Establishing Localized Safe Zones to Mitigate Against the Spread of a Pandemic

Michael Palmer

April 4, 2020

Introduction

Background

The global coronavirus pandemic of 2020 has caused widespread infection of people, along with many deaths. One strategy that appears to have been at least partially effective in combatting the spread of the virus has been to issue 'stay at home' orders. While these are effective, the toll on businesses and the people that own or work at them has been great. Businesses have been forced to close, resulting in record unemployment levels in the United States, along with the resulting effects on people's incomes and their mental health.

On April 4, 2020, President Trump said in a press conference that it was imperative to get to a point where life can get back to normal. He highlighted restaurants, sports venues and other social gathering locations in particular as being desirable to reopen as soon as possible.

This study will look into the feasibility of establishing 'safe zones' – parts of the country that are - and can be kept - free of the virus. People who have tested negative for the virus could be moved from nearby hot zones into the safe zones, and in the event that someone in the safe zone test positive, they can be moved out of the safe zone.

Obviously, such a proposal has implications beyond the data, not least the ethical and legal aspects, but I leave such issues to the philosophers and lawmakers.

Problem

While the idea may sound simple, there are many factors that need to be taken into account in deciding where a safe zone can be established. It should be a location currently free of the virus (or with very few cases) where the population numbers and density, availability of temporary housing for people coming into the safe zone, availability of businesses to support the population, recreational facilities and healthcare services will all play a factor.

Within a safe zone, it would be desirable for life to go on as normal, as much as possible. Ideally, a safe zone would allow people to go to work and enjoy social and recreational activities as they would normally. It will include healthcare facilities such as hospitals, and other essential businesses including grocery stores.

Interest

The results of this study will be of interest to state and federal governments and may help to design a template for dealing with future pandemics, or at least a talking point to help design plans for the future.

The Data

Data Sources

Studying and modelling such a proposition would be impossible without data to draw insights from, and the current COVID-19 pandemic allows us to access real pandemic data. In this study, I will be utilizing the following data:

For the numbers of confirmed cases by US county, I have chosen the time_series_covid19_confirmed_US.csv dataset maintained by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) at https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series.

Social and economic data is available from the US Census Bureau at https://www.census.gov/data/ in both downloadable datasets and via API access. This will provide data on population numbers, population density and other economic factors.

Official US health, local government and public safety datasets are available at https://www.data.gov/. These will give insights on the availability and capabilities of hospitals, other healthcare services and public safety services such as emergency services.

Local venue and neighborhood data will be accessed via the Foursquare API to allow exploration of potential safe zones for the number of venues such as stores, cinemas, restaurants and coffee shops.

Data on temporary housing such as RV parks and hotels will be accessed via the Yelp API.

Data Acquisition and Cleaning

I have downloaded the *time_series_covid19_confirmed_US.csv* dataset for April 5, 2020 maintained by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) from https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series. The data contains a large number of columns:

```
Index(['UID', 'iso2', 'iso3', 'code3', 'FIPS', 'Admin2', 'Province_State',
       'Country_Region', 'Lat', 'Long_', 'Combined_Key', '1/22/20', '1/23/20',
       '1/24/20', '1/25/20', '1/26/20', '1/27/20', '1/28/20', '1/29/20',
       '1/30/20', '1/31/20', '2/1/20', '2/2/20', '2/3/20', '2/4/20', '2/5/20',
       '2/6/20', '2/7/20', '2/8/20', '2/9/20', '2/10/20', '2/11/20', '2/12/20',
       '2/13/20', '2/14/20', '2/15/20', '2/16/20', '2/17/20', '2/18/20',
       '2/19/20', '2/20/20', '2/21/20', '2/22/20', '2/23/20', '2/24/20',
       '2/25/20', '2/26/20', '2/27/20', '2/28/20', '2/29/20', '3/1/20',
       '3/2/20', '3/3/20', '3/4/20', '3/5/20', '3/6/20', '3/7/20', '3/8/20',
       '3/9/20', '3/10/20', '3/11/20', '3/12/20', '3/13/20', '3/14/20',
       '3/15/20', '3/16/20', '3/17/20', '3/18/20', '3/19/20', '3/20/20',
       '3/21/20', '3/22/20', '3/23/20', '3/24/20', '3/25/20', '3/26/20',
       '3/27/20', '3/28/20', '3/29/20', '3/30/20', '3/31/20', '4/1/20',
       '4/2/20', '4/3/20', '4/4/20', '4/5/20'],
```

For the purposes of this study, all that was required was *Admin2* (the county name), *Province_State* (the state name), and most recent number of cases in *4/5/20*. I copied those columns to a new dataframe and renamed them.

	county	state	cases
0	NaN	American Samoa	0
1	NaN	Guam	112
2	NaN	Northern Mariana Islands	6
3	NaN	Puerto Rico	475
4	NaN	Virgin Islands	42

For data on hospitals, I downloaded the *Hospital_General_Information.csv* dataset from Data.gov. Again, this contained several columns that were not required:

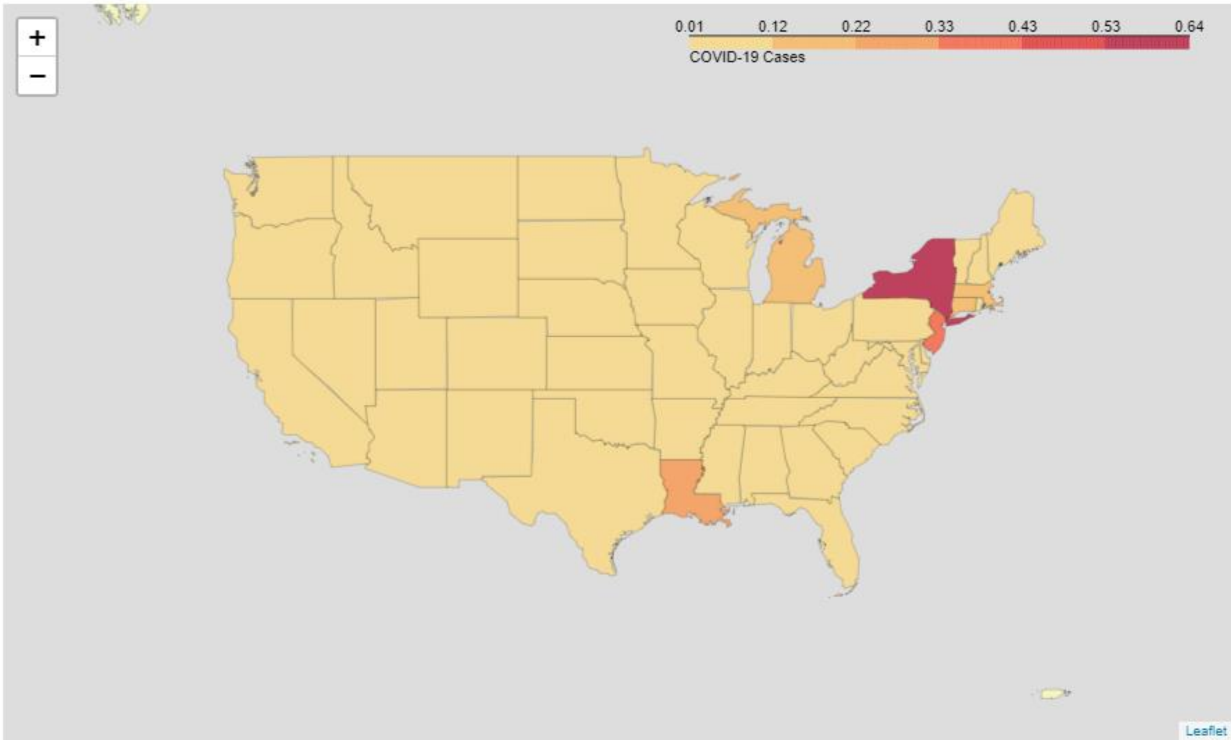
```
Index(['Facility ID', 'Facility Name', 'Address', 'City', 'State', 'ZIP Code',
       'County Name', 'Phone Number', 'Hospital Type', 'Hospital Ownership',
       'Emergency Services', 'Meets criteria for meaningful use of EHRs',
       'Hospital overall rating', 'Hospital overall rating footnote',
       'Mortality national comparison',
       'Mortality national comparison footnote',
       'Safety of care national comparison',
       'Safety of care national comparison footnote',
       'Readmission national comparison',
       'Readmission national comparison footnote',
       'Patient experience national comparison',
       'Patient experience national comparison footnote',
       'Effectiveness of care national comparison',
       'Effectiveness of care national comparison footnote',
       'Timeliness of care national comparison',
       'Timeliness of care national comparison footnote',
       'Efficient use of medical imaging national comparison',
       'Efficient use of medical imaging national comparison footnote',
       'Location'],
```

I copied the Facility Name, Address, City, State, ZIP Code, County Name, Hospital Type, Emergency Services, and Location to another dataframe.

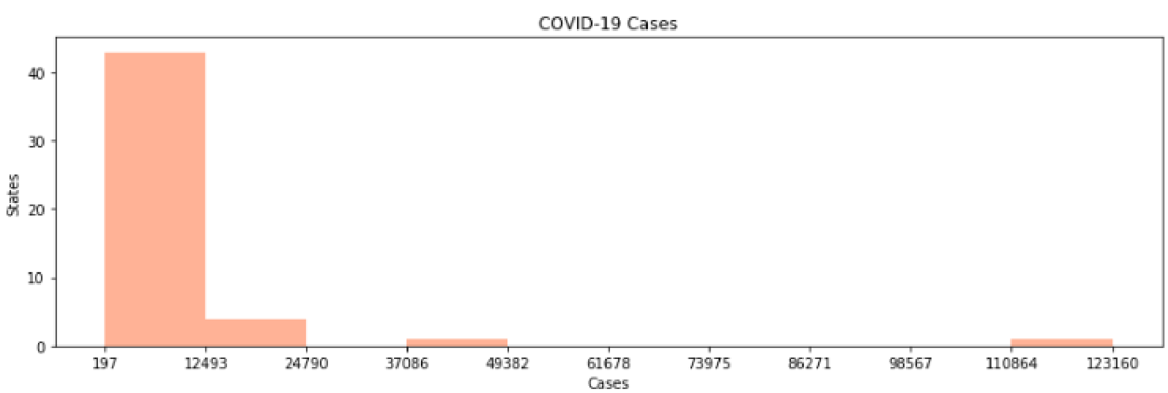
Methodology

Looking at the data for COVID-19 cases, I could see that it includes off-shore locations, which can be discounted for the purpose of this study. I removed American Samoa, Northern Mariana Islands, Virgin Islands, Guam, Puerto Rico and Hawaii. I also removed Diamond Princess, Grand Princess and Alaska for the same reason.

Grouping the COVID-19 cases by state to see which states have the lowest number of cases, there was a wide range of cases from 197 to 123,160, so I decided to look at cases by percentage of the state population. To do this, I downloaded population figures for each state from the US Census Bureau via an API call. I then calculated the number of cases as a percentage of the population and created a new dataframe and mapped the results on a Choropleth map.



This showed that most of the states had less than 0.12% of the population infected, so I binned the data and plotted to see where most of the figures lay.

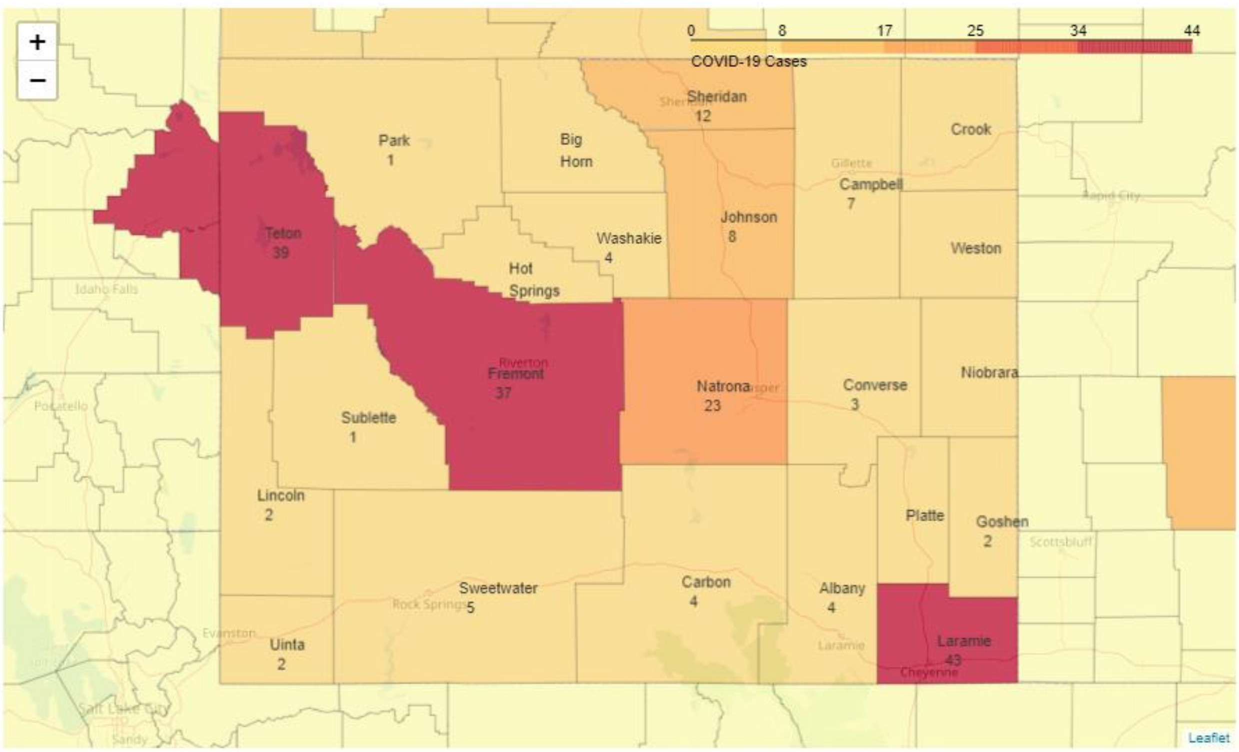


	cases	pop	percentage
count	49.000000	4.900000e+01	49.000000
mean	6851.469388	6.654941e+06	0.079796
std	18155.580829	7.428653e+06	0.107442
min	197.000000	5.787590e+05	0.020000
25%	837.000000	1.934408e+06	0.030000
50%	1855.000000	4.648794e+06	0.040000
75%	5675.000000	7.614893e+06	0.080000
max	123160.000000	3.951222e+07	0.630000

Looking at the data, 25% of states have less than 0.03% of the population affected, so I created a data frame showing those states:

	state	cases	pop	percentage
0	Wyoming	197	578759	0.03
1	North Dakota	207	762062	0.03
2	South Dakota	240	884659	0.03
3	Montana	286	1068778	0.03
4	West Virginia	324	1792147	0.02

Wyoming had the lowest number of cases so I theorized it might be a good state. I used the data from the original COVID-19 dataframe to map the counties and number of cases in each.



Although there were a few counties with zero cases, those counties did not contain major cities, so I chose Uinta, Sweetwater and Albany for comparison.

One of the criteria for establishing a county-level safe zone is the availability of shops, restaurants and other venues to support a normal way of life. To gauge this, I used the Foursquare API to check the major cities in each county: Rock Springs in Sweetwater County, Evanston in Uinta County, and Laramie in Albany County. Foursquare returned just 4 venues for Rock Springs, 25 for Evanston and 63 for Laramie. This suggested that Uinta and Albany seemed to be the better counties but needed to match the next criteria – availability of hospitals.

Taking the *Hospital_General_Information.csv* data, I filtered it to the three counties of interest.

	Facility Name	Address	City	State	ZIP Code	County Name	Hospital Type	Emergency Services	Location
1229	MEMORIAL HOSPITAL SWEETWATER COUNTY	1200 COLLEGE DRIVE	ROCK SPRINGS	WY	82901	SWEETWATER	Acute Care Hospitals	True	POINT (-109.233578 41.58406)
1882	WYOMING STATE HOSPITAL	831 HWY 150 SOUTH	EVANSTON	WY	82931	UINTA	Psychiatric	False	POINT (-110.946273 41.258885)
2095	ASPEN MOUNTAIN MEDICAL CENTER	4401 COLLEGE DRIVE	ROCK SPRINGS	WY	82901	SWEETWATER	Acute Care Hospitals	False	POINT (-109.247917 41.603187)
3726	IVINSON MEMORIAL HOSPITAL	255 N 30TH	LARAMIE	WY	82072	ALBANY	Acute Care Hospitals	True	POINT (-105.557512 41.31139)
4933	EVANSTON REGIONAL HOSPITAL	190 ARROWHEAD DR	EVANSTON	WY	82930	UINTA	Acute Care Hospitals	True	POINT (-110.991336 41.242214)

Sweetwater looked promising with two acute care hospitals, but it had the fewest Foursquare venues. Both Uinta and Albany have a hospital providing emergency services and a good number of venues.

To further narrow down the choice, I used the Yelp API to look at temporary housing like hotels and RV parks. That showed that Uinta has 4 RV parks and 32 hotels, and Albany has 1 RV parks and 39 hotels.

Results

Having narrowed it down, both Uinta County and Albany County, Wyoming, are candidates for a county-level safe zone.

Discussion

In this study I looked at the available data on COVID-19 cases, and where it may be possible to establish a county-level safe zone.

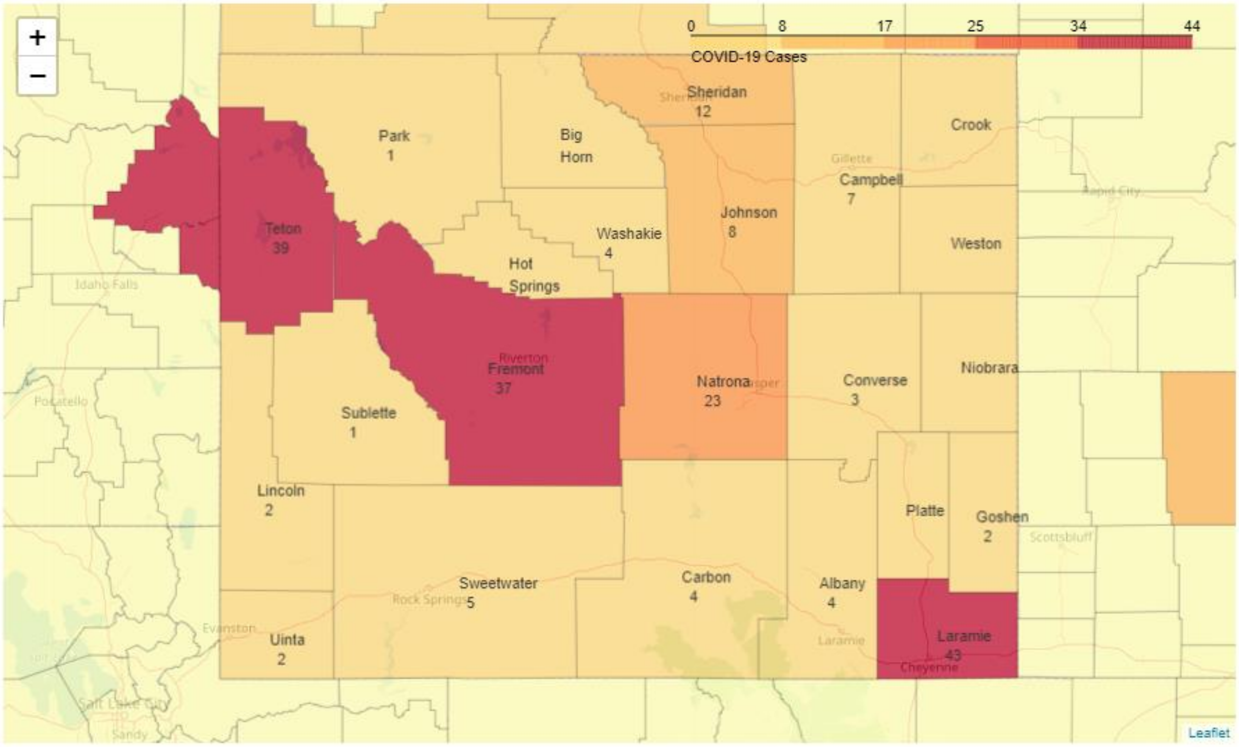
The criteria I chose were fairly easy to meet, and much stricter criteria could be introduced using the same methodology.

Likewise, although I looked only at one state – Wyoming – the process could be repeated for any – or every – state. The available data allows in-depth investigation and could be taken to further levels if desired. And although I used my eyes to see that Albany County is next to Laramie county with 43 cases, that could have been determined programmatically.

The study itself, while a theoretical exercise, could be adapted to develop a 'playbook' for not only COVID-19, but for future pandemics or other national or statewide emergencies, such as wildfires.

Conclusion

In the end, I have chosen Uinta County as the Wyoming Safe Zone. Although Albany County has more venues listed in Foursquare, it is next to Laramie County which has 43 confirmed cases of COVID-19.



0 8 17 25 34 44
COVID-19 Cases
Sheridan
12
Park
1
Big
Horn
Crook
Campbell
7
Johnson
8
Teton
39
Washakie
4
Hot
Springs
Weston
Fremont
37
Natrona
23
Converse
3
Niobrara
Sublette
1
Lincoln
2
Platte
Goshen
2
Sweetwater
5
Carbon
4
Albany
4
Laramie
43
Uinta
2
Leaflet