Colorado School of Mines

Trump Rallies and COVID Cases: An Investigative Approach

Jordan Tehranchi, Joline Sikora, Walden Ruemmele, Jacky Wang

CSCI 403 Project 9

December 9, 2020

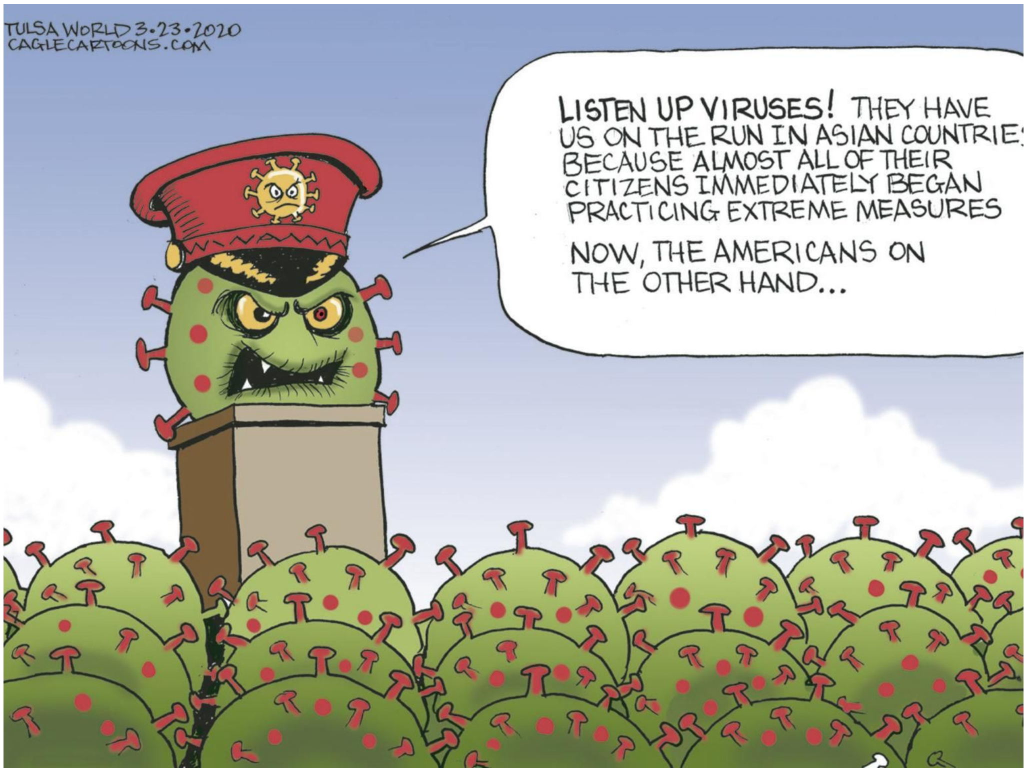

TULSA WORLD 3•23•2020
CAGLECARTOONS.COM
LISTEN UP VIRUSES! THEY HAVE US ON THE RUN IN ASIAN COUNTRIE BECAUSE ALMOST ALL OF THEIR CITIZENS IMMEDIATELY BEGAN PRACTICING EXTREME MEASURES
NOW, THE AMERICANS ON THE OTHER HAND...

Figure 1: Sourced from Bruce Plante cartoon: General COVID-19

1 Introduction

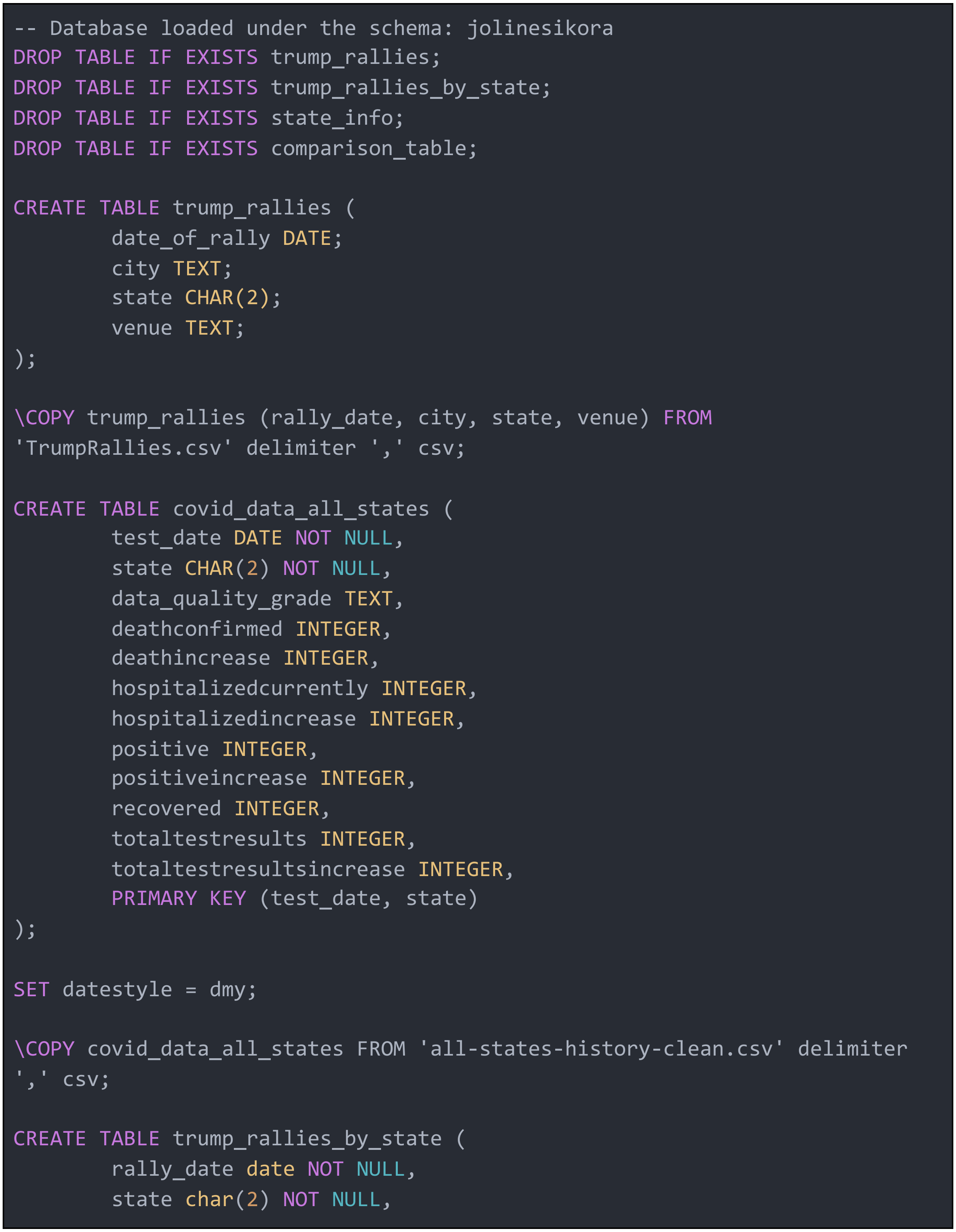
Right now we are engrossed within an era of the COVID-19 pandemic. While there are many questions, concerns, and changes going on, there is a lot of data and questions the technical database management team or individual might ask about the pandemic in relation to other events. In our case, we were curious to find out how Trump rallies that have been ongoing within the last few months have influenced COVID-19 cases. Some of these influences could include: if the cases started to peak during these rallies, the state where they occurred within the United States, the amount of COVID-19 recoveries after being exposed, and much more.

Because COVID-19 has been such a major topic within the past year and especially within the US, there is a lot of data that has been collected during this timeframe. Within the data set that we obtained from the COVID tracking website, there was a ton of data and data columns. Because of this, our team decided to trim out some of the extraneous variables that we did not want to necessarily focus on during the course of this project. Some of these variables were things like positive and negative test results, positive and negative antibody tests, whether or not people were on ventilators and more. In the end, our team decided to stay with the following columns: data, state, quality of the data, confirmed deaths, increase in deaths on that day, number of hospitalizations there currently are, increase in hospitalizations on that day, number of positive cases, increase in positive case on that day, total recoveries, total test results, and increase in test results on that day. This table information was obtained from the COVID tracking website on November 23, 2020. This combined with our other table would allow us to see the effects of Trump rallies on COVID and propose a variety of different queries. The Trump rallies table consists of the following columns: the rally date, city and state where the rally was held, and the specific venue it was held at. The Trump data was created from a wikipedia database.

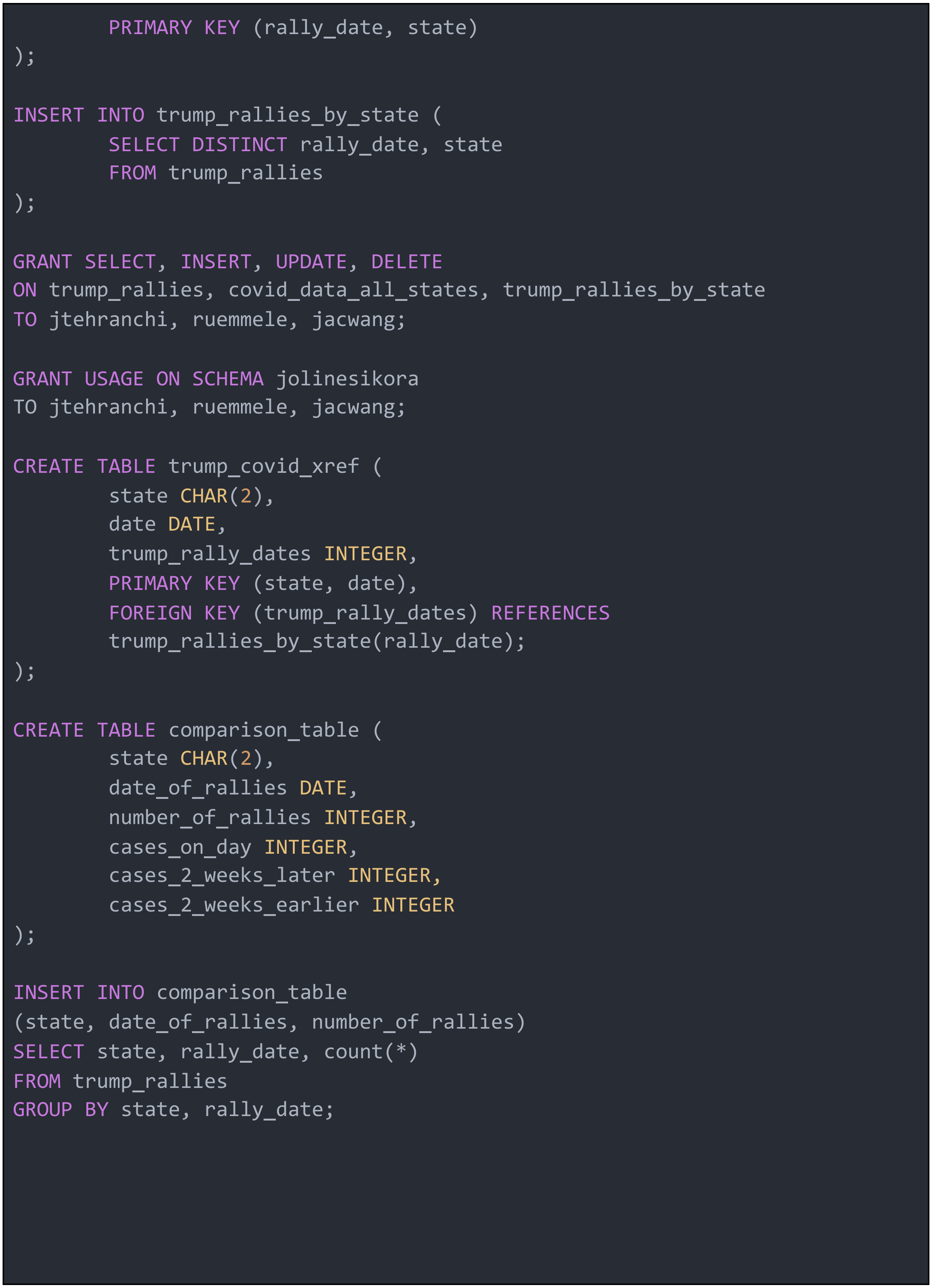
There are some potential issues that we could run into that we need to address before diving into the various queries and potential conclusions we could draw from this project. Since we are just looking at a very narrow view of the variables affecting COVID-19 data, the data will be very skewed and the rallies could potentially have no bearing on the data at all. There are many environmental and surrounding factors related to this but we will not consider them and these two tables will be our whole “environment”.

Our group split the project up in order to optimize workflow. Joline worked to create/upload all of the tables into her schema, she also worked on the SQL for the comparison table and found the source Trump Rally data. Walden worked on the write up and SQL queries, as well as finding the source covid data for the states. Jordan worked on the creation of the introduction, conclusion and ERD diagram representing the created COVID Data/Rally database. Jacky worked on the write up, formatting, editing, and the SQL queries as well.

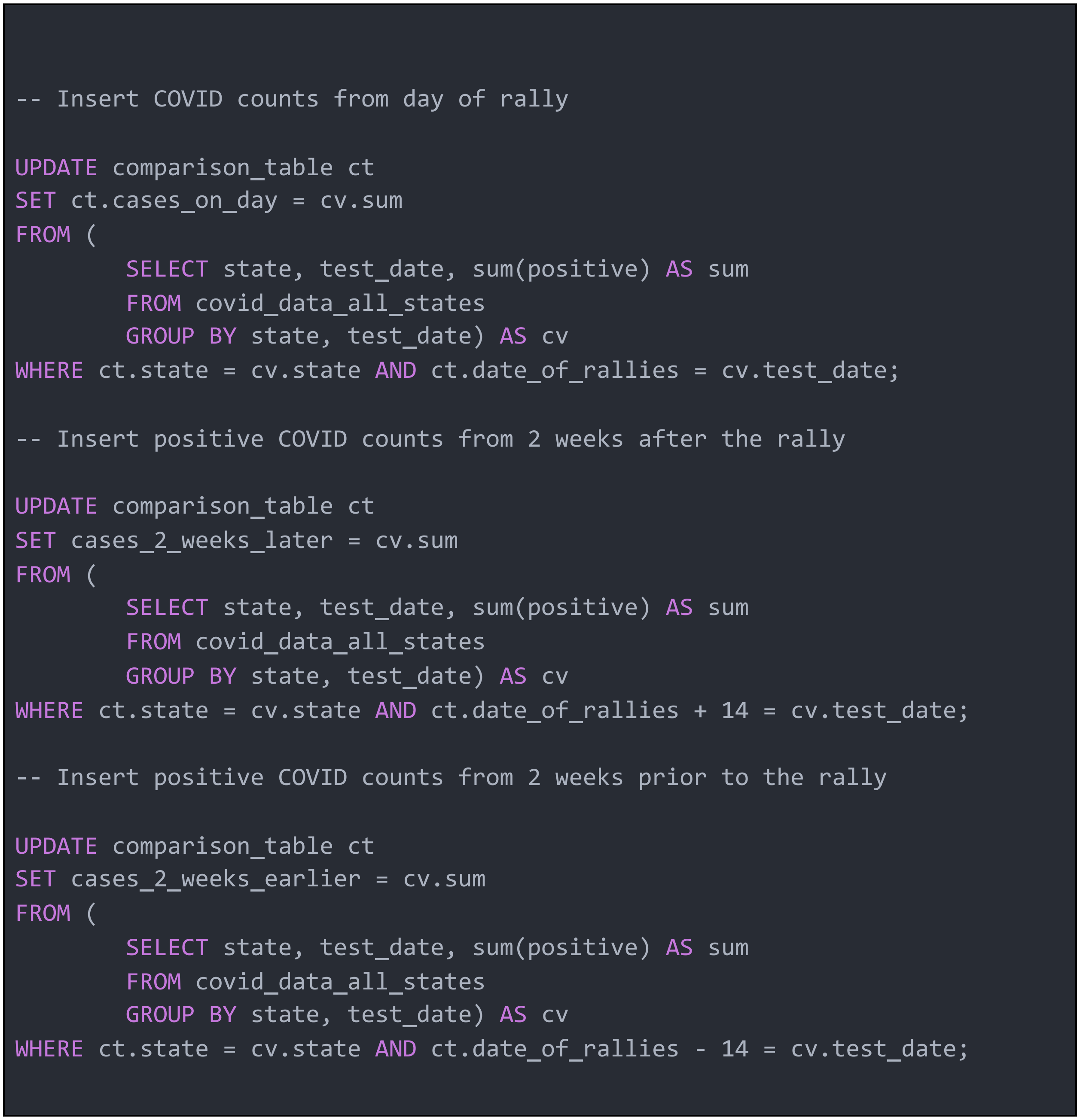
2 SQL Code for Database Creation



```
-- Database loaded under the schema: jolinesikora
DROP TABLE IF EXISTS trump_rallies;
DROP TABLE IF EXISTS trump_rallies_by_state;
DROP TABLE IF EXISTS state_info;
DROP TABLE IF EXISTS comparison_table;

CREATE TABLE trump_rallies (
        date_of_rally DATE;
        city TEXT;
        state CHAR(2);
        venue TEXT;
);

\COPY trump_rallies (rally_date, city, state, venue) FROM
'TrumpRallies.csv' delimiter ',' csv;

CREATE TABLE covid_data_all_states (
        test_date DATE NOT NULL,
        state CHAR(2) NOT NULL,
        data_quality_grade TEXT,
        deathconfirmed INTEGER,
        deathincrease INTEGER,
        hospitalizedcurrently INTEGER,
        hospitalizedincrease INTEGER,
        positive INTEGER,
        positiveincrease INTEGER,
        recovered INTEGER,
        totaltestresults INTEGER,
        totaltestresultsincrease INTEGER,
        PRIMARY KEY (test_date, state)
);

SET datestyle = dmy;

\COPY covid_data_all_states FROM 'all-states-history-clean.csv' delimiter
',' csv;

CREATE TABLE trump_rallies_by_state (
        rally_date date NOT NULL,
        state char(2) NOT NULL,
```



```
        PRIMARY KEY (rally_date, state)
);

INSERT INTO trump_rallies_by_state (
        SELECT DISTINCT rally_date, state
        FROM trump_rallies
);

GRANT SELECT, INSERT, UPDATE, DELETE
ON trump_rallies, covid_data_all_states, trump_rallies_by_state
TO jtehranchi, ruemmele, jacwang;

GRANT USAGE ON SCHEMA jolinesikora
TO jtehranchi, ruemmele, jacwang;

CREATE TABLE trump_covid_xref (
        state CHAR(2),
        date DATE,
        trump_rally_dates INTEGER,
        PRIMARY KEY (state, date),
        FOREIGN KEY (trump_rally_dates) REFERENCES
        trump_rallies_by_state(rally_date);
);

CREATE TABLE comparison_table (
        state CHAR(2),
        date_of_rallies DATE,
        number_of_rallies INTEGER,
        cases_on_day INTEGER,
        cases_2_weeks_later INTEGER,
        cases_2_weeks_earlier INTEGER
);

INSERT INTO comparison_table
(state, date_of_rallies, number_of_rallies)
SELECT state, rally_date, count(*)
FROM trump_rallies
GROUP BY state, rally_date;
```



```sql
-- Insert COVID counts from day of rally

UPDATE comparison_table ct
SET ct.cases_on_day = cv.sum
FROM (
        SELECT state, test_date, sum(positive) AS sum
        FROM covid_data_all_states
        GROUP BY state, test_date) AS cv
WHERE ct.state = cv.state AND ct.date_of_rallies = cv.test_date;

-- Insert positive COVID counts from 2 weeks after the rally

UPDATE comparison_table ct
SET cases_2_weeks_later = cv.sum
FROM (
        SELECT state, test_date, sum(positive) AS sum
        FROM covid_data_all_states
        GROUP BY state, test_date) AS cv
WHERE ct.state = cv.state AND ct.date_of_rallies + 14 = cv.test_date;

-- Insert positive COVID counts from 2 weeks prior to the rally

UPDATE comparison_table ct
SET cases_2_weeks_earlier = cv.sum
FROM (
        SELECT state, test_date, sum(positive) AS sum
        FROM covid_data_all_states
        GROUP BY state, test_date) AS cv
WHERE ct.state = cv.state AND ct.date_of_rallies - 14 = cv.test_date;
```

3 Visualization

For the visualization portion of this project report, we created an ERD to show the relationships between the two data sets that we referenced. Based on this ERD diagram, we were able to implement the above SQL code to create the different tables that are used in the queries and code portions of the project. The relationship between Trump rallies by state and the COVID data for all states was assumed to be a many-to-many relationship, where there were many rallies in multiple states and the data for each state was spread out over many different days. This ERD along with the assumptions made in the introduction allowed us to create the different SQL tables.

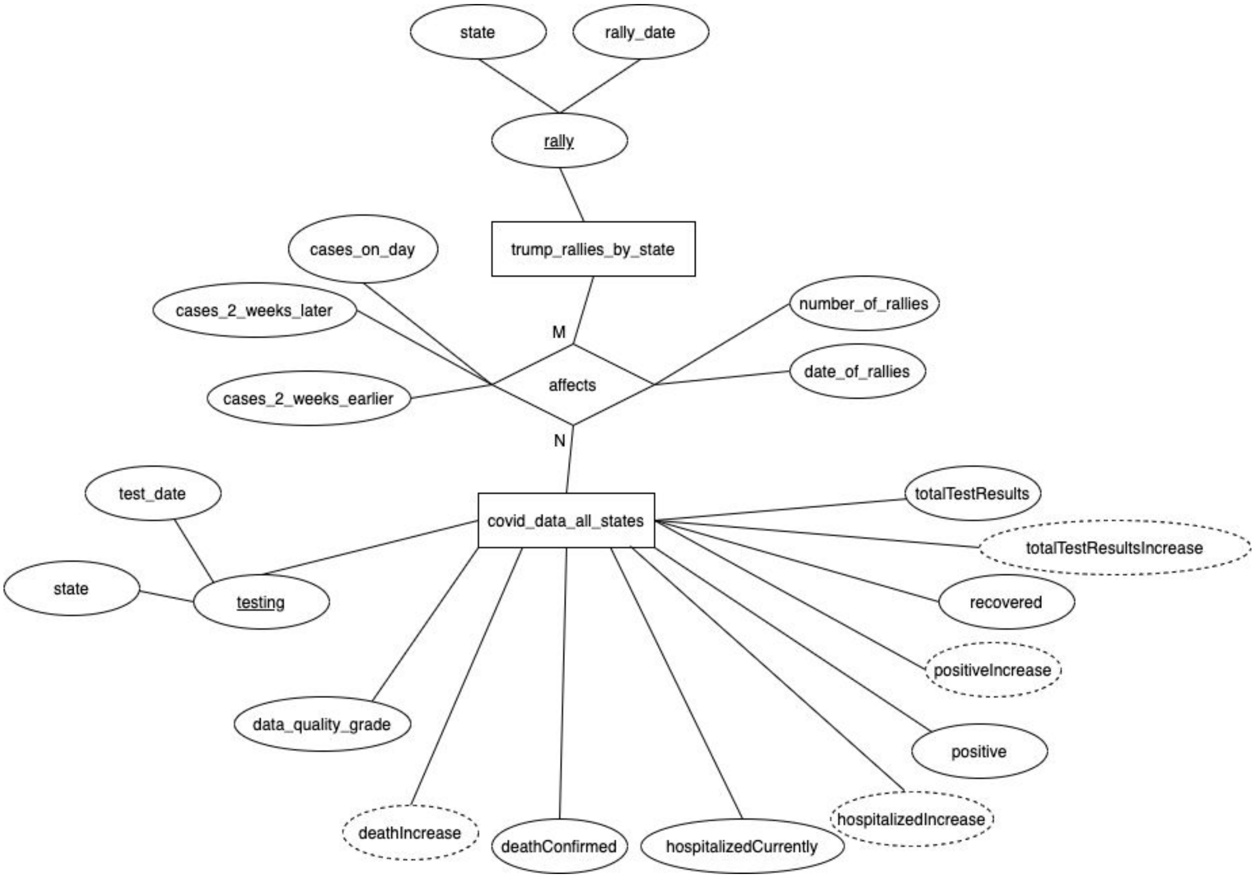
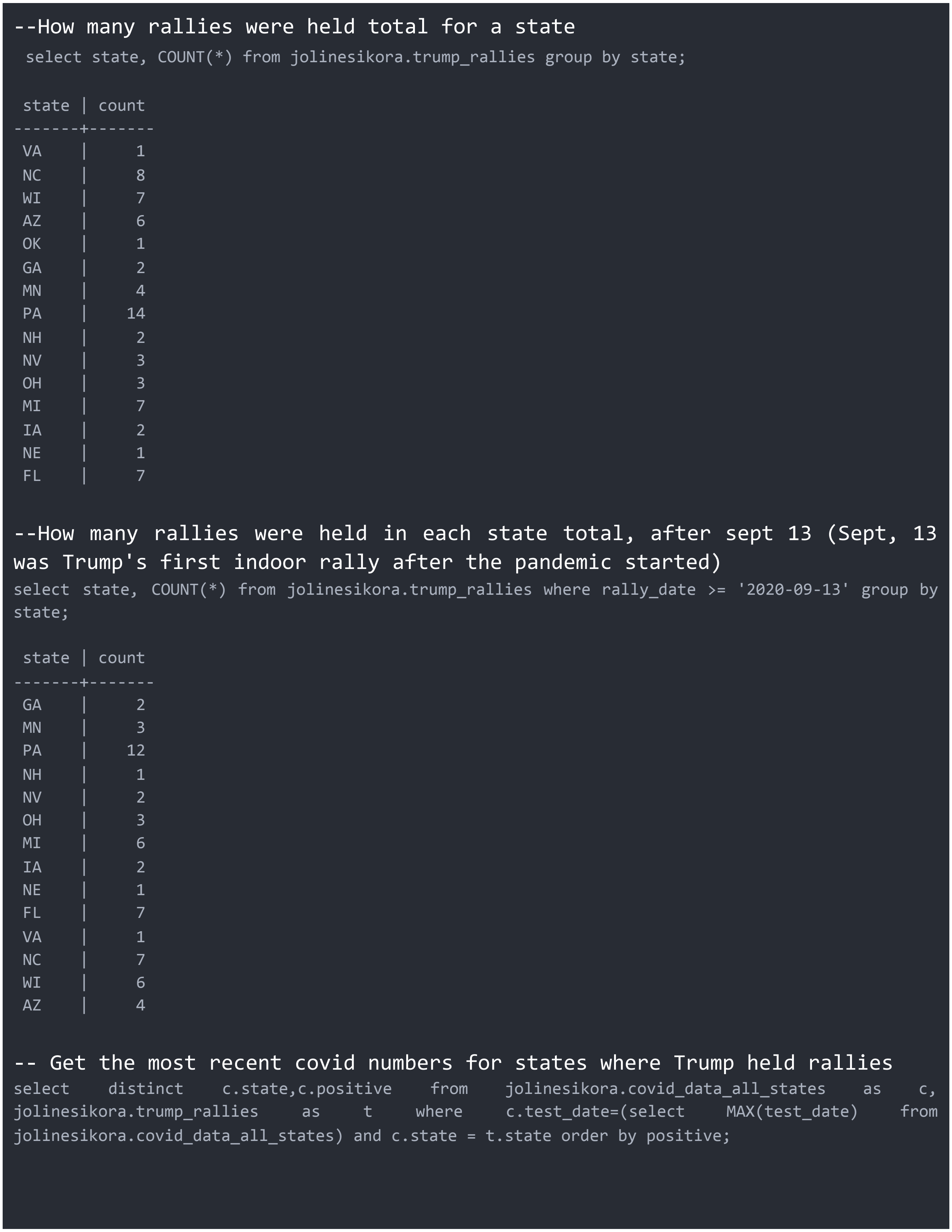


Figure 3.1: ERD Diagram Relating Trump Rallies by State to COVID Data Within All States

4 Queries

Below are some of the queries and what they returned when put into SQL:



```
--How many rallies were held total for a state
 select state, COUNT(*) from jolinesikora.trump_rallies group by state;

 state | count
-------+-------
 VA    |     1
 NC    |     8
 WI    |     7
 AZ    |     6
 OK    |     1
 GA    |     2
 MN    |     4
 PA    |    14
 NH    |     2
 NV    |     3
 OH    |     3
 MI    |     7
 IA    |     2
 NE    |     1
 FL    |     7

--How many rallies were held in each state total, after sept 13 (Sept, 13
was Trump's first indoor rally after the pandemic started)
select state, COUNT(*) from jolinesikora.trump_rallies where rally_date >= '2020-09-13' group by
state;

 state | count
-------+-------
 GA    |     2
 MN    |     3
 PA    |    12
 NH    |     1
 NV    |     2
 OH    |     3
 MI    |     6
 IA    |     2
 NE    |     1
 FL    |     7
 VA    |     1
 NC    |     7
 WI    |     6
 AZ    |     4

-- Get the most recent covid numbers for states where Trump held rallies
select distinct c.state,c.positive from jolinesikora.covid_data_all_states as c,
jolinesikora.trump_rallies as t where c.test_date=(select MAX(test_date) from
jolinesikora.covid_data_all_states) and c.state = t.state order by positive;
```

```
 state | positive
-------+----------
 NH    |    18042
 NE    |   114061
 NV    |   136227
 OK    |   177874
 IA    |   189856
 VA    |   221038
 MN    |   276500
 AZ    |   302324
 PA    |   314401
 NC    |   339194
 MI    |   340964
 OH    |   363304
 WI    |   379693
 GA    |   406220
 FL    |   930728
```

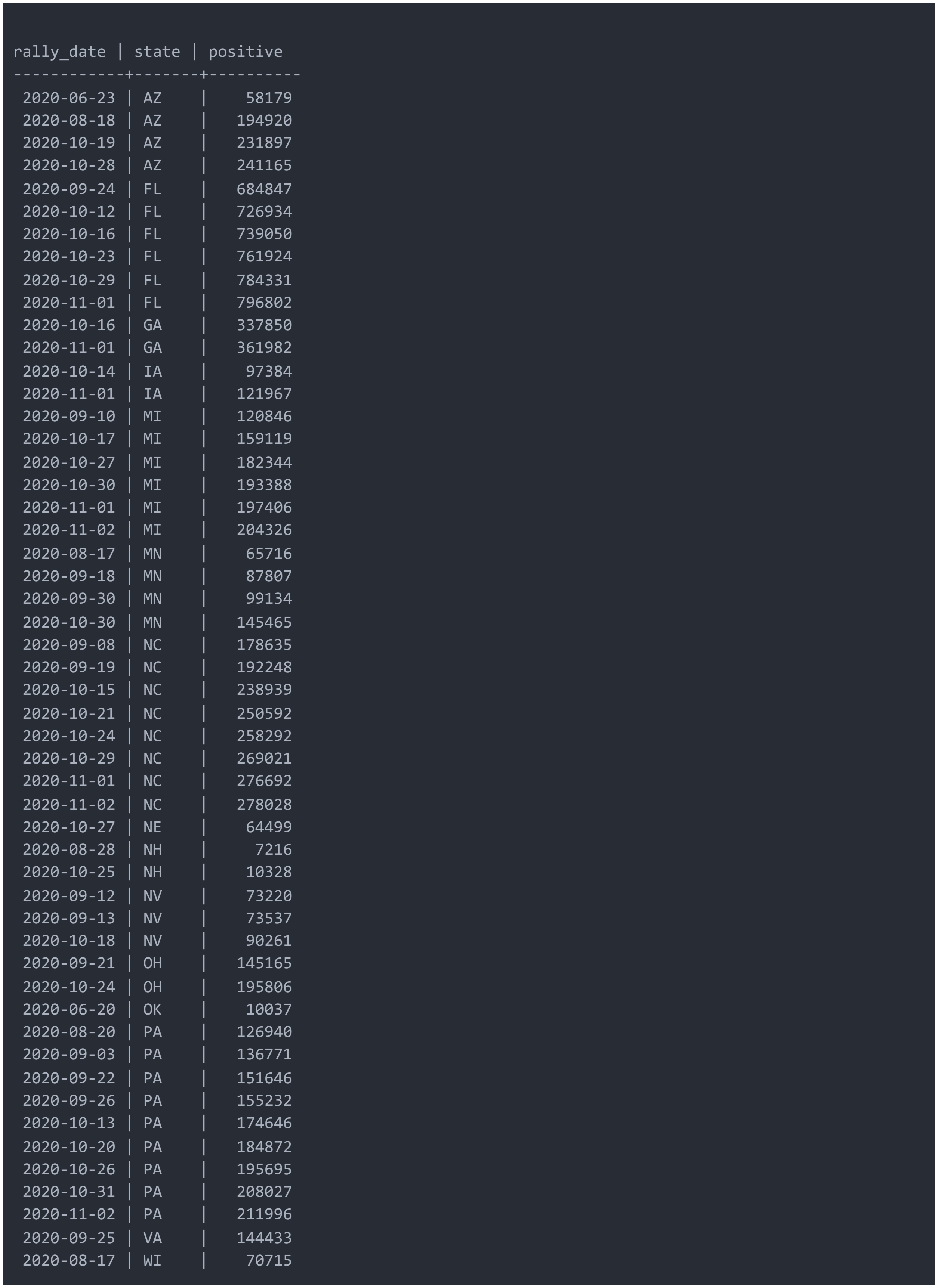
```
-- What was the increase in COVID cases 2 weeks after each rally? 2 weeks earlier?
select state, date_of_rallies, number_of_rallies, cases_on_day, cases_2_weeks_later, cases_2_weeks_earlier
from jolinesikora.comparison_table;
```

```
state | date_of_rallies | rallies | cases_on_day | cases_2_weeks_later | cases_2_weeks_earlier
-------+-----------------+---------+--------------+---------------------+-----------------------
 AZ    | 2020-06-23      |       1 |        58179 |              105094 |                 28296
 AZ    | 2020-08-18      |       1 |       194920 |              202342 |                180505
 AZ    | 2020-10-19      |       2 |       231897 |              248139 |                221070
 AZ    | 2020-10-28      |       2 |       241165 |              265163 |                227635
 FL    | 2020-09-24      |       1 |       684847 |              717148 |                647318
 FL    | 2020-10-12      |       1 |       726934 |              771989 |                692962
 FL    | 2020-10-16      |       1 |       739050 |              789714 |                703212
 FL    | 2020-10-23      |       2 |       761924 |              821526 |                720001
 FL    | 2020-10-29      |       1 |       784331 |              851825 |                735685
 FL    | 2020-11-01      |       1 |       796802 |              872810 |                745492
 GA    | 2020-10-16      |       1 |       337850 |              358225 |                320634
 GA    | 2020-11-01      |       1 |       361982 |              386949 |                340558
 IA    | 2020-10-14      |       1 |        97384 |              113042 |                 85596
 IA    | 2020-11-01      |       1 |       121967 |              167282 |                101960
 MI    | 2020-09-10      |       1 |       120846 |              132337 |                110343
 MI    | 2020-10-17      |       1 |       159119 |              197406 |                141271
 MI    | 2020-10-27      |       1 |       182344 |              245252 |                152862
 MI    | 2020-10-30      |       1 |       193388 |              268362 |                159119
 MI    | 2020-11-01      |       1 |       197406 |              275792 |                159119
 MI    | 2020-11-02      |       2 |       204326 |              288954 |                164123
 MN    | 2020-08-17      |       1 |        65716 |               75864 |                 56560
 MN    | 2020-09-18      |       1 |        87807 |              101366 |                 78966
 MN    | 2020-09-30      |       1 |        99134 |              115943 |                 85813
 MN    | 2020-10-30      |       1 |       145465 |              207339 |                119396
 NC    | 2020-09-08      |       1 |       178635 |              195549 |                157741
 NC    | 2020-09-19      |       1 |       192248 |              216886 |                175815
 NC    | 2020-10-15      |       1 |       238939 |              269021 |                212909
 NC    | 2020-10-21      |       1 |       250592 |              282802 |                222969
 NC    | 2020-10-24      |       1 |       258292 |                 291245 |                    229752
 NC    | 2020-10-29      |       1 |       269021 |                 303454 |                    238939
 NC    | 2020-11-01      |       1 |       276692 |                 312235 |                    246028
 NC    | 2020-11-02      |       1 |       278028 |                 314207 |                    247172
 NE    | 2020-10-27      |       1 |        64499 |                  85551 |                     52839
 NH    | 2020-08-28      |       1 |         7216 |                   7620 |                      6921
 NH    | 2020-10-25      |       1 |        10328 |                  12488 |                      9092
 NV    | 2020-09-12      |       1 |        73220 |                  78355 |                     68461
 NV    | 2020-09-13      |       1 |        73537 |                  78728 |                     68908
 NV    | 2020-10-18      |       1 |        90261 |                 101479 |                     82100
 OH    | 2020-09-21      |       2 |       145165 |                 159964 |                    131336
 OH    | 2020-10-24      |       1 |       195806 |                 245727 |                    167458
 OK    | 2020-06-20      |       1 |        10037 |                  15645 |                      7003
 PA    | 2020-08-20      |       1 |       126940 |                 136771 |                    116521
 PA    | 2020-09-03      |       1 |       136771 |                 147923 |                    126940
 PA    | 2020-09-22      |       1 |       151646 |                 165243 |                    140359
 PA    | 2020-09-26      |       1 |       155232 |                 171050 |                    143805
 PA    | 2020-10-13      |       1 |       174646 |                 198446 |                    157814
 PA    | 2020-10-20      |       1 |       184872 |                 214871 |                    165243
 PA    | 2020-10-26      |       3 |       195695 |                 234296 |                    173304
 PA    | 2020-10-31      |       4 |       208027 |                 259938 |                    180943
 PA    | 2020-11-02      |       1 |       211996 |                 269613 |                    183315
 VA    | 2020-09-25      |       1 |       144433 |                 156649 |                    131640
 WI    | 2020-08-17      |       1 |        70715 |                  80568 |                     59401
 WI    | 2020-09-17      |       1 |       100574 |                 132123 |                     82922
 WI    | 2020-10-17      |       1 |       175227 |                 237870 |                    138002
 WI    | 2020-10-24      |       1 |       205139 |                 277503 |                    155602
 WI    | 2020-10-27      |       1 |       217429 |                 293812 |                    163759
 WI    | 2020-10-30      |       1 |       232062 |                 318023 |                    175227
 WI    | 2020-11-02      |       1 |       244928 |                 334562 |                    183142
(58 rows)
```

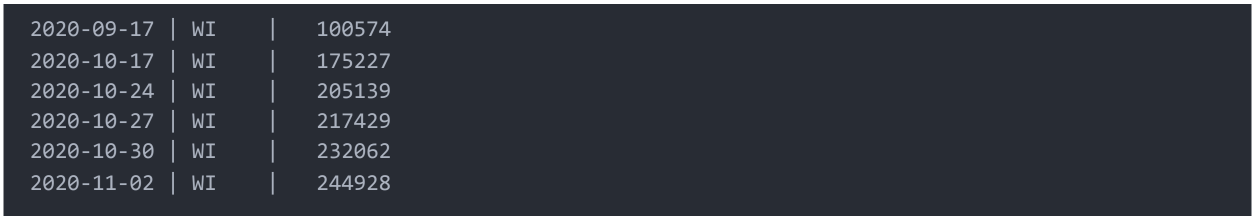
```
-- Which states have a data quality rating of C for November, 23rd, 2020?
select state from jolinesikora.covid_data_all_states where data_quality_grade = ‘C’ and test_date = ‘2020-11-23’ group by state;
```

```
-- What is the state that was hit the hardest in terms of increase in positive tests on November, 23rd, 2020?
select state from jolinesikora.covid_data_all_states where positiveincrease = (select max(positiveincrease) from jolinesikora.covid_data_all_states where test_date = '2020-11-23');

 state
-------
 MI
(1 row)
```

```
-- What was the baseline number of positive test cases at the time of a trump rally?
SELECT TR.rally_date, TR.state, CD.positive
FROM jolinesikora.covid_data_all_states CD, jolinesikora.trump_rallies_by_state TR
WHERE CD.test_date = TR.rally_date
AND CD.state = TR.state
ORDER BY TR.state, rally_date;
```



```
rally_date | state | positive
------------+-------+----------
 2020-06-23 | AZ    |    58179
 2020-08-18 | AZ    |   194920
 2020-10-19 | AZ    |   231897
 2020-10-28 | AZ    |   241165
 2020-09-24 | FL    |   684847
 2020-10-12 | FL    |   726934
 2020-10-16 | FL    |   739050
 2020-10-23 | FL    |   761924
 2020-10-29 | FL    |   784331
 2020-11-01 | FL    |   796802
 2020-10-16 | GA    |   337850
 2020-11-01 | GA    |   361982
 2020-10-14 | IA    |    97384
 2020-11-01 | IA    |   121967
 2020-09-10 | MI    |   120846
 2020-10-17 | MI    |   159119
 2020-10-27 | MI    |   182344
 2020-10-30 | MI    |   193388
 2020-11-01 | MI    |   197406
 2020-11-02 | MI    |   204326
 2020-08-17 | MN    |    65716
 2020-09-18 | MN    |    87807
 2020-09-30 | MN    |    99134
 2020-10-30 | MN    |   145465
 2020-09-08 | NC    |   178635
 2020-09-19 | NC    |   192248
 2020-10-15 | NC    |   238939
 2020-10-21 | NC    |   250592
 2020-10-24 | NC    |   258292
 2020-10-29 | NC    |   269021
 2020-11-01 | NC    |   276692
 2020-11-02 | NC    |   278028
 2020-10-27 | NE    |    64499
 2020-08-28 | NH    |     7216
 2020-10-25 | NH    |    10328
 2020-09-12 | NV    |    73220
 2020-09-13 | NV    |    73537
 2020-10-18 | NV    |    90261
 2020-09-21 | OH    |   145165
 2020-10-24 | OH    |   195806
 2020-06-20 | OK    |    10037
 2020-08-20 | PA    |   126940
 2020-09-03 | PA    |   136771
 2020-09-22 | PA    |   151646
 2020-09-26 | PA    |   155232
 2020-10-13 | PA    |   174646
 2020-10-20 | PA    |   184872
 2020-10-26 | PA    |   195695
 2020-10-31 | PA    |   208027
 2020-11-02 | PA    |   211996
 2020-09-25 | VA    |   144433
 2020-08-17 | WI    |    70715
```



```
2020-09-17 | WI    |   100574
2020-10-17 | WI    |   175227
2020-10-24 | WI    |   205139
2020-10-27 | WI    |   217429
2020-10-30 | WI    |   232062
2020-11-02 | WI    |   244928
```

5 Challenges

One of the challenges we faced while doing this project was how we could interpret the data and come up with ways to utilize the two tables we created. From the start, our team focused on seeing if we could draw any conclusions. Therefore we had to decide on queries which would show any possible correlations between our tables. This search for general conclusions allowed us to look at the data in many different ways. While at the same time, it made us not have a specific target, meaning we had to find our own ways to interpret the data. We chose to use coronavirus' general 2 week incubation period to determine whether or not Trump rallies had affected the communities where they had been held.

Another challenge we faced was the initial form of the data. It required us to change it into a UNIX encoded file format and on the COVID data the date was in the form dd-mm-yy, to fix this Joline, who loaded the data, changed her datastyle to 'dmy'

A very small challenge we encountered was being able to access the tables, this was solved quickly by allowing access to both the table and the schema. We also learned how to add days to a date type in SQL, this was used for populating the comparison table.

We also ran into general challenges of working with the SQL language and how to interpret some of the questions we proposed and therefore convert them into the correct queries used. This was seen in every aspect of the project, from creating the tables and working with them to ensure the compatibility of the ERD with the SQL tables, to the queries themselves and making sure the result they returned was accurate.

6 Conclusion

After looking through the different queries that our group tried and the results from the tables we got, there was a definite increase in the number of positive cases after each of the different rallies. In 40 out of the 58 instances, rates of increase were higher 2 weeks after the rally versus 2 weeks before, with 24 instances being between 1-5% higher, 6 instances between 6-10% higher, and 10 instances with 11% or higher rates 2 weeks after the rally versus 2 weeks before. We determined this by looking at the baseline for the number of cases in each state before the rally and 2 weeks afterwards. We also examined the states that were outliers in terms of new cases. One of the query results, the state that was hit the hardest, Michigan, seems to demonstrate that the rallies did have a major effect towards increasing the new cases in each state. However, this result cannot be fully attributed to the rallies that were held, as cases everywhere in the US have been growing fairly rapidly due to other causes, such as small gatherings. Because of this, although our data seems to support the idea, it is in no way conclusive.

Some other ways to expand the project if given the opportunity, would be to include additional queries, visualizations, and potentially other data points or even additional information regarding the rallies table. Additional queries would've allowed the group to address more questions and potentially branch out to other conclusions that could have been made after seeing the data pool. Some visualizations the group talked about were graphs and scatter plots that could have been created to visually see the impact that the rallies had on the test case numbers. Similarly to the example report given, an idea was proposed that we could show a "heat map" of the increase in test cases over a 2 week period. We could then use this info to see if a huge increase in new cases directly correlated with states that had a rally or even multiple rallies. An additional data column that could have been added but would have involved much more research would be the number of people that attended each of the rallies. This way, the group could see if this number scaled with the number of new test cases in each of the states that the rallies were held in. This would give additional information and support to the main idea that we were looking for. One last way to expand this project would be to see how extraneous variables affected the COVID data table, in terms of some of the columns we eliminated early on. For example, one of the columns was the number of people on ventilators and the number of people that recovered. Availability of resources to each state would be interesting to relate to these ideas and many different conclusions could be drawn.

Overall, this was an interesting experience to go through. Sorting through the vast amounts of information and data out there was a very unique way to familiarize ourselves with how databases work and all the nuances of how a database is constructed effectively. Given how much COVID-19 has affected everyone's lives and sense of normality, understanding a small portion of the data that is out there really brought this whole class to a satisfying conclusion.

Bibliography (MLA)

"Donald Trump 2020 Presidential Campaign." *Wikipedia*, Wikimedia Foundation, 20 Nov. 2020, en.wikipedia.org/wiki/Donald_Trump_2020_presidential_campaign.

"Our Data." *The COVID Tracking Project*, covidtracking.com/data.

World, Bruce PlanteTulsa. *Bruce Plante Cartoon: General COVID-19*. 24 Mar. 2020, theeagle.com/opinion/cartoons/bruce-plante-cartoon-general-covid-19/article_d58b25b9-950f-5714-a3fe-83c3d4fed62f.html.