Marc Laugharn

Covid-19 Lockdown Impact Model

I was interested in finding out the impact of lockdown on the acceleration of the daily cases per state in the United States (and Washington DC). I collected timeseries data from a wide variety of sources:

- Apple's COVID-19 Mobility Trends Data, specifically change in percent of driving from baseline, as recorded by Apple Maps usage
- Google's COVID-19 Mobility Trends Data, as percent change from baseline:
    - Retail and recreation
    - Grocery and pharmacies
    - Parks
    - Transit stations
    - Workplaces
    - Residential areas
- COVID Act Now's daily data, including:
    - Hospital beds required and in use
    - ICU Beds capacity and in use
    - Ventilators capacity and in use
    - Real time estimate of R_0 value
    - Cumulative dead, infected
    - Cumulative positive, negative tests
- Lockdown level by state, where:
    0. No or few containment measures
    1. Ban on public gatherings, cancellation of major events
    2. Schools and universities closed
    3. Nonessential shops, restaurants and bars closed
    4. Night curfew/partial lockdown
    5. All-day lockdown: shelter in place order, citizens allowed to leave home
    6. Harsh lockdown: citizens not allowed to leave home

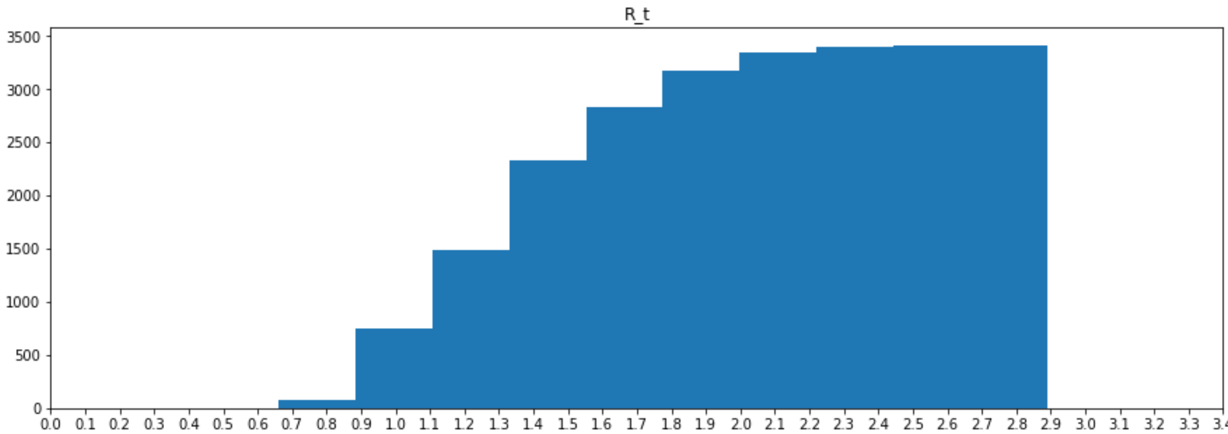
I also collected various non-time-series data per region:

- US Census health and poverty data (including obesity rate, smoking rate, income, poverty rate, healthcare access)
- Density per square mile
- 2018 population
- Percentage of people living in urban environments

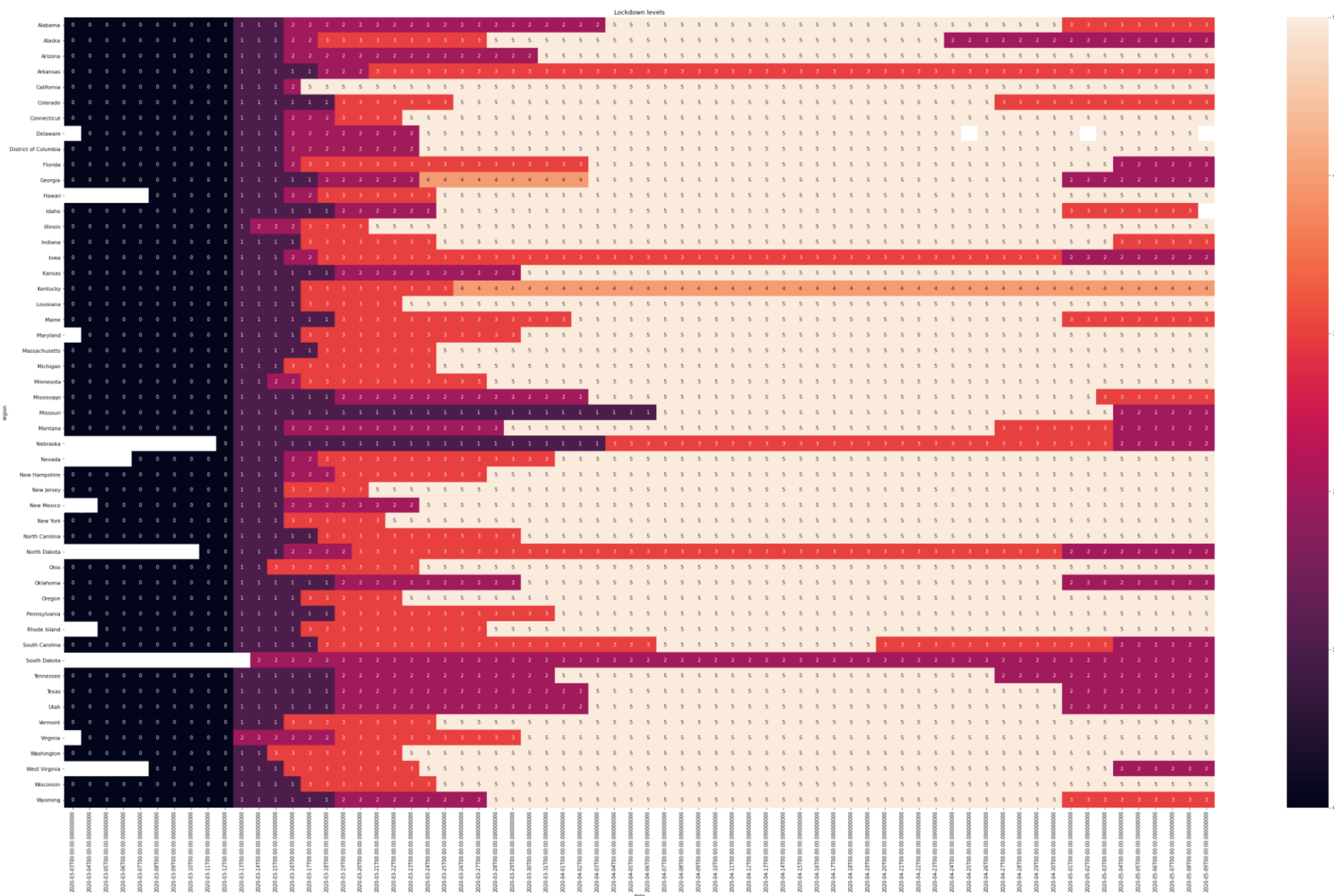
To produce the daily case numbers, I took both the 2nd order differences of the cumulative infected and their 7-day exponentially-weighted moving average. I took the differences of those, and used those as the infection acceleration target variable.

EDA:

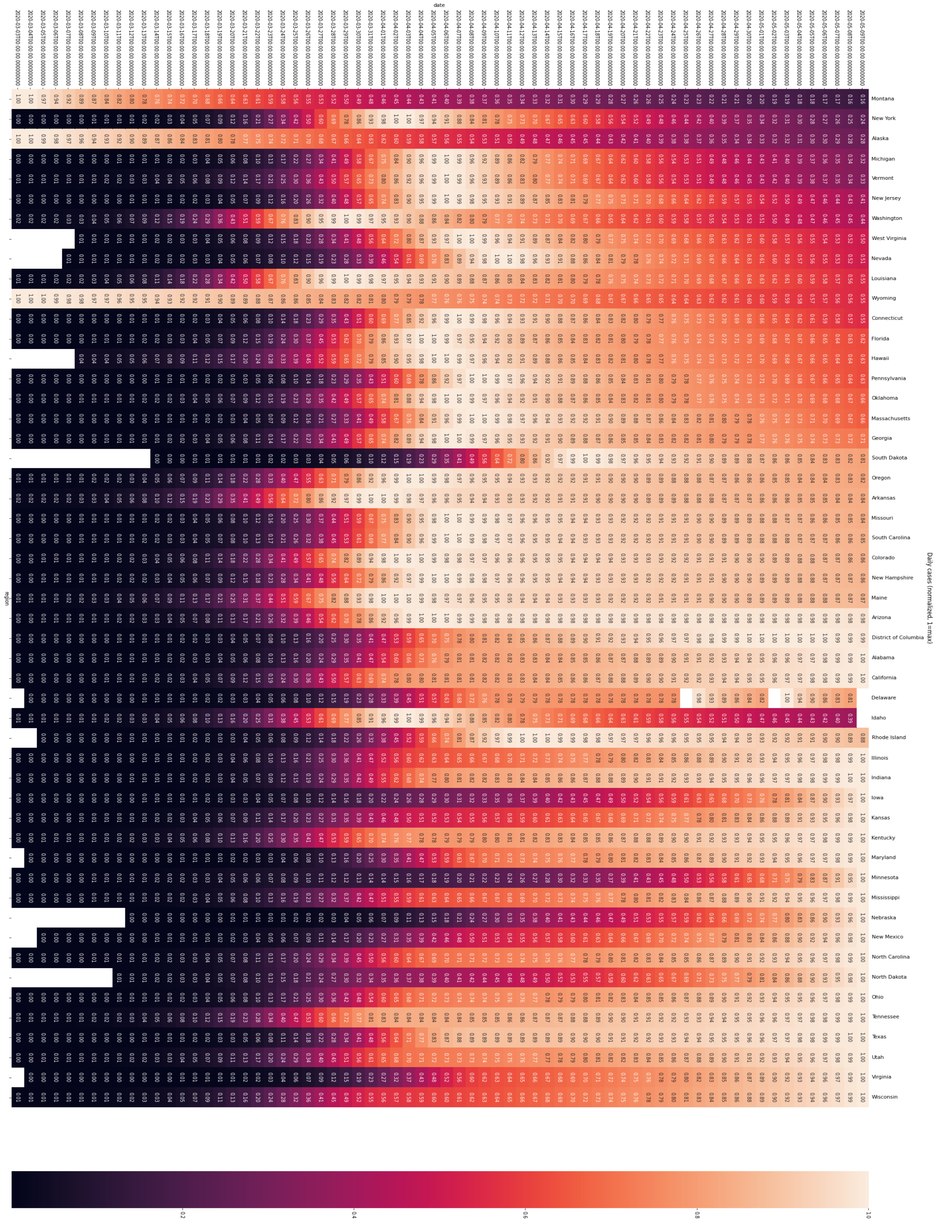
I found that, for the vast majority of days, the R_t value was greater than 1.0:



I was able to get 55 days of data spanning all states across all my datasets.



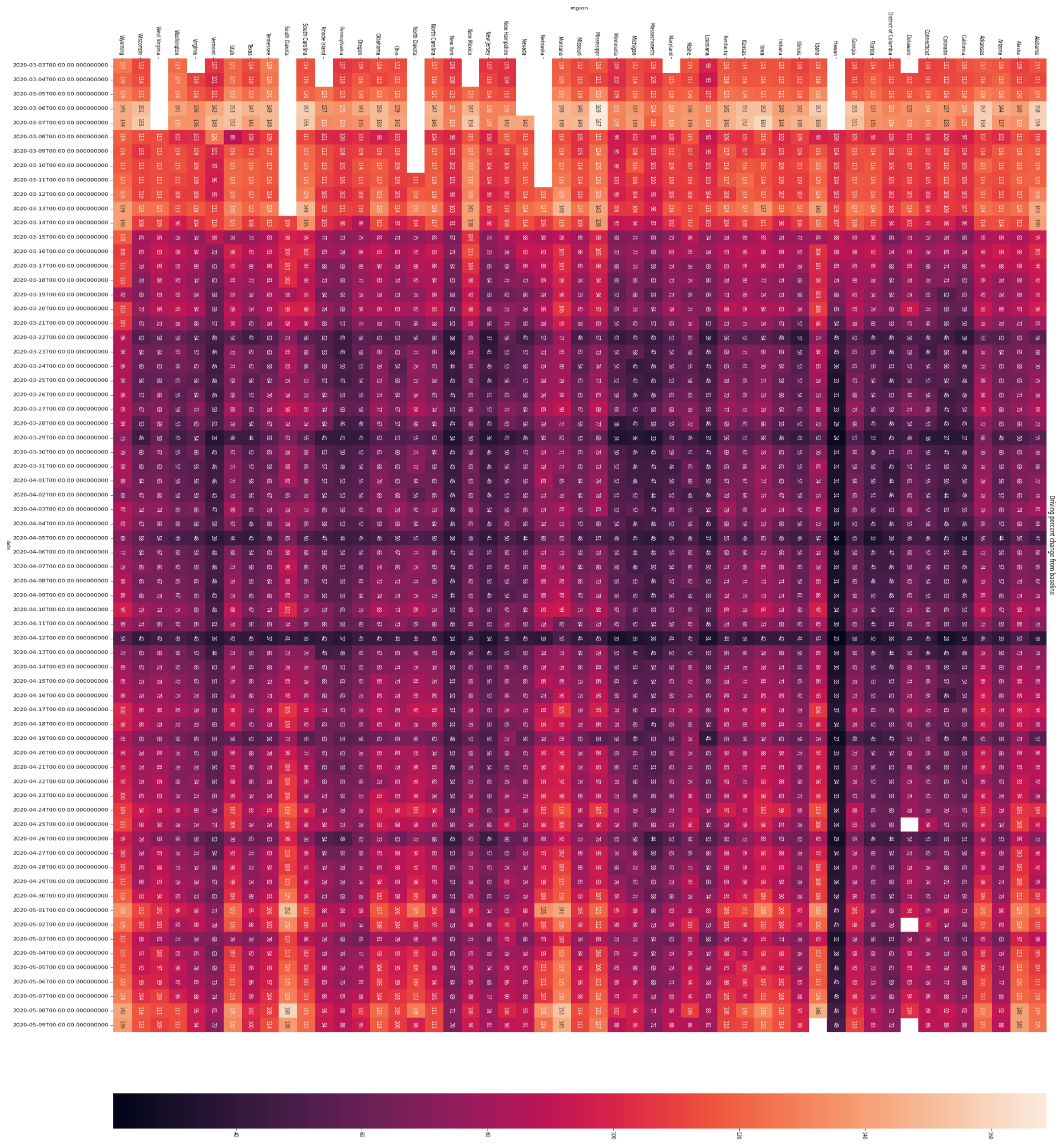
I plotted the lockdowns over time. The majority of states are at, or have been until very recently, a level 5 lockdown, though it is quite evident some states are reopening at the moment.



date
2020-03-03T00:00:00.000000000
2020-03-04T00:00:00.000000000
2020-03-05T00:00:00.000000000
2020-03-06T00:00:00.000000000
2020-03-07T00:00:00.000000000
2020-03-08T00:00:00.000000000
2020-03-09T00:00:00.000000000
2020-03-10T00:00:00.000000000
2020-03-11T00:00:00.000000000
2020-03-12T00:00:00.000000000
2020-03-13T00:00:00.000000000
2020-03-14T00:00:00.000000000
2020-03-15T00:00:00.000000000
2020-03-16T00:00:00.000000000
2020-03-17T00:00:00.000000000
2020-03-18T00:00:00.000000000
2020-03-19T00:00:00.000000000
2020-03-20T00:00:00.000000000
2020-03-21T00:00:00.000000000
2020-03-22T00:00:00.000000000
2020-03-23T00:00:00.000000000
2020-03-24T00:00:00.000000000
2020-03-25T00:00:00.000000000
2020-03-26T00:00:00.000000000
2020-03-27T00:00:00.000000000
2020-03-28T00:00:00.000000000
2020-03-29T00:00:00.000000000
2020-03-30T00:00:00.000000000
2020-03-31T00:00:00.000000000
2020-04-01T00:00:00.000000000
2020-04-02T00:00:00.000000000
2020-04-03T00:00:00.000000000
2020-04-04T00:00:00.000000000
2020-04-05T00:00:00.000000000
2020-04-06T00:00:00.000000000
2020-04-07T00:00:00.000000000
2020-04-08T00:00:00.000000000
2020-04-09T00:00:00.000000000
2020-04-10T00:00:00.000000000
2020-04-11T00:00:00.000000000
2020-04-12T00:00:00.000000000
2020-04-13T00:00:00.000000000
2020-04-14T00:00:00.000000000
2020-04-15T00:00:00.000000000
2020-04-16T00:00:00.000000000
2020-04-17T00:00:00.000000000
2020-04-18T00:00:00.000000000
2020-04-19T00:00:00.000000000
2020-04-20T00:00:00.000000000
2020-04-21T00:00:00.000000000
2020-04-22T00:00:00.000000000
2020-04-23T00:00:00.000000000
2020-04-24T00:00:00.000000000
2020-04-25T00:00:00.000000000
2020-04-26T00:00:00.000000000
2020-04-27T00:00:00.000000000
2020-04-28T00:00:00.000000000
2020-04-29T00:00:00.000000000
2020-04-30T00:00:00.000000000
2020-05-01T00:00:00.000000000
2020-05-02T00:00:00.000000000
2020-05-03T00:00:00.000000000
2020-05-04T00:00:00.000000000
2020-05-05T00:00:00.000000000
2020-05-06T00:00:00.000000000
2020-05-07T00:00:00.000000000
2020-05-08T00:00:00.000000000
2020-05-09T00:00:00.000000000
region
Montana
New York
Alaska
Michigan
Vermont
New Jersey
Washington
West Virginia
Nevada
Louisiana
Wyoming
Connecticut
Florida
Hawaii
Pennsylvania
Oklahoma
Massachusetts
Georgia
South Dakota
Oregon
Arkansas
Missouri
South Carolina
Colorado
New Hampshire
Maine
Arizona
District of Columbia
Alabama
California
Delaware
Idaho
Rhode Island
Illinois
Indiana
Iowa
Kansas
Kentucky
Maryland
Minnesota
Mississippi
Nebraska
New Mexico
North Carolina
North Dakota
Ohio
Tennessee
Texas
Utah
Virginia
Wisconsin
Daily cases (normalized, 1=max)
0.2
0.4
0.6
0.8
1.0

I plotted a heatmap of the daily cases over time, where each state's cases have been normalized to their maximum. As you can see, the vast majority of states are still above half of their max daily cases. Currently, 20 states are actually right now at their peak in daily cases.

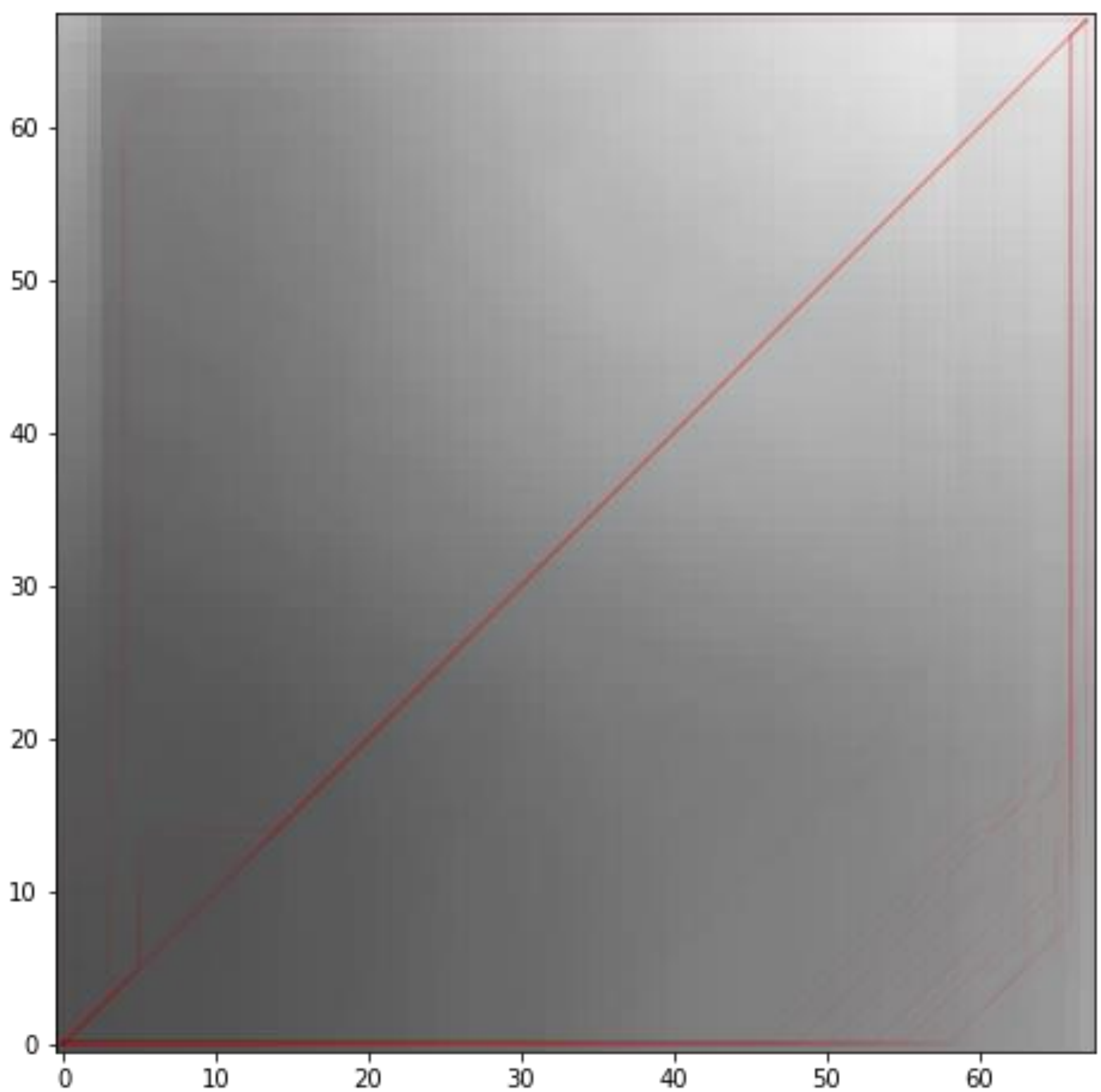
I also wanted to visualize the impact of the lockdowns on driving behavior.



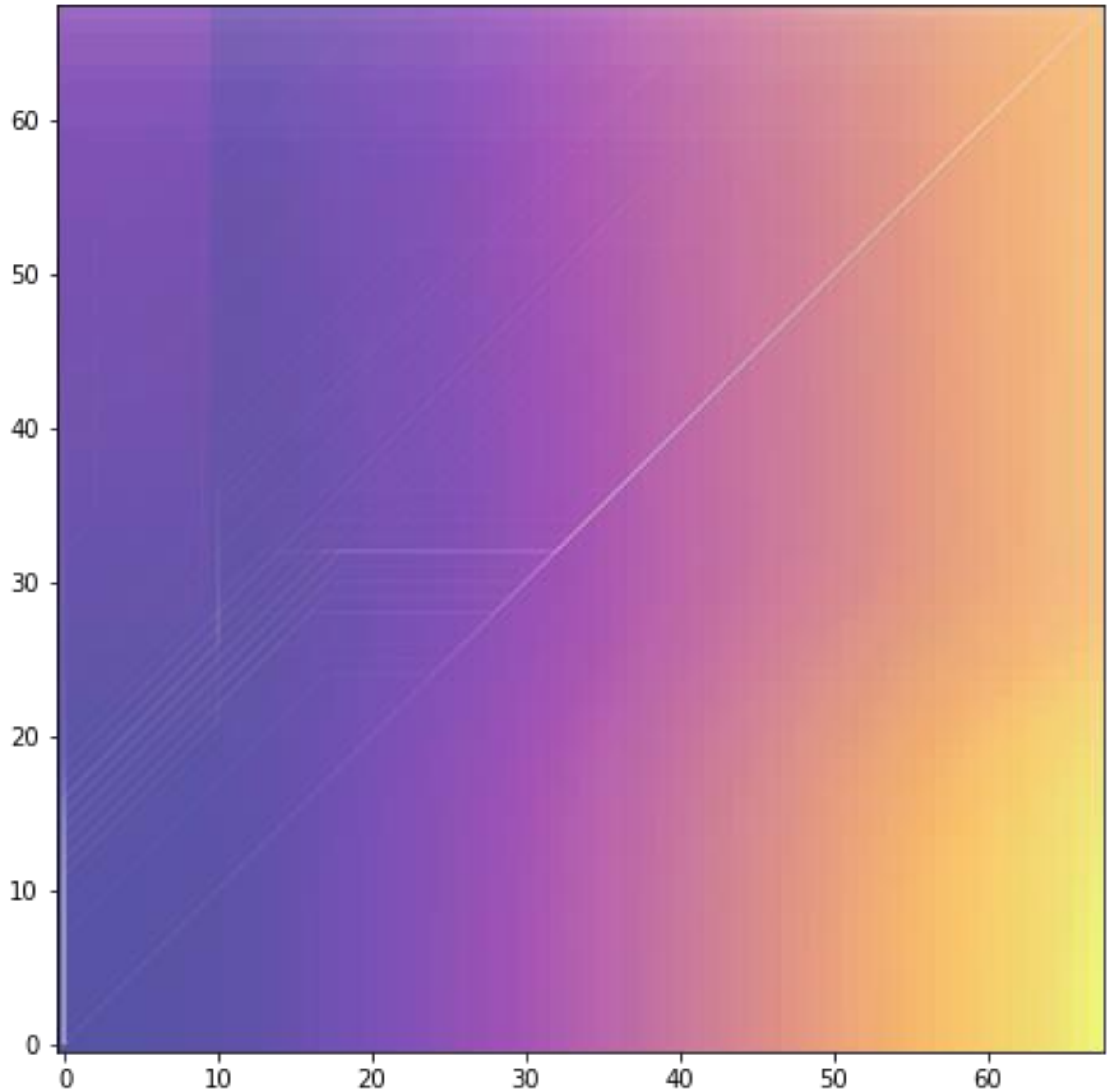
It is quite apparent from this that people significantly reduced their driving. It also seems people ramped up their driving in preparation of the lockdowns, driving quite significantly more than usual for a week or so before the lockdowns. It also seems that March 15th was the tipping point for driving percentage for the majority of states. It appears as though people were voluntarily altering their behavior before the harshest lockdown measures had arrived. Also, this heatmap clearly shows that April 12th was the lowest driving percentage day of all days of the crisis- April 12th , Easter Sunday this year, famously being President Trump's initially-hoped day for the re-opening of the country.

I also tried to do some dynamic-time-warping alignments between driving percent and daily cases, lockdown level and daily cases and driving percent and lockdown level, but they were fairly opaque for concluding things apart from that in the majority of cases, the best alignment of sequences was linear:

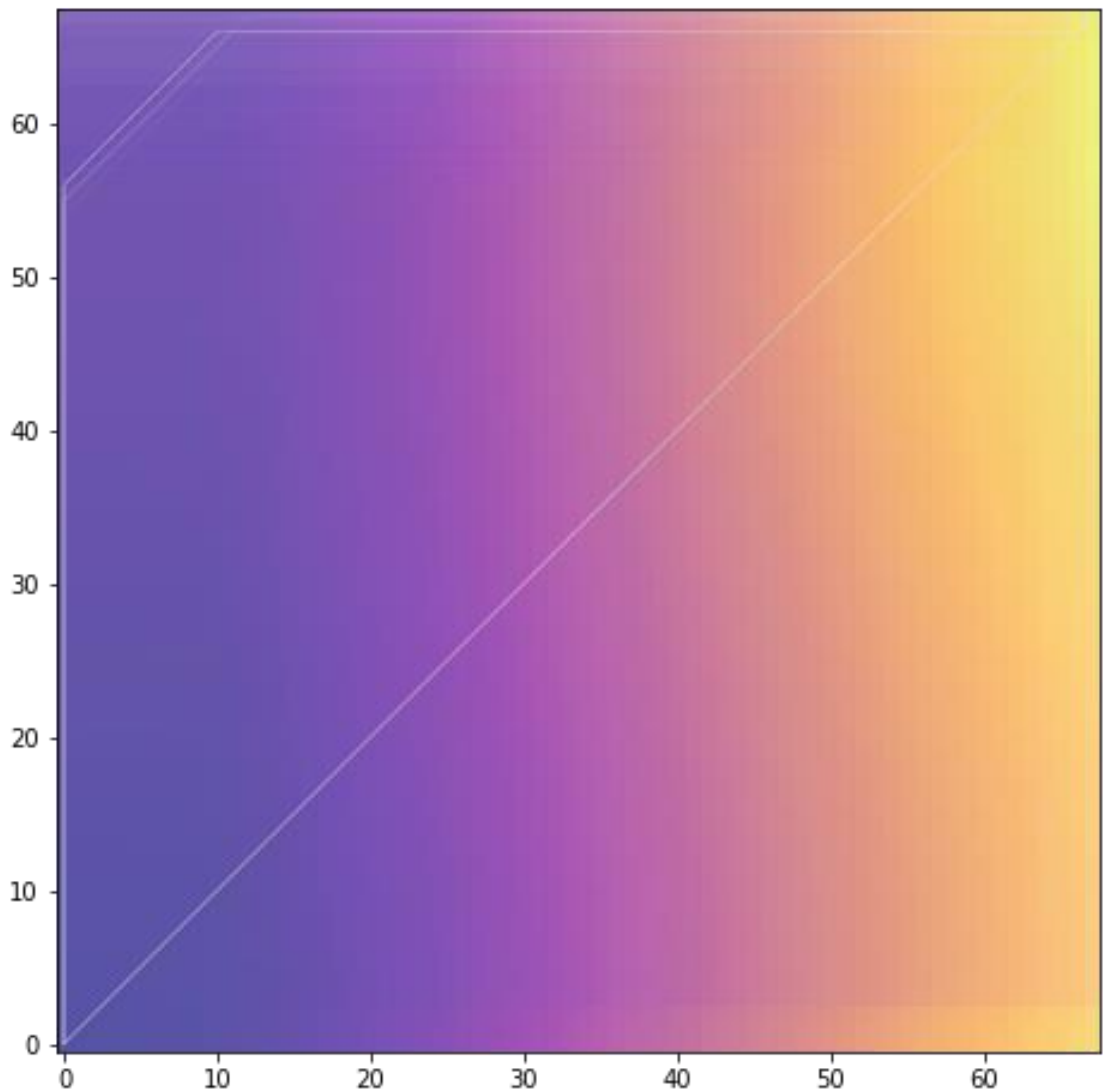
Driving percent (X) vs Daily Cases (Y) min-cost alignment curves (mean accumulated cost matrix is background)



Lockdown level vs daily cases alignment curves on top of mean accumulation matrix costs



Lockdown level vs driving percent mean alignment curves on top of mean accumulated cost of alignment



I wanted to see how good a model I could get from these data. At first I tried doing a regression for the R_t number, but this was a failure. I also tried to do an ordinal regression on R_t classes; e.g. predict 0 if $Rt < 0.8$, predict 1 if $0.8 < R\_t < 1$, predict 2 if $1 < Rt < 1.2$, etc until predict 5 for $1.7 < Rt$, but this was not easy either.

Then I tried to do an LSTM model, which worked alright, but it wasn't that good. I did models at both the regional and the national level.

The last model that I came up with was a combination of a Temporal Convolutional Network (which I learned is a good architecture for timeseries) given the timeseries data as inputs, and another set of inputs, the features that are not time-varying in my dataset (e.g. the obesity level as measured in 2010, or the urbanization percent for the state)

The inputs were as follows:

Time-varying inputs:
```
[
'driving_percent',
'retail_and_recreation_percent_change_from_baseline',
'grocery_and_pharmacy_percent_change_from_baseline',
'transit_stations_percent_change_from_baseline',
'workplaces_percent_change_from_baseline',
'residential_percent_change_from_baseline',
'lockdown_level_0',
'lockdown_level_1',
'lockdown_level_2',
'lockdown_level_3',
'lockdown_level_4',
'lockdown_level_5',
'lockdown_level_6',
'lockdown_level_0_sum',
'lockdown_level_1_sum',
'lockdown_level_2_sum',
'lockdown_level_3_sum',
'lockdown_level_4_sum',
'lockdown_level_5_sum',
'lockdown_level_6_sum',
'hospitalBedsRequired',
'ICUBedsInUse',
'positive_test_rate',
'cumulativeDeaths', 'cumulativeInfected',
'daily_cases',
'daily_cases_ewm_avg',
'RtIndicator',
'cumulativePositiveTests', 'cumulativeNegativeTests'
]
```

Non-time-varying inputs:
```
['lat', 'long',
'hospitalBedCapacity',
'ICUBedCapacity',
'poverty', 'age', 'income',
'healthcare', 'healthcareLow', 'healthcareHigh',
'obesity', 'smokes', 'smokesLow', 'smokesHigh',
'Density per square mile of land area',
'2018 Population',
'urbanization']
```

The target variable was whether the last 7-days-exponentially-weighted daily cases was accelerating or not.

The inputs 'lockdown_level_i' were indicator variables. I made lockdown level 0 equal to 1 if no lockdown measures were in place. If for e.g. lockdown level 4 were in place, I made lockdown level 0 equal 0, and made every lockdown_level_i for 1<=i<=4 equal to 1. Lockdown_level_i_sum is the number of days that region has been at that lockdown level over time.

Here is my model architecture:

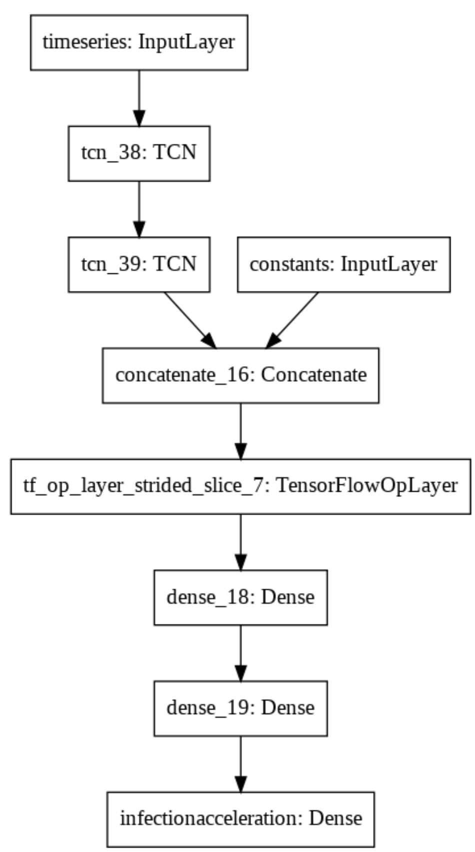
For temporal inputs:

1. TCN layer with kernel size = 7, dilations = [1,2,3,4,5,8,16,32], 2 stacks, returning sequences
2. TCN layer with kernel size = 3, dilations = [1,2,4,8,16,32,64], 1 stack, returning sequences

Then I concatenated the output of the last TCN layer with the constant features into one tensor. Then with that vector I:

1. Took the last element of the sequence vector (aka index [:, -1, :])
2. Fed it through 100-neuron fully-connected layer with ReLU activation
3. Fed it through 100-neuron fully-connected layer with ReLU activation again
4. Lastly ran through a 1-neuron fully-connected output layer with sigmoid activation

I used binary cross entropy to train the model.



I fed it sequences of 5 timesteps of data at a time. I held out the most recent 10 days' worth of data as a test set. At first, I used 10% of the train set for model validation purposes, but at test time then just used the entire training set. This meant roughly 45 days for training, and of the 10 days of the test set, a sequence of 5 days was used for data, and 10-5-1 (minus 1 day because of predicting the output *after* the sequence input) = 4 days were predicted.

I used the Adam optimizer with a learning rate of 3e-3. I tracked the accuracy, precision, and recall per epoch, for 100 epochs. The precision is (true positive)/(actual results) and the recall is (true positive)/(predicted results).

It was important to me to have a high recall, because it would indicate the ratio of correct infection acceleration predictions to false infection deceleration predictions would be high. It is better to tell people it is not quite safe yet when it is actually safe, than to tell them it is now safe when it is actually dangerous.

Note: I initially mistakenly trained the model on batches on size 1, greatly hampering its performance.

Here are the performance metrics:

Lastly I grouped the states by their latest lockdown level. Per lockdown level, I determined which were predicted to have decelerating daily cases, and which were predicted to have accelerating daily cases.

The accuracy on the test set was 63.7%, precision 62.9%, and recall 77.7%. Here were the states' predicted daily case acceleration outcomes, indexed by by lockdown level and then by result, where the result is mapped by {0: decelerating, 1: accelerating}:

```
{0: {0: [], 1: []},
 1: {0: [], 1: []},
 2: {0: ['South Carolina', 'Montana', 'West Virginia'],
     1: ['Texas',
         'North Dakota',
         'Missouri',
         'Oklahoma',
         'South Dakota',
         'Nebraska',
         'Florida',
         'Utah',
         'Georgia',
         'Iowa',
         'Tennessee',
         'Alaska']},
 3: {0: ['Maine', 'Wyoming', 'Idaho'],
     1: ['Arkansas', 'Colorado', 'Indiana', 'Alabama', 'Mississippi']},
 4: {0: [], 1: ['Kentucky']},
 5: {0: [],
     1: ['Virginia',
         'Illinois',
         'Louisiana',
         'Oregon',
         'Hawaii',
         'Maryland',
         'New Jersey',
         'Pennsylvania',
```

```
         'Kansas',
         'Vermont',
         'Nevada',
         'Rhode Island',
         'Massachusetts',
         'Ohio',
         'Michigan',
         'District of Columbia',
         'Arizona',
         'New York',
         'Washington',
         'Wisconsin',
         'North Carolina',
         'Delaware',
         'Connecticut',
         'Minnesota',
         'California',
         'New Mexico',
         'New Hampshire']},
 6: {0: [], 1: []}}
```

This paints a fairly bleak picture in my view, but I definitely think these results are to be taken with a grain of salt.. Also I don't know if this model lets us say anything about lockdown level and its impact on case acceleration, other than most states are in quite strict lockdown, and it may be that things will get worse for them in the short term.