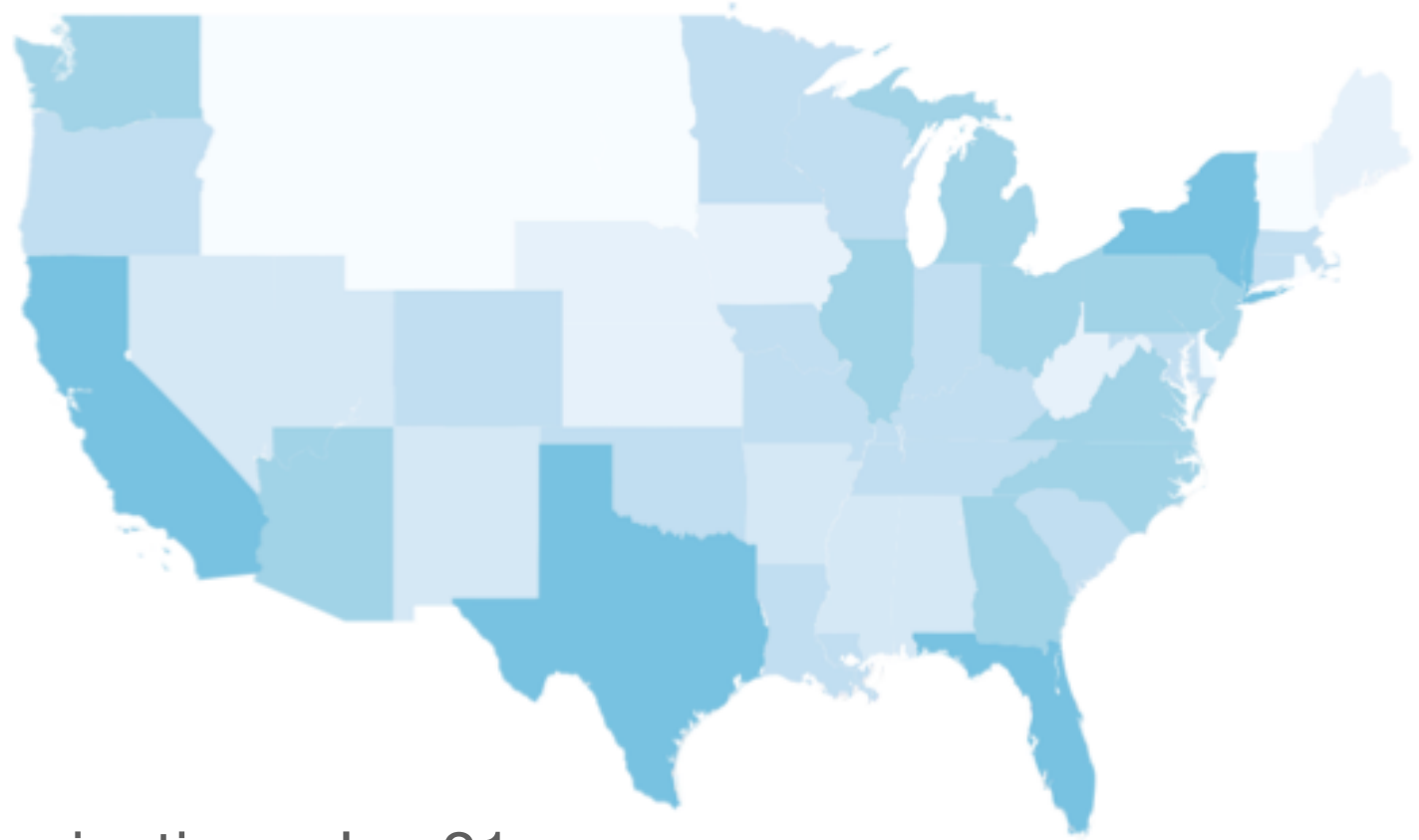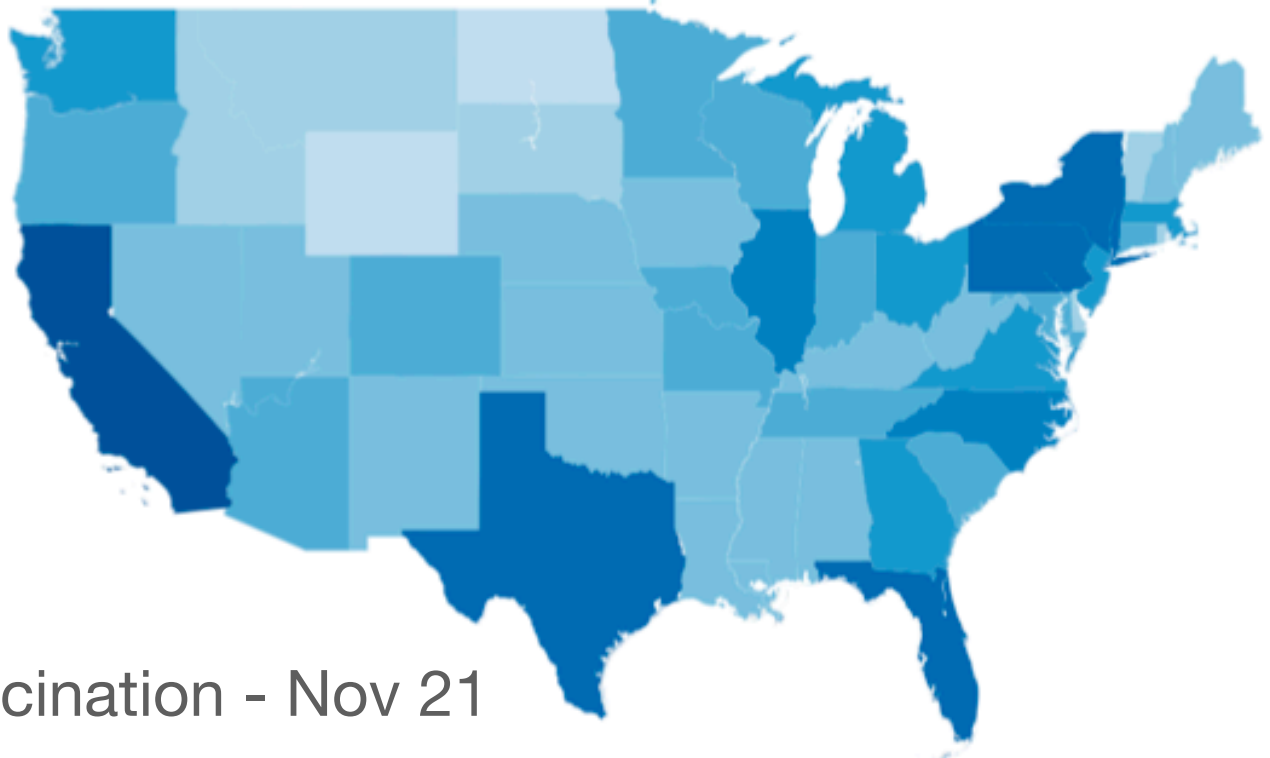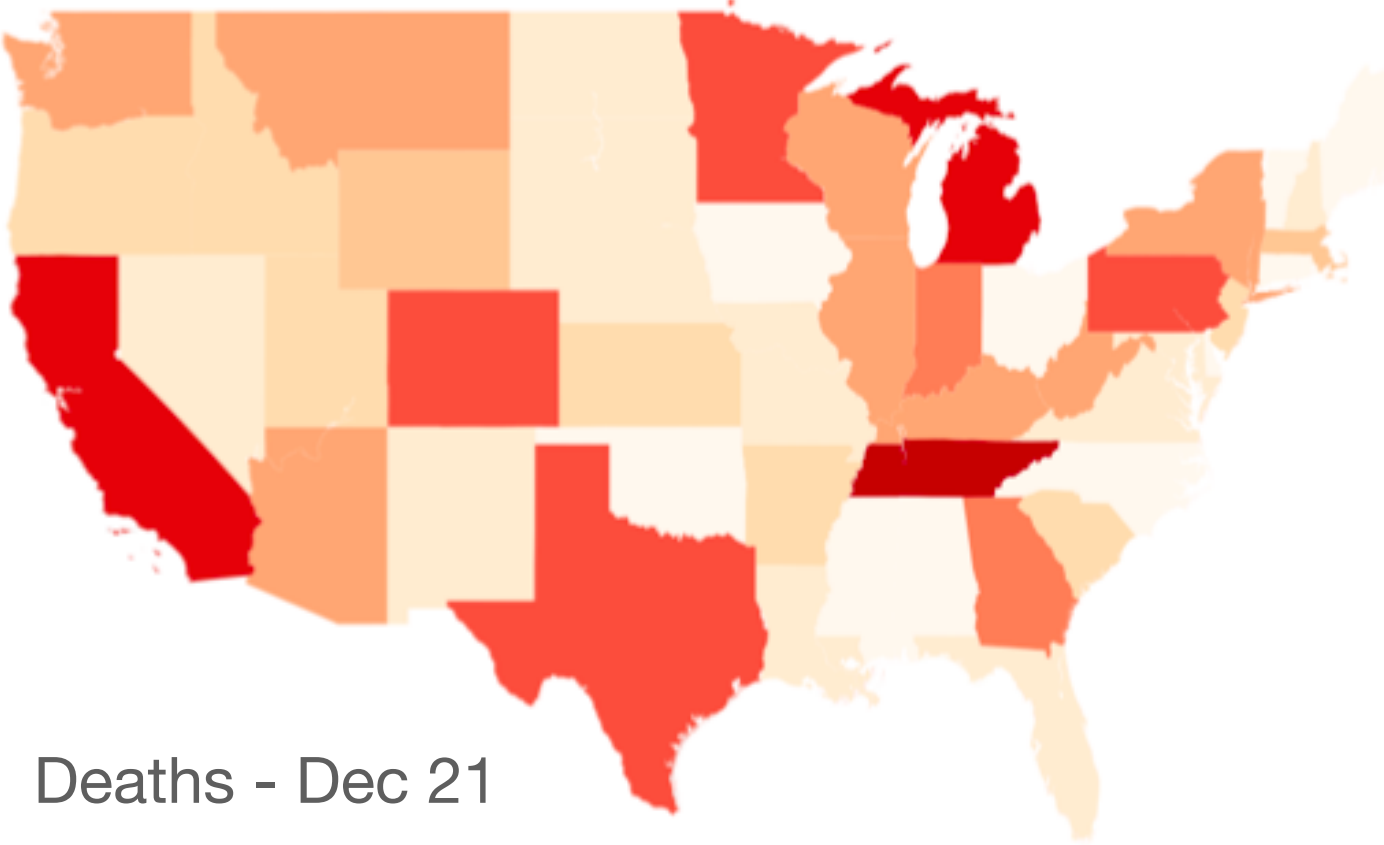

Vaccination - Jan 21
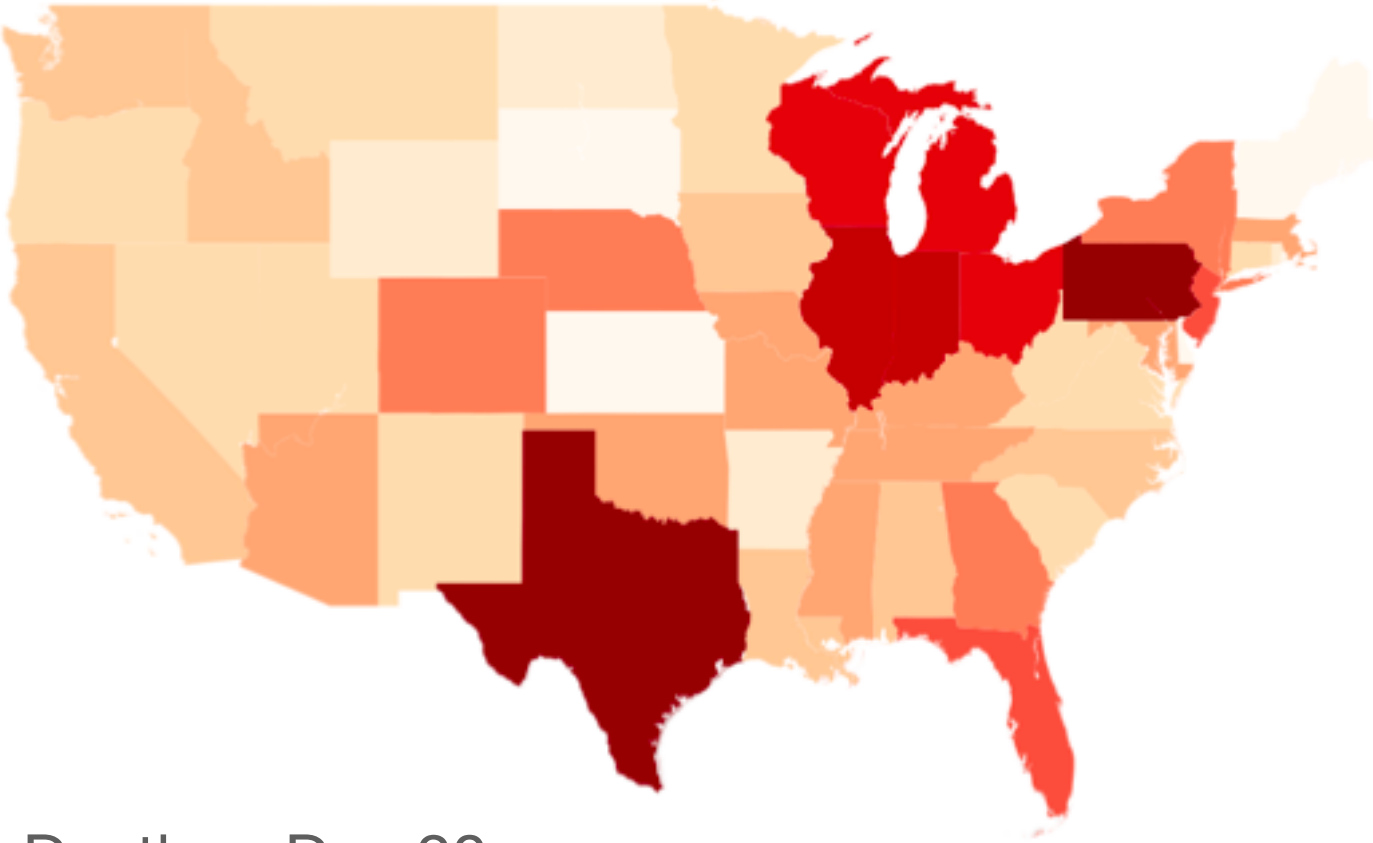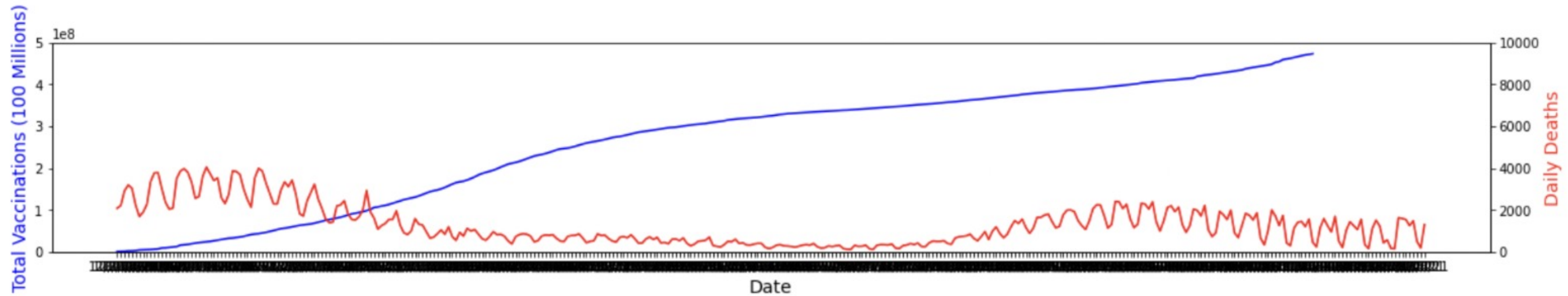
Deaths - Dec 20

Vaccination - July 21

Death - July 21

Vaccination - Nov 21

Deaths - Dec 21

Vaccination vs Deaths Relation - US

Summary

- Although vaccination works, the effect is transient

- North-west region of the US is lagging in the vaccination count

- Relation between COVID deaths and vaccinations is

$$deaths \propto \frac{1}{vaccination}$$

- Booster shots and other subsequent shots in future are going to be the key in fight against corona virus pandemic

Acknowledgements

- We would like to thank Professor Tang for giving us an opportunity to build this project

- We would also like to extend our gratitude to CDC and Oxford for making such excellent set of resources available for general audience, and HPC for all their support

References

- <u>Proceedings of the National Academy of Science of the USA</u>

- <u>Our World in Data</u>

- <u>CDC</u>

Thank you!

Oh! and get vaccinated.