Project-DATA604

By: TEAM 6

United States COVID-19 Cases and Deaths by State over Time

Team 6:

Chirag Jadav (chiragj1@umbc.edu)

Adishree Pandey (p108@umbc.edu)

Harshil Prajapati (h148@umbc.edu)

DATASET:

https://healthdata.gov/dataset/United-States-COVID-19-Cases-and-Deaths-by-State-o/hiyb-zgc2

Project-DATA604

By: TEAM 6

INDEX

Project-DATA604

By: TEAM 6

Problem Statement

CDC reports aggregate counts of COVID-19 cases and death numbers daily online. Data on the COVID-19 website and CDC's COVID Data Tracker are based on these most recent numbers reported by states, territories, and other jurisdictions. This data set of "United States COVID-19 Cases and Deaths by State over Time" combines this information. However, data are dependent on jurisdictions' timely and accurate reporting.

Separately, CDC also regularly reports provisional death certificate data from the National Vital Statistics System (NVSS) on data.cdc.gov. Details are described on the NCHS website. Reporting the number of deaths by using death certificates ultimately provides more complete information but is a longer process; therefore, these numbers will be less than the death counts on the COVID-19 website.

Approach

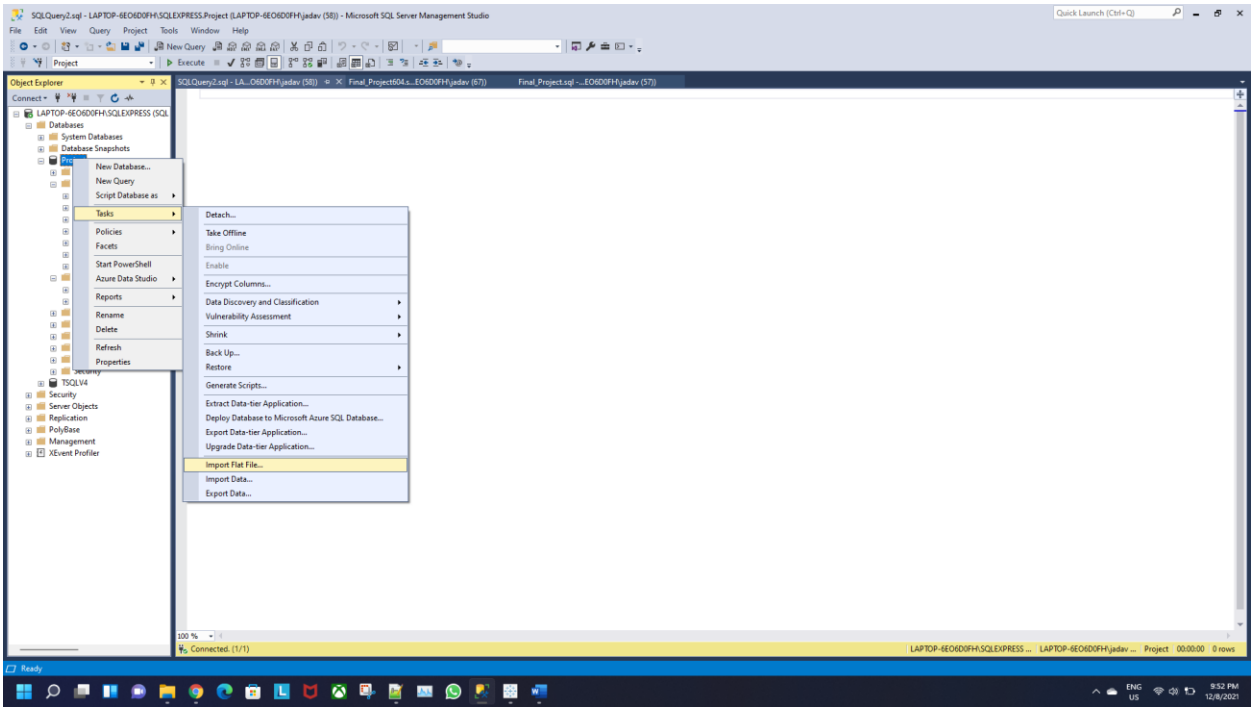
The idea of this project is to implement and visualize the COVID-19 cases, We have exported the US covid data and from the admin file we have segregated it further into three tables covering- US Covid Consent, US Covid Deaths, US Covid cases. After careful consideration of all tables and entities from the mentioned CSV files we have uploaded all the three CSV files on AWS S3 bucket under bucket named DATA604-Project and then created tables for each CS files.

The entities included in US Covid Consent are: submission_date, state, created_at, consent cases, consent_deaths. The entities included in US Covid Deaths are: submission_date, state, tot_death, conf_death, prob_death, new_death pnew_death. The entities included in US Covid cases are :submission_date, state tot_cases, conf_cases, prob_cases, new_case, pnew_case. We will have Serial Number as our Primary Key which will be used as a joining condition to display and visualize the data of US Covid-19. To perform visualization of data we will use Tableau tool.

Project-DATA604

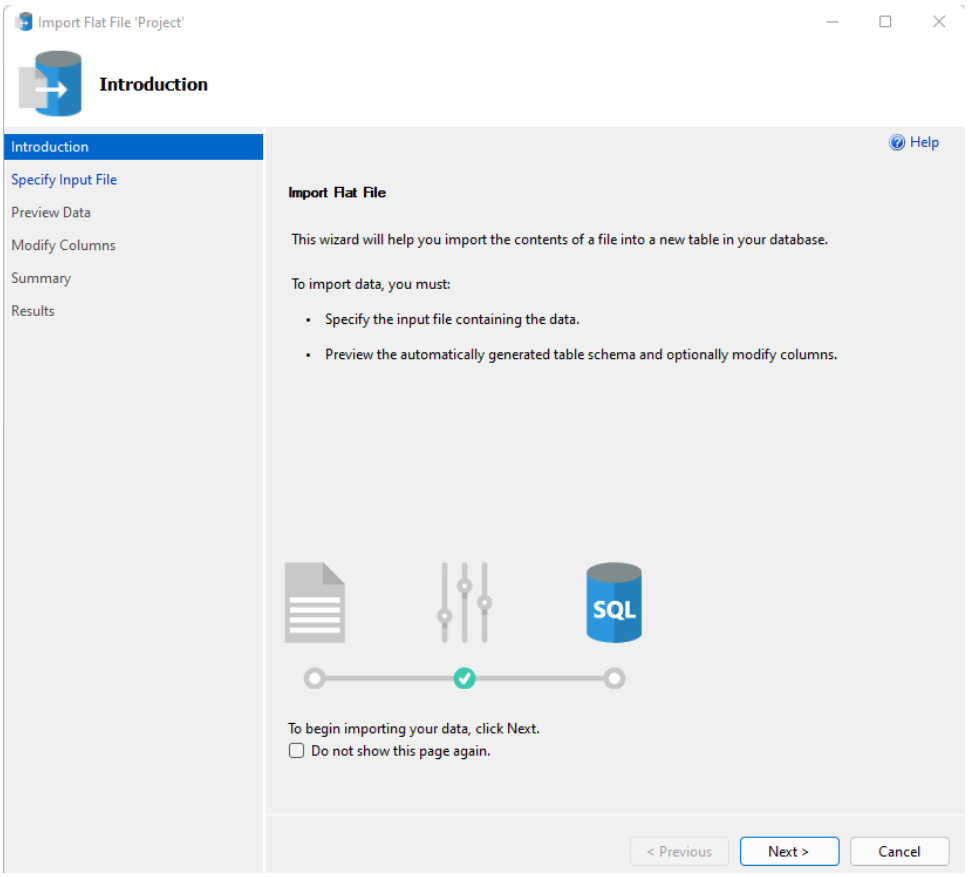
By: TEAM 6

Created the Database- "**Project**" on Microsoft SQL Server and importing manually the flat .csv files:



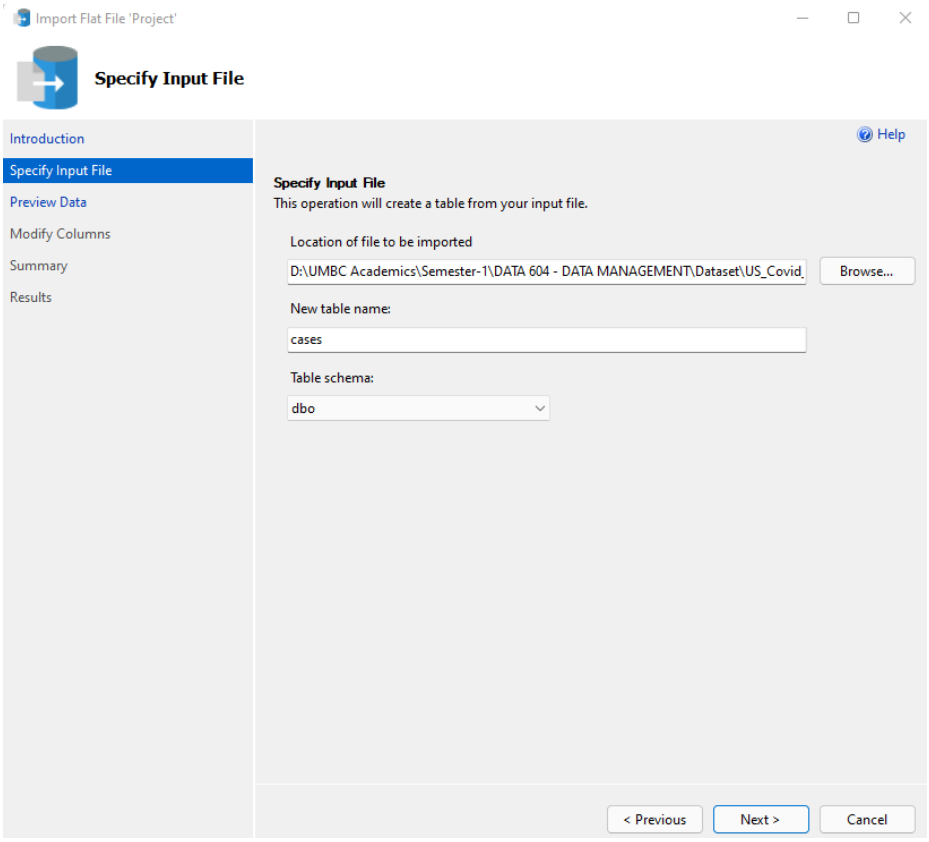
Project-DATA604

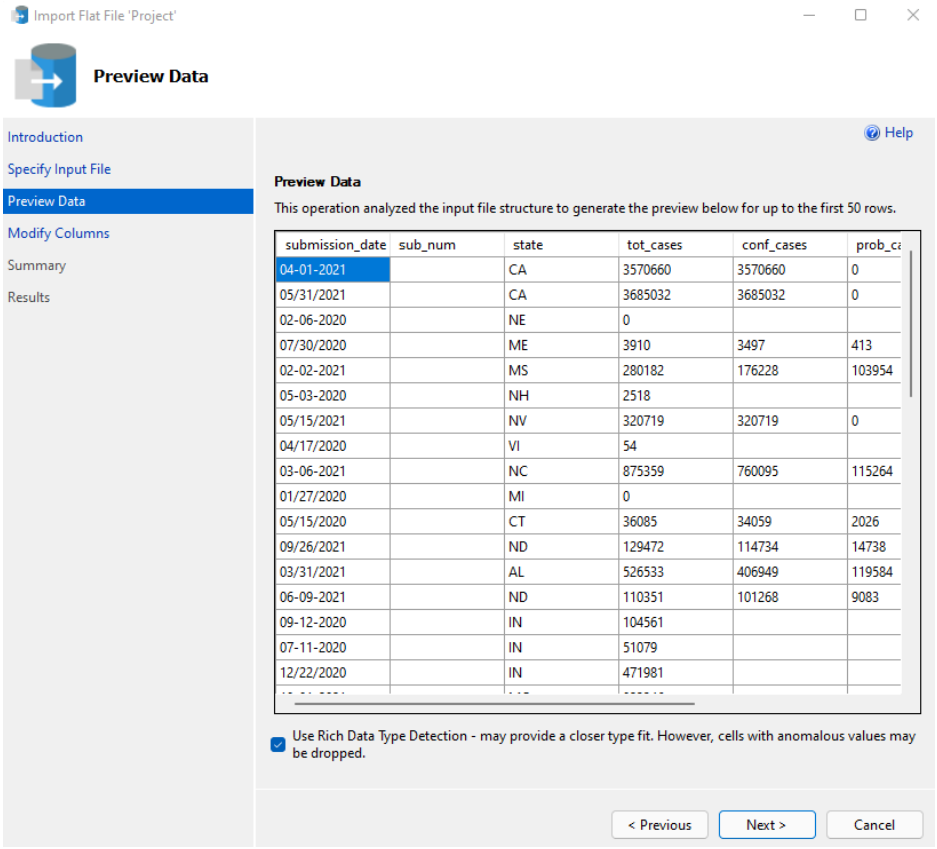
By: TEAM 6



Project-DATA604

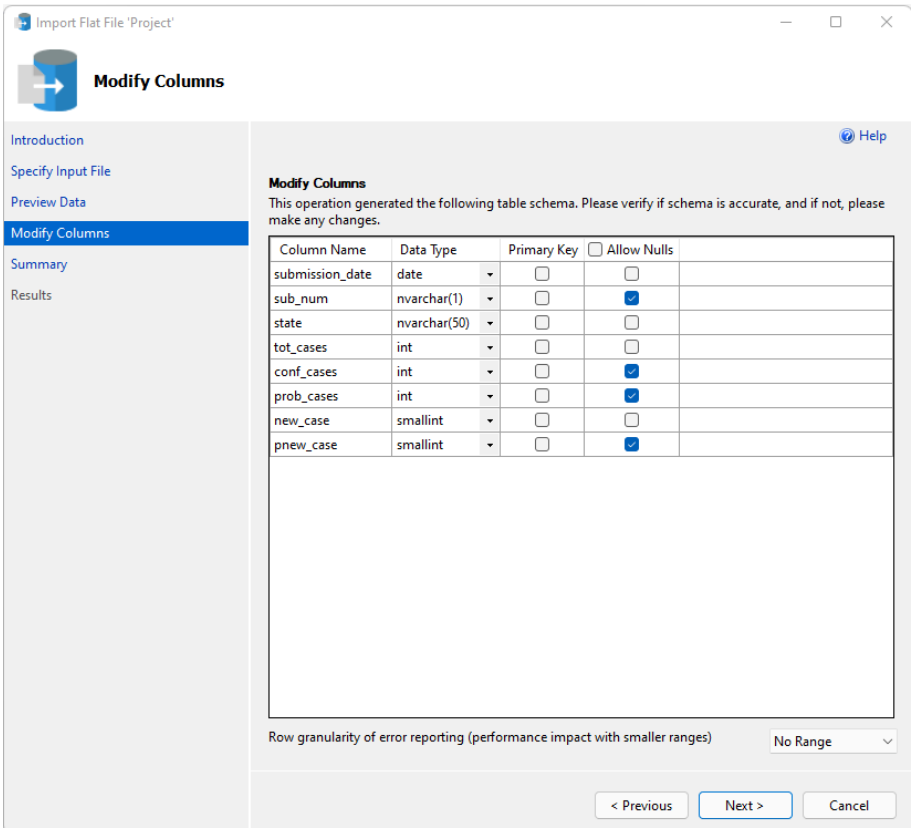
By: TEAM 6

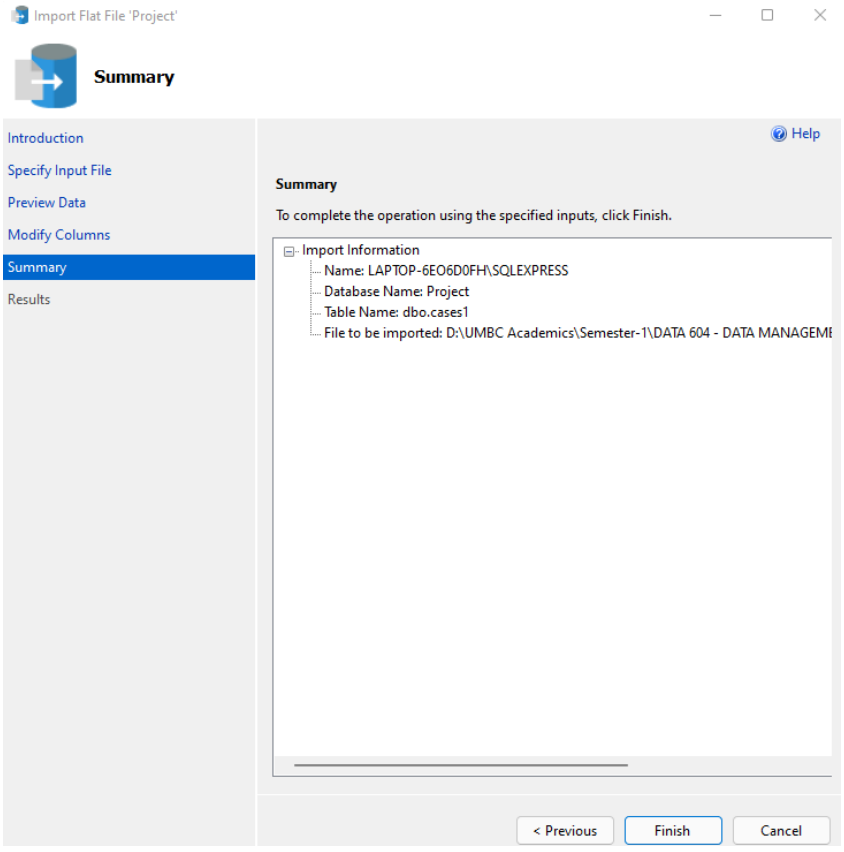




Project-DATA604

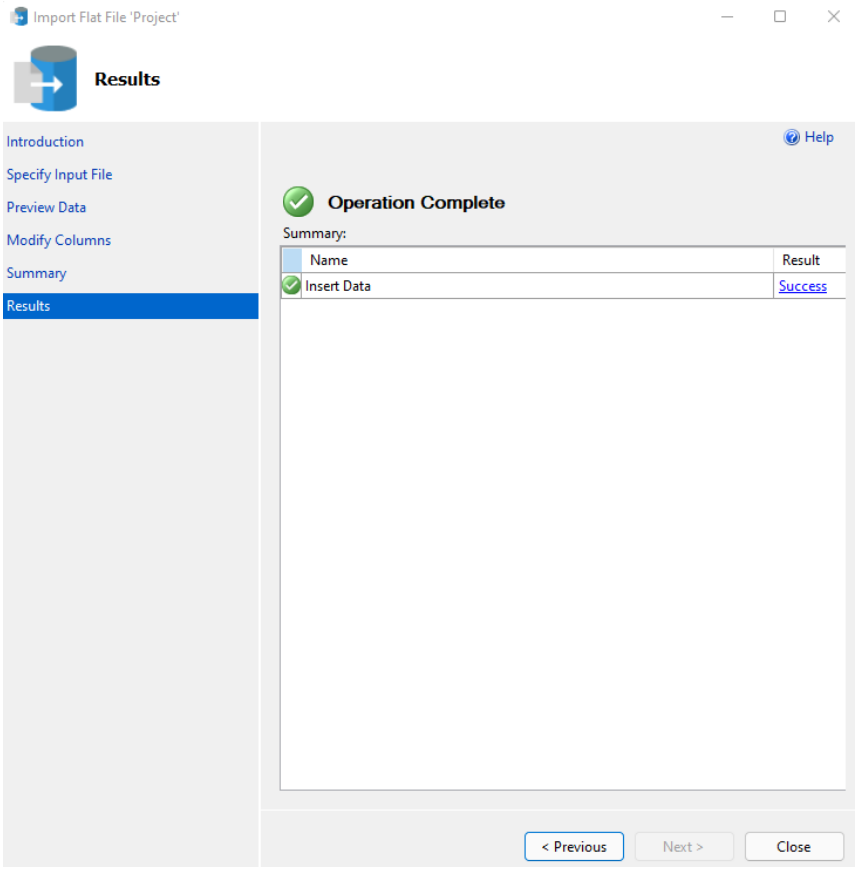
By: TEAM 6





Project-DATA604

By: TEAM 6



Similarly, we have created tables for – consent and death datasets.

Now, Altering the column data types:

```
ALTER TABLE project.dbo.cases ALTER COLUMN sub_num int;
ALTER TABLE Project.dbo.cases ALTER COLUMN tot_cases int;
ALTER TABLE Project.dbo.cases ALTER COLUMN conf_cases int;
ALTER TABLE Project.dbo.cases ALTER COLUMN prob_cases int;
ALTER TABLE Project.dbo.cases ALTER COLUMN new_case int;
ALTER TABLE Project.dbo.cases ALTER COLUMN pnew_case int;

ALTER TABLE project.dbo.consent ALTER COLUMN sub_num int;

ALTER TABLE project.dbo.death ALTER COLUMN sub_num int;
ALTER TABLE Project.dbo.death ALTER COLUMN tot_death int;
ALTER TABLE Project.dbo.death ALTER COLUMN conf_death int;
ALTER TABLE Project.dbo.death ALTER COLUMN prob_death int;
ALTER TABLE Project.dbo.death ALTER COLUMN new_death int;
ALTER TABLE Project.dbo.death ALTER COLUMN pnew_death int;
```

```
00 %
Messages
Commands completed successfully.

Completion time: 2021-12-08T22:32:04.1610312-05:00
```

Project-DATA604

By: TEAM 6

Information Schema to check the data types:

```
SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, COLUMN_NAME, ORDINAL_POSITION, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM Project.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'cases';

SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, COLUMN_NAME, ORDINAL_POSITION, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM Project.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'consent';

SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, COLUMN_NAME, ORDINAL_POSITION, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM Project.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'death';
```

100 %

Results | Messages

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
1	Project	dbo	cases	submission_date	1	date	NULL
2	Project	dbo	cases	sub_num	2	int	NULL
3	Project	dbo	cases	state	3	nvarchar	50
4	Project	dbo	cases	tot_cases	4	int	NULL
5	Project	dbo	cases	conf_cases	5	int	NULL
6	Project	dbo	cases	prob_cases	6	int	NULL
7	Project	dbo	cases	new_case	7	int	NULL
8	Project	dbo	cases	pnew_case	8	int	NULL

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
1	Project	dbo	consent	submission_date	1	date	NULL
2	Project	dbo	consent	sub_num	2	int	NULL
3	Project	dbo	consent	state	3	nvarchar	50
4	Project	dbo	consent	consent_cases	4	nvarchar	50
5	Project	dbo	consent	consent_deaths	5	nvarchar	50

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
1	Project	dbo	death	submission_date	1	date	NULL
2	Project	dbo	death	sub_num	2	int	NULL
3	Project	dbo	death	state	3	nvarchar	50
4	Project	dbo	death	tot_death	4	int	NULL
5	Project	dbo	death	conf_death	5	int	NULL
6	Project	dbo	death	prob_death	6	int	NULL
7	Project	dbo	death	new_death	7	int	NULL
8	Project	dbo	death	pnew_death	8	int	NULL

Total Data counts:

Project-DATA604

By: TEAM 6

```
select count(1) as datacount_cases from project.dbo.cases;
select count(1) as datacount_consent from project.dbo.consent;
select count(1) as datacount_death from project.dbo.death;
```

.00 %

Results | Messages

	datacount_cases
1	39960

	datacount_consent
1	39960

	datacount_death
1	39960

Creating the **temporary** table and using the **rank** function to order the data based on submission date and state:

By: TEAM 6

```
drop table if exists temp_update_cases
SELECT
submission_date
,state
,ROW_NUMBER() OVER (ORDER BY SUBMISSION_DATE, STATE) AS RN1
INTO temp_update_cases
FROM project.dbo.cases;

select * from temp_update_cases;
```

Results | Messages

submission_date	state	RN1
2020-01-22	AK	1
2020-01-22	AL	2
2020-01-22	AR	3
2020-01-22	AS	4
2020-01-22	AZ	5
2020-01-22	CA	6
2020-01-22	CO	7
2020-01-22	CT	8
2020-01-22	DC	9
2020-01-22	DE	10
2020-01-22	FL	11
2020-01-22	FSM	12
2020-01-22	GA	13
2020-01-22	GU	14
2020-01-22	HI	15
2020-01-22	IA	16
2020-01-22	ID	17
2020-01-22	IL	18
2020-01-22	IN	19
2020-01-22	KS	20

Implementing the similarly changes for – consent and death datasets.

Project-DATA604

By: TEAM 6

Using the **update** function to assign the ranked values to submission_num for cases dataset:

```
UPDATE p1
SET p1.sub_num = t1.RN1
FROM project.dbo.cases as p1
INNER JOIN temp_update_cases as t1
ON p1.submission_date = t1.submission_date and p1.state = t1.state;

select * from project.dbo.cases order by submission_date, state;
```

100 %

Results | Messages

	submission_date	sub_num	state	tot_cases	conf_cases	prob_cases	new_case	pnew_case
1	2020-01-22	1	AK	0	NULL	NULL	0	NULL
2	2020-01-22	2	AL	0	0	0	0	0
3	2020-01-22	3	AR	0	NULL	NULL	0	NULL
4	2020-01-22	4	AS	0	NULL	NULL	0	NULL
5	2020-01-22	5	AZ	0	NULL	NULL	0	NULL
6	2020-01-22	6	CA	0	0	0	0	0
7	2020-01-22	7	CO	0	NULL	NULL	0	NULL
8	2020-01-22	8	CT	0	NULL	NULL	0	NULL
9	2020-01-22	9	DC	0	NULL	NULL	0	NULL
10	2020-01-22	10	DE	0	NULL	NULL	0	NULL
11	2020-01-22	11	FL	0	NULL	NULL	0	NULL
12	2020-01-22	12	FSM	0	NULL	NULL	0	NULL
13	2020-01-22	13	GA	0	NULL	NULL	0	NULL
14	2020-01-22	14	GU	0	NULL	NULL	0	NULL
15	2020-01-22	15	HI	0	NULL	NULL	0	NULL
16	2020-01-22	16	IA	0	NULL	NULL	0	NULL
17	2020-01-22	17	ID	0	NULL	NULL	0	NULL
18	2020-01-22	18	IL	0	NULL	NULL	0	NULL
19	2020-01-22	19	IN	0	NULL	NULL	0	NULL
20	2020-01-22	20	KS	0	NULL	NULL	0	NULL
21	2020-01-22	21	KY	0	NULL	NULL	0	NULL
22	2020-01-22	22	LA	0	NULL	NULL	0	NULL
23	2020-01-22	23	MA	0	NULL	NULL	0	NULL

Implementing the similarly changes for – consent and death datasets.

Project-DATA604

By: TEAM 6

Creating the View - **US_Covid19** while using the inner-join on cases, consent, and death tables. We have used **submission_num** as a **primary key**.

```
USE Project;
GO
DROP VIEW IF EXISTS US_Covid19;

USE Project;
GO
CREATE VIEW US_Covid19 AS
SELECT
c.sub_num as submission_num
,c.submission_date
,c.state
,a.tot_cases as total_cases
,a.conf_cases as confirmed_cases
,a.prob_cases as probable_cases
,a.new_case as new_cases
,a.pnew_case as probable_new_cases
,b.consent_cases
,b.consent_deaths
,c.tot_death as total_deaths
,c.conf_death as confirmed_deaths
,c.prob_death as probable_deaths
,c.new_death as new_deaths
,c.pnew_death as probable_new_deaths
from
project.dbo.cases a
inner join project.dbo.consent b on a.sub_num = b.sub_num
inner join project.dbo.death c on b.sub_num = c.sub_num;

select v.*
from Project.dbo.US_Covid19 v;
```

Results | Messages

	submission_num	submission_date	state	total_cases	confirmed_cases	probable_cases	new_cases	probable_new_cases	consent_cases	consent_deaths	total_deaths	confirmed_deaths	probable_deaths	new_deaths	probable_new_deaths
1	26106	2021-04-01	CA	3570660	3570660	0	2234	0	Agree	Agree	58090	58090	0	154	0
2	29706	2021-05-31	CA	3685032	3685032	0	644	0	Agree	Agree	62011	62011	0	5	0
3	934	2020-02-06	NE	0	NULL	NULL	0	NULL	Agree	Agree	0	NULL	NULL	0	NULL
4	11425	2020-07-30	ME	3910	3497	413	22	2	Agree	Agree	123	122	1	2	0
5	22650	2021-02-02	MS	280182	176228	103954	1059	559	Agree	Agree	6730	4739	1991	13	7
6	6155	2020-05-03	NH	2518	NULL	NULL	89	0	Not agree	Not agree	86	NULL	NULL	2	0
7	28778	2021-05-15	NV	320719	320719	0	180	0	Agree	NULL	5530	NULL	NULL	0	0
8	28879	2021-05-17	IN	735999	NULL	NULL	537	0	Not agree	Agree	13460	13044	416	6	0
9	24062	2021-02-26	AL	492393	388404	103989	903	256	Agree	Agree	10523	8396	2127	29	8
10	24572	2021-03-06	NC	875359	760095	115264	1614	450	Agree	Agree	12363	10933	1430	16	2
11	326	2020-01-27	MI	0	NULL	NULL	0	NULL	Agree	Agree	0	NULL	NULL	0	NULL
12	23348	2021-02-14	CT	267337	250915	16422	0	0	Agree	Agree	7381	6049	1332	0	0
13	14338	2020-09-16	WI	98440	92712	5728	1502	94	Agree	Agree	1237	1228	9	8	0

We have found few rows in the consent_cases and consent_deaths with the data as **N/A**.

Using the update function to assign the **NULL** values to those records having values as **N/A**.

```
Update v
SET consent_cases = NULL
,consent_deaths = NULL
FROM Project.dbo.US_Covid19 v
where consent_cases = 'N/A' or consent_deaths = 'N/A';
```

Project-DATA604

By: TEAM 6

Checking the datatype of the View- **US_Covid19**:

```
--INFORMATION SCHEMA:
SELECT TABLE_CATALOG, TABLE_SCHEMA, TABLE_NAME, COLUMN_NAME, ORDINAL_POSITION, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH
FROM Project.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'US_Covid19';
```

100 %

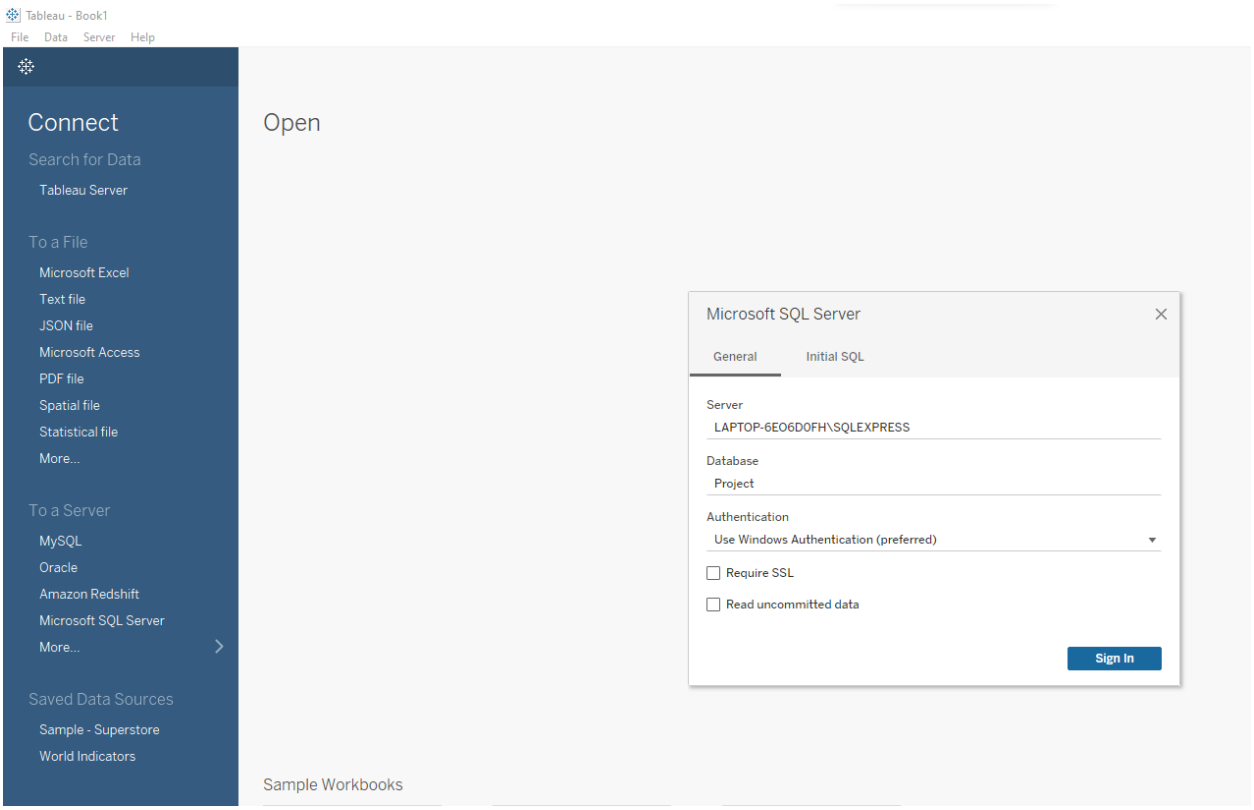
Results | Messages

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
1	Project	dbo	US_Covid19	submission_num	1	int	NULL
2	Project	dbo	US_Covid19	submission_date	2	date	NULL
3	Project	dbo	US_Covid19	state	3	nvarchar	50
4	Project	dbo	US_Covid19	total_cases	4	int	NULL
5	Project	dbo	US_Covid19	confirmed_cases	5	int	NULL
6	Project	dbo	US_Covid19	probable_cases	6	int	NULL
7	Project	dbo	US_Covid19	new_cases	7	int	NULL
8	Project	dbo	US_Covid19	probable_new_cases	8	int	NULL
9	Project	dbo	US_Covid19	consent_cases	9	nvarchar	50
10	Project	dbo	US_Covid19	consent_deaths	10	nvarchar	50
11	Project	dbo	US_Covid19	total_deaths	11	int	NULL
12	Project	dbo	US_Covid19	confirmed_deaths	12	int	NULL
13	Project	dbo	US_Covid19	probable_deaths	13	int	NULL
14	Project	dbo	US_Covid19	new_deaths	14	int	NULL
15	Project	dbo	US_Covid19	probable_new_dea...	15	int	NULL

Connecting the Tableau with the Microsoft SQL Server:

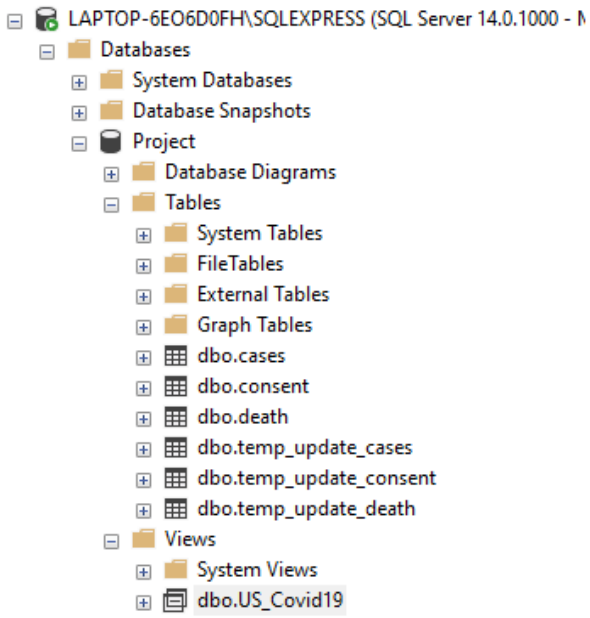
Project-DATA604

By: TEAM 6



All the tables and views which are present in the Microsoft SQL Server as now imported to Tableau:

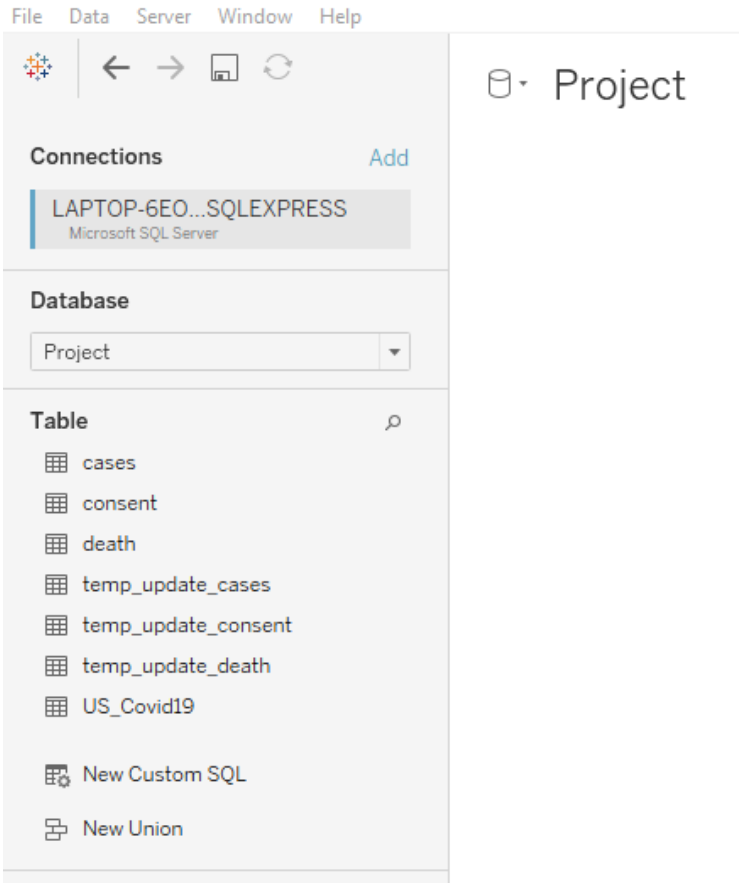
Microsoft SQL Server



Project-DATA604

By: TEAM 6

Tableau

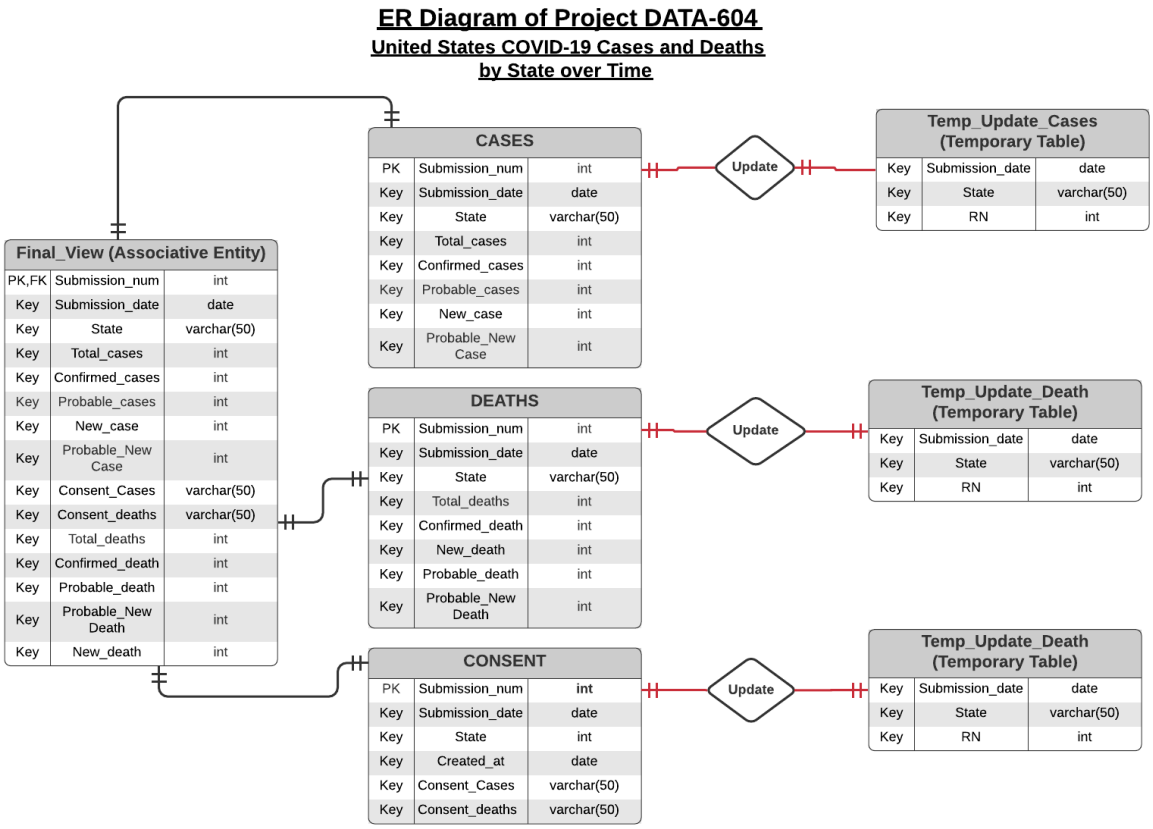


Project-DATA604

By: TEAM 6

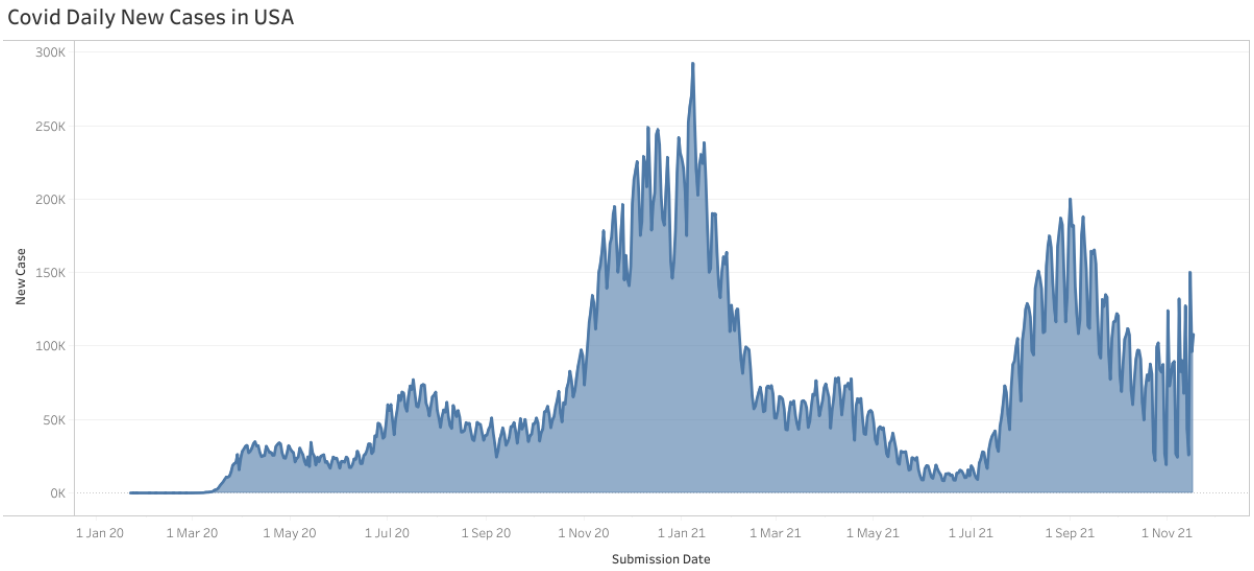
ER- Diagram: United States COVID-19 Cases and Deaths by State over Time

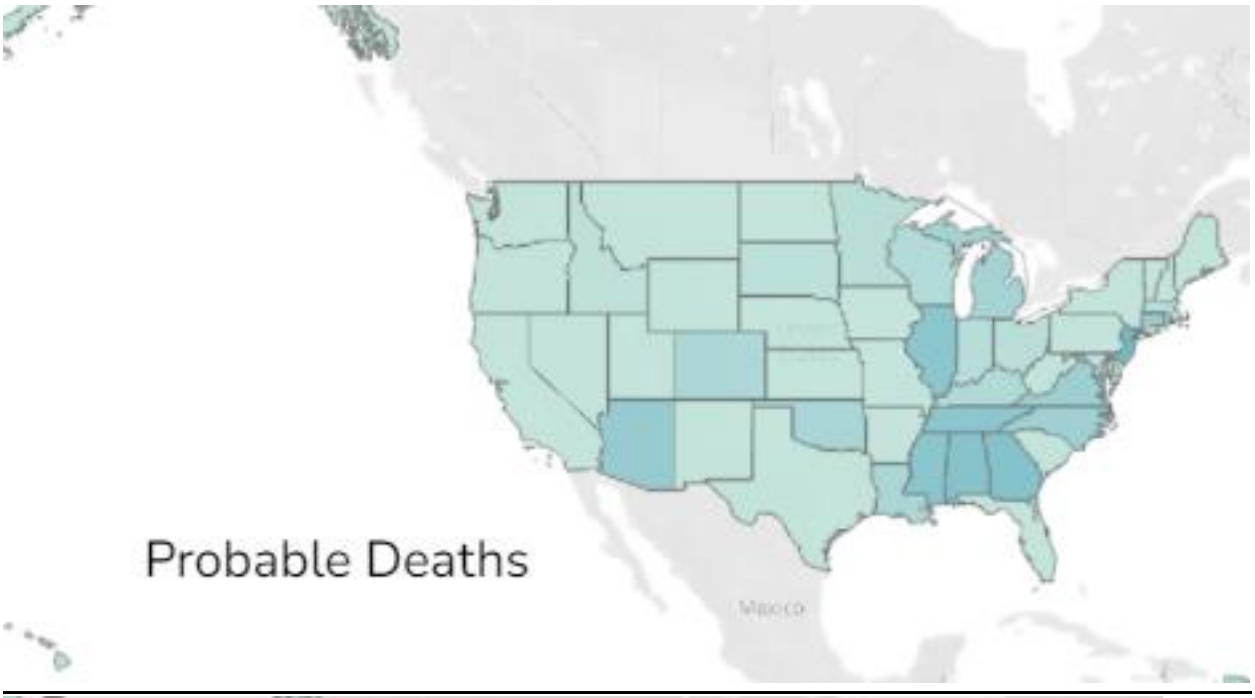
yeah same name alexa further

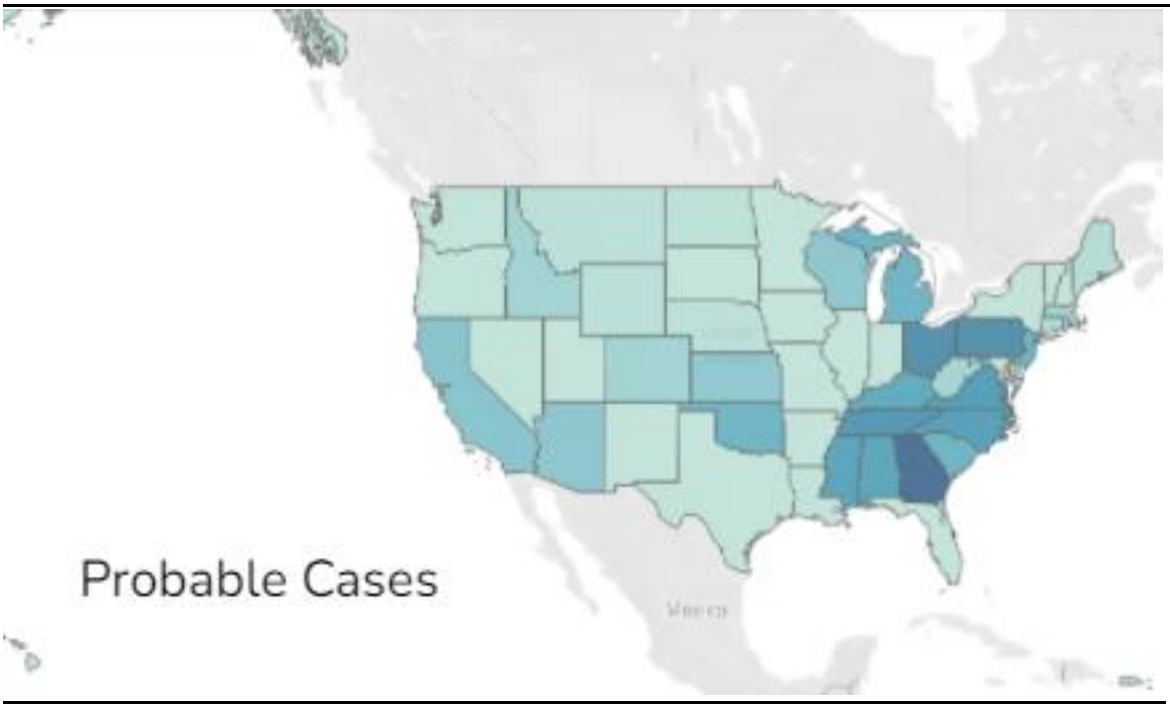


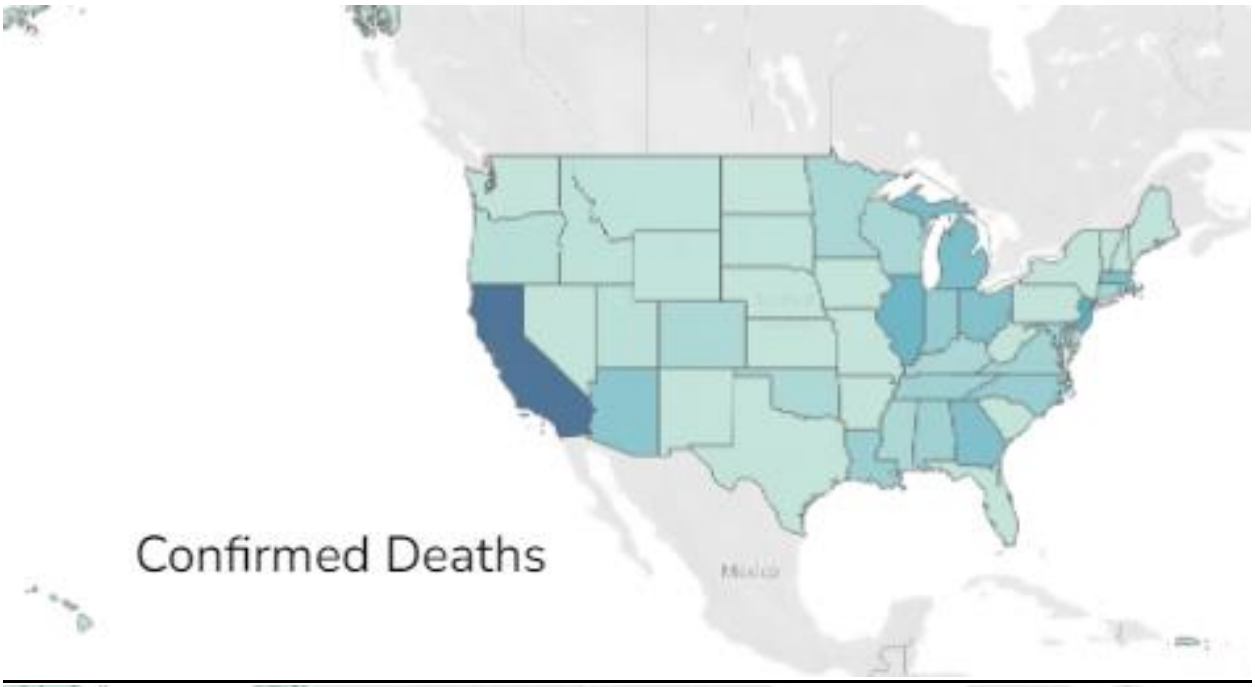
Project-DATA604

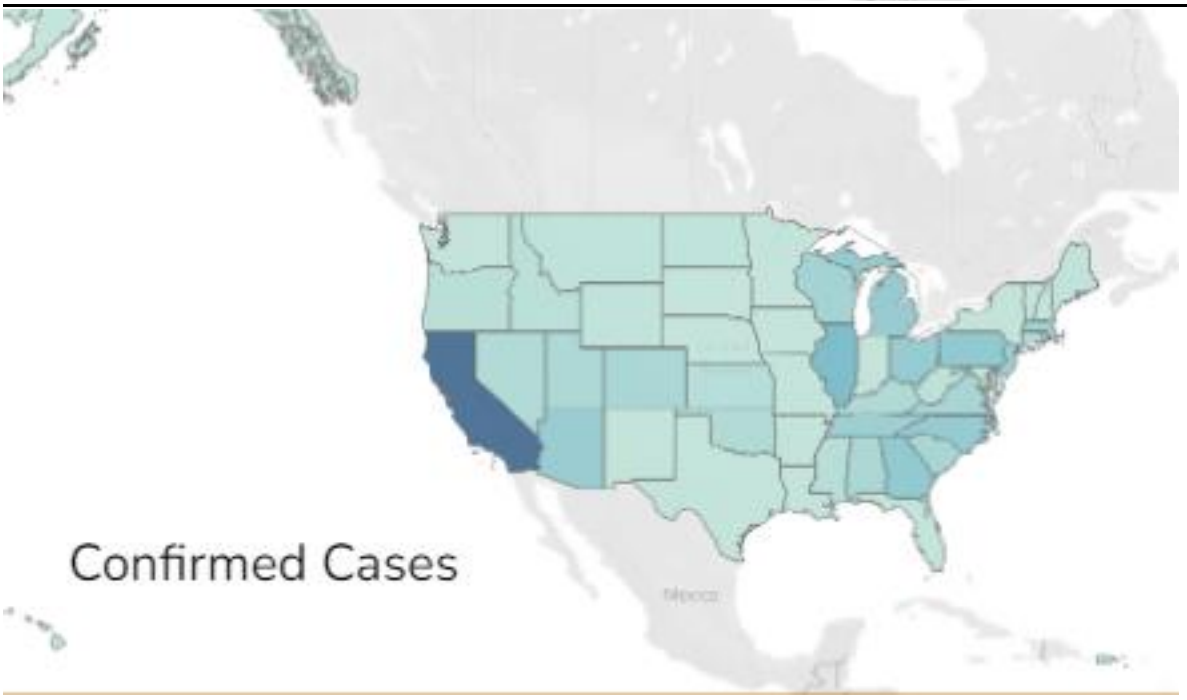
By: TEAM 6




Probable Deaths


Probable Cases


Confirmed Deaths


Confirmed Cases

Project-DATA604

By: TEAM 6