Protest Activity During the Year of 2020

Background

On May 25th, George Floyd was murdered by police officer Derek Chauvin while he was being accused of trying to pass a counterfeit twenty dollar bill. The video of the incident went viral and this killing sparked massive protests across the United States. In the two weeks following George Floyd's death, polls reported that between 15 and 26 million people participated in protests related to police misconduct (link).

Given that background, we explored the circumstances surrounding the protests in the year of 2020. We looked at the frequency of protests over the year, and found a spike and lasting effect after May 2020. Inspired by this observation, we have conducted an exploratory study of these protests and our results are described in this report.

Datasets

The Crowd Counting Consortium: Collects publicly available data on political crowds reported in the United States. Consortium that was originally established to generate accurate counts of the pro-Washington. It is maintained by a team of researchers and volunteers. Dataset contains 67 fields and 84,662 records. 42,221 records

US Census Bureau: We looked at state and county population data. Joins were possible using state acronyms or county FIPS codes.

US Bureau of Labour Statistics: We looked at unemployment data. Data is available by state/territory and by year. Since we looked at 2020, this secondary dataset had just over 50 records.

Exploratory Questions

We seeked to answer the following questions by joining and analyzing the datasets listed above. The following four question categories guided us through this report.

1. How were protests distributed across the United States?
    a. In size (protest turnout)
    b. In frequency (number of protests per day)
    c. Geographically
2. Is there a difference in the geographic distribution of protests of opposite political valence?
3. Is there an association between:
    a. The frequency of protests and unemployment rates?
    b. The size of protests and unemployment rates?
    c. The frequency of protests and population?
    d. The size of protests and population?

Variables & Sanity Checks

After our join but before our analysis, we limited our protests dataset to those protests that happened within our time interval of interest (the year of 2020). Then we selected our variables of interest (listed below) and we performed some sanity checks to confirm this information.

The variables explored in this project were:

- "date": Datetime of protest event. Date range is only within 2020.
- "locality": The name of the city where the protest took place.
- "state": Abbreviations. There are 54 because of some US territories.
- "valence": Indicates whether the left-leaning ("1") or right-leaning ("2").
- "issues": Overarching issue topic that is a theme of the protest.
- "claims": Details about the protester's stance on the cited issue.
- "size_mean": An estimate of the amount of participants in the protest.
- "size_cat": Categorizes "size_mean" into three size categories and a category for no data.
- "lat": Latitude value of the location in decimal degrees.
- "lon": Longitude value of the location in decimal degrees.
- "county_y": The name of the local county.
- "fips_code": This is a 4-5 digit code (string) that indicates the local county.
- "state_unemp_rate": Values are for state and percentages rounded are to the nearest tenth.

In addition, we also learned from our sanity checks that protests that care about the same issue can be either left-leaning or right-leaning. For example, claims that include "Black Lives Matter" and "Blue Lives Matter" have opposite valence values but cite the same issue "policing".

Definitions and terminology

- ➢ Political Valence: Indicates whether a reported protest was for a left wing or right wing cause.
- ➢ Frequency of Protest: The number of days for which a protest was reported in the primary dataset.
- ➢ Protest Size: The number of people that turned up to a protest is a suitable measure for the size of the protest, i.e. protests with more participants are considered to be larger.
- ➢ Adjusted for Population: Means that the value was divided by its corresponding value for population.

Assumptions

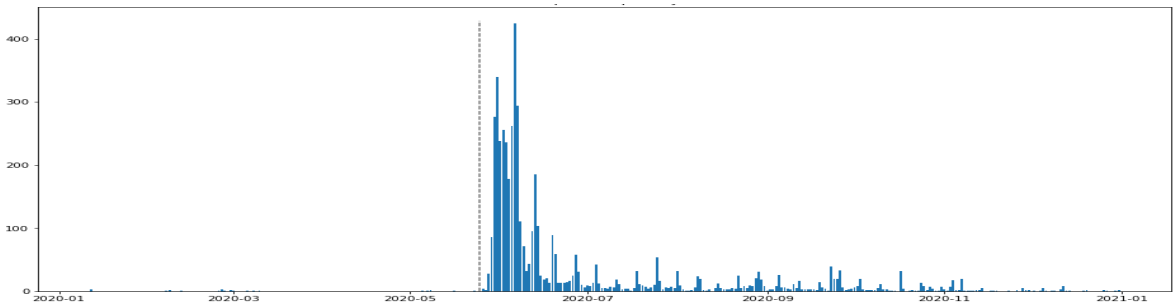
- Considered protests in the year 2020 only, where the reported issues included either racism or policing
- Each protest reported lasted only one day
- We considered protests where only 100 people or more attended

- We used the size_mean variable to represent the number of people who turned up to each protest (the dataset provided upper and lower estimates for event crowd size)
- Considered only annual unemployment rates per state (despite fluctuations over the year)
- We considered state and county population figures for the year, despite fluctuations throughout the year
- Accuracy of the CCC data collection methods

How did the frequency of protests in the United States change throughout the year?

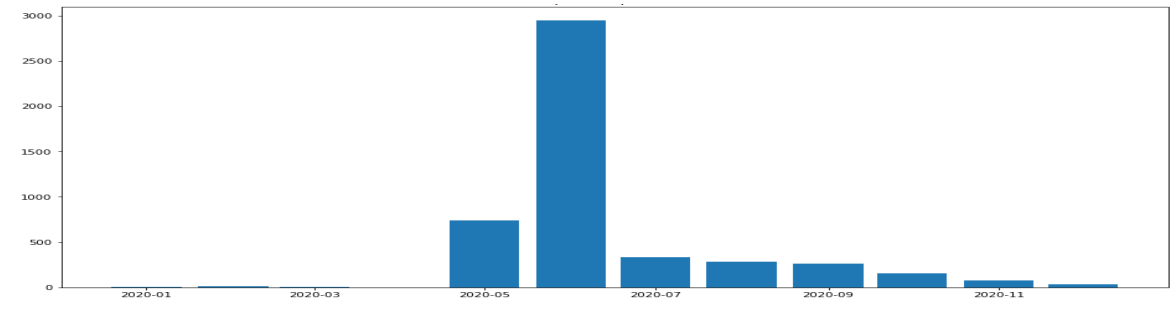
There is a clear pattern in the number of pretests where policing and racism were stated as core issues. The number of protests seems to be flat (near 0) between January 1st 2020 and May 26th 2020, with few incidents reported. After May 26th and through June, there is a sharp increase in the number of reported protests. Following June, the number of protests per day (and per month), fall gradually, but never return to level of protests reported before

Protests per day - 2020



The Grey line on the chart indicates the date of George Flyod's death. We believe the increase in the number of protests following this date are related to this event. We speculate that the relatively low number of protests before this date may be due to covid restrictions and social distancing measures that discouraged gatherings of large groups.

Protests per Month - 2020



The impact of this event seems to have continued throughout the year as the number of protests has not settled back to the previous baseline. This is more apparent in the chart above.

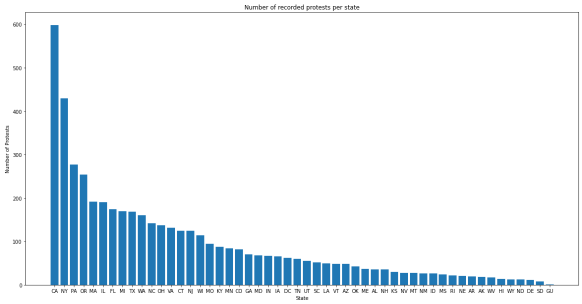
How did the frequency of protests in the United States change throughout the year? By Month-County & Year

The following GIF shows the frequency of protest by county. The frequency of protest was clearly highest in May and June. During May, most of the protests were in the Southwest and Northeast. During June, the Oregon and Florida areas became quite active.

When this data is aggregated for the whole year, we can identify the highest protest frequencies.

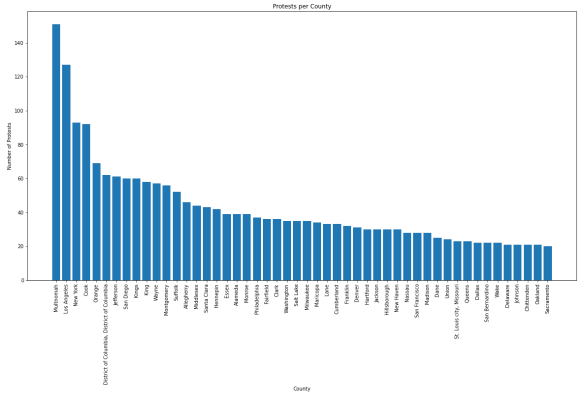
Total Recorded Protest Events for the Year of 2020

Recorded protests per state



The 10 states with the highest number of protests in 2020 were: California, New York, Pennsylvania, Oregon, Massachusetts, Illinois, Florida, Michigan, Texas and Washington.

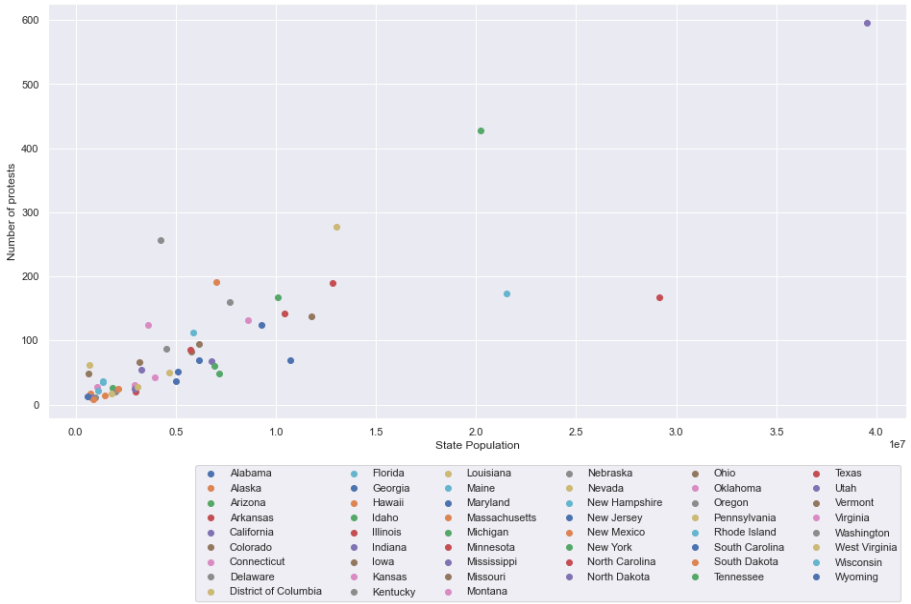
Recorded protests per county (top 50)



The top ten counties with the highest number of protests: Multnomah (OR), Los Angeles (CA), New York (NY), Cook (IL), Orange (CA), DC (WDC), Jefferson, San Diego (CA), Kings (NY), and King (WA).

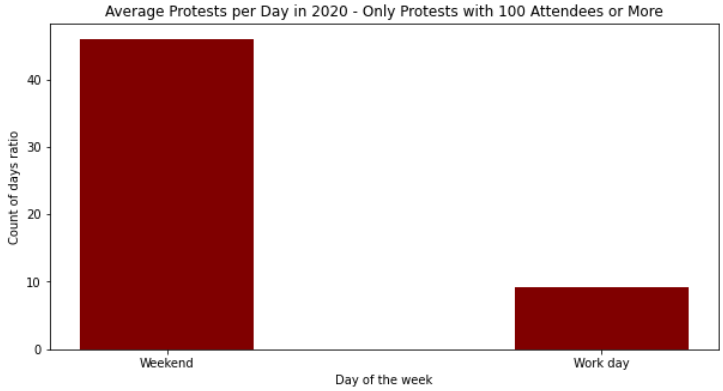
Was there a relationship between the frequency of protests and state population?

We hypothesized that the larger protests would occur where state populations are larger. Upon observation of the data, there does not seem to be a correlation between the protest size and state population. However, our research does show a significant connection between the state population and the frequency of protests (pearson, r = 0.84). For example, our data shows that California (598) had the highest frequency of protests.



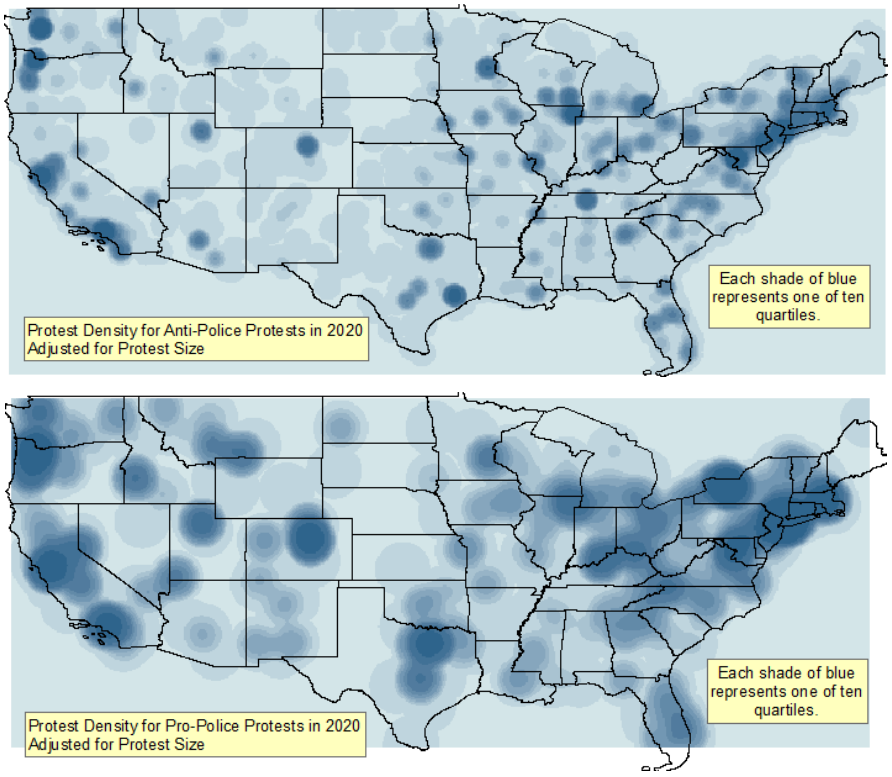
Are protests more likely to occur on weekends or workdays?

We observed that in the chart showing the number of protests reported per day (page 3), there are semi-regular peaks in the number of protests per day, from June through to November. We hypothesized that those peaks corresponded with weekends. Our analysis confirms this hypothesis with the chart seen below. We found that protests were approximately 5 times more likely to occur on weekends as compared with weekdays.



What was the distribution of right-wing vs. left-wing protests?

In order to analyze this distribution, we split our data by valence and used the coordinate points (provided by longitude and latitude values) in order to run a kernel density function. This allowed us to picture a likely distribution of protestors with these political alignments. Surprisingly, we found that there was not much difference in geographic distribution. We had assumed that right-wing protests would be more concentrated in the south, but we found the distribution to be more uniform than expected. We theorize that this may have to do with either the dynamic of anti-protest-protests or that when people go out to protest, they probably choose to protest at specific places where they feel they will be seen.



Where in the United States were there the largest protests of 2020?

In the following chart, we are able to see some differences across the top ten largest protests per capita, per state, by comparing the protest turnout (size mean) and the protest turnout over the state population (adjusted population).

Counties in Texas, Pennsylvania and California appear within the top 10 for raw protest attendance, while completely different county states such as Virginia, Maine, Vermont, appear in the top ten when we adjust for population. This suggests that some counties in smaller states have been able to achieve relatively high participation considering their state populations.

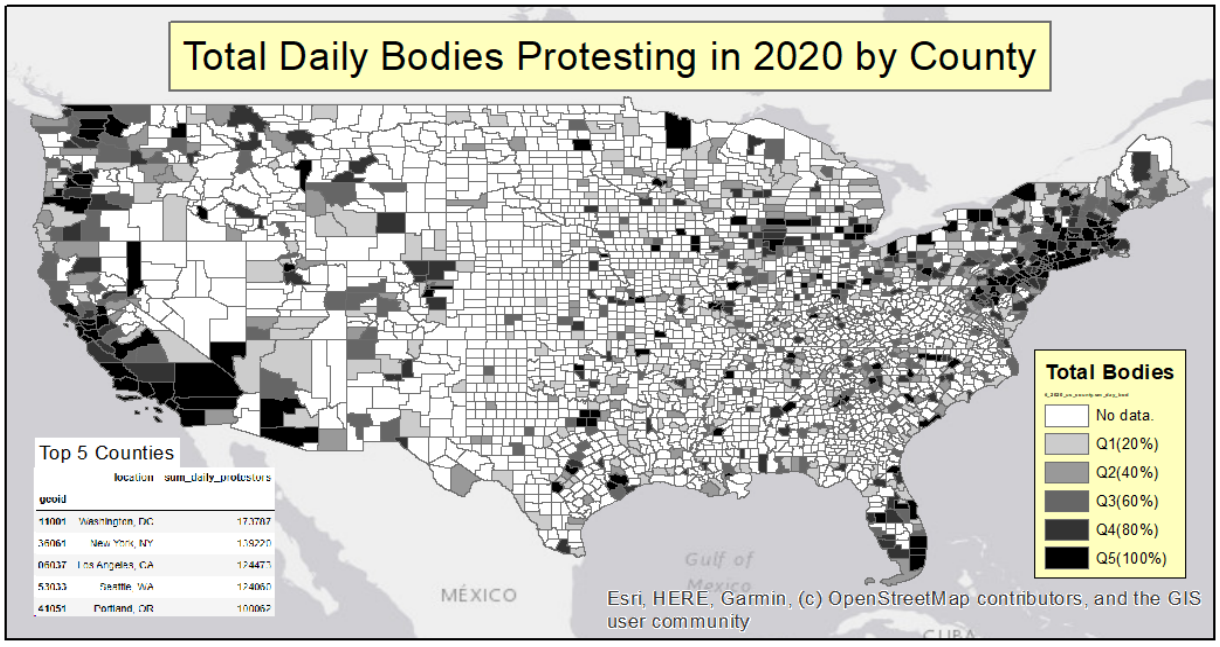
The population of the state may be a confounding variable when studying the factors that influence protest turnout. Therefore, it might be interesting to do case studies of the places that had higher turnout relative to their population, instead of doing case studies of just where the protests were largest. Focusing on higher percent turnout as opposed to the size of the city could help to learn how to effectively organize political protests.

date	state	county_y	size_mean
2020-06-06	DC	District of Columbia, District of Columbia	110000.0
2020-06-12	WA	King	60000.0
2020-06-02	TX	Harris	60000.0
2020-06-06	PA	Philadelphia	44000.0
2020-08-28	DC	District of Columbia, District of Columbia	35000.0
2020-06-07	CA	Los Angeles	30000.0
2020-06-14	CA	Los Angeles	25000.0
2020-06-06	IL	Cook	25000.0
2020-06-28	NY	New York	20000.0
2020-06-04	TN	Davidson	20000.0

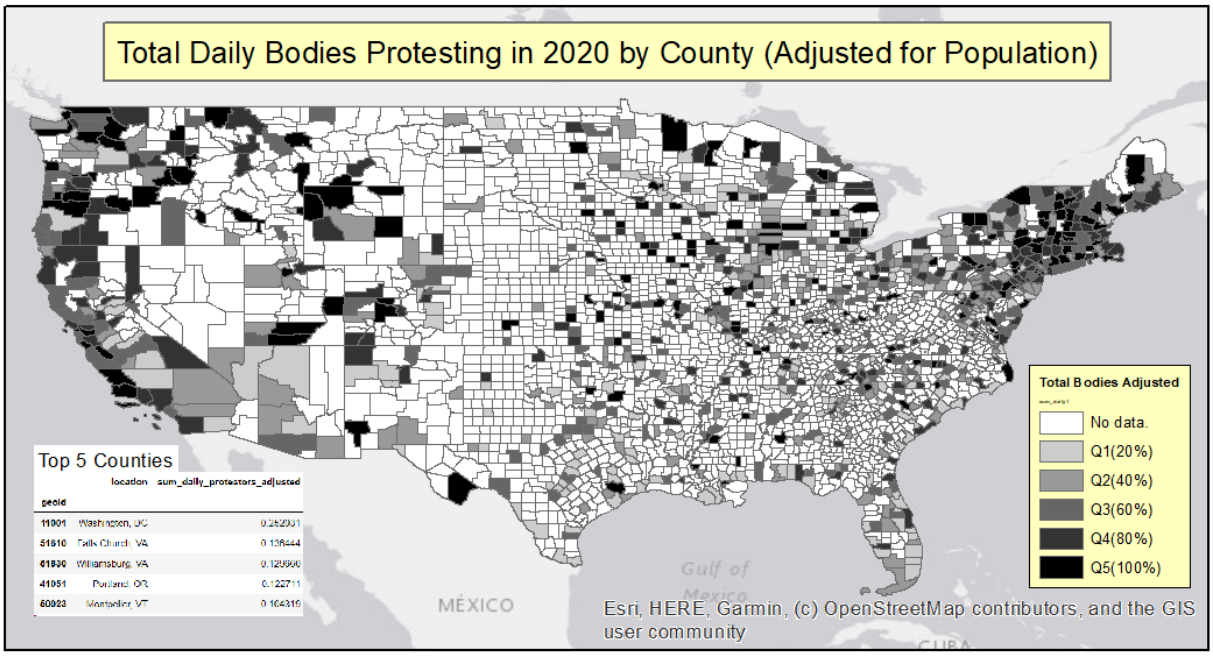
date	state	county_y	_over_pop
2020-06-06	DC	District of Columbia, District of Columbia	0.159525
2020-06-04	VA	Falls Church city, Virginia	0.136444
2020-06-05	ME	Piscataquis	0.101190
2020-06-07	VT	Washington	0.083602
2020-08-17	MT	Carbon	0.083262
2020-06-03	AK	Haines Borough, Alaska	0.072115
2020-06-07	NY	Rensselaer	0.068268
2020-06-01	AK	Hoonah-Angoon Census Area, Alaska	0.063425
2020-06-05	OK	Pontotoc	0.052542

Which counties had the highest number of protesting bodies in 2020?

To get another measure of protest activity, we summed the daily protest sizes for the whole year, by county. That data is displayed in the map directly below. We are able to see a higher amount of total bodies along the west coast, in the northeast, and in the Florida area. We also observe that although the density of protest activity is less as you move close to the central US, there are still a few counties with high total body counts scattered throughout that area.



On the map below, we have adjusted for each county's corresponding population. In essence, we are trying to identify counties whose population might be more politically mobilized than others. When we do this, we can see that the values on the map shift such that the counties in the top 20th percentile (Q5) are more spread out than before. We also notice that when we display the data this way, there seems to be more protest activity in the north-west region of the US and less protest activity both in the So-Cal and Florida regions of the US.



Counties with the Most Daily Bodies Protesting

District of Columbia, District of Columbia
New York County, New York
Los Angeles County, California
King County, Washington
Multnomah County, Oregon

Adjusted for County Population

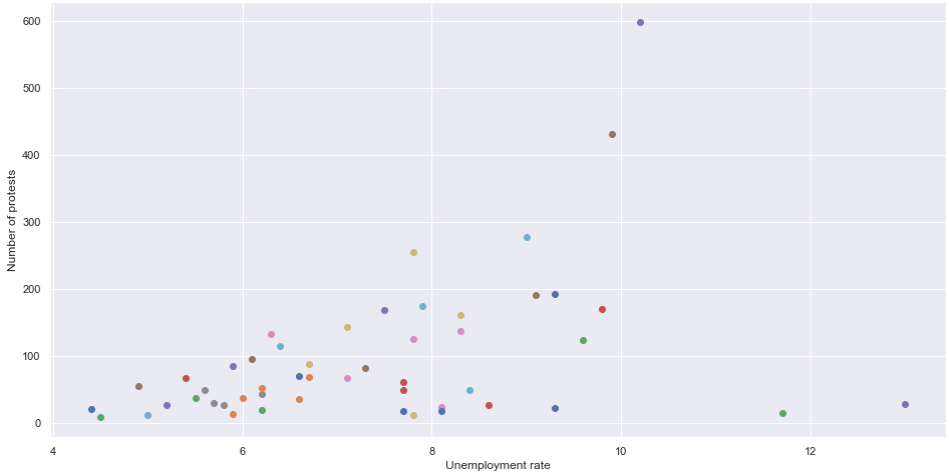
District of Columbia, District of Columbia
Williamsburg city, Virginia
Falls Church city, Virginia
Multnomah County, Oregon
Washington County, Vermont
Piscataquis County, Maine
Carbon County, Montana
New York County, New York
Rensselaer County, New York
Benton County, Oregon

In the charts to the left, we note the top 5 counties in our first map and the top 10 counties in the second map.

We believe the smaller counties in the second category would be more interesting to follow up with case studies.

Is there an association between unemployment rates and total protesting bodies?

We hypothesized that perhaps states experiencing higher rates of unemployment would have more participants in protests during the year of 2020. Although the relationship between these variables was not as detectable as we had expected, we did find a moderate correlation (~0.4).



	num_protests	sum_daily_protestors	sum_daily_protestors_adjusted	rate
num_protests	1.000000	0.953266	-0.410871	0.413076
sum_daily_protestors	0.953266	1.000000	-0.335938	0.414457
sum_daily_protestors_adjusted	-0.410871	-0.335938	1.000000	-0.390873
rate	0.413076	0.414457	-0.390873	1.000000

Summary of Findings

1. How were protests distributed across the United States?
   Without an adjusted population, we are able to see a higher amount of total bodies along the west coast, in the northeast, and in the Florida area. With a value adjusted population, it seems that there were more protest activities in the north-west region of the US and less protest activity both in the So-Cal and Florida regions of the US.
2. Is there a difference in the geographic distribution of protests of opposite political valence?
   According to the data, there was not a major difference of geographic distribution in terms of political opposition. That could mean that both pro and anti protests tended to take place by the same territory, for the most part.
3. Were there any associations between variables?
   Strongest correlations between:
   a. *State population and number of protests reported (pearson, r = 0.84)*
   b. *State unemployment rate and the number of protests reported (pearson, r = 0.41)*
4. Additional findings to note.

- *We found that the pattern for the number of protests per month follows similar patterns for all states sampled. This suggests that similar sentiment across the states sampled with regards to the issues of policing and racism.*
- *Counties with the highest number of protests correspond to the state with the highest number of reported protests.*
- *Most protests were not organized. They were general protestors who were not associated with an organization like BLM.*
- *Of the police/racism related protests, 6.27 % of them were pro-police or anti-anti racism.*
- *Of the 4845 protests with relevant data, the group that organized the most protests was BLM, with a total of only 37 BLM-organized protests reported in 2020.*

Possible Sources of Error

1. Size of protest estimates - the number of protestors is based on a calculation of the mean value of the upper and lower estimates of reported values. The mean range of these estimates was 150 protesters.

2. Missing data for protests size - Some records in the dataset were missing values for protest size. If the sizes were not very negligible, this could've had a large impact on findings.

3. Protests that span over a single day - Protests that span over a single day may be reported multiple times in the dataset leading to some duplicates in the final observations.

4. If it is the case that there are many protests where participants commute from outside the protest's county, our method for adjusting for the local population may not be as reliable.

Concluding Thoughts

Through this exploratory analysis we have managed to describe key features of this extremely unique event - an eruption of protests in the midst of a pandemic. Due to the uniqueness of the environment in 2020, we advise that these general conclusions not be used to predict future events. Rather, these conclusions help us paint a detailed picture of what actually happened. This is even more valuable in regards to a year full of events that may have further polarized the people of the United States. An average citizen's perceptions of where these protests were happening and their corresponding crowd sizes, depended on that citizen's media source. Similarly, the screentime a protest received depended on the media source as well. Thus, this analysis is our best effort at an unbiased description of the 2020 protests and should be used to either inform history or inspire case studies of what happened in particular regions of the country.