

How State and Protester Violence Affect Protest Dynamics

Zachary C. Steinert-Threlkeld*, Alexander Chan†, and Jungseock Joo‡

Abstract

How do state and protester violence affect whether protests grow or shrink? Previous research finds conflicting results for how violence affects protest dynamics. This paper argues that expectations and emotions should generate an n-shaped relationship between the severity of state repression and changes in protest size the next day. Protester violence should reduce the appeal of protesting and increase the expected cost of protesting, decreasing subsequent protest size. Since testing this argument requires precise measurements, a pipeline is built that applies convolutional neural networks to images shared in geolocated tweets. Continuously valued estimates of state and protester violence are generated per city-day for 24 cities across five countries, as are estimates of protest size and the age and gender of protesters. The results suggest a solution to the repression-dissent puzzle and join a growing body of research benefiting from the use of social media to understand subnational conflict.

Keywords. Protest, repression, social media, computer vision.

Supplementary material for this article is available in the appendix in the online addition.

Replication files are available in the JOP Data Archive on Dataverse.

This version is accepted at the *Journal of Politics* as of 05.05.2021. Estimated publication date is April 2022.

*zst@luskin.ucla.edu

†alexander.chan@ucla.edu

‡jjoo@comm.ucla.edu

Though formal and empirical research has established the importance of large protests (DeNardo, 1985, Wouters and Walgrave, 2017), their dynamics remain less understood. For example, Biggs (2003) argues for a positive feedback loop but does not specify when an initial protest is more likely to generate that process, and built environments or electoral fraud encourage protest participation (Tucker, 2007, Zhao, 1998). Empirical investigations have generated contradictory results for decades, making repression and protest dynamics an enduring puzzle (Davenport, 2007). Studying these dynamics has been difficult because existing methodologies generate coarse (ordinal) estimates of violence and have difficulty measuring the size of protests, possibly contributing to the literature’s conflicting results.

How violence affects protest dynamics depends on its source and severity. When it comes from the state, low amounts of violence mobilize more protesters while high amounts demobilize them, creating an n-shaped relationship between state violence and subsequent protest size. Lower than expected costs to protest and emotional backlash generate the increase in protest size, while higher than expected costs and dispiriting emotions shrink it. Protester violence always leads to smaller protests because it decreases the appeal of protesting and increases the probability of state violence.

In addition to explaining the importance of the source and severity of violence, this paper improves the measurement of violence and protest size with new methodology (computer vision) applied to large data (millions of geolocated tweets containing images). A convolutional neural network (CNN) is developed to recognize protest images. Of 42.6 million tweets from protest waves across five countries, 4.6 million contain an image. Approximately 115,000 of these images likely contain a protest. A second CNN generates binary and continuous measures of state and protester violence; these classifiers outperform Google Vision, a third-party CNN. This scene classifier is complemented with a third CNN, a face classifier. This classifier estimates the gender and age of each face, allowing us to control for well-known correlations between these demographic features and protest participation (Nordås and Davenport, 2013, Schaftenaar, 2017). Summing faces in protest images generates estimates of

protest size, and extensive corroboration shows that these estimates are consistent with actual protest size. This pipeline generates daily estimates of the size of protest, the severity of state and protester violence continuously valued from $[0, 1]$, and potential confounds for twenty-four cities in five countries

The next section explains how state and protester violence should affect protest size. After, detail on measuring and validating the concepts is provided. Then we explain how the concepts are measured and validate the classifier estimates. The penultimate section presents results and a battery of robustness checks. The conclusion suggests directions for future research.¹

Protest Dynamics

Protesters want to join a large protest since they have larger net benefits because the probability of suffering repression is lower and the probability of policy change higher. Costs and benefits are uncertain *a priori*, however, so individuals use realized state and protester violence to calculate their payoffs (Shadmehr and Bernhardt, 2011). State violence causes different responses depending on its severity, generating an n-shaped relationship between it and subsequent protest size. Protester violence should always decrease protest size because it decreases the consumption benefit of protest while increasing the probability of repression.

The Importance of Large Protests

Three assumptions lead to the conclusion that large protests are more likely to change policy than small ones. If (1) the purpose of a protest is to convince political leaders to change a policy, (2) a leader cares about the median voter (Downs, 1957) or his or her winning coalition exhibits some response to the median person (Bueno de Mesquita et al., 2003), and (3) a large protest's policy preference is closer to the median individual than a small one's, then a large protest is more likely to change policy than a small one.

Protest size matters regardless of a country's political institutions. In democracies, voting

¹ The empirical data has been successfully replicated by the JOP replication analyst.

is the most common method of policy change but occurs infrequently. Protests, however, can occur at any time and usually have a clear policy goal (Battaglini, 2017), and they provide an additional outlet for revealing pressing sentiments. While protest is unlikely to change an autocrat’s policy, it nonetheless provides a key signal of discontent to which a government can respond (Bratton and Walle, 1992). This signal is especially pertinent if opinion polling is unreliable (Robertson, 2007) or the media are not free (Qin, Strömberg and Wu, 2017).

This argument holds without assuming a leader aims for the median individual’s policy preference. If a leader only desires to stay in power and a large protest means the probability of remaining in power is lower than the leader previously believed, a large protest is still more likely to lead to policy change than a small one. Though a regime should become less responsive to protests as its selectorate shrinks (Bueno de Mesquita et al., 2003), the importance of large protests increases inversely with the size of the selectorate since other mechanisms of policy change are foreclosed. This theory should therefore apply in a wide range of regimes such as rentier states, single-party regimes, and those with low Polity scores. While a large protest is not necessarily successful, almost all successful protests are large.

State Violence

State violence at protests should generate an n-shaped relationship via two mechanisms.² First, how protesters and bystanders respond to repression depends on its severity relative to expectations, not an absolute level. If individuals protest when the expected benefit outweighs the expected cost and the probability of suffering repression affects expected cost (Tullock, 1971), then individuals decide to mobilize based on some expectation of harm. If the realized probability of harm is lower than the value an individual used in their personal calculation, then there will be bystanders who now expect to benefit from protesting. So long as the repression is not larger than expected for a larger number of those already protesting, repression can generate growth in protest size. If, however, repression is more severe than expected, more protesters will demobilize than bystanders mobilize, and protest

² “State violence” refers to what others call “protest policing” (Earl et al., 2013).

size decreases. This relative expectation mechanism operates similarly to expectations about protester identity in information signalling models (Lohmann, 1994).

Second, repression triggers reflexive emotional responses that can cause more or fewer people to mobilize, depending on the emotions triggered (Jasper, 2011). When repression generates emboldening emotions such as anger, joy, or pride, protesters are more likely to persist and bystanders are more likely to mobilize. Anger incites a desire to confront the agents committing repression and facilitates blame attribution against the state. Joy reflects pleasure as protesters sense progress towards their policy goal. Pride enhances feelings of self-worth and belonging, especially in contrast to “bad” state agents (Pearlman, 2013).

Repression can also trigger dispiriting emotions such as fear, sadness, and shame, causing protest size to decrease. A fearful protester is more likely to cease protesting in order to escape the threat of repression. A sad one has determined that the current policy cannot be changed by individual action. Others may feel shame, the belief that they have personally failed. In these cases, protesters demobilize and bystanders continue standing by: protests shrink in size. In addition, these emotions change individuals’ risk aversion, perceived repression risk, and strategic consideration about other potential protesters, making individuals less likely to join a protest. (Young, 2019).

Emotions are especially useful for understanding how low amounts of repression can cause larger protests even when following no repression. If a protest at t_0 has p_0 participation and no repression, any repression at t_1 should cause $p_1 < p_0$, *ceteris paribus*. But if repression induces feelings of anger, joy, or pride in more people than it creates fear, sadness, or shame, $p_1 > p_0$. This backlash effect is in addition to any increase in size due to the first mechanism, difference in expectations. As Siegel (2011) shows:

Less obvious is that, if anger is strong enough, participation levels can be higher under repression than absent it. Individual anger at local repression endogenously enables aggregate backlash. Further, for weak repression, comparatively little anger is needed to achieve backlash against the repressive entity.

Emotions do not transform a calculating individual into an instinctual one. They do not

replace calculation: they are inputs to that process.

Finally, we assume that a low level of repression triggers emboldening emotions while severe repression induces the dispiriting ones. If expectations and emotions influence protesters as described and the relationship between repression and triggered emotions follows the just stated assumption, then:

H1: There exists an n-shaped relationship between the severity of state repression and the size of a protest the next day.

This relationship should hold in democracies and autocracies. For example, 2005 protests against chemical plants in Huashi, Zhejiang drew much greater participation after an initial attempt to remove protesters' encampments led to hundreds of injuries; authorities eventually closed the plants (O'Brien and Deng, 2015). The Occupy Wall Street movement in the United States experienced a surge in participants after New York City police arrested over 700 protesters marching on the Brooklyn Bridge, generating anger. In Egypt, the protests starting on January 25, 2011 were met with initial state resistance and some casualties; 18 days later, the Armed Forces forced President Hosni Mubarak to abdicate. Two years later, the Armed Forces launched a coup against the elected president, Mohamed Morsi. Large pro-Morsi protests erupted and continued for six weeks. The Armed Forces' initial attempts to demobilize them were counterproductive; finally, morning massacres on August 14 at the two main encampments killed at least 1,000 protesters, injured even more, and heralded the return of military rule (Shakir, 2014).³

³ The expectation of backlash conditional on repression severity is consistent with Francisco (2004)'s argument for a strictly positive relationship between the two. Backlash in that theory encompasses non-protest actions such as strikes, building occupations, or guerrilla action and allows for this substitution to occur much later (Moore, 2000). The tests of that backlash theory also find protesters initially demobilize in response to severe repression, in line with this paper's argument.

Protester Violence

Protesters can also engage in violence, so a theory of protest dynamics should take their action into account. Protester violence should always decrease the size of protests for two reasons: it decreases the number of bystanders to whom protesting appeals and increases the cost of protesting to the remaining bystanders not deterred by protester violence.

One method by which bystanders determine whether to mobilize is to compare protesters' ideological distance to their own (Lohmann, 1994). Since most individuals do not support violence or receive consumption value from it (Muñoz and Anduiza, 2019, Simpson, Willer and Feinberg, 2018), protester violence signals that protesters, and therefore the policy changes for which they agitate, are likely not near the median policy preference. Being far from the mainstream, bystanders continue to stand by because the new policy the violent protesters seek is inferred to not reflect non-protesters' preferred policy.

Protester violence decreases the likelihood of regime defections, further decreasing the pool of potential protesters. Peaceful protest convinces regime agents of their physical safety should they defect, increasing the probability that police, members of the armed forces, or legislators switch allegiances (Stephan and Chenoweth, 2008). Violent protesters, however, induce fear in these agents that they will meet the same fate if they do not remain loyal. Violence therefore reduces the pool of those willing to protest, making the state stronger than if facing an otherwise equivalent peaceful protest.

Protests containing protester violence are seen as less legitimate than peaceful ones (Bashir et al., 2013), increasing the probability of repression and therefore the cost of protest. Peaceful protests enjoy high domestic and international legitimacy, so state violence against them risks generating a backlash that increases subsequent protests' size. But since the state can frame violent protesters as rioters, terrorists, or foreign agitators (Benford and Snow, 2000), bystanders are more supportive of repressing violent protests than nonviolent ones (Murdie and Purser, 2017). For the same reasons, the state is also less likely to receive international sanction when repressing violent protests.

Conversely, protester non-violence increases the probability that a protest grows in size, especially when states repress. Because non-violence increases the legitimacy of protests, it decreases the probability that a state represses, as the state will pay large reputation costs. The lower probability of repression induces more bystanders to mobilize, generating a positive feedback loop (Biggs, 2003). In Morocco, for example, attempts to repress non-violent protesters at the start of the Arab Spring led to larger crowd sizes (Lawrence, 2017), and government violence in Tunisia did not prevent the spread of those protests.

Since protester violence alienates bystanders and potential regime defectors and increases the cost of protesting, it should be that:

H2: There exists a negative relationship between the severity of protester violence and the size of a protest the next day.

Methodology

Research finds that state repression decreases protest (Moore, 2000, Olzak, Beasley and Olivier, 2003), increases it (Davenport and Armstrong II, 2004, Gurr and Moore, 1997, Hess and Martin, 2006), or has no effect (Gupta, Singh and Sprague, 1993, Ritter and Conrad, 2016). While this paper is not the first to suggest repression severity generates an n-shaped relationship, previous studies have relied on coarse measures such as ordinal variables or annual number of political detainees at the country level. By generating a more precise measure of the severity of repression, this paper resolves these contradictory findings.

Since the effect of state violence should vary based on its severity, precise measures of it are required, and image analysis enables this precision in two ways. First, convolutional neural networks will generate continuously valued estimates from $[0, 1]$. Current leading datasets, by contrast, map violence onto different discrete categories (values of an ordinal variable). For example, the Social Conflict Analysis Database (SCAD), Urban Social Disorder, and Armed Conflict Locations and Event Data (ACLED) datasets record repression as a binary variable (Raleigh et al., 2010, Salehyan et al., 2012, Urdal and Hoelscher, 2012). Repression elsewhere

is coded as ordinal or nominal, (Goldstein, 1992, Stephan and Chenoweth, 2008, Clark and Regan, 2016), including in machine-coded event data (Boschee et al., 2015). Second, whether one text, e.g. a newspaper article, shows that a protest is violent will depend on the written language used and the research teams’ interpretation of that language. Researchers thus necessarily have to use coarse measures for large-n analysis. The closest continuously valued measure of repression is fatalities, which is not always recorded and is the most severe type of state violence. When focusing on violence in one setting, scholars have been more successful at disambiguating it to generate ordinal measures (Khawaja, 1993, Olivier, 1991).

This paper develops three classifiers based on convolutional neural networks to automatically code the variables of interest: one identifies protest images, a scene classifier extracts data from them, and a face classifier generates size estimates and demographic controls.⁴

The pipeline outlined in Table 1 resembles the approach described in Zhang and Pan (2019). Step 1 uses keywords, including hard negatives like “concert” or “stadium”, and Google Image Search to acquire training images. Step 2 then trains a CNN on these images. Step 3 uses this classifier to identify protest images from just under 43 million geolocated tweets, resulting in 40,764 protest images. In Step 4, workers from Amazon Mechanical Turk label these Twitter images. To measure state and protester violence, annotators are presented pairs of images and asked which is more violent, and the Bradley-Terry model generates continuous estimates from the resulting ordering (Bradley and Terry, 1952). Each image is coded by at least two individuals; a third is used to break ties. These labels then train the second classifier (Step 5), and this classifier provides the estimates for state and protester violence. Separately, in Step 6, a third classifier assigns gender and age estimates to identified faces (Kärkkäinen and Joo, 2019); the count of faces provides an estimate of

⁴ The first two classifiers are in fact partially combined in implementation such that one integrated classifier can generate two sets of outputs, although they differ conceptually. This is called multi-task learning (Girshick, 2015). We discuss two classifiers separately because they are trained on different data and used in different steps.

protest size, and the demographic estimates generate control variables. These steps result in 40,764 tweets with new data about the demographics of individuals and severity and type of violence recorded in each image.

Figure 1 shows sample images of protest (top), state violence (middle), and protester violence (bottom) the pipeline identifies. For an overview of CNNs, see Section S1. Section S2 provides additional detail on our pipeline and validation of the results. Section S3 shows manual validation from a team of research assistants specifically trained for this project.

Table 1: Protest Data Pipeline

Steps	Input	Source	Output
<i>Collecting Images for Training Set</i>			
1. Image search	Keywords	Google	100,000 images
2. Train a protest image classifier	Images from Step 1	Self	Initial CNN
3. Protest images from Twitter corpus	Model from Step 2	Twitter	40,764 images
<i>Developing Protest and Scene Classifier</i>			
4. Manual annotation	Images from Step 3	Amazon Turk	13 ground-truth labels
5. Train a CNN	Training data from Step 4	Self	Protest and scene classifier
<i>Face Attribute Classification</i>			
6. Face classification	-	Kärkkäinen and Joo (2019)	Gender, age, and size estimates

Note: The protest data pipeline encompasses six primary steps. Section S2 provides more detail.

Data

We identify five protest periods from polities with diverse population, income, and institutional characteristics, mitigating the risk that subsequent findings arise from underlying similarities in the cases. These polities are Hong Kong, Pakistan, South Korea, Spain, and Venezuela. The primary criteria is to construct a sample from different types of regimes, which we measure with the Polity4 score. Hong Kong has a score of -4; Venezuela, 4, Pakistan, 7, South Korea, 8, and Spain 10. We also consulted the Varieties of Democracy dataset to ensure these countries contain different media environments (`e.v2xme_altinf_5C`) and civil society freedom (`e.v2xcs_ccsi_5C`). With varying amounts of freedom of assembly, the

Figure 1: Sample Images and Their Classifier Outputs



Note: The top panel shows sample images and the protest classifier’s rating of them. The use of hard negatives in the training set ensures that scenes that contain crowds (bottom row, left), individuals walking on streets (top row, third), or a non-protest sign (bottom row, third) are not included in analysis. The middle panel shows protest images with their state violence rating and the bottom shows protest images’ protester violence rating. Labels contain each image’s city and label probability.

press, and civil society, studying these countries minimizes, though it does not eliminate, the possibility that results derive from case selection.

Table 2 details the cities included from these countries, the issues driving protest, and the frequency of protest images per city. For each period, we searched from one week prior to the first reported protest and one week after the last one. This process identifies 42,579,188 tweets containing 4,456,981 images. Keeping only tweets whose images generate a protest score of at least .849 results in 26,142 tweets with images.⁵ These tweets are the data used

⁵ This threshold is the value which maximizes recall with .85 precision.

for regressions.

We then aggregate tweets to their city of origin and the day they were created. Cities are kept for analysis when at least $\frac{1}{7}$ of their days contain a protest image. Table 2 shows these 24 cities, which account for 6,303 protest images. (Most images do not have a location resolution more precise than the country.) These 6,303 protest images spread across 4,143 city days. We treat missing dates as true zeroes, and a robustness check shows that this interpolation does not change results. Some of these images are duplicates, but later deduplication shows they do not affect inference.

Using these data introduces two ethical concerns. First, it is possible that minors are part of this study, as Twitter does not perform age verification for accounts and minors could appear in others' photos. Though many protests, such as Hong Kong's protests or the 2019 protests in Chile, feature prominent actors under the age of 18, we have only used faces from individuals estimated to be at least 20 years old. Second, protesters may not be as anonymous as they think. Though these data are observational and publicly available, individuals in photographs may not have consented to appear in those photographs. Authorities could monitor images shared on social media to identify people who protested; many already do (Purdy, 2018).⁶ To prevent the identification of individuals in our data, we have released only the aggregated city-day data.

Operationalization

The dependent variable is $\text{Log}_{10}(\text{Protest Size})_{i,t}$, the logarithm of the sum of the number of faces in all protest photos from city i on day t . Because the resulting numbers are certainly lower than the true protest size (the largest protest in our dataset contains 627 faces), five checks are performed to provide confidence that this operationalization actually measures protest dynamics. The estimates generated are consistent with others' and record actual events.

⁶ At the same time, however, shared protest images can identify incriminating state behavior that would otherwise be denied (Lim, 2013).

Table 2: Protest Periods

	City	Country	Start	End	Issue	Protest Images/Day	Protest Images/Day if >0
1	Central	Hong Kong	2014.09.18	2014.12.23	China reforms	1.96	5.00
2	Kowloon	Hong Kong	2014.09.18	2014.12.23	China reforms	1.29	2.92
3	Lahore	Pakistan	2017.11.07	2017.11.23	Blasphemy	.18	1
4	Kimhae	South Korea	2016.10.20	2017.03.14	Anti-incumbency	.47	1.92
5	Seoul	South Korea	2016.10.20	2017.03.14	Anti-incumbency	2.40	3.76
6	Citutat Vella	Spain	2017.09.01	2017.12.31	Secession	.94	4.95
7	Barcelona	Spain	2017.09.01	2017.12.31	Secession	3.07	11.60
8	Girona	Spain	2017.09.01	2017.12.31	Secession	1.10	3.26
9	Granera	Spain	2017.09.01	2017.12.31	Secession	.62	2.33
10	Granollers	Spain	2017.09.01	2017.12.31	Secession	.23	1.25
11	Lleida	Spain	2017.09.01	2017.12.31	Secession	.42	1.88
12	Mataro	Spain	2017.09.01	2017.12.31	Secession	.51	2.33
13	Reus	Spain	2017.09.01	2017.12.31	Secession	.35	1.68
14	Sabadell	Spain	2017.09.01	2017.12.31	Secession	.96	2.66
15	St. Cugat d. Valles	Spain	2017.09.01	2017.12.31	Secession	.31	2.06
16	St. Feliu d. Pallerols	Spain	2017.09.01	2017.12.31	Secession	.61	2.19
17	St. Salvador d. Guardiola	Spain	2017.09.01	2017.12.31	Secession	.48	2.15
18	Tarragona	Spain	2017.09.01	2017.12.31	Secession	.57	1.94
19	Terrassa	Spain	2017.09.01	2017.12.31	Secession	.57	2.22
20	Boca del Rio	Venezuela	2014.03.27	2017.12.17	Anti-Maduro	.26	1.34
21	Caracas	Venezuela	2014.03.27	2017.12.17	Anti-Maduro	4.82	7.63
22	Caucagua	Venezuela	2014.03.27	2017.12.17	Anti-Maduro	.53	1.72
23	Maracaibo	Venezuela	2014.03.27	2017.12.17	Anti-Maduro	.39	1.49
24	Valencia	Venezuela	2014.03.27	2017.12.17	Anti-Maduro	.41	1.62

Note: The last column is the average number of photos for days containing a protest photo.

The first check manually validates the face counts per photo. To complement the manual validation from Amazon Mechanical Turk that Figures A3 through A5 show, we trained a team of three students to count faces in images and label whether they contain state or protester violence; we then compare their coding to our classifier’s estimates. The number of faces the human coders identify closely matches the classifier’s face count, and images humans label that contain state or protester violence receive much higher classifier estimates for those labels than those that do not. Section S3 presents measures of intercoder reliability, and Figure A10 shows this comparison.

Second, size estimates for large protests could be biased upwards if the number of images and faces per image increases with protest size. Figure A11 shows no linear relationship between the size of a protest and the number of faces per photo. City-days with larger protests are therefore driven by the production of more protest images, not the sharing of crowded images. This result matches other work finding that counting faces in protest images generates accurate estimates of protest size variation (Sobolev et al., 2020).

Next, we consult other sources’ estimates of protest size in Barcelona, Caracas, Seoul, and Hong Kong. For Caracas, we use crowd density estimates of protest images from Venezuelan newspapers (Rodríguez, 2020); this methodology is used widely because it generates accurate estimates of large crowds without directly counting every participant (McPhail and McCarthy, 2004).⁷ For Barcelona and Hong Kong, we trained a team of undergraduates to follow the coding procedure and sources of Weidmann and Rod (2018). That approach did not generate enough size estimates for Seoul, so we use police and activist size reports provided by Wikipedia.⁸ Table 3 shows this correlation for matched events and the results of two residual tests. Though higher correlations are preferable, the residual plots of Figure A12 show that $\text{Log}_{10}(\text{Protest Size})_{i,t}$ is not biased as a function of the protest size recorded

⁷ We also tried the Mass Mobilization in Autocracies dataset (Weidmann and Rod, 2018), but it recorded numeric estimates of protest size for only four events. Correlation with those is greater than .9.

⁸ https://ko.wikipedia.org/wiki/박근혜_대통령_퇴진_운동

in other sources. $\text{Log}_{10}(\text{Protest Size})_{i,t}$ is therefore a noisy but consistent estimate of protest size available from other sources.⁹

Table 3: Verifying Protest Size Estimate Using Other Sources

City	Source	Matched Events	Correlation	S-W	K-S
Barcelona	AFP, BBC, AP	5	.7980	.0397	.4714
Caracas	Rodríguez (2020)	18	.4101	.5975	.7974
Hong Kong	AFP, BBC, AP	11	.3477	.8135	.7967
Seoul	Wikipedia: Police	12	.3686	.7294	.8785
Seoul	Wikipedia: Activists	21	.4349	.0491	.7121

Note: Taking the log of the sum of faces in protest photos correlates with logged estimates from newspapers (Barcelona, Hong Kong), Wikipedia (Seoul), and crowd density estimates (Caracas). The S-W column shows the p-value from the Shapiro-Wilks test, and the K-S column is for the Kolmogorov-Smirnov test; both are conducted on the residuals from regressing $\text{Log}_{10}(\text{Protest Size})_{i,t}$ on the log of the reported protest size. Lahore is not shown because we found no newspaper estimates of protest size from there.

Fourth, this pipeline recovers a very large percentage of protests identified in other event datasets for these countries, as Table 4 shows. The top rows show the number of city-days with protest observed using geolocated images for each city, region, or country. These records are compared to three leading event datasets; in each cell, the number is the number of events that dataset records and the percentage is the percent of those this methodology captures. These results are in line with what Zhang and Pan (2019) finds for Sina Weibo in China: their protest detection pipeline finds 52% of the events ICEWS does, 56% of GDELT’s, and 88% of WiseNews. Note as well that most events we record are not recorded in the other event datasets, in line with similar comparisons in Chile, China, and Venezuela (Steinert-Threlkeld and Joo, 2020, Steinhardt and Goebel, 2019).¹⁰

Finally, the temporal variation of protest size is consistent with actual events. Figure 2 shows this result: the correspondence is consistent and appears to be more than chance. We

⁹ These other sources are not used for size estimates because they contain too many false negatives.

¹⁰ These two articles do provide exact comparisons but show that social media generates data on many more events than MMAD or ICEWS in Venezuela, about the same as many as ACLED in Chile, and much more than domestic and international organizations in China.

Table 4: Verifying Protest Coverage Using Other Event Data

Source Data		Catalonia	Hong Kong	Lahore, Pakistan	South Korea	Venezuela
Images	Twitter	1,421	137	17	232	2,336
ACLED	Local & international news; reports.	—	—	9, 33%	—	—
ICEWS	Local & international news	54, 94%	49, 59%	2, 0%	99, 82%	365, 49%
MMAD	AFP, BBC, AP	—	105, 91%	—	—	4, 100%

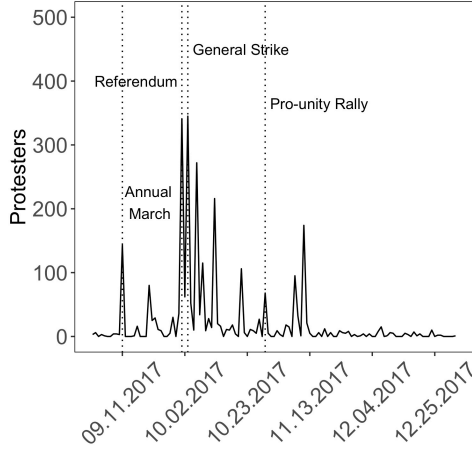
Note: The top row is the number of events in the Twitter data used for regressions. Each subsequent row shows (# events in dataset, % match with Twitter); for example, ICEWS records 54 events in Catalonia, 94% of which the Twitter data contain. All ICEWS events with a city of 'nan' are dropped. MMAD contains more events from Hong Kong here than in Table 3 because this table does not restrict events to those with size estimates.

did not attempt to label every peak, but other work has found that social media records protests at least as well as other event datasets (Dowd et al., 2020).

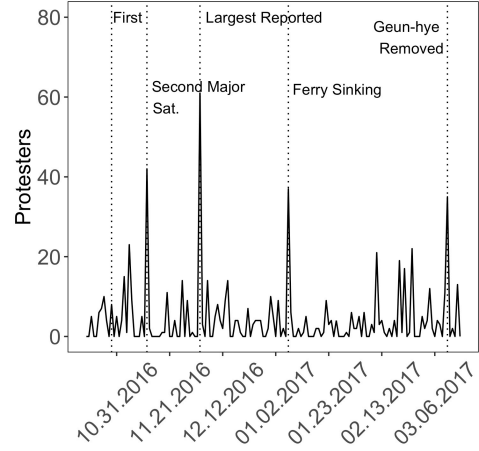
The violence variables to test Hypothesis 1 are *Perceived State Violence* $_{i,t-1}$ and its square. For H2, *Perceived Protester Violence* $_{i,t-1}$. These are the average of the classifier estimate for all protest images per city-day. Figure A13 in Section S5 shows a time series of the mean state violence recorded in protest images for the same four cities.

We describe the violence variables as “perceived” for three reasons. First, the true amount of violence is unknown because violence is a latent concept, not a physical entity, like temperature or pressure, directly measurable. Second, the images people share may be strategically chosen. This possible selection effect is true of any event data that relies on secondary sources, which is to say almost all event data. For discussion and analysis of bias that these measures may introduce, see Section S8. Third, the main analysis does not deduplicate images, meaning images which are shared often will have a greater impact on people’s decision making process than those only tweeted once. Deduplicating images to more closely approximate the “true” violence at events does not change results, as Table 6 shows.

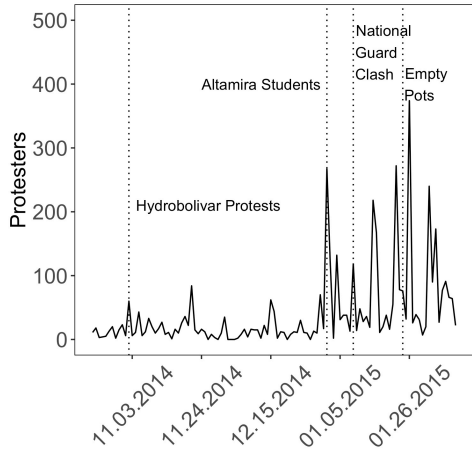
Figure 2: Verifying Protest Size, Time Series



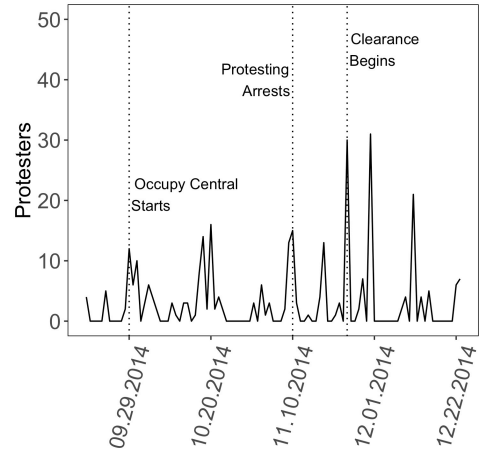
(a) Barcelona



(b) Seoul



(c) Caracas



(d) Hong Kong

Note: Measuring protest size using faces in images records changes in protest size that match reported dynamics.

Model

In addition to the operationalizations detailed in the previous section, we include six control variables. Two are demographic: the gender and age attributes of protesters. A society with greater gender equality is more likely to see nonviolent than violent action (Schaffenaar, 2017), and the same is true at the movement level (Asal et al., 2013). The percent of protesters who are male, $Male\ Percent_{i,t-1}$, is therefore a variable for which we control. Since youth often spearhead mass protests and these effects are amplified when there are many of them,

the percent of participants aged 20-29, *Young Adult Percent*_{*i,t-1*}, is a variable for which we control (González, 2020, Urdal, 2006).¹¹

We also generate two binary variables, *Police*_{*i,t-1*} and *Fire*_{*i,t-1*}. The police and fire variables are the sum of images containing a police officer or fire, respectively, based on the thresholds identified in Table A2. *Police*_{*i,t-1*} provides another estimate of repression, while *Fire*_{*i,t-1*} approximates protester violence (Figure A14 shows that fire and protester violence have the highest city-day correlation of any two variables). Fifth, *Tweets*_{*i,t-1*} is the number of lagged protest images per country-day and controls for any protest information not directly measured, such as tactical knowledge about a protest (Little, 2015). Sixth, we include a lagged dependent variable to account for autocorrelation as well as any regression to the mean. Table 5 provides descriptive statistics of these variables.

Table 5: Summary Statistics for Regression Variables

Statistic	N	Mean	St. Dev.	Min.	Max.
Protest Size _{<i>i,t</i>}	4,143	2.57	15.36	0	627
Perceived Protester Violence _{<i>i,t-1</i>}	4,121	0.03	0.12	0.00	1.00
Perceived State Violence _{<i>i,t-1</i>}	4,121	0.02	0.08	0.00	0.94
Police _{<i>i,t-1</i>}	4,121	0.001	0.04	0.00	1.00
Fire _{<i>i,t-1</i>}	4,121	0.08	0.42	0.00	7.00
Male Percent _{<i>i,t-1</i>}	4,121	0.03	0.11	0.00	1.00
Young Adult Percent _{<i>i,t-1</i>}	4,121	0.02	0.08	0.00	1.00
Tweets _{<i>i,t-1</i>}	4,121	1.52	7.02	0.00	238.00

Note: Summary statistics for the regression variables.

We build three models. The first uses only covariates that measure violence. The second focuses on the demographic control variables. The final model combine the two sets of variables. All independent variables are lagged one day. All models include city fixed effects and city-clustered standard errors. To guard against overfitting, we use five-fold cross-validation. Ordinary least squares is the estimator.

¹¹ 20 is the the oldest age of legal adulthood (Japan) of which we are aware, so we use it to be cautious.

Results

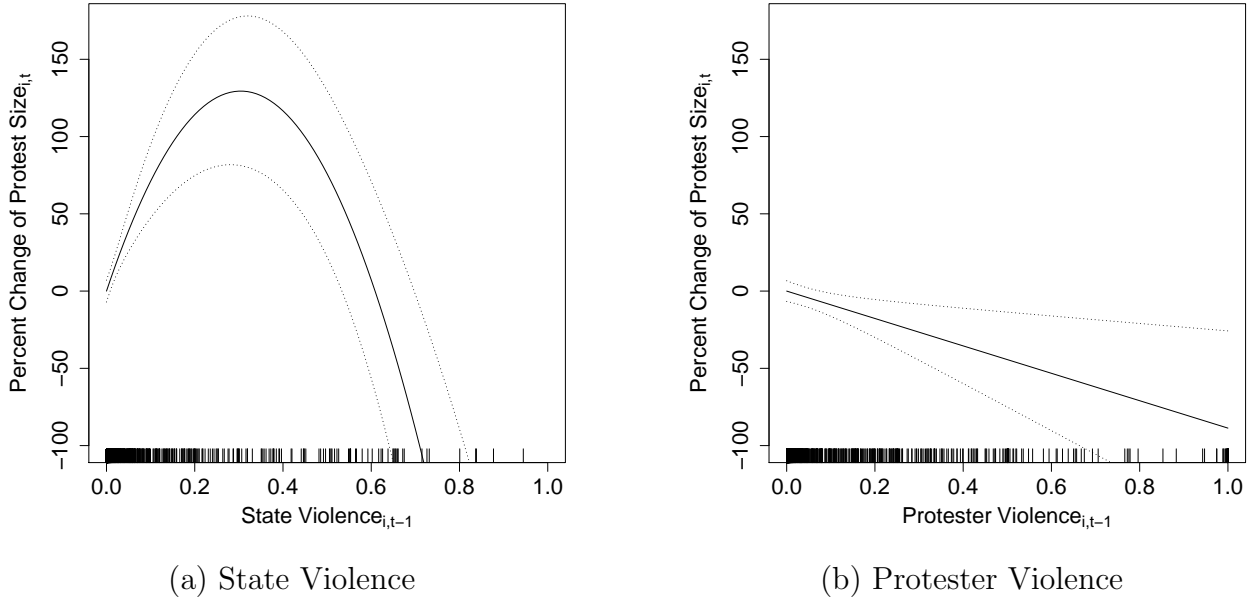
Results match expectations: low amounts of state violence correlate with larger subsequent protests ($p < .001$), though severe enough violence with subsequently smaller protests ($p < .001$). When protesters engage in violence, recorded subsequent protest size is smaller ($p < .05$). In addition, protester violence has a much smaller coefficient than either state violence variable, with the largest magnitude occurring when states engage in high levels of violence. Table A5 in Section S7 shows these results.

Figure 3 shows marginal effects of state and protester violence from that model. From values of $[0-.3)$, state violence correlates with larger subsequent protest. At that amount of violence, protest size the next day is 137% higher than if there was no state violence. Moreover, state repression usually leads to larger protests: only 77 of 1,467 city-days of protest contain average state violence greater than .3. Increased protester violence always correlates with subsequently smaller protest. The change, however, is much smaller than for state violence: moving from no protester violence to its mean (.035) correlates with a 2% smaller protest, while the difference between state violence and its mean is an increase of just over 17%. A one standard deviation increase of state violence from 0 correlates with a 63% increase in protest size; a one standard deviation increase in protester violence from zero correlates with a 12% smaller protest.

Robustness Checks

A series of robustness checks on state violence confirms its n-shaped relationship with subsequent protest size. First, images with more violence could contain fewer faces, causing the regression results to be driven by measurement problems. While lagging the independent variables mitigates this concern, Figure 4 shows that no relationship appears to exist between the violence an image records and the number of faces contained therein. To the extent that one does, it is actually slightly positive, biasing against finding a negative relationship. Second, the n-shaped relationship between state violence and the next day's protest size could

Figure 3: Marginal Effects

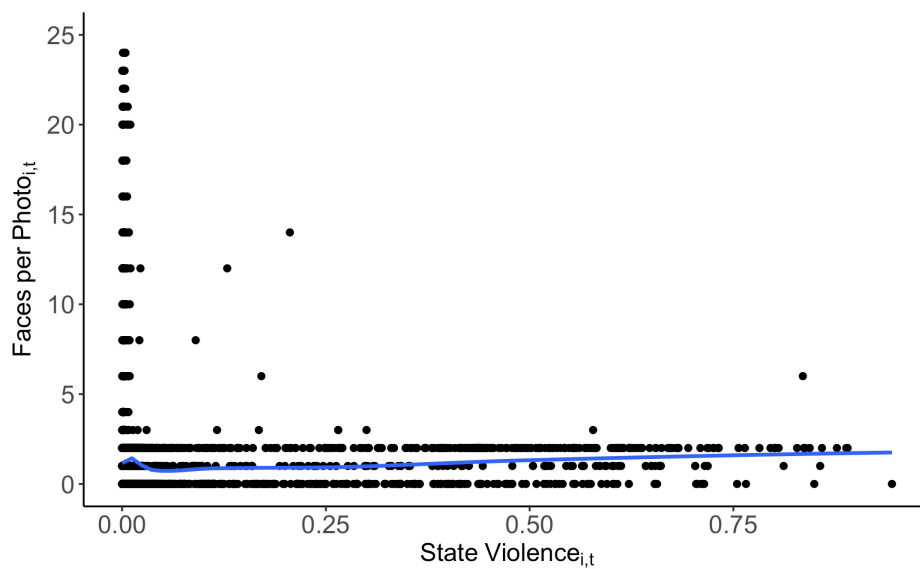


Note: Marginal effects of *Perceived State Violence* $_{i,t-1}$ and *Perceived Protester Violence* $_{i,t-1}$ from Model 3 of Table A5. State violence exhibits an n-shaped relationship with subsequent protest size while protester violence always correlates with smaller subsequent protests.

be an artifact of fitting a parametric model with a square term. Figure 5 shows the results of tests demonstrating the persistence of this relationship. Whether fitting a local average of the relationship between *Perceived State Violence* $_{i,t-1}$ and $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$ or a spline with 50 knots, binning *Perceived State Violence* $_{i,t-1}$ into ten evenly spaced groups, or regressing *Perceived State Violence* $_{i,t-1}$ on partial residuals, the n-shaped relationship between state violence and subsequent protest size holds.

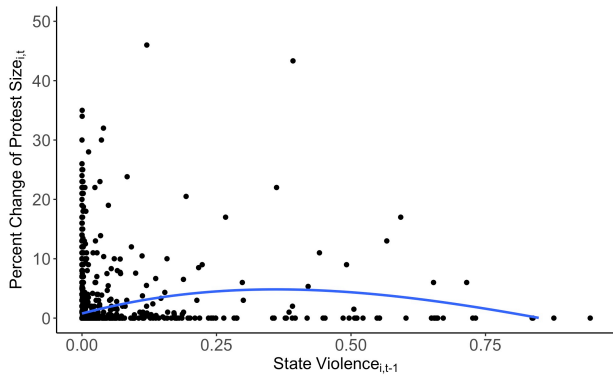
Strategic behavior of protesters and state agents may endogenously cause the observed correlations. Protesters often seek international attention because it increases their legitimacy and raises the cost of repression (Bruns, Highfield and Burgess, 2013), and protesters are strategic about the language in which they make these appeals (Driscoll and Steinert-Threlkeld, 2020, Metzger, Nagler and Tucker, 2015). These appeals may downplay protester violence and exaggerate the size of protests. To deter protester coordination, state actors will emphasize protester and state violence as well as downplay the size of crowds.

Figure 4: State Violence Does Not Cause Fewer Faces per Photo

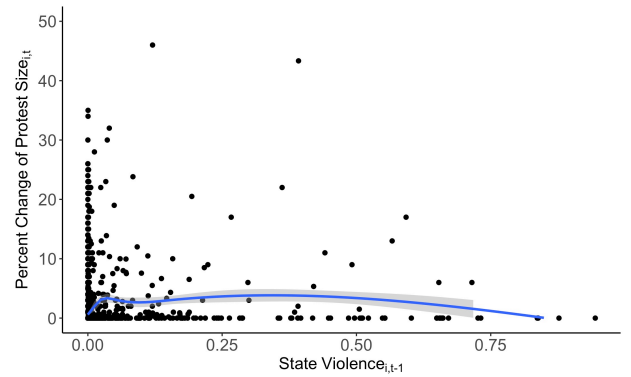


Note: No correlation exists between state violence in an image and the number of faces. A linear fit suggests a slight positive relationship, and restricting the relationship to images with fewer than ten faces does not change results.

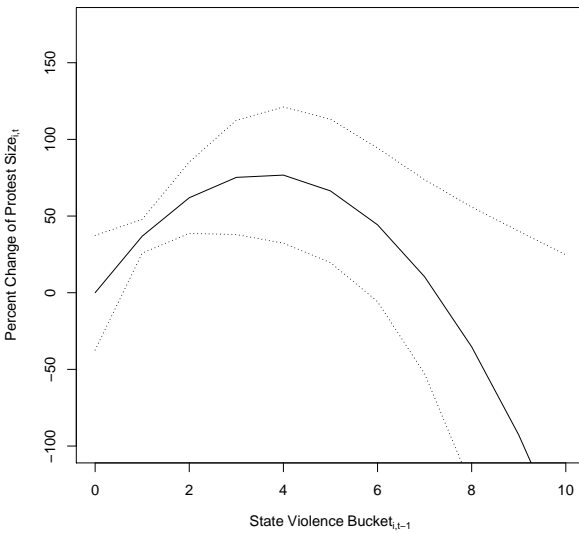
Figure 5: State Violence Results Remain in Flexible Operationalizations



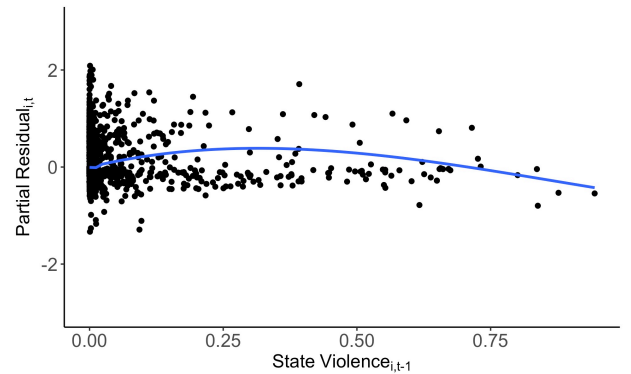
(a) LOESS (Span = .2)



(b) Spline, 50 Knots



(c) Binned Marginal Effects



(d) LOESS on Partial Residuals

Note: The n-shaped relationship holds in a non-parametric relationship (a, b). Generating fixed effects for state violence in bins of width of .1 finds the same relationship (c). Regressing state violence on the partial residuals of Model 3 from Table A5.

Table 6 shows that tests for this strategic behavior do not change inference. Model 1 restricts analysis to those tweets in the country’s *lingua franca*, as they are the ones most relevant for local actors. Model 2 drops tweets from bots (accounts controlled by computer code) since actors use them to strategically amplify messages.¹² Since strategic actors may emphasize particular protest features by repeatedly sharing photos, Model 3 keeps only the first occurrence of an image.¹³ Strategic actors are also more likely to have very many followers (authority figures) or very few (bots), so Model 4 restricts analysis to only tweets from accounts within the 25th-75th percentile of their country’s follower distribution. Model 5 focuses on attempts to manipulate estimates of protest size by making the dependent variable equal the logarithm of the number of unique accounts that share protest images. This new dependent variable is not affected by images chosen for the number of protesters they show or strategic actors tweeting frequently.

Across all models save the last, the results for state violence stay the same; Model 5 also records an n-shaped relationship, but the coefficient on the squared term is too small to be statistically significant. (Table A8 shows that the raw number of users recreates the n-shaped relationship.) Inferences about protester violence also do not change except for one model, Model 3; its sign is still the same, and results weighted by the number of tweets per city-day (Table A9) make the variable significant again.

Finally, Table 7 shows that accounting for more complicated time dynamics does not change inferences about state violence, though it weakens the negative correlation of protester violence and subsequent protest size. A partial autocorrelation plot suggests 15 lags of the

¹² We submit every user to the Botometer service and remove tweets with a complete automation probability $\geq .4$, the threshold which that produces the most accurate classification of bots (Varol et al., 2017). Table A13 shows that no more than 10.8% of tweets in any city are from bots.

¹³ The data do not contain retweets because Twitter does not assign coordinates to retweets. They contain replies, and replies contain the image of the original tweet. Section S11 explains the deduplication methodology, and Table A14 show the percent of tweets per city that are duplicates.

Table 6: Robust to Strategic Behaviors

	Country Language	No Bots	Deduplicated Images	IQR Users	DV: $\text{Log}_{10}(\text{Number of Users}_{i,t})$
	(1)	(2)	(3)	(4)	(5)
Perceived Prtstr. Violence $_{i,t-1}$	-.1982*** (.0387)	-.1438** (.0704)	-.1301 (.0853)	-.1530*** (.0446)	-.1784*** (.0495)
Perceived Stt. Violence $_{i,t-1}$	1.5665*** (.4553)	1.2236*** (.3349)	1.2138** (.5284)	1.1894*** (.3333)	.3938* (.2340)
Perceived Stt. Violence $^2_{i,t-1}$	-2.5718*** (.8864)	-2.0184*** (.5787)	-2.0015** (.9100)	-2.0125*** (.5509)	-.6597 (.4044)
Police $_{i,t-1}$.5070 (.3090)	.6357* (.3712)	.9419* (.5190)	.8552** (.3763)	.2250 (.2720)
Fire $_{i,t-1}$.0753*** (.0249)	.0912*** (.0202)	.0594 (.0507)	.0576* (.0316)	.0667*** (.0152)
Male Percent $_{i,t-1}$	-.0617 (.1264)	-.1848* (.1078)	-.0833 (.1201)	-.0195 (.0458)	-.1068** (.0543)
Young Adult Percent $_{i,t-1}$.0173 (.0845)	.2166** (.0905)	.2848** (.1370)	.2584*** (.0953)	.1162** (.0466)
Tweets $_{i,t-1}$.0249*** (.0047)	.0119** (.0047)	.0162** (.0069)	.0148** (.0060)	.0037* (.0021)
DV $_{i,t-1}$.1145 (.0772)	.1766** (.0782)	.1490* (.0844)	.1047*** (.0363)	.4062*** (.1004)
Intercept	.0694*** (.0138)	.1186*** (.0157)	.1319*** (.0172)	.1270*** (.0160)	.1027*** (.0105)
N	3,481	4,121	2,533	3,462	4,121
Adjusted R ²	.3144	.2732	.2223	.1897	.4125
Cluster SE	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y

*p < .1; **p < .05; ***p < .01

Model 1 keeps only tweets in each country's *lingua franca*. Model 2 drops all tweets from accounts identified as bots. Model 3 removes duplicate images. Model 4 keeps only tweets from users within the interquartile range of their country's follower distribution. Model 5's dependent variable is the logarithm transformation of the number of users tweeting protest images. Standard errors are clustered by city.

dependent variable, which Model 1 includes. Model 2 includes weekend fixed effects, and Model 3 includes weekday ones. Model 4 includes a control for the protest duration and the number of consecutive days of state repression. In all specifications, the n-shaped relationship of state violence and subsequent protest remains statistically significant. Models 1 and 4 no longer find significant results for protester violence, suggesting that protesters may become more violent the longer a protest lasts. Model 4 also shows that protests decrease in size over time, though persistent state violence correlates with larger protests.

Section S8 presents two checks of the data generating process to address concerns about selection bias. Users who share protest images may differ from those who share non-protest

Table 7: Time Effects Do Not Change Results

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$			
	15 Lags	Weekend FE	Weekday FE	Duration Controls
	(1)	(2)	(3)	(4)
Perceived Prtstr. Violence $_{i,t-1}$	-.0366 (.0831)	-.1543* (.0830)	-.1571* (.0831)	-.0489 (.0775)
Perceived Stt. Violence $_{i,t-1}$.7348** (.3226)	1.2716*** (.3722)	1.2864*** (.3727)	.8831*** (.3283)
Perceived Stt. Violence $^2_{i,t-1}$	-1.2775** (.5291)	-2.0853*** (.6515)	-2.1055*** (.6528)	-1.5302*** (.5435)
Police $_{i,t-1}$.6381* (.3864)	.7512 (.4584)	.7553* (.4570)	.5867* (.3391)
Fire $_{i,t-1}$.0229 (.0334)	.0995*** (.0360)	.0998*** (.0360)	.0127 (.0313)
Male Percent $_{i,t-1}$	-.1952*** (.0735)	-.1805** (.0734)	-.1789** (.0731)	-.1800*** (.0674)
Young Adult Percent $_{i,t-1}$.2127* (.1093)	.1930* (.1030)	.1927* (.1032)	.1575 (.1009)
Tweets $_{i,t-1}$.0059*** (.0021)	.0088*** (.0030)	.0087*** (.0031)	.0055*** (.0020)
DV $_{i,t-1}$.0864*** (.0308)	.1998*** (.0352)	.2002*** (.0352)	.1141*** (.0294)
Protest Days $_{i,t}$				-.0006*** (.0001)
Consec. Stt. Violence $_{i,t}$.0638*** (.0064)
Intercept	.0278** (.0130)	.1161*** (.0184)	.1260*** (.0218)	.2655*** (.0271)
N	3,777	4,121	4,121	4,121
Adjusted R ²	.3489	.2684	.2686	.3505
Cluster SE	Y	Y	Y	Y
City FE	Y	Y	Y	Y

*p < .1; **p < .05; ***p < .01

The n-shaped relationship of state violence and subsequent protest size holds when controlling for time dynamics. Protest Days $_{i,t}$ refers to the length of a protest, and Consec. Stt. Violence $_{i,t}$ the number of consecutive days of state violence.

images; Figure A15 shows, and t-tests confirm, these users do not differ. Users may strategically choose the level of geographic specificity to assign to a tweet depending on the tweet content. Figure A16 shows the distribution of estimates for number of faces and perceived state and protester violence by country and four levels of geographic specificity. Most tweets are with neighborhood or city specificity, and the classifier results do not systematically vary

by geographic level within a country.

Section S9 presents nine additional robustness checks. Section S9.1 performs the same non-parametric examinations of *Perceived Protester Violence* $_{i,t-1}$ that were performed on *Perceived State Violence* $_{i,t-1}$. Section S9.2 presents the results from a vector autoregression, the results of which are consistent with Table 7. Section S9.3 aggregates tweets to their state or country to see if individuals selecting the tweet’s geographic level biases results. Section S9.4 confirms that modeling days without protest images as days of no protest does not affect results. Section S9.5 shows that different transformations of the dependent variable do not change results. Section S9.6 weights city-days by number of tweets and their inverse; results do not change, suggesting these findings are not an artifact of Twitter prevalence in certain locales. Section S9.7 analyzes tweets most likely to originate at a protest: those from mobile phones or during protest hours. Section S9.8 shows results controlling for tweets about protest but without images. Section S9.9 disaggregates results by country. In all models, the protester and state violence variables exhibit the same relationships found in the other results, though protester violence’s coefficient occasionally becomes close enough to zero to be indistinguishable from it.¹⁴

Discussion

The results presented here suggest that state violence generates an n-shaped correlation with subsequent protest size while protester violence is consistently negative. While emphasizing the severity of repression during protest policing is not new (Khawaja, 1993, Muller, 1985), measuring non-lethal repression as a continuous variable is. This insight is not measurable without applying recent advances in computer vision to large datasets of individual behavior, in this case geolocated images shared on Twitter.

Synthesizing this paper’s results with others’ suggests protest dynamics work as follows. Preventative repression, such as arresting a group’s leaders or seizing their offices, makes it harder for protests to start (Danneman and Ritter, 2013, Sutton, Butcher and Svensson, 2014,

¹⁴ We also reran the main model but with $\text{Log}(\text{Tweets}_{i,t-1})$, The results do not change.

Sullivan, 2016). Once started, protester violence decreases support for protest and therefore its subsequent size (Wasow, 2020). Repression produces differential effects depending on its severity: light state violence generates backlash while severe state violence “works”.

Future research should focus on several areas. It is possible that when in a protest cycle repression occurs correlates with its severity; future work should interact state violence with when in a protest cycle that violence occurs. Mechanisms connecting state and protester violence to changes in protest size also require investigation; emotions can be measured from social media text and images (Steinert-Threlkeld and Joo, 2020), and expectations can be measured with longitudinal surveys (Cantoni et al., 2019). In addition, global event datasets could be constructed using this methodology. An advantage of using images is that they are closer to a universal language than text (Graber, 1996). Classifiers generated from images are therefore less context dependent than text ones, so they can be applied across settings and time periods. For more detail on the advantages of images for generating protest event data, see Steinert-Threlkeld (2019).

Social media make valuable contributions to understanding subnational conflict. Metzger, Nagler and Tucker (2015) and Driscoll and Steinert-Threlkeld (2020) use social media as a quasi-poll to understand mass protests in Ukraine. Steinert-Threlkeld (2017) shows that coordinating activity correlates with more protests when that activity comes less influential members of society. Larson et al. (2019) studies 130,000,000 protesters in France, finding that protesters are embedded in denser social networks than bystanders. Muchlinski et al. (2020) measures electoral violence using CNNs and tweets. Social media, moreover, often generates more extensive records of protest than wire reports or newspapers, the primary source for existing event datasets (Steinert-Threlkeld and Joo, 2020, Steinhardt and Goebel, 2019). These data are not a fad or detached from reality: they offer unprecedented insights into human behavior and should continue to grow in importance to political science research.

If a picture is worth 1,000 words, then it would require approximately two kilobytes of storage (Jagenstedt, 2008). Images from consumer cell phones and digital cameras, however,

require at least three megabytes of storage, usually more. Even images shared on social media platforms, which are compressed from their original size, require hundreds of kilobytes of space. A picture, in other words, is worth anywhere from 50,000 (100 kilobytes) to 1,500,000 words (3 megabytes).¹⁵ A picture is actually worth a book, and there are vast libraries waiting to be explored.

¹⁵ This estimate is poetic. Another way to think of images is that they have high entropy, meaning they cannot be compressed as much as text. The greater size of images reflects this greater difficulty of compressing them, not necessarily a true quantum of information.

Acknowledgments. We would like to thank the editor and three reviewers for their excellent and punctual feedback. Justin Jerro, Xiaofeng Lin, Jun Luo, Jack Schafer, Donghyeon Won, and Cindy Yuan provided essential research assistance. Alexei Abrahams, Konstantin Ash, Michael Chwe, Emilio Ferrara, Tim Groeling, Susan Hyde, Jason Jones, Andrew Little, Susanne Lohmann, Brandon Merrell, Tamar Mitts, David Muchlinski, Kevin O’Brien, Jennifer Pan, Margaret Roberts, Jacob Shapiro, Joshua Tucker, Austin Wright, Nils Weidmann, Thomas Zeitzoff, and Yuri Zhukov have provided valuable feedback throughout. Comments at APSA, EPSA, IC2S2, ISA, Peace Science, Politics and Computational Social Science, Political Methodology, and SPSA also refined the paper. Research costs money, and support from the NSF (SBE-SMA #1831848), California Center for Population Research, and internal UCLA grants made this work possible. All remaining errors are unintentional.

References

- Asal, Victor, Richard Legault, Ora Szekely and Jonathan Wilkenfeld. 2013. "Gender ideologies and forms of contentious mobilization in the Middle East." *Journal of Peace Research* 50(3):305–318.
- Bashir, Nadia Y., Penelope Lockwood, Alison L. Chasteen, Daniel Nadolny and Indra Noyes. 2013. "The ironic impact of activists: Negative stereotypes reduce social change influence." *European Journal of Social Psychology* 43(7):614–626.
- Battaglini, Marco. 2017. "Public Protests and Policy Making." *Quarterly Journal of Economics* 132(1):485–549.
- Benford, Robert D. and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology* 26:611–639.
- Biggs, Michael. 2003. "Positive feedback in collective mobilization: The American strike wave of 1886." *Theory and Society* 32:217–254.
- Boschee, Elizabeth, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz and Michael Ward. 2015. "ICEWS Coded Event Data." <http://dx.doi.org/10.7910/DVN/28075> .
- Bradley, Ralph Allan and Milton E Terry. 1952. "Rank analysis of incomplete block designs: I. The method of paired comparisons." *Biometrika* 39(3/4):324–345.
- Bratton, Michael and Nicolas Van De Walle. 1992. "Popular Protest and Political Reform in Africa." *Comparative Politics* 24(4):419–442.
- Bruns, Axel., T. Highfield and J. Burgess. 2013. "The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks." *American Behavioral Scientist* 57(7):871–898.

- Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge: MIT Press.
- Cantoni, Davide, David Y. Yang, Noam Yuchtman and Y. Jane Zhang. 2019. “Protests as Strategic Games: Experimental Evidence from Hong Kong’s Anti-Authoritarian Movement.” *The Quarterly Journal of Economics* 134(2):1021–1077.
- Clark, David H. and Patrick M. Regan. 2016. “Mass Mobilization.” <https://www.binghamton.edu/massmobilization/about.html> .
- Danneman, Nathan and Emily H. Ritter. 2013. “Contagious Rebellion and Preemptive Repression.” *Journal of Conflict Resolution* 58(2):254–279.
- Davenport, Christian. 2007. “State Repression and Political Order.” *Annual Review of Political Science* 10(1):1–23.
- Davenport, Christian and David A. Armstrong II. 2004. “Democracy and the Violation of Human Rights: A Statistical Analysis from 1976 to 1996.” *American Journal of Political Science* 48(3):538–554.
- DeNardo, James. 1985. *Power in Numbers: The Political Strategy of Protest and Rebellion*. Princeton: Princeton University Press.
- Dowd, Caitriona, Patricia Justino, Roudabeh Kishi and Gauthier Marchais. 2020. “Comparing ‘New’ and ‘Old’ Media for Violence Monitoring and Crisis Response in Kenya.” *Research and Politics* 7(3):1–9.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York City: Harper and Row.
- Driscoll, Jesse and Zachary C. Steinert-Threlkeld. 2020. “Social media and Russian territorial irredentism: some facts and a conjecture conjecture.” *Post-Soviet Affairs* 36(2):101–121.

- Earl, Jennifer, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan and Ashley Arlotti. 2013. "This Protest Will Be Tweeted: Twitter and protest policing during the Pittsburgh G20." *Information, Communication & Society* 16(4):459–478.
- Francisco, Ronald A. 2004. "After the Massacre: Mobilization in the Wake of Harsh Repression." *Mobilization: An International Journal* 9(2):107–126.
- Girshick, Ross. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. pp. 1440–1448.
- Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36(2):369–385.
- González, Felipe. 2020. "Collective action in networks: Evidence from the Chilean student movement." *Journal of Public Economics* 188:104220.
- Graber, Doris A. 1996. "Say It with Pictures." *The ANNALS of the American Academy of Political and Social Science* 546(1):85–96.
- Gupta, Dipak K., Harinder Singh and Tom Sprague. 1993. "Government Coercion of Dissidents: Deterrence or Provocation?" *Journal of Conflict Resolution* 37(2):301–339.
- Gurr, Ted Robert and Will H. Moore. 1997. "Ethnopolitical Rebellion: A Cross-Sectional Analysis of the 1980s with Risk Assessments for the 1990s." *American Journal of Political Science* 41(4):1079–1103.
- Hess, David and Brian Martin. 2006. "Repression, Backfire, and the Theory of Transformative Events." *Mobilization: An International Journal* 11(2):249–267.
- Jagenstedt, Philip. 2008. "How much a thousand words are worth." <https://blog.foolip.org/2008/05/17/how-much-a-thousand-words-are-worth/>.
- Jasper, James M. 2011. "Emotions and Social Movements: Twenty Years of Theory and Research." *Annual Review of Sociology* 37(1):285–303.

- Kärkkäinen, Kimmo and Jungseock Joo. 2019. "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age." *arXiv preprint arXiv:1908.04913* .
- Khawaja, Marwan. 1993. "Repression and Popular Collective Action: Evidence from the West Bank." *Sociological Forum* 8(1):47–71.
- Larson, Jennifer M, Jonathan Nagler, Jonathan Ronen and Joshua A Tucker. 2019. "Social Networks and Protest Participation: Evidence from 130 Million Twitter Users." *American Journal of Political Science* 63(3):690–705.
- Lawrence, Adria K. 2017. "Repression and Activism among the Arab Spring's First Movers: Evidence from Morocco's February 20th Movement." *British Journal of Political Science* 47(3):699–718.
- Lim, Merlyna. 2013. "Framing Bouazizi: 'White lies', hybrid network, and collective/connective action in the 2010-11 Tunisian uprising." *Journalism* 14(7):921–941.
- Little, Andrew T. 2015. "Communication Technology and Protest." *Journal of Politics* 78(1):152–166.
- Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91." *World Politics* 47(1):42–101.
- McPhail, Clark and John McCarthy. 2004. "Who Counts and How: Estimating the Size of Protests." *Contexts* 3(3):12–18.
- Metzger, Megan, Jonathan Nagler and Joshua a. Tucker. 2015. "Tweeting Identity? Ukrainian, Russian, and #Euromaidan." *Journal of Comparative Economics* 44(1):16–40.
- Moore, Will H. 2000. "The Repression of Dissent: A Substitution Model of Government Coercion." *Journal of Conflict Resolution* 44(1):107–127.

- Muchlinski, David, Xiao Yang, Sarah Birch, Craig MacDonald and Iadh Ounis. 2020. "We Need to Go Deeper: Measuring Electoral Violence using Convolutional Neural Networks and Social Media." *Political Science Research and Methods* pp. 1–18.
- Muller, Edward N. 1985. "Income Inequality, Regime Repressiveness, and Political Violence." *American Sociological Review* 50(1):47–61.
- Muñoz, Jordi and Eva Anduiza. 2019. "‘If a fight starts, watch the crowd’: The effect of violence on popular support for social movements." *Journal of Peace Research* 56(4):485–498.
- Murdie, Amanda and Carolin Purser. 2017. "How protest affects opinions of peaceful demonstration and expression rights." *Journal of Human Rights* 16(3):351–369.
- Nordås, Ragnhild and Christian Davenport. 2013. "Fight the Youth: Youth Bulges and State Repression." *American Journal of Political Science* 57(4):926–940.
- O’Brien, Kevin J. and Yanhua Deng. 2015. "Repression Backfires: tactical radicalization and protest spectacle in rural China." *Journal of Contemporary China* 24(93):457–470.
- Olivier, Johan L. 1991. "State repression and collective action in South Africa, 1970 – 84." *South African Journal of Sociology* 22(4):109–117.
- Olzak, Susan, Maya Beasley and Johan Olivier. 2003. "The Impact of State Reforms on Protest Against Apartheid in South Africa." *Mobilization* 8(1):27–50.
- Pearlman, Wendy. 2013. "Emotions and the Microfoundations of the Arab Uprisings." *Perspectives on Politics* 11(02):387–409.
- Purdy, Chase. 2018. "China is launching a dystopian program to monitor citizens in Beijing." <https://qz.com/1473966/china-is-starting-a-big-brother-monitoring-program-in-beijing/>.

- Qin, Bei, David Strömberg and Yanhui Wu. 2017. “Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda.” *Journal of Economic Perspectives* 31(1):117–140.
- Raleigh, Clionadh, Andrew Linke, Havard Hegre and Joakim Karlsen. 2010. “Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature.” *Journal of Peace Research* 47(5):651–660.
- Ritter, Emily Hencken and Courtenay R. Conrad. 2016. “Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression.” *American Political Science Review* 110(1):85–99.
- Robertson, Graeme B. 2007. “Strikes and Labor Organization in Hybrid Regimes.” *American Political Science Review* 101(04):781–798.
- Rodríguez, Francisco. 2020. *Things Fall Apart: Nicolás Maduro and the Unraveling of Venezuela’s Populist Experiment 2012-2020*. Unpublished Manuscript.
- Salehyan, Idean, Cullen Hendrix, Jesse Hammer, Christina Case, Christopher Linebarger, Emily Stull and Jennifer Williams. 2012. “Social Conflict in Africa: A New Database.” *International Interactions* 38(4):503–511.
- Schaftenaar, Susanne. 2017. “How (wo)men rebel: Exploring the effect of gender equality on nonviolent and armed conflict onset.” *Journal of Peace Research* 54(6):762–776.
- Shadmehr, Mehdi and Dan Bernhardt. 2011. “Collective Action with Uncertain Payoffs: Coordination, Public Signals, and Punishment Dilemmas.” *American Political Science Review* 105(04):829–851.
- Shakir, Omar. 2014. All According to Plan: The Rab’a Massacre and Mass Killings of Protesters in Egypt. Technical report Human Rights Watch.

URL: <https://www.hrw.org/report/2014/08/12/all-according-plan/raba-massacre-and-mass-killings-protesters-egypt>

Siegel, David A. 2011. “When Does Repression Work? Collective Action in Social Networks.” *The Journal of Politics* 73(04):993–1010.

Simpson, Brent, Robb Willer and Matthew Feinberg. 2018. “Does Violent Protest Backfire? Testing a Theory of Public Reactions to Activist Violence.” *Socius: Sociological Research for a Dynamic World* 4:237802311880318.

Sobolev, Anton, Jungseock Joo, Keith Chen and Zachary C. Steinert-Threlkeld. 2020. “News and Geolocated Social Media Accurately Measure Protest Size Variation.” *American Political Science Review* pp. 1–9.

Steinert-Threlkeld, Zachary C. 2017. “Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring.” *American Political Science Review* 111(02):379–403.

Steinert-Threlkeld, Zachary C. 2019. “Comment: The Future of Event Data is Images.” *Sociological Methodology* 49.

Steinert-Threlkeld, Zachary C and Jungseock Joo. 2020. “Protest Event Data from Geolocated Social Media Content.” *APSA Preprints*: <https://doi.org/10.33774/apsa-2020-mjz4s>

Steinhardt, H. Christoph and Christian Goebel. 2019. “Better coverage, less bias: Using social media to measure protest in authoritarian regimes.” https://www.researchgate.net/publication/332112415_Better_coverage_less_bias_Using_social_media_to_measure_protest_in_authoritarian_regimes .

Stephan, Maria J. and Erica Chenoweth. 2008. “Why Civil Resistance Works.” *International Security* 33(1):7–44.

- Sullivan, Christopher M. 2016. "Political Repression and the Destruction of Dissident Organizations." *World Politics* 68(4):645–676.
- Sutton, Jonathan, Charles R. Butcher and Isak Svensson. 2014. "Explaining political jiu-jitsu: Institution-building and the outcomes of regime violence against unarmed protests." *Journal of Peace Research* 51(5):559–573.
- Tucker, Joshua A. 2007. "Enough! Electoral Fraud, Collective Action Problems, and Post-Communist Colored Revolutions." *Perspectives on Politics* 5(03):535.
- Tullock, Gordon. 1971. "The Paradox of Revolution." *Public Choice* 11:89–99.
- Urdal, Henrik. 2006. "A Clash of Generations? Youth Bulges and Political Violence." *International Studies Quarterly* 50:607–629.
- Urdal, Henrik and Kristian Hoelscher. 2012. "Explaining Urban Social Disorder and Violence: An Empirical Study of Event Data from Asian and Sub-Saharan African Cities." *International Interactions* 38(4):512–528.
- Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer and Alessandro Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." *arXiv preprint arXiv:1703.03107v2*.
- Wasow, Omar. 2020. "Agenda Seeding: How 1960s Black Protests Moved Elites, Public Opinion and Voting." *American Political Science Review* 114(3):638–659.
- Weidmann, Nils B. and Espen Geelmuyden Rod. 2018. Coding Protest Events in Autocracies. In *The Internet and Political Protest in Autocracies*. Oxford University Press chapter Chapter 4.
- Wouters, Ruud and Stefaan Walgrave. 2017. "Demonstrating Power: How Protest Persuades Political Representatives." *American Sociological Review* 82(2):361–383.

- Young, Lauren E. 2019. "The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe." *American Political Science Review* 113(1):140–155.
- Zhang, Han and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." *Sociological Methodology* 49:1–48.
- Zhao, Dingxin. 1998. "Ecologies of Social Movements: Student Mobilization during the 1989 Prodemocracy Movement in Beijing." *American Journal of Sociology* 103(6):1493–1529.

Biographical Statements. Zachary C. Steinert-Threlkeld is an assistant professor at UCLA, Los Angeles, CA 90095. Alexander Chan was a graduate research assistant at UCLA, Los Angeles, CA 90095 and is now an Infrastructure Data Scientist at Facebook. Jungseock Joo is an assistant professor at UCLA, Los Angeles, CA 90095.

Supplementary Materials for

How State and Protester Violence Affect Protest Dynamics

S1 Detail

S1.1 Convolutional Neural Networks

We use convolutional neural networks (CNN) to identify and analyze protest images. A CNN is a type of artificial neural network, a machine learning algorithm inspired by the human brain (Rosenblatt, 1958), that has gained widespread adoption in the field of computer vision. It has been successful in various applications including face recognition (Sun et al., 2014, Parkhi et al., 2015, Baltrušaitis, Robinson and Morency, 2016), object detection (Girshick et al., 2014, Ren et al., 2015, Redmon et al., 2016), and self-driving cars (Huval et al., 2015, Bojarski et al., 2016, Xu et al., 2017). For methodological detail on computer vision for political scientists, see Joo and Steinert-Threlkeld (2018), Cantu (2019), and Williams, Casas and Wilkerson (2020).

A CNN is a function whose outputs are computed through a series of sequential operations from the input values. For example, in image classification, the input is an image (i.e., an array of color intensities at pixels) and the output is the class that the image belongs to, e.g., an object category such as “police” or “sign”. A CNN transforms the given input through many operations until it reaches the final step which produces the output.

Each operation is also called a layer, and a CNN usually has multiple convolutional layers. A convolutional layer performs convolution, which consists of an element-wise multiplication between pixel values and filter values (connection strengths) and a summation over adjacent pixels. A CNN contains many layers and the output of a layer becomes the input of the next layer. The input to the first layer is the original input image’s pixel intensities. For non-first layers, their inputs are given from nodes on two dimensional grid in the previous layer, not from the image pixels.

A CNN, as well as other artificial neural networks, is trained to minimize a loss function, a measure of difference between the model prediction and ground truth label. This optimization is typically done by stochastic gradient descent.

Each CNN is defined by its architecture – the structural configuration specifying the number of layers, the order of their placement, and the types of non-linear transformations used. There exist many different CNN architectures with different properties. The architecture of our model is a “Residual Network” (ResNet) (He, Zhang, Ren and Sun, 2016) and has 50 convolutional layers. ResNet has been used in many of the state-of-the-art computer vision applications such as object detection (Ren et al., 2015) and human pose estimation (Güler, Neverova and Kokkinos, 2018). We use a ResNet model pre-trained on ImageNet data and finetune it with our data.

S2 Measurement Using Social Media Images

S2.1 Protest Classification by Weakly-supervised Learning

Step 1. We use a combination of weakly-supervised and supervised learning. In weakly-supervised learning, the ground-truth labels on the target variable are not directly available but can be inferred from other variables (Bergamo and Torresani, 2010). For instance, we can use any online image search service to query images with a particular keyword (e.g., “protest”), and this step will furnish a large quantity of relevant images. While this sample set will contain some noisy data, it is still useful to train a rough initial model which can be used to fetch better samples. These samples can then be manually annotated as in typical supervised learning.

Specifically, we first collected about 10,000 protest images from Google Image Search by using manually selected keywords such as “protest,” “riot,” “Venezuela Protest,” “Hong Kong protest” and many others, as well 90,000 non-protest, hard-negative images by using keywords including “concert,” “stadium,” or “airport crowd.” These negative examples are called hard-negatives because they look similar to protest images, and classifiers can easily misassign their labels. Since these images are simply outputs of search queries, their assigned labels are not accurate. For example, the query of “protest” may return some photographs of politicians. However, we did not verify the correct classification labels of these images because the main purpose of this first step is to train a rough classifier with the assumption that the majority of labels are still correct.

Step 2. Using these data, we trained a CNN whose only output denotes whether an image captures a protest event or not. We then applied this classifier to geolocated images from Twitter and obtained the classification scores. Each score can be considered as the confidence about the output, the probability of the input image containing protesters. For more details of how CNNs work, see Joo and Steinert-Threlkeld (2018), Cantu (2019), and Williams, Casas and Wilkerson (2020).

Step 3. Twitter provides tweets in real time through its streaming application programming interface (API). We have collected these tweets in real-time, approximately five million per day, since August 26, 2013. For more information on working with Twitter data, see [Steinert-Threlkeld \(2018\)](#). We then query the stored tweets to extract those from countries and days of interest and next select only those tweets that contain images. These images form the raw material from which we generate our protest data.

We apply the protest classifier to images from periods and countries during which protest occurred. These periods, shown in [Table 2](#), generated 42,579,188 tweets containing 4,456,981 images. The classifier is applied to all 4.46 million plus the 100,000 from Google, and all images with a classification score less than .6 are dropped as they are most likely non-protest images. The remaining 115,060 potential protest images were then stratified based on their classification scores and sampled to ensure that the chosen images capture diverse visual features. This process resulted in 40,764 images that form our training set.

S2.2 Protest and Scene Classification

Step 4. Amazon Mechanical Turk provided the labor to manually annotate these 40,764 images. We asked the workers to identify the twelve non-face features shown in [Table A1](#).

[Figure A1](#) provides examples of the AMT annotation pages. In the first task, each annotator was presented with an image and asked to judge if the image captures a protest. We assigned two workers to each image, and if the two workers did not agree, the image was sent to a third judge for a final verification. 11,659 of the training images contain a protest. Similarly, in the second task, annotators label the attributes listed in [Table A1](#) that are not related to faces or violence, such as “police”, “fire”, “children”, or “flag”.

As violence is a subjective and continuous variable, we used pairwise comparison annotation to generate an estimate of the *perceived violence* in an image. Among the 11,659 protest images, we randomly sampled image pairs such that each image is paired ten times, creating 58,295 ($11,659 \times 10 \div 2$) pairs to annotate. We then assigned ten workers for each pair and asked them to select which image looks more violent than the other. To assign the continuous

violence score to each image, we use the Bradley-Terry model (Bradley and Terry, 1952) and scaled the scores to the range of $[0, 1]$. Such a pairwise comparison method requires more annotations but can produce more reliable and consistent ratings for subjective assessment of photographs (Kovashka, Parikh and Grauman, 2012, Joo et al., 2014).

Step 5. The 40,764 annotated images train a CNN which produces outputs for the twelve labels. We used 80% of the images as the training set and the rest as the validation set. For the labels that are not face or violence related, we use a binary cross entropy loss function; for violence, which is continuously valued, we use mean squared error. For more technical details in model training, see Won, Steinert-Threlkeld and Joo (2017).

S2.3 Measuring Faces

Step 6. We use the FairFace model developed by Kärkkäinen and Joo (2019) to classify gender and age of people in images. It captures these attributes better than other models, such as FaceNet (Schroff, Kalenichenko and Philbin, 2015) or Face++, because it is trained on a large corpus of images of varying resolution, perspective, and lighting, the YFCC100M dataset (Thomee et al., 2016). This dataset is in contrast to other datasets whose images tend to be high quality, well-lit, and from the same perspective (Liu et al., 2015). Figure A2 shows an image from South Korea from our Twitter corpus with the face classifier applied. This step also provides the count of the number of faces per photo.

S2.4 Classifier Calibration

For binary variables in our analysis, we need to transform continuous outputs from the CNN to binary values by choosing a decision threshold such that we can assert an image contains the variable of interest. The optimal decision threshold needs to be chosen to balance true positive and true negative rates, evaluated on the target data distribution. To this end, we chose 3,000 protest images from additional random samples from our Twitter pipeline and used Amazon Mechanical Turk to annotate them. We then generated a precision-recall curve

for each attribute and chose the threshold at the minimum precision of .845.¹ For each image and each attribute, our model therefore produces a probability estimate (a real number) via the classifier as well as a binary output (0 or 1). Table A2 shows the twelve attributes and their thresholds. Figure A3 shows the precision-recall curve for each attribute, providing the threshold value for each.


Table A1: List of visual attributes.

Attribute	Description
1. Protester Violence	How violent protesters are.
2. State Violence	How violent the state is.
3. Police	Police or troops are present in the scene.
4. Fire	There is fire or smoke in the scene.
5. Gender	Is the face male or female?
6. Age	0-2, 3-9, 10-19, . . . , 70+
7. Face	Presence of a face.
8. Race	Is the face White, Middle Eastern, East Asian, Southeast Asian, Black, Indian, or Latino?
9. Group 20	There are roughly more than 20 people in the scene.
10. Group 100	There are roughly more than 100 people in the scene.
## Children	Children are in the scene.
## Shout	One or more people shouting.
## Photo	Protesters holding signs or a photograph of a person (politicians or celebrities).
## Flag	There are flags in the scene.
## Night	It is at night.
## Sign	Protesters holding a visual sign (on paper, panel, or wood).

NB: Attributes without numbers could not be classified precisely enough to consider for inclusion in regression. Attributes in **bold** are generated using the face classifier.


¹ One could also use another method such as F-measure to choose the optimal decision threshold. We choose to maintain the minimum precision (true positive rate) at a high point for every attribute, rather than trying to detect more relevant images while making more mistakes.

Figure A1: Examples of Our Annotation Interface (in Amazon Mechanical Turk)




Q1. Does this image contain a scene of protest?

Protest Uncertain Not Protest



Q2. Does this image contain a scene of protest?


Protest Uncertain Not Protest



Q3. Does this image contain a scene of protest?

Protest Uncertain Not Protest

Q0. Please answer all the questions for the below image.



Protesters (or a protester) holding **visual signs** (on a paper, panel, or wood).
 Yes Not sure No

There is **fire or smoke** in the scene.
 Yes Not sure No

Children (or a child) are in the scene.
 Yes Not sure No

There are roughly **more than 100 people** in the scene.
 Yes Not sure No

It is at **night**.
 Yes Not sure No

Protesters (or a protester) holding signs of a **photograph** of a person (politicians or celebrities).
 Yes Not sure No

Police or troops are present in the scene.
 Yes Not sure No


There are roughly **more than 20 people** in the scene.
 Yes Not sure No

There are **flags** in the scene.
 Yes Not sure No

There is one or more people **shouting**.
 Yes Not sure No

Q0. Choose the image that you feel is more violent.

Image 1



Similar

Image 2




Figure A2: Example Results of the FairFace Model



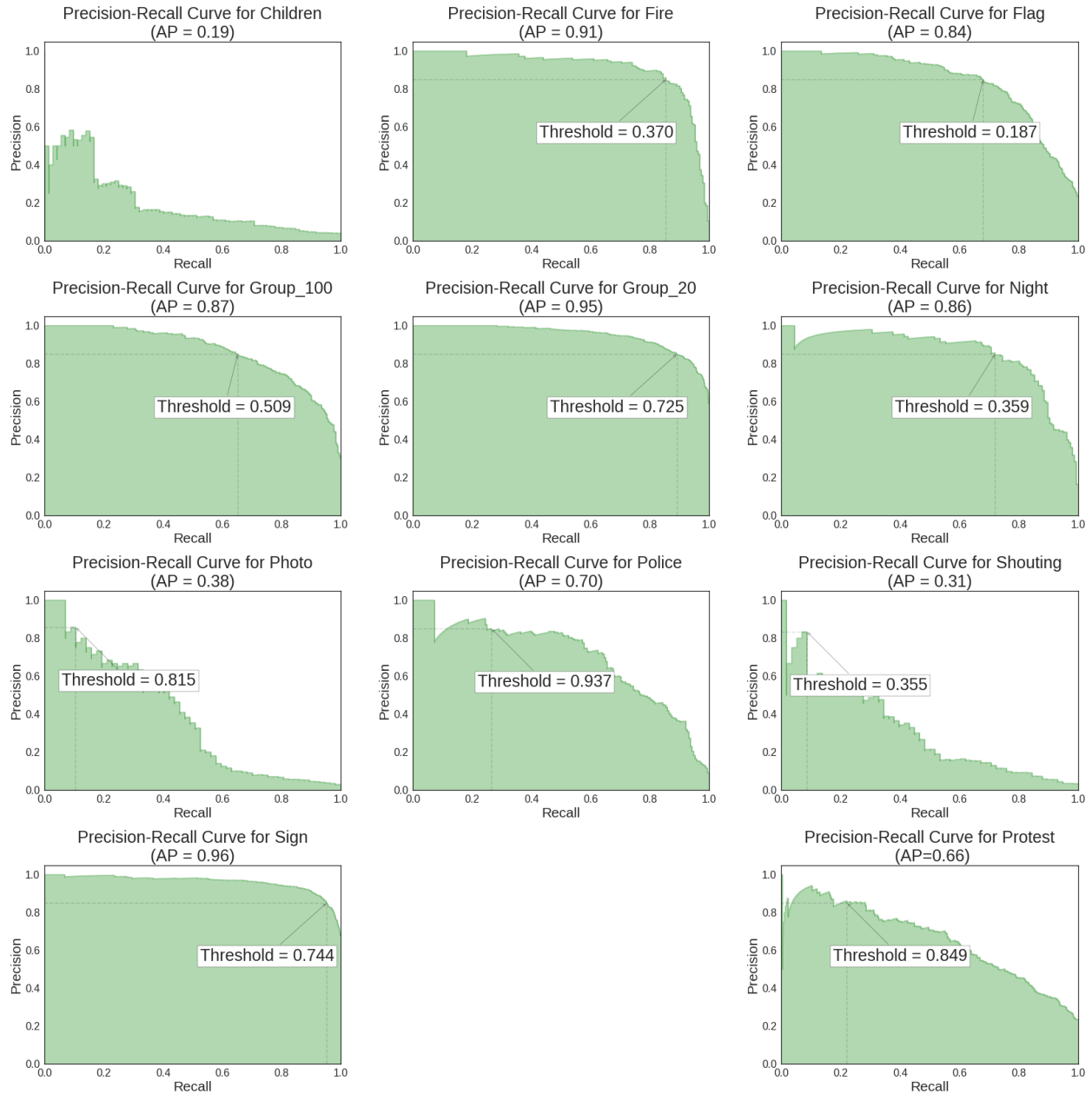
Note: The FairFace model accurately retrieves facial features of protests.

Table A2: Attributes and Thresholds

Attribute	Threshold
Protester Violence	.021
State Violence	.01
Police	.937
Fire	.37
Child	.15
Small Group	.725
Large Group	.509
Shout	.355
Photo	.815
Flag	.187
Night	.359
Sign	.744

Note: If a classifier generates a value equal to or above these thresholds, the image contains that attribute.

Figure A3: Precision-Recall Curves For Binary Attributes.



Note: AP stands for average precision, which is the standard accuracy measure for binary classification. AP is also equal to the area under the precision-recall curve.

S2.5 Evaluating the CNN

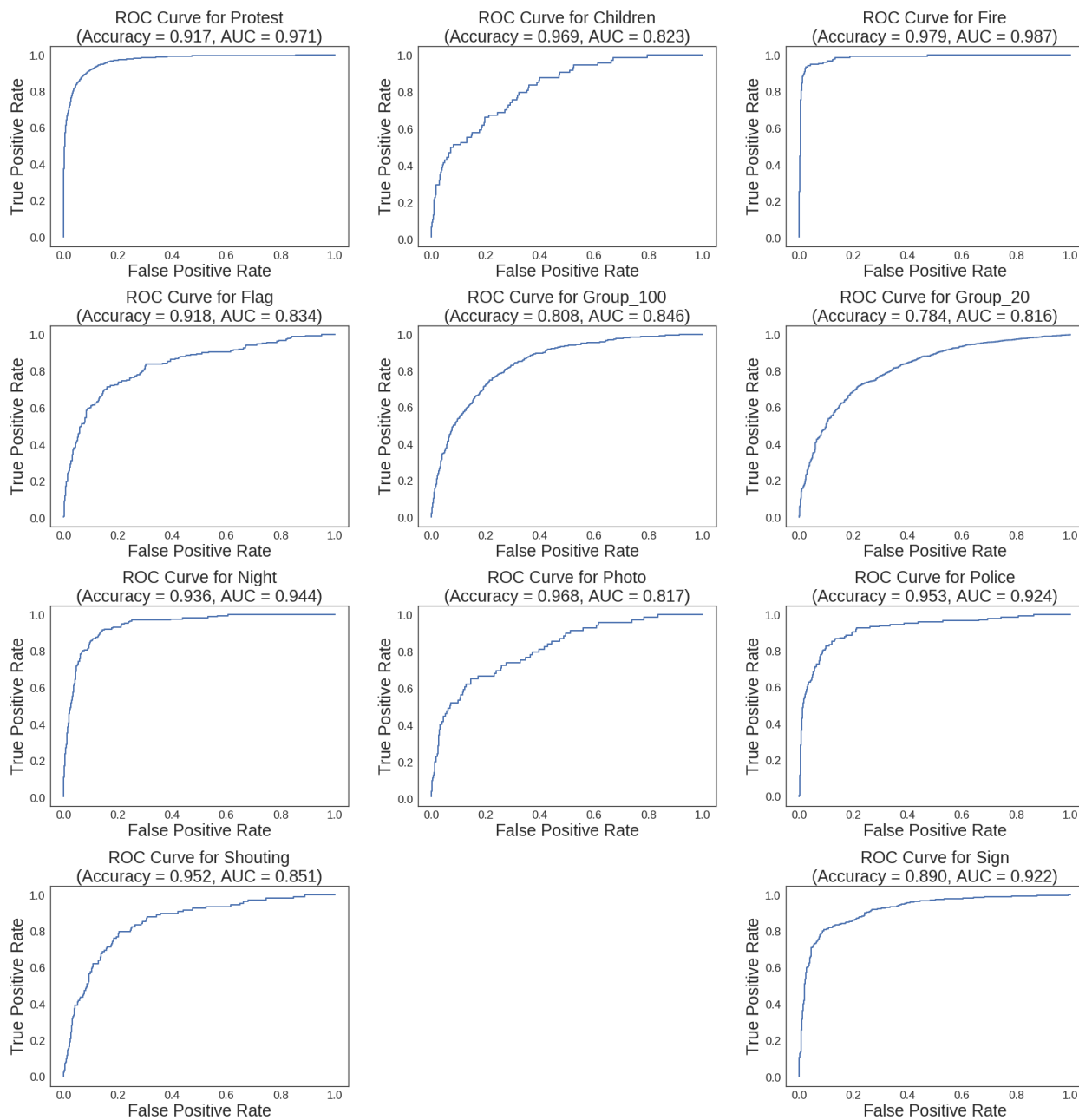
This section shows results from the manual validation performed by workers from Amazon Mechanical Turk. Figure A4 shows the model performance measured on the validation set. The Receiver-Operator Curve (ROC) documents the relationship between false-positive and true-positives, with a higher area-under-curve (AUC) corresponding to a better accuracy. Generating it requires combining the classifier outputs with hand-labeled data. Visually, the closer the curve is to the upper-left corner, the better the classifier for that label.

Figure A5 shows a scatterplot of the classifier’s output for violence against the rating recovered from the Bradley-Terry model. It also shows the ROC curve for protester violence and state violence. Like the previous figure, it can only be generated using manual verification. All three subfigures demonstrate strong performance of our classifier’s ability to measure perceived violence as determined by human coders from Amazon Mechanical Turk.

To intuitively visualize how the classifier works, we use Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2016). Grad-CAM highlights important regions for classifying the concept in an image. It does so by tracing back the classification outcome to the input image through passing gradients. The results are shown in Figure A6, with red color indicating more important regions. For technical details, see Selvaraju et al. (2016).

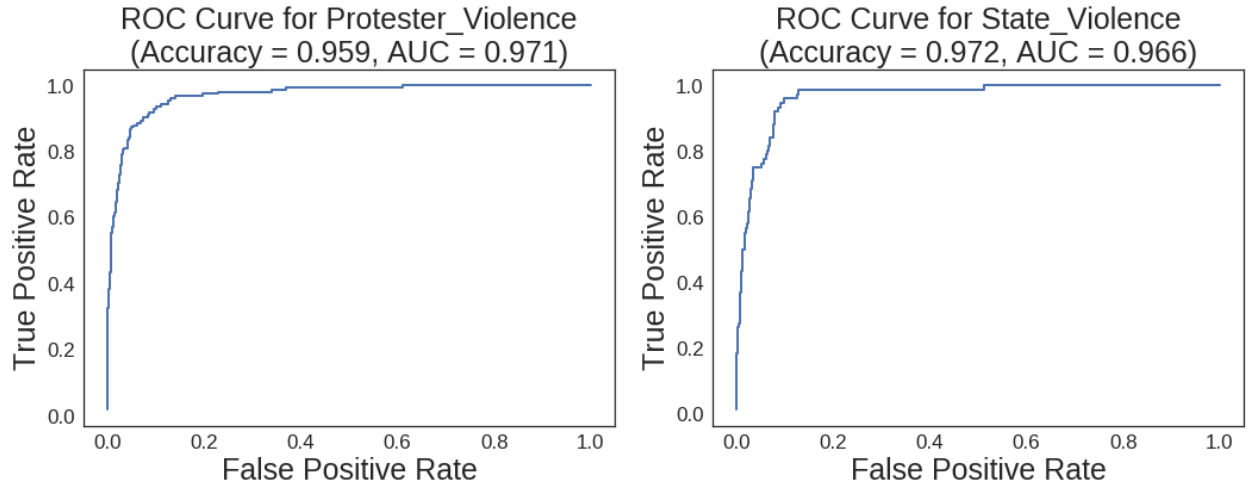
Figure A7 arrays images from each category by the classification scores from the CNN. As the classification score approaches 1 for each category, images more closely exhibit the visual concept. This manual validation also suggests that

Figure A4: Model Performance



Note: ROC curves generated from Amazon Mechanical Turk manual validation of classifier output.

Figure A5: Validating Violence Measurement



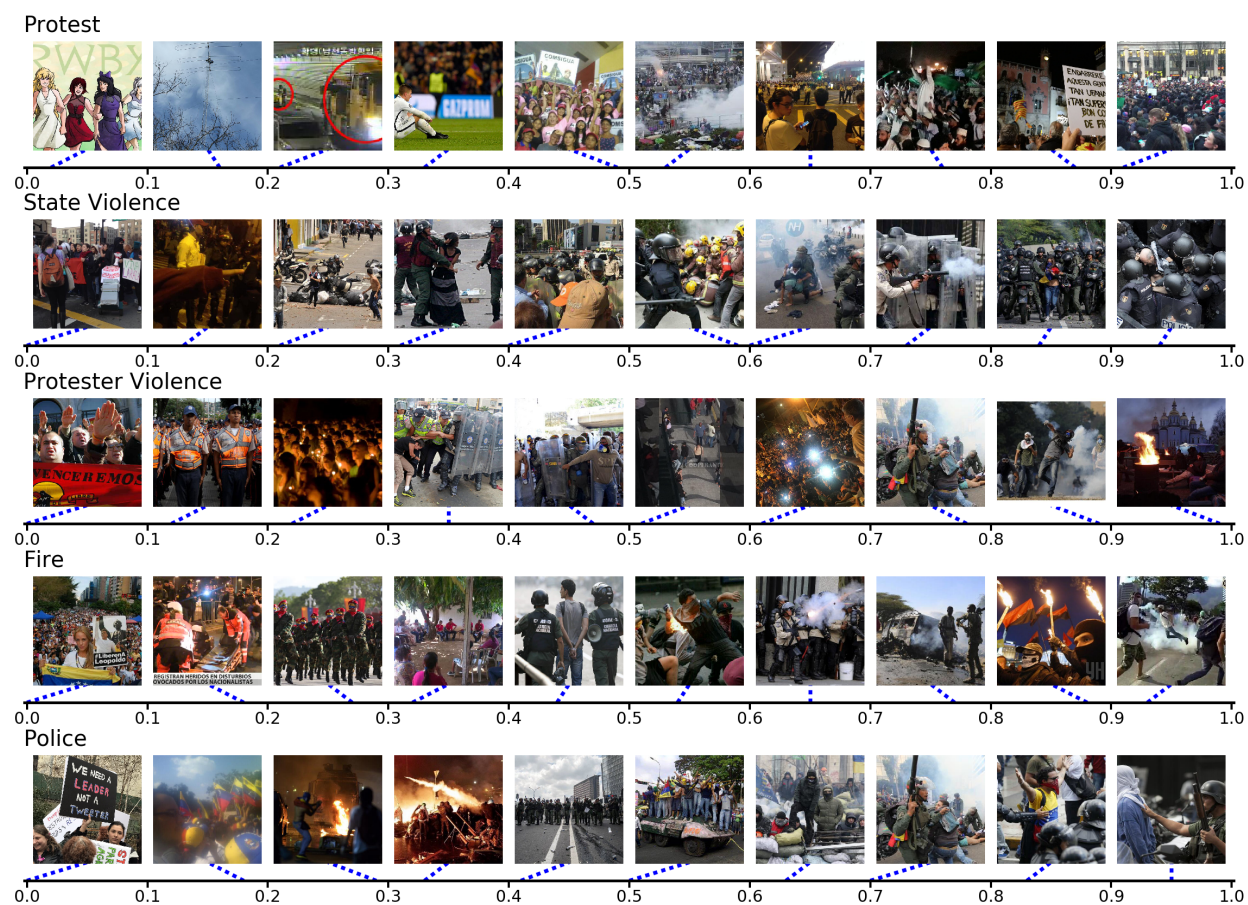
Note: ROC curves generated from Amazon Mechanical Turk manual validation of classifier output.

Figure A6: Visualization of Region Importance in Classification Using Grad-CAM: Important regions which more contribute to the classification for each attribute are highlighted in red.



Note: The closer to red a pixel is colored, the more it contributes to that label.

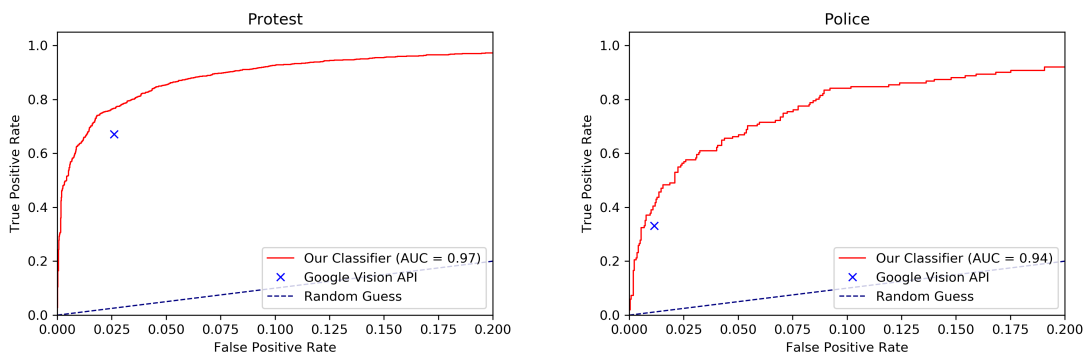
Figure A7: Sample Classifier Estimates by Category.



Note: Images ordered by their classification scores. The dotted lines mark the exact classification scores of corresponding image.

Finally, to compare the classification performance of existing commercial classifiers against our own classifier, we use Google Vision’s label detection on the test set of UCLA Protest Image Database (Won, Steinert-Threlkeld and Joo, 2017) and measure the classification accuracy. That dataset has 11,000 test images with various labels related to protest activity. Since Google’s label detection automatically identifies visual concepts and objects in many categories, including protest and police, from an input image, we directly compared its accuracy with our model accuracy. As shown in Figure A8, the protest and scene models classified protest and police more accurately than the Google Vision API. The superior result is most likely due to the fact that we specifically collected diverse protest images and hard-negatives from many sources. The Google Vision API may perform better on general image classification and can be very useful when one does not have any training data.

Figure A8: Classification performance comparison between our model and the public model from Google’s Vision API.



Note: Our CNN outperforms Google Vision for identifying protest and police.

S3 Manual Validation of Classifier

S3.1 Inter-coder reliability from Amazon Mechanical Turk

We used Fleiss’ Kappa to measure the inter-coder reliability of our training image annotations. In many cases, the inter-coder reliability is typically measured on the coding data on which the actual analysis is conducted. In our study, the manual coding was performed on the training data, and the reliability was measured for the annotations to ensure that

the models are trained in a consistent manner. Table A3 shows the estimated reliability statistics, and the level of agreement is sufficiently strong (Koo and Li, 2016).

Table A3: Inter-coder reliability

Label	Kappa
Perceived Violence	.316
Perceived Protester Violence	.566
Perceived State Violence	.473
Large Group	.434
Small Group	.388
Police	.564
Fire	.702
Child	.457

Note: Intercoder reliability for manual validation performed on Amazon Mechanical Turk. Agreement is high (Koo and Li, 2016).

S3.2 Local Validation on Out-of-sample Images

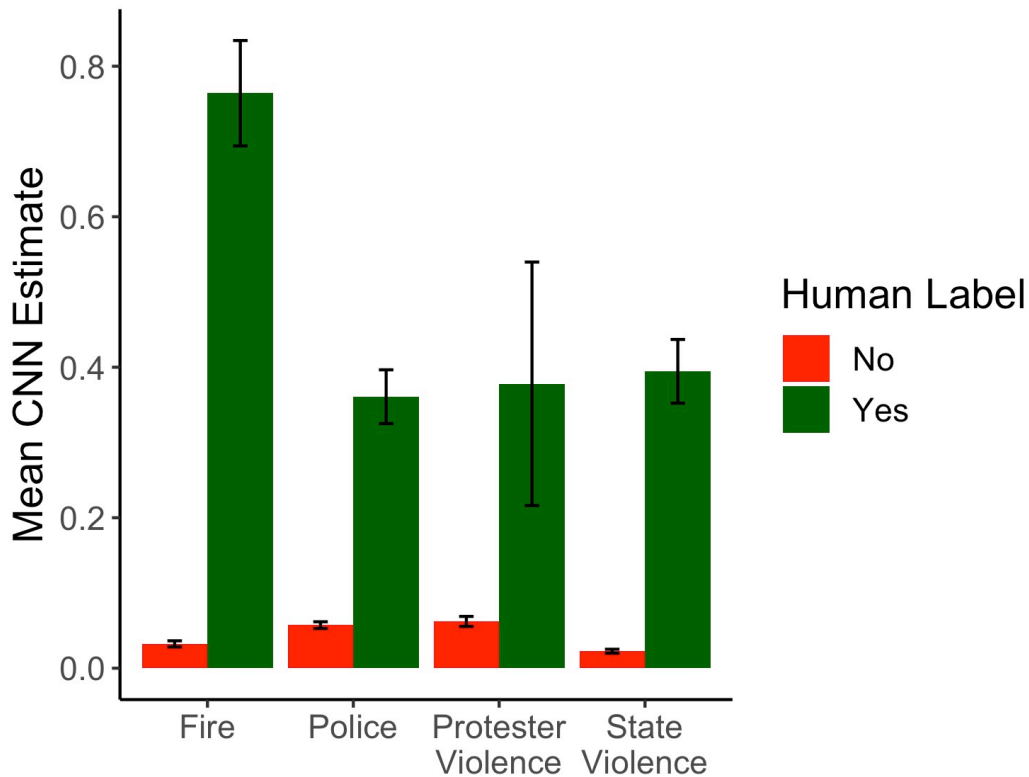
In addition to the manual validation conducted by Amazon Mechanical Turk workers that Section S2.5 shows, we conducted validation checks using a team of three research assistants. This section describes their intercoder reliability and results. The team was trained on 102 images, and each member then coded 300. This second round of manual validation reinforces shows that the convolutional neural networks we use to identify state and protester violence as well as other scene attributes measures those attributes, not noise.

To measure intercoder reliability for the number of faces in an image, a continuous measure, we use the intraclass correlation coefficient. It is .836, which is considered very good (Koo and Li, 2016). Converting the estimates to ordinal values in increments of five gener-

ates a Fleiss' kappa of .61, with agreement particularly high for images containing 0-4 and 5-9 faces. The Fleiss' kappa for State Violence is .69; for Police, .97; and for Fire, .87. For Protester Violence it is -.004, though that is because only one coder labeled one image as containing protester violence.

Finally, the classifier ratings for containing state violence, protester violence, police, or fire is much higher for those images human coders also identify as containing one of those. Figure A9 shows these results.

Figure A9: The Classifier Generates Significantly Higher Ratings for Images Humans Identify Containing that Label



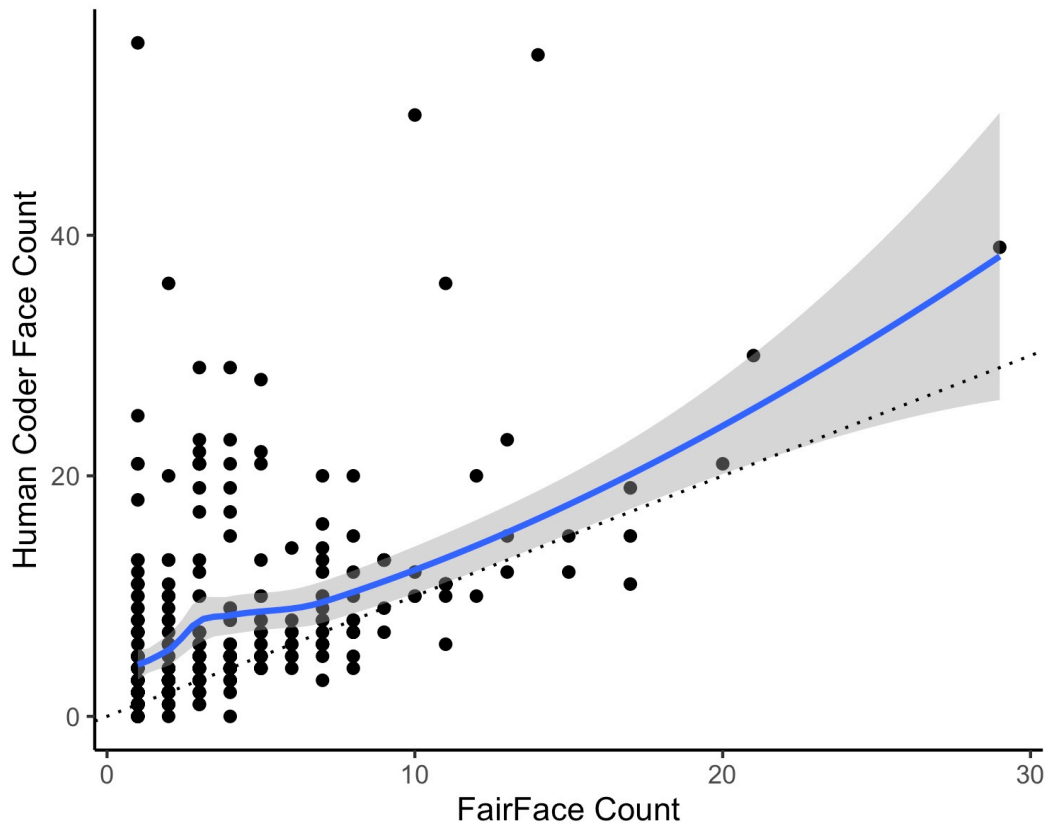
Note: The convolutional neural network records much higher estimates (y-axis) for images humans identify as containing that label (x-axis).

S4 Verifying Using Faces to Measure Protest Size

We also compare the number of faces FairFace records with those the annotators count. Figure A10 shows that FairFace slightly undercounts the number of faces per protest image

but does so consistently. A linear regression of the manual count on the FairFace shows that one could add 3.4 faces to each FairFace estimate without concern for the number of faces in the image, as the slope on the FairFace estimate is .991 with $t = 10.652$.

Figure A10: Face Count Concords with Manual Face Count

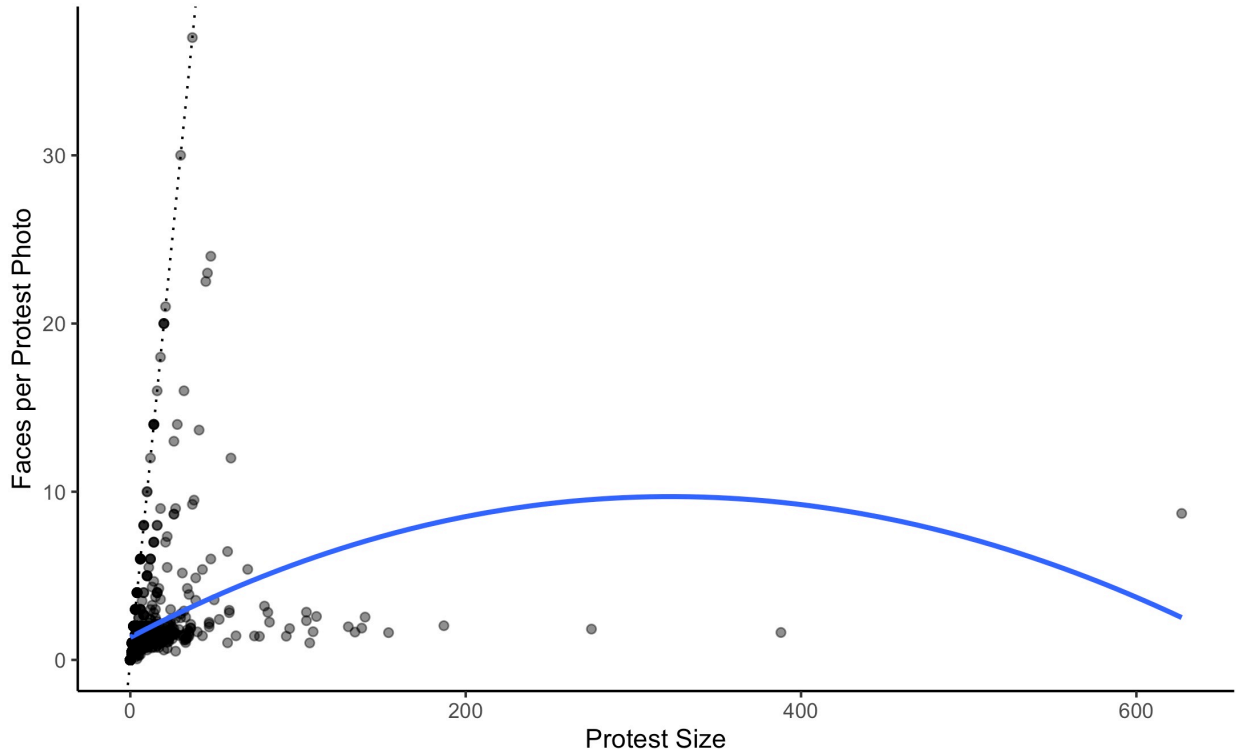


Note: FairFace, an automatic facial classifier, and humans code consistent numbers of faces in protest images.

One concern is that the number of faces per photo will increase as protests become larger, inflating upwards resulting size estimates. Figure A11 shows there exists no clear relationship between the size of a protest and the number of faces per photo from that protest.

To more directly verify the face counting approach, we look at the residuals of regression regressing $\text{Log}_{10}(\text{Protest Size})_{i,t}$ on estimates from newspapers, activists, and police for matched events. Figure A12 shows these results. No clear pattern emerges, suggesting the Twitter image approach records protest size consistently across the range of others' reported size. The Twitter image approach is biased downwards but consistent across the range of

Figure A11: Protest Size and Faces per Photo



Note: The dotted line has a slope of 1.

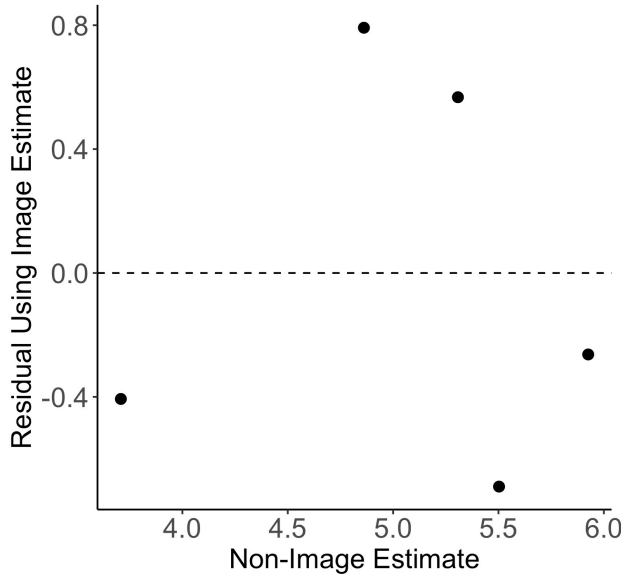
other reports of protest size.

S5 Verifying Using Images to Measure State Violence

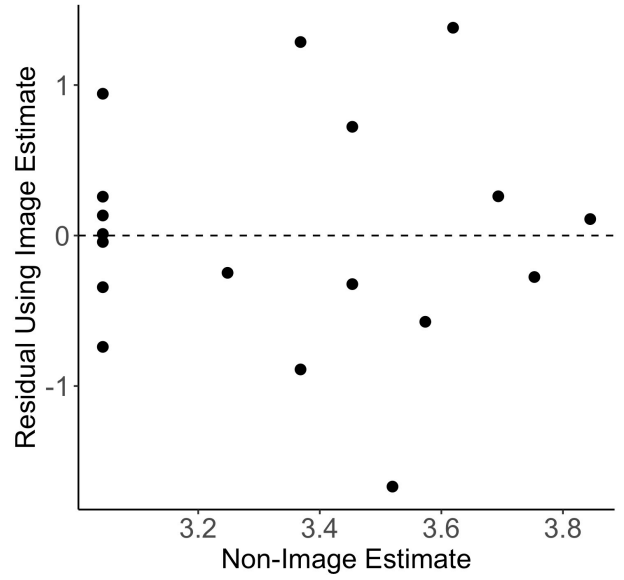
Figure A13 shows the mean daily state violence recorded in protest photos for Barcelona, Seoul, Caracas, and Hong Kong. The cross-city trends match expectations: Caracas and Hong Kong show the most frequent repression activity, followed by Seoul and Barcelona. Spanish police, which are federal, did often employ violence, to the surprise of the international community.

We also regress the onset of repression on protest size. Previous research suggests that repression becomes more likely as protests become larger (Moore, 2000, Francisco, 2004), so our measures of state violence and protest size should capture this dynamic. Table A4 shows that the onset of repression becomes more likely on the days of large protest but not after large protests. This corroboration further supports the operationalization of state violence

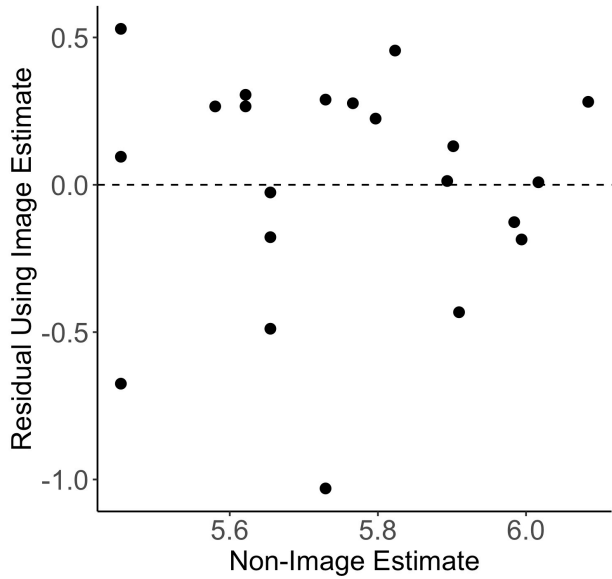
Figure A12: Verifying Protest Size, Residual Distribution



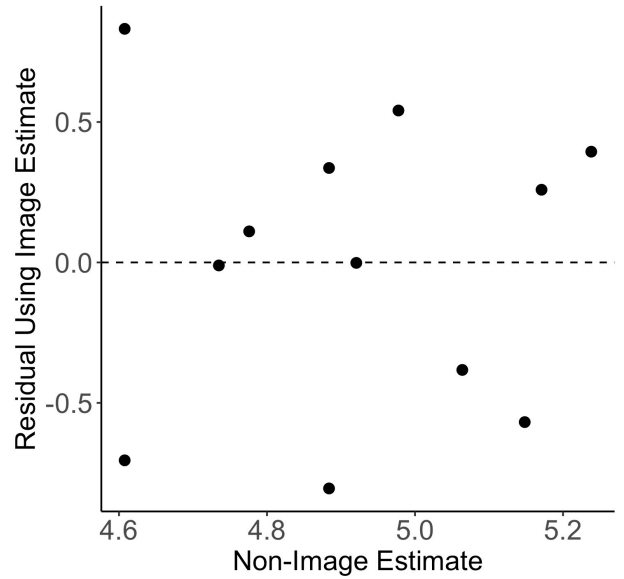
(a) Barcelona



(b) Hong Kong



(c) Seoul, Activist



(d) Seoul, Police

and protest size.

S6 Variable Correlation

Figure A14 shows the correlation between variables used in the main models.

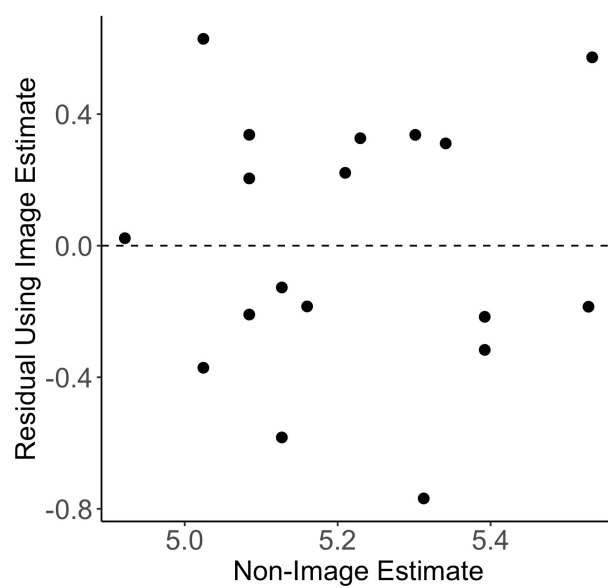
Table A4: Larger Protests Are More Likely to Induce Repression

	DV: Repression Onset	
	Logistic	OLS
	(1)	(2)
$\text{Log}_{10}(\text{Sum of Faces})_{i,t}$	5.3799*** (1.2714)	.4923*** (.0642)
$\text{Log}_{10}(\text{Sum of Faces})_{i,t-1}$.6508 (1.6250)	.0199 (.0711)
Intercept	-4.0234** (1.7507)	.0993 (.1218)
N	293	293
Adjusted R ²		.3256
Log Likelihood	-36.7280	
Cluster SE	Y	Y
City FE	Y	Y

*p < .1; **p < .05; ***p < .01

The image pipeline finds that states become more likely to repress as protests increase in size. This result matches others' findings, lending additional support to our operationalization strategy. The sample size is smaller than later regressions because city-days after repression onset are dropped.

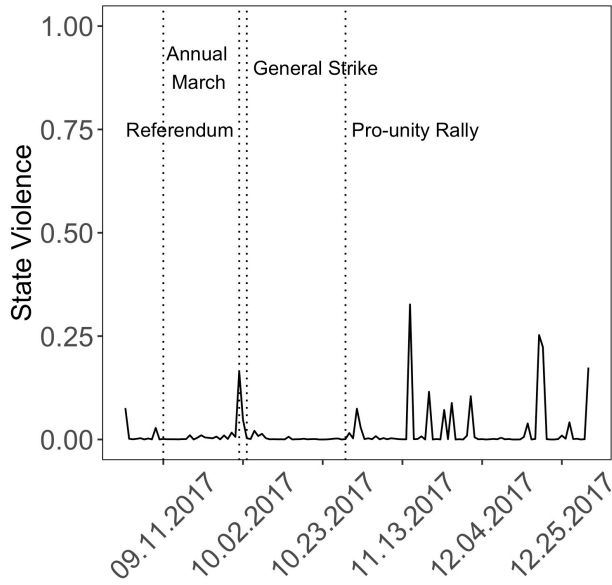
Figure A12: Verifying Protest Size, Residual Distribution



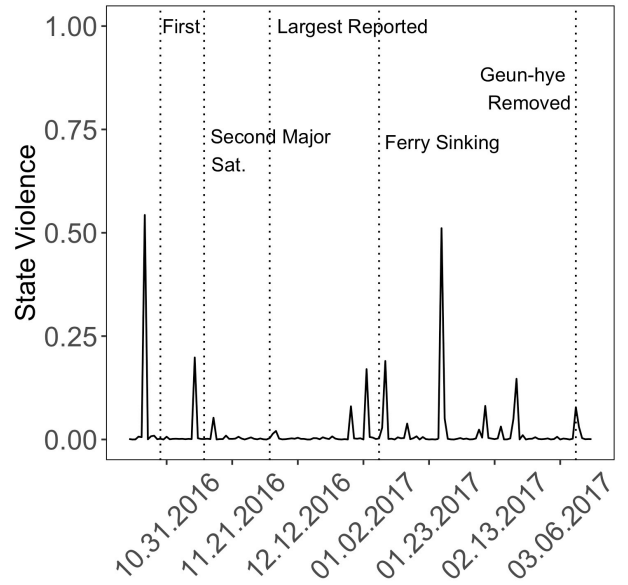
(e) Caracas

Note: These figures show the residuals from regressing our paper's protest size estimate on matched events' size reported in newspapers or Wikipedia. The residuals lack bias, as the tests in Table 3 show.

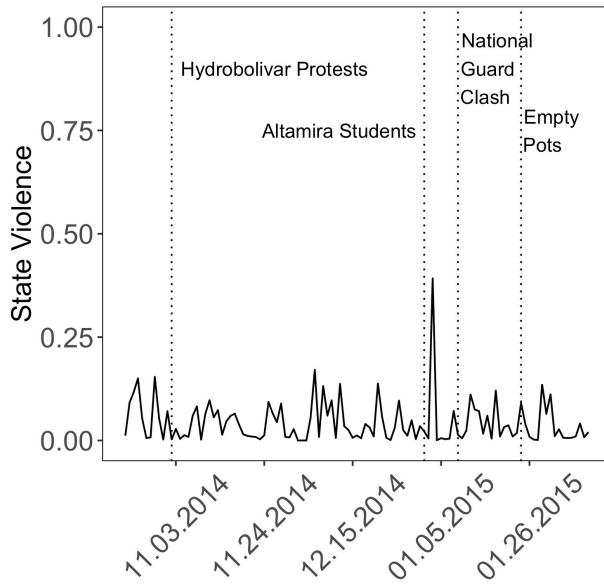
Figure A13: State Violence Time Series



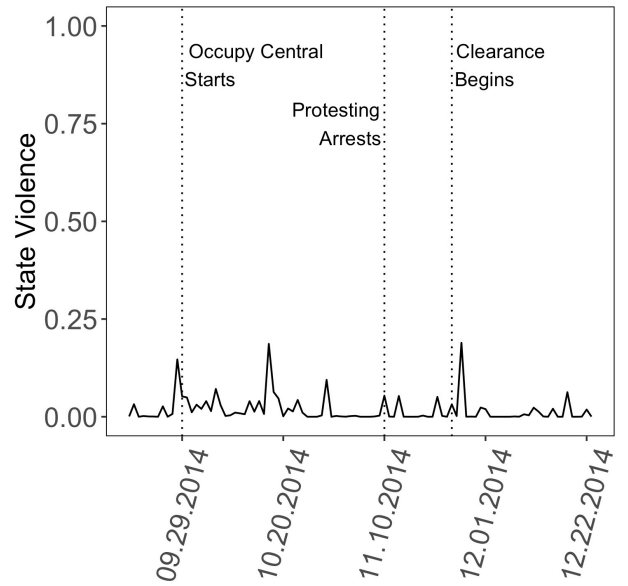
(a) Barcelona



(b) Seoul



(c) Caracas



(d) Hong Kong

Figure A14: Correlation of Regression Variables

Log(Protest Size _{i,t-1})	-0.46	0.15	0.2	0.06	0.36	0.38	0.3	1
Perc Yng. Adlt. _{i,t-1}	-0.17	0.03	0.06	0.04	0.09	0.39	1	0.3
Perc. Male _{i,t-1}	-0.14	0.02	0.07	0.02	0.08	1	0.39	0.38
Fire _{i,t-1}	-0.28	0.54	0.15	0.02	1	0.08	0.09	0.36
Police _{i,t-1}	-0.07	0.07	0.29	1	0.02	0.02	0.04	0.06
Perc. Stt. Violence _{i,t-1}	-0.09	0.45	1	0.29	0.15	0.07	0.06	0.2
Perc. Prtstr. Violence _{i,t-1}	0.1	1	0.45	0.07	0.54	0.02	0.03	0.15
Log(Protest Size _{i,t})	1	0.1	0.09	0.07	0.28	0.14	0.17	0.46
	Log(Protest Size _{i,t})	Perc. Prtstr. Violence _{i,t-1}	Perc. Stt. Violence _{i,t-1}	Police _{i,t-1}	Fire _{i,t-1}	Perc. Male _{i,t-1}	Perc Yng. Adlt. _{i,t-1}	Log(Protest Size _{i,t-1})

Note: Correlation of regression variables at the city-day level.

S7 Original Regression Results

Table A5: Main Result

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$		
	Violence (1)	Demographics (2)	Combined (3)
Perceived Prtstr. Violence $_{i,t-1}$	-.1549** (.0737)		-.1537** (.0745)
Perceived Stt. Violence $_{i,t-1}$	1.2684*** (.3183)		1.2688*** (.3352)
Perceived Stt. Violence $^2_{i,t-1}$	-2.0895*** (.6002)		-2.0845*** (.3657)
Police $_{i,t-1}$.7573* (.4540)		.7457* (.4441)
Fire $_{i,t-1}$.1014*** (.0151)		.0989*** (.0159)
Male Percent $_{i,t-1}$		-.1939* (.1101)	-.1816 (.1194)
Young Adult $_{i,t-1}$.2255** (.1089)	.1941** (.0894)
Tweets $_{i,t-1}$.0089*** (.0032)	.0102*** (.0039)	.0088*** (.0032)
DV $_{i,t-1}$.1882*** (.0699)	.2331*** (.0869)	.2001*** (.0806)
Intercept	.1216*** (.0160)	.1259*** (.0161)	.1225*** (.0159)
N	4,121	4,121	4,121
Adjusted R ²	.2656	.2510	.2679
City FE	Y	Y	Y
Cluster SE	Y	Y	Y

*p < .1; **p < .05; ***p < .01

City-clustered standard errors shown in parentheses.

S8 Bias

In the United States, Twitter users who geotag are richer, more likely to live in cities, young, and non-white than the overall population (Malik et al., 2015). In the United Kingdom, Twitter users are younger, more educated, more likely to be male, and more politically engaged (but less likely to vote) than others (Mellon and Prosser, 2017). Once on Twitter, geotagging users are slightly older than non-geotaggers, there is some difference in rates of geotagging across profession, and there is large variation by tweet language in the percentage of users who geotag (a low of 0.4% for Arabic accounts to a high of 8.3% for Turkish, with an average of 3.1%) (Sloan and Morgan, 2015).

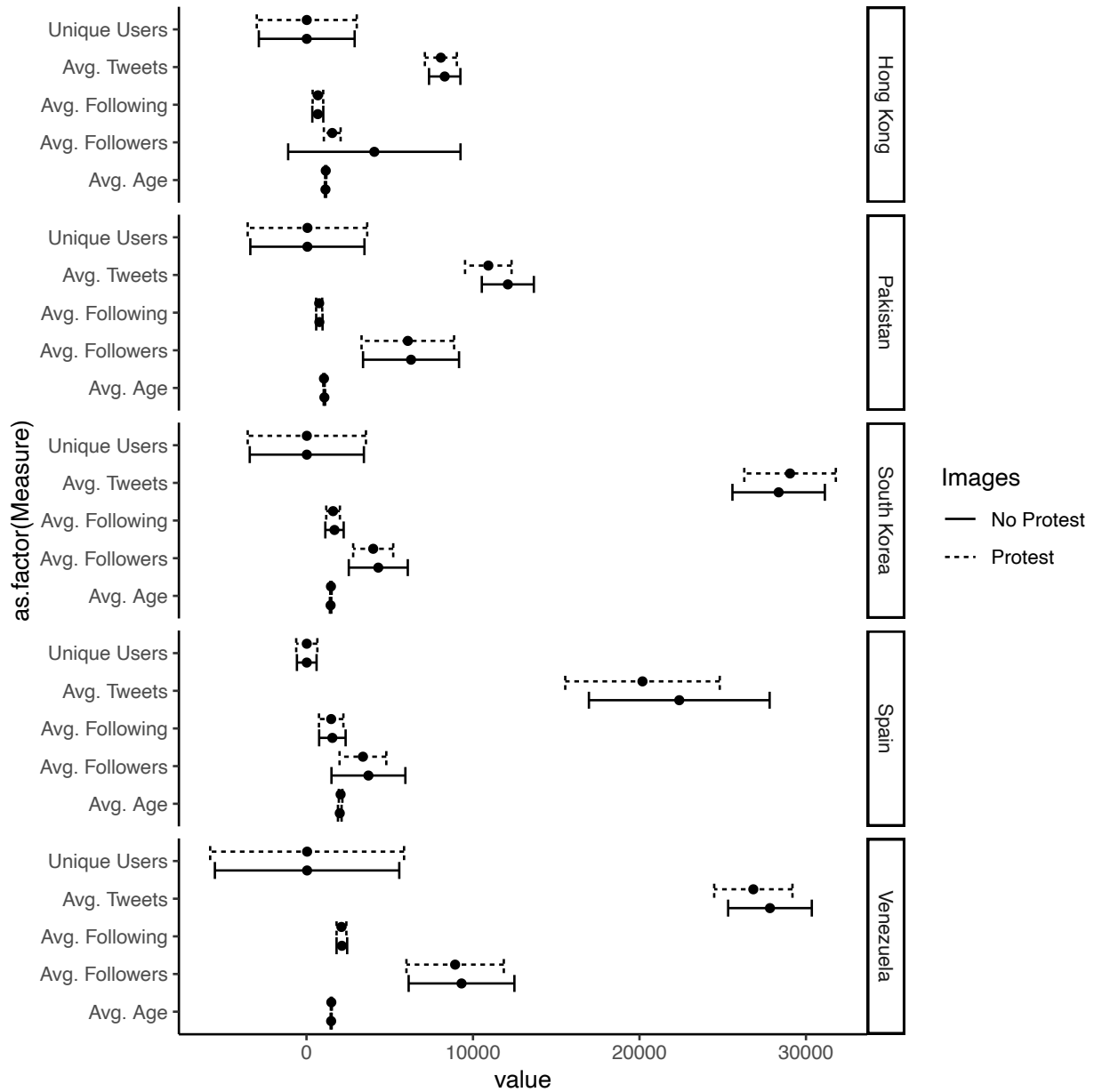
To investigate protest image and regular image sharing in geolocated tweets, we analyze account covariates by country-photo type (protest or not protest). We randomly selected 20,000 tweets with protest images and 20,000 with non-protest images. For each set of photos, we kept each user only once, which discarded about 40% of tweets. The account characteristics we analyze are the median number of followers, following, statuses, and account age (days on Twitter). An account can appear in both samples. We also count the number of unique users on each country-day by image type. The points estimates and 95% confidence intervals are shown in Figure [A15](#).

A series of paired t-tests by country confirms there exists essentially no difference between accounts which share protest and non-protest images. The only statistically significant difference is median account age in South Korea, where the median age of accounts tweeting protest photos is 40 days older than those not. If using the average age of accounts, those from Catalonia are also older. The average number of statuses of accounts tweeting protest photos is also statistically significantly lower in Venezuela, Spain, and Pakistan. Though the difference in number of users is never statistically significant, in each country more users tweet protest photos than non-protest photos. The minor differences that do exist therefore suggest older, less active accounts tweet protest photos.

While there is not systematic evidence that accounts which share protest and non-protest images differ, we still cannot be sure that accounts which share geotagged protest photos share photos representative of a protest. This paper therefore assumes that the protest images convey a representative sample of protest activity, and we now proceed to investigate that assumption three ways.

First is by analyzing the classifier estimates by level of geographic aggregation. After composing a tweet but before sending it, Twitter autosuggests the city name for location but allows the user to change the location to be more specific (neighborhood) or less (state, country). Twitter users may behave strategically when choosing which geographic specificity — neighborhood, city, state, or country — to assign to their tweet. For example, a protester

Figure A15: Users Tweeting Protests vs. Not

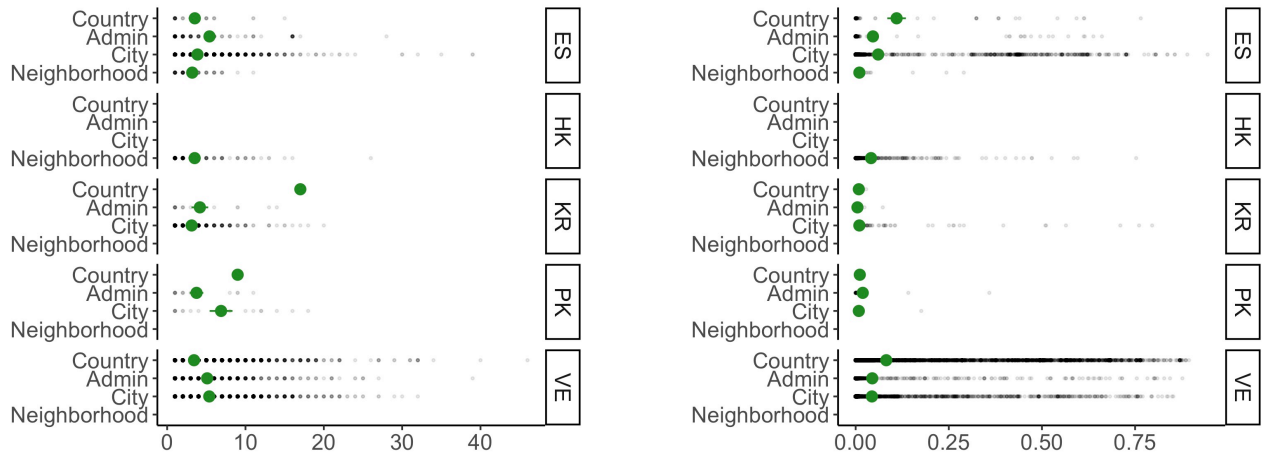


may tag an image with a high amount of protester violence at the country level so as to minimize negative responses that could occur if people knew where the violence occurred.

To understand if images vary by the specificity of geographic resolution, we show the distribution of tweets by country and geographic level for the number of faces per photo, state violence, and protester violence. Figure A16 shows this exploration: neighborhood and

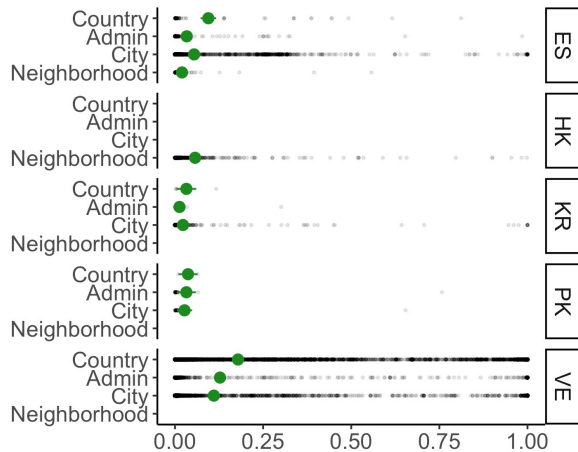
city are the most common levels of aggregation, though Venezuela has many tweets at the country level. No systematic differences emerge based on the geographic specificity users assign to tweets.

Figure A16: Quantities by Geographic Specificity



(a) Faces

(b) State Violence



(c) Protester Violence

Note: Distribution of number of faces (a), perceived state violence (b), and perceived protester violence (c) per protest image. Tweets are grouped by country and level of geographic specificity. “Admin” is equivalent to a state. The green dot is the mean.

Second, since the tweets analyzed in this paper all contain coordinates, we turn to another dataset to compare tweets with and without coordinates. We download 7,665,497 tweets DocNow identifies as about the 2017 Charlottesville Unite the Right Rally and find the

almost 700,000 that contain an image.² We apply the CNNs developed for this paper to those images as well as to just over 7,681 Charlottesville tweets with images collected from data we already had. For the images whose protest probability is greater than .6, we analyze their protest, state violence, and protester violence scores as they vary by level of geographic aggregation. Crucially, most of the tweets do not contain GPS coordinates, so we can observe how sharing patterns vary between users who assign location to tweets and those who do not. We also split the non-GPS tweets based on whether or not the user provides a string to the location field of their profile. Figure A17 shows these results.³

While each of the three labels contain statistically significant variation by level of aggregation, that variation does not follow patterns suggestive of strategic behavior. For example, if individuals are concerned about sharing protest images, tweets with no location should have the highest protest classifier rating, but they are no different than tweets tagged to the neighborhood level, the most precise level possible.. More generally, higher levels of geographic aggregation should have higher average classifier scores: no location more than profile location, profile location more than country, country more than admin, and so on. Only for perceived state violence, however, do tweets possibly follow this trend, though even in that category tweets at the country level do not follow the expected pattern.

Finally, Table A6 shows the main regressions after aggregating all tweets to the state and country level. See Section S9.3 for a longer discussion of these results.

S9 Additional Robustness Checks

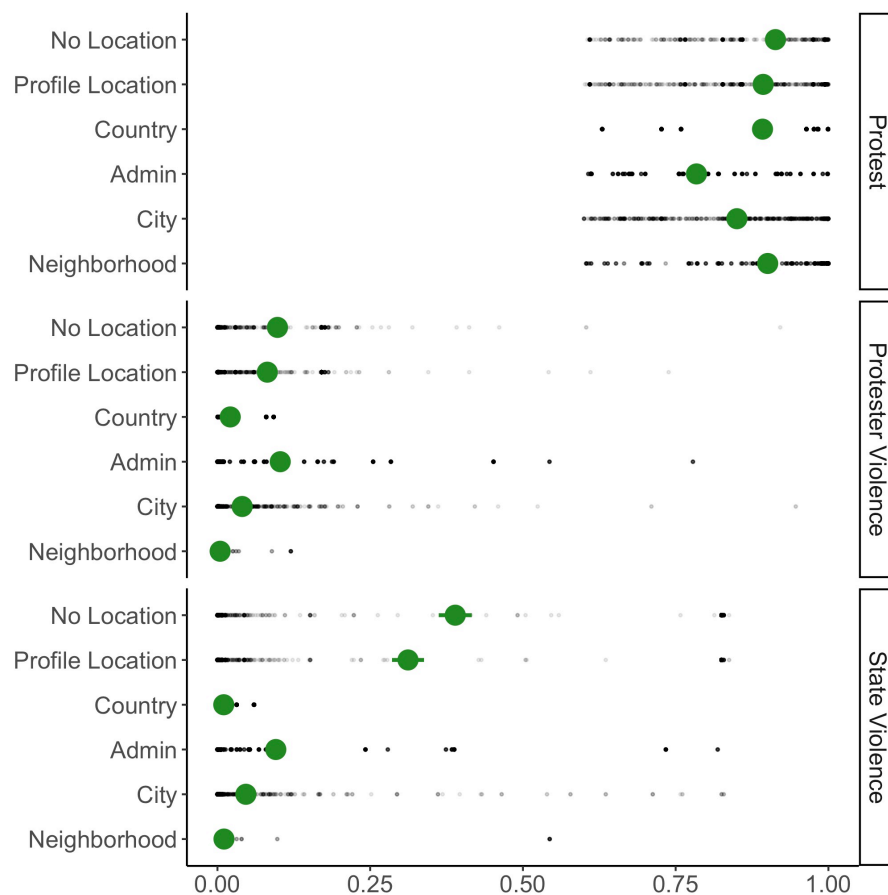
S9.1 Confirming Protester Violence Correlation

Figures A18 and A19 show that the results for *Perceived Protester Violence* _{$i,t-1$} are robust to flexible forms. Regression results (not shown) introducing a square term confirm that only a linear correlation is statistically significant.

² <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DVLJTO>

³ Keep in mind that these classifiers were not trained on these images and have not been tested for their applicability outside of the paper’s protests.

Figure A17: Quantities by Geographic Specificity for the 2017 Unite the Right Rally in Charlottesville, VA

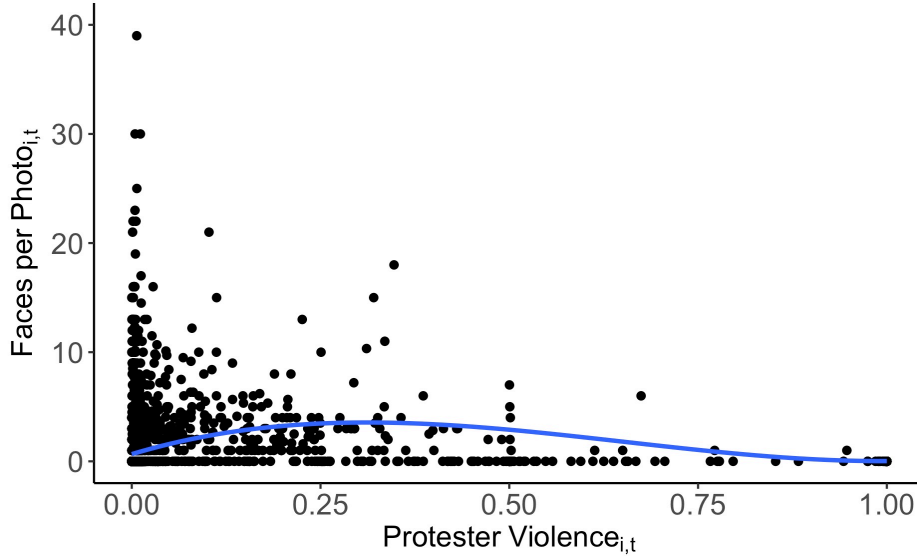


Note: Distribution of protest image probability, perceived state violence, and perceived protester violence for tweets about the 2017 Unite the Right rally in Charlottesville, VA. “No Location” refers to tweets where a user does not assign a location to the tweet or report their location in their profile (“Profile Location”). All other categories are based on GPS location; “Admin” is equivalent to a state. The green dot is the mean.

S9.2 Time Trends

A vector autoregressive (VAR) model may better capture complex temporal dynamics (Zeit-zoff, 2011), so Model 3 of Table A5 is replicated with one, using 41 lags. Figure A20 shows the resulting impulse response plot: the n-shaped relationship holds. Protester violence is not significant with that many lags.

Figure A18: Protester Violence and Faces per Photo



Note: Regression results introducing a square term confirm that only a linear correlation is statistically significant.

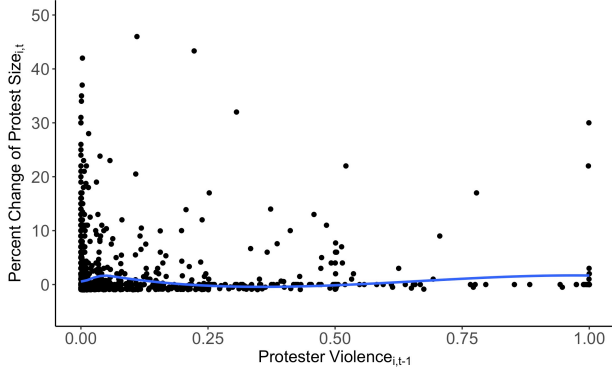
S9.3 Geographic Aggregation

Table A6 shows results when aggregating tweets to their state (Model 1) or country (Model 2). The results are broadly consistent: state violence exhibits the same n-shaped relationship, though it is not significant in the country model, and protester violence is no longer statistically significant and is much closer to zero.

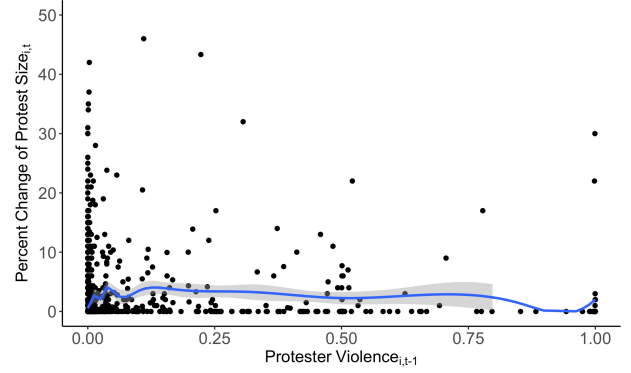
S9.4 Accounting for Days with No Protests

Table A7 shows attempts to account for days with no protest. Model 1 drops all days with no protest images. Model 2 is a Poisson model. Model 3 is a negative binomial, and Model 4 is a zero-inflated negative binomial model. To converge, Model 4 excludes city fixed effects and clustered standard errors; it does use country fixed effects and models the zero and count components using the same variables. Results for state violence do not change; protester violence remains negative but is no longer statistically significant in two models.

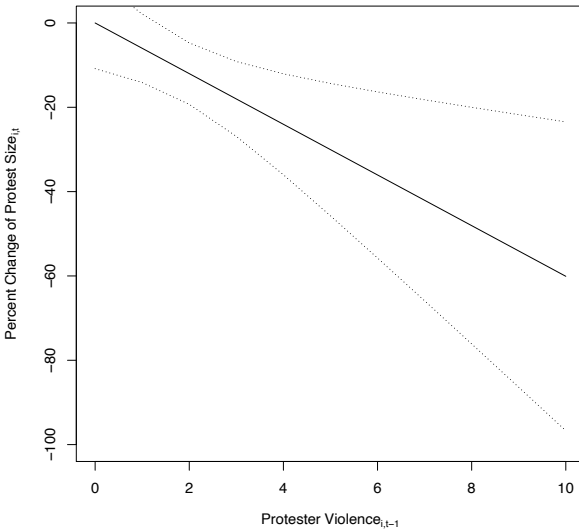
Figure A19: Protester Violence Results Remain in Flexible Operationalizations



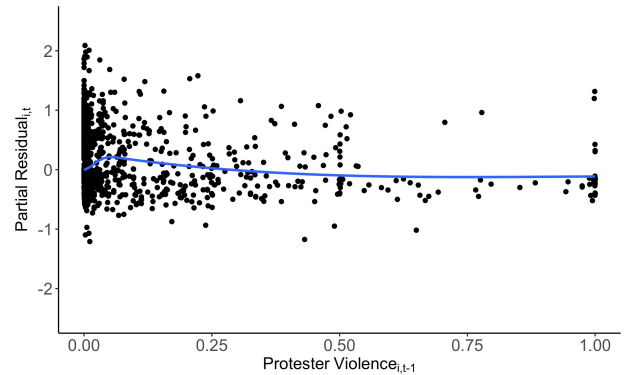
(a) LOESS (Span = .7)



(b) Spline, 50 Knots



(c) Binned Marginal Effects



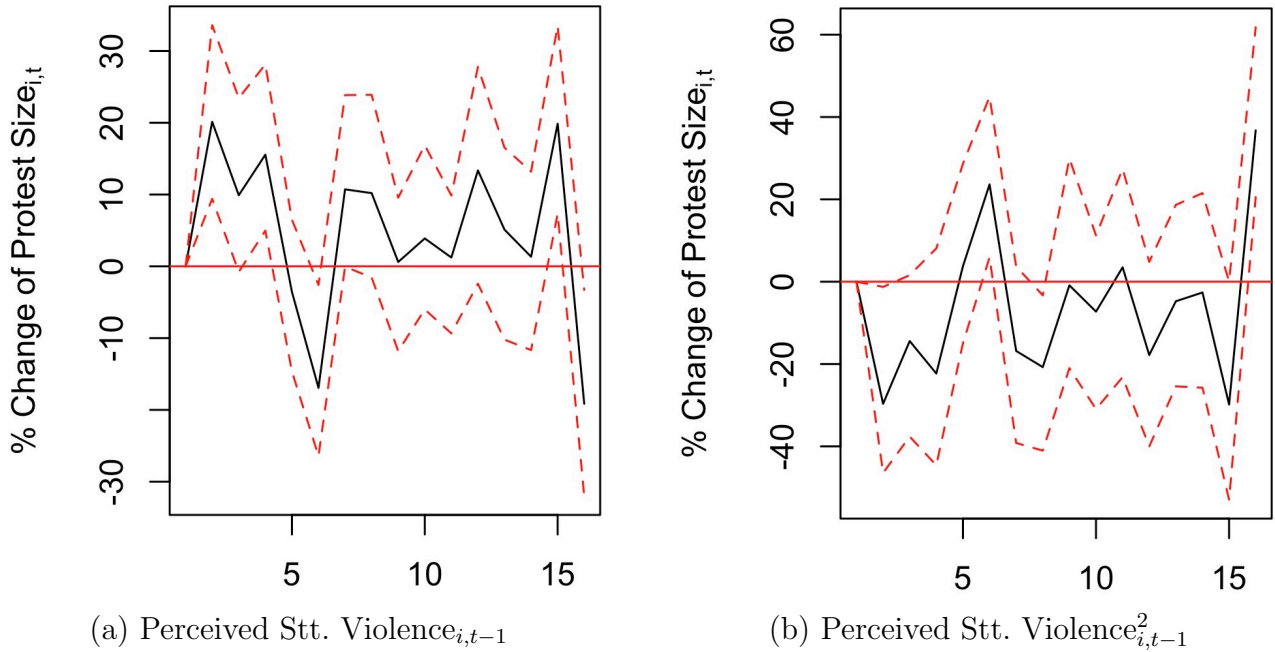
(d) LOESS on Partial Residuals

Note: The negative relationship broadly holds in non-parametric forms (a, b). Generating fixed effects for protester violence in bins of width of .1 finds the same relationship (c). Regressing protester violence on the partial residuals of Model 3 from Table A5 shows the same negative relationship (d).

S9.5 Different Dependent Variable

Table A8 repeats the main model using four different operationalizations of the dependent variable. The first does not log-transform the number of protesters. The second measures the size of protest using the number of users who share a protest photo per city-day. This quantity is smaller than the number of protest photos per day because users often tweet multiple times per day. To lessen noise in the estimates, the third makes the dependent

Figure A20: Vector Auto Regression, $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$ Response to State Violence



Note: Impulse response plots using Model 3 from Table A5 shows no change in interpretation of the state violence correlation. 41 lags are included, and results do not change if one or seven lags are used instead. The x-axis is the number of days after the state violence.

variable logged and ordinal while the fourth rounds the size to the nearest ten. Results do not change.

S9.6 Weighted Results

Table A9 shows that weighting each city-day observation by the number of protest photos shared from it strengthens the paper's results substantially. The violence coefficients are much larger than the unweighted models, and model fit is almost 50% better than the paper's main models. $\text{Male Percent}_{i,t=1}$ becomes statistically significant. Model 1 shows these results. Weighting by the inverse number of tweets produces a much worse fitting model than the original, and $\text{Perceived Protester Violence}_{i,t-1}$ loses statistical significance but remains negatively signed. Model 2 shows these results.

S9.7 Most Likely Protest Tweets

These two models select tweets based on features that increase the probability they come from a protest. Model 1 restricts tweets to those only from mobile devices, based on the

Table A6: Results With Coarser Geographic Aggregation

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$	
	Unit: State	Unit: Country
	(1)	(2)
Perceived Prtstr. Violence $_{i,t-1}$.0062 (.0462)	-.0360 (.1127)
Perceived Stt. Violence $_{i,t-1}$.9445*** (.2973)	.4352 (.5914)
Perceived Stt. Violence $^2_{i,t-1}$	-1.8371*** (.5329)	-1.4061*** (.5914)
Police $_{i,t-1}$.3183*** (.0435)	.2841*** (.0435)
Fire $_{i,t-1}$.0202 (.0159)	.0026 (.0131)
Male Percent $_{i,t-1}$	-.0336 (.0655)	-.0506* (.0270)
Young Adult $_{i,t-1}$.1052*** (.0672)	.2462* (.1407)
Tweets $_{i,t-1}$.0050 (.0035)	.0022* (.0012)
DV $_{i,t-1}$.2523*** (.0732)	.3100*** (.0956)
Intercept	.0412*** (.0146)	.7028*** (.0590)
N	10,344	1,076
Adjusted R ²	.4225	.3696
Cluster SE	State	Country
Fixed Effect	State	Country

*p < .1; **p < .05; ***p < .01

source field Twitter provides with each tweet. If that field contains “Twitter Web Client” or “Hootsuite”, the tweet is discarded. The mobile model fits the data less than half as well as the full model. Model 2 keeps only tweets issued between 10 a.m. and 10 p.m., the most likely protest windows. Results do not change.

S9.8 Protest Text Tweets

Because protest information can be contained in tweets without images that are nonetheless about a protest, we build a topic model to identify such tweets and include them in our main

Table A7: Count Models

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$		DV: $\text{Sum of Faces}_{i,t}$	
	No Zero Days	Poisson	Negative Binomial	Zero-inflated Negative Binomial
	(1)	(2)	(3)	(4)
Perceived Prtstr. Violence $_{i,t-1}$	-0.1832 (0.1430)	-1.4194*** (0.1191)	0.5736 (0.6359)	-0.9433* (0.4996)
Perceived Stt. Violence $_{i,t-1}$	1.0935** (0.4820)	9.6831*** (0.3289)	7.7665*** (2.5288)	7.1163*** (2.0728)
Perceived Stt. Violence $^2_{i,t-1}$	-1.7337** (0.7531)	-17.4641*** (0.7703)	-12.3918*** (4.3438)	-12.0539*** (3.6066)
Police $_{i,t-1}$	1.5004*** (0.1921)	2.8867*** (0.0545)	1.0149 (1.7536)	2.4718** (1.1504)
Fire $_{i,t-1}$	0.0644** (0.0353)	0.3002*** (0.0102)	0.1258 (0.1749)	0.3850*** (0.1173)
Male Percent $_{i,t-1}$	-0.2400** (0.1216)	0.0467 (0.0761)	0.5523 (0.5728)	-0.1215 (0.4156)
Young Adult $_{i,t-1}$	0.1910 (0.1357)	0.6778*** (0.0840)	0.6311 (0.7673)	-0.0561 (0.4350)
Tweets $_{i,t-1}$	0.0052*** (0.0019)	0.0068*** (0.0006)	0.1007*** (0.0140)	0.0135 (0.0103)
DV $_{i,t-1}$	0.1852*** (0.0432)	0.0045*** (0.0003)	-0.0009 (0.0059)	0.0055* (0.0033)
Intercept	0.4896*** (0.5552)	1.2485*** (0.0244)	0.8570*** (0.1756)	2.0955*** (0.1154)
N	1,354	4,121	4,121	4,121
Adjusted R ²	0.2019			
Log Likelihood		-19,774.2000	-4,249.5310	-4,078.6570
Cluster SE	Y	Y	Y	N
City FE	Y	Y	Y	N

*p < .1; **p < .05; ***p < .01

regression model. To identify protest tweets, we use a dictionary of keywords that vary by country. Any tweet containing that keyword is classified as a protest tweet. We then count the number of protest tweets per city-day and include its lag as a control variable. Table [A11](#) shows this result.

The n-shaped relationship between state violence and subsequent protest size remains statistically significant. Protester violence remains negative but becomes statistically indistinguishable from zero. All other coefficients retain the same direction and significance. Model fit increases by .1.

Table A8: Different Measures of Protest Size

	No Log	Number Users	Ordinal (Log)	Ordinal (Tens)
	(1)	(2)	(3)	(4)
Perceived Prtstr. Violence $_{i,t-1}$	-12.4349*** (4.2307)	-3.7820** (1.7493)	-.2150** (.1078)	-12.8400*** (4.2606)
Perceived Stt. Violence $_{i,t-1}$	94.2101** (38.9302)	25.1615* (13.6320)	1.9251*** (.4295)	93.5590** (38.4090)
Perceived Stt. Violence $^2_{i,t-1}$	-178.2995** (81.66)	-46.5605* (26.09)	-3.0325*** (.7527)	-175.7669** (80.4525)
Police $_{i,t-1}$	126.1253* (75.3122)	31.8432 (19.8364)	.8192* (.4925)	124.5473* (74.5913)
Fire $_{i,t-1}$	4.6883*** (.8754)	1.3978** (.6733)	.1419*** (.0279)	4.8113*** (.8719)
Male Percent $_{i,t-1}$	-2.1077 (3.0608)	-.6492 (.6364)	-.2381 (.1634)	-1.8765 (3.0732)
Young Adult $_{i,t-1}$	3.6088*** (1.2864)	2.0567* (1.1131)	.3385** (.1638)	2.8223** (1.2919)
Tweets $_{i,t-1}$.1870 (.1287)	.0230 (.0527)	.0118*** (.0047)	.1883 (.1213)
DV $_{i,t-1}$.1354*** (.0390)	.2633*** (.0280)	.1811*** (.0690)	.1350*** (.0458)
Intercept	3.4836*** (.7012)	1.5182*** (.1798)	1.0035*** (.0318)	3.3766*** (.0160)
N	4,121	4,121	4,121	4,121
Adjusted R ²	.1915	.2749	.2497	.1888
Cluster SE	Y	Y	Y	Y
City FE	Y	Y	Y	Y

*p < .1; **p < .05; ***p < .01

S9.9 Models by Country

Table A12 repeats the main model but by country. Pakistan is not included because it has too few observations. The violence correlations are consistent across countries.

Table A9: Results Weighted by Protest Tweets per City

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$	
	# Tweets	$\frac{1}{\# \text{ Tweets}}$
	(1)	(2)
Perceived Prtstr. Violence $_{i,t-1}$	-.3666** (.1442)	-.0710 (.0753)
Perceived Stt. Violence $_{i,t-1}$	2.0692*** (.3862)	.6184** (.2870)
Perceived Stt. Violence $^2_{i,t-1}$	-3.8405*** (.7090)	-.8720* (.4708)
Police $_{i,t-1}$	1.3179*** (.1482)	.0885 (.2211)
Fire $_{i,t-1}$.0145 (.0144)	.0625*** (.0385)
Male Percent $_{i,t-1}$	-.5328*** (.0919)	-.0576 (.0681)
Young Adult $_{i,t-1}$.3339*** (.1091)	.0836 (.0902)
Tweets $_{i,t-1}$.0027*** (.0004)	.0316*** (.0044)
DV $_{i,t-1}$.2642*** (.0286)	.0366 (.0285)
Intercept	.4282*** (.0494)	.0832*** (.0128)
N	4,121	4,121
Adjusted R ²	.5398	.0950
Cluster SE	Y	Y
City FE	Y	Y

*p < .1; **p < .05; ***p < .01

Table A10: Most Likely Protest Tweets

	DV: $\log_{10}(\text{Sum of Faces})_{i,t}$	
	Source Mobile	Protest Time
	(2)	(3)
Perceived Prtstr. Violence $_{i,t-1}$	-.0109 (.1020)	-.2099*** (.0477)
Perceived Stt. Violence $_{i,t-1}$.6146** (.2991)	1.4624*** (.4582)
Perceived Stt. Violence $^2_{i,t-1}$	-.9819** (.4505)	-2.3707*** (.8031)
Police $_{i,t-1}$.0001 (.1359)	.7266* (.3905)
Fire $_{i,t-1}$.0454*** (.0175)	.1140*** (.0250)
Male Percent $_{i,t-1}$	-.0326 (.0944)	-.1327** (.0586)
Young Adult $_{i,t-1}$.1208 (.1106)	.1405* (.0854)
Tweets $_{i,t-1}$.0097*** (.0014)	.0096*** (.0034)
DV $_{i,t-1}$.1014 (.0640)	.1597** (.0756)
Intercept	.1382*** (.0015)	.1323*** (.0164)
N	4,091	4,004
Adjusted R ²	.1298	.2282
Cluster SE	Y	Y
City FE	Y	Y

*p < .1; **p < .05; ***p < .01

Table A11: Controlling for Protest Text Tweets

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$ Protest Text Tweets (2)
Perceived Prtstr. Violence $_{i,t-1}$	-.0946 (.0829)
Perceived Stt. Violence $_{i,t-1}$	1.1456*** (.3715)
Perceived Stt. Violence $^2_{i,t-1}$	-1.9334*** (.6571)
Police $_{i,t-1}$.7598 (.4689)
Fire $_{i,t-1}$.0584* (.0352)
Male Percent $_{i,t-1}$	-.1585** (.0727)
Young Adult $_{i,t-1}$.1997* (.1035)
Tweets $_{i,t-1}$.0080*** (.0025)
Protest Tweets $_{i,t-1}$.0044*** (.0010)
DV $_{i,t-1}$.1820*** (.0329)
Intercept	.1287*** (.0186)
N	4,121
Adjusted R ²	.2783
Cluster SE	Y
City FE	Y

*p < .1; **p < .05; ***p < .01

Results hold when controlling for the number of tweets with protest keywords.

Table A12: Tables by Country

	DV: $\text{Log}_{10}(\text{Sum of Faces})_{i,t}$				
	Spain	Hong Kong	South Korea	Venezuela 2014-2015	Venezuela 2017
	(1)	(2)	(3)	(4)	(5)
Perceived Prtstr. Violence $_{i,t-1}$.3575* (.1934)	-.4625 (.7133)	-.6048 (.5097)	.0856 (.1240)	-.0988 (.0681)
Perceived Stt. Violence $_{i,t-1}$.6581* (.3415)	3.9860* (2.0648)	-.0055 (1.5260)	.9679* (.5368)	.7031** (.3533)
Perceived Stt. Violence $^2_{i,t-1}$	-1.5309*** (.5184)	-10.8883 (6.6133)	-.9196 (3.3142)	-1.3262 (.8925)	-1.0177* (.6112)
Police $_{i,t-1}$	1.2167*** (.1948)			.0819 (.4728)	-.1233 (.2207)
Fire $_{i,t-1}$.0182 (.0485)	.0547 (.1191)	.2862* (.1596)	-.0420* (.0241)	.0365 (.0316)
Male Percent $_{i,t-1}$	-.3593*** (.1184)	-.3017 (.2406)	.3481* (.1792)	-.1619 (.1118)	-.0837 (.0768)
Young Adult $_{i,t-1}$.3878*** (.1275)	.4702* (.2487)	-.0395 (.1631)	.0166 (.1746)	-.0564 (.1270)
Tweets $_{i,t-1}$.0044*** (.0011)	.0032 (.0087)	.0031 (.0035)	.0135*** (.0034)	.0171*** (.0038)
DV $_{i,t-1}$.0941*** (.0356)	.0668 (.1153)	-.2125*** (.0797)	.1407*** (.0523)	.0226 (.0287)
Intercept	.5956*** (.0367)	.2666*** (.0616)	-.0013 (.0351)	.0343 (.0359)	.0166 (.0168)
N	1,412	136	231	573	1,752
Adjusted R ²	.2936	.0902	.3017	.6067	.0972
Cluster SE	N	N	N	N	N
City FE	Y	Y	Y	Y	Y

*p < .1; **p < .05; ***p < .01

S10 Bot Distribution

Table A13 compliments Table 6. No more than 13.3% of accounts are from bots, and no more than 10.8% of tweets.

Table A13: Distribution of Bots by City

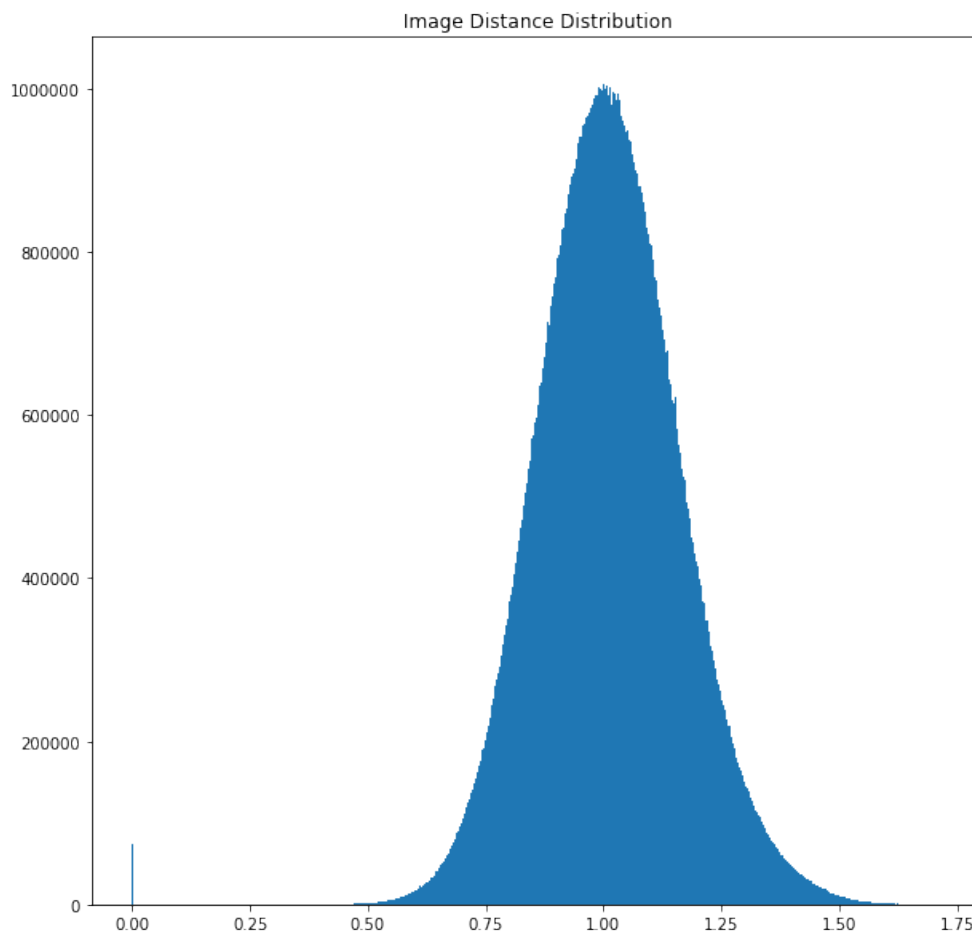
City	Avg. Bot Score	Max. Bot Score	SD Bot Score	Percent Tweets from Bots	Percent Accounts of Bots
Ciutat Vella	0.131	0.948	0.297	0.108	0.040
Lahore	0.067	0.559	0.173	0.100	0.143
Sant Salvador de Guardiola	0.066	0.611	0.155	0.083	0.094
Granera	0.053	0.812	0.156	0.054	0.065
Valencia	0.063	0.921	0.145	0.052	0.058
Tarragona	0.058	0.829	0.172	0.052	0.079
Central	0.050	0.845	0.156	0.051	0.133
Maracaibo	0.050	0.685	0.115	0.039	0.037
Seoul	0.145	0.905	0.170	0.035	0.056
Barcelona	0.029	0.939	0.110	0.024	0.022
Caucagua	0.054	0.905	0.118	0.020	0.027
Boca del Rio	0.032	0.829	0.117	0.020	0.047
Girona	0.034	0.637	0.088	0.019	0.038
Sant Feliu de Pallerols	0.040	0.661	0.086	0.018	0.048
Caracas	0.043	0.942	0.103	0.010	0.025
Granollers	0.011	0.084	0.019	0	0
Kimhae	0.052	0.054	0.009	0	0
Kowloon	0.021	0.355	0.062	0	0
Lleida	0.018	0.355	0.058	0	0
Mataró	0.007	0.054	0.009	0	0
Reus	0.018	0.297	0.051	0	0
Sabadell	0.018	0.355	0.044	0	0
Sant Cugat del Vallè	0.015	0.270	0.047	0	0
Terrassa	0.005	0.030	0.006	0	0

S11 Deduplicating Images

To deduplicate images, we extracted 1,000 features from a pre-trained ResNet50 model (He, Zhang, Ren and Shun, 2016). Each image was resized to 256 x 256 pixels. Then, a center-crop of 224 x 224 pixels was performed. Finally, the cropped images were normalized to the mean and standard deviation of the ImageNet dataset (Deng et al., 2009). The L2 distance among the normalized data is computed, and images are considered matches if the distance is less than a threshold of 0.2. The histogram of the distribution of distances is shown in Figure A21.

Two manual checks verify these results. First, the largest 90 clusters were manually in-

Figure A21: Distribution of Pairwise Image Distances



spected and no images were misidentified as duplicates. Second, the 220 most common images identified as duplicates, shared 2,500 times, were inspected, and none were misidentified as duplicates.

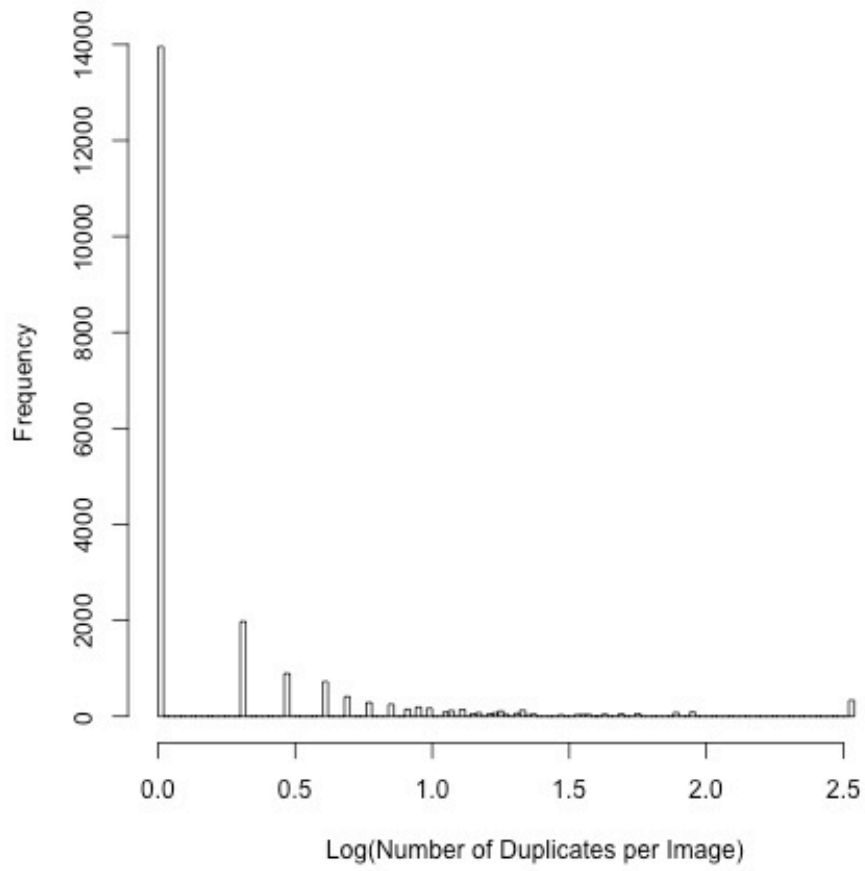
Table [A14](#) shows the percentage of tweets per city that are duplicates.

Figure [A22](#) shows the number of times each image appears in the dataset.

Table A14: Duplicate Images

City	Percentage Duplicates	Total Tweets
Kimhae	0.957	46
Sant Feliu de Pallerols	0.621	58
Girona	0.500	108
Caracas	0.463	2,105
Mataró	0.425	40
Sant Cugat del Vallè	0.424	33
Caucagua	0.281	224
Tarragona	0.274	62
Valencia	0.269	167
Sant Salvador de Guardiola	0.250	40
Maracaibo	0.243	152
Terrassa	0.237	59
Boca del Rio	0.227	66
Sabadell	0.187	75
Reus	0.184	38
Barcelona	0.182	1,338
Granera	0.154	65
Granollers	0.150	20
Lleida	0.116	43
Ciutat Vella	0.100	40
Seoul	0.052	326
Central	0	41
Kowloon	0	66
Lahore	0	5

Figure A22: Distribution of Number of Duplicates



Appendix References

- Baltrušaitis, Tadas, Peter Robinson and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE pp. 1–10.
- Bergamo, Alessandro and Lorenzo Torresani. 2010. Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. In *Advances in neural information processing systems*. pp. 181–189.
- Bojarski, Mariusz, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Praseem Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang et al. 2016. “End to end learning for self-driving cars.” *arXiv preprint arXiv:1604.07316* .
- Boschee, Elizabeth, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz and Michael Ward. 2015. “ICEWS Coded Event Data.”
URL: <http://dx.doi.org/10.7910/DVN/28075>
- Bradley, Ralph Allan and Milton E Terry. 1952. “Rank analysis of incomplete block designs: I. The method of paired comparisons.” *Biometrika* 39(3/4):324–345.
- Cantu, Francisco. 2019. “The Fingerprints of Fraud: Evidence From Mexico’s 1988 Presidential Election.” *American Political Science Review* 113(3):710–726.
- Clark, David H. and Patrick M. Regan. 2016. “Mass Mobilization.”
URL: <https://www.binghamton.edu/massmobilization/about.html>
- Deng, Jia, Wei Dong, Richard Socher, Li-jia Li, Kai Li and Li Fei-fei. 2009. ImageNet : A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 248–255.
- Francisco, Ronald A. 2004. “After the Massacre: Mobilization in the Wake of Harsh Repression.” *Mobilization: An International Journal* 9(2):107–126.
- Girshick, Ross. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. pp. 1440–1448.
- Girshick, Ross, Jeff Donahue, Trevor Darrell and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 580–587.
- Goldstein, Joshua S. 1992. “A Conflict-Cooperation Scale for WEIS Events Data.” *Journal of Conflict Resolution* 36(2):369–385.
- Güler, Rıza Alp, Natalia Neverova and Iasonas Kokkinos. 2018. “DensePose: Dense Human Pose Estimation In The Wild.” *arXiv preprint arXiv:1802.00434* .
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Shun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.

- He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778.
- Huval, Brody, Tao Wang, Sameep Tandon, Jeff Kiske, Will Song, Joel Pazhayampallil, Mykhaylo Andriluka, Pranav Rajpurkar, Toki Migimatsu, Royce Cheng-Yue et al. 2015. “An empirical evaluation of deep learning on highway driving.” *arXiv preprint arXiv:1504.01716* .
- Joo, Jungseock, Weixin Li, Francis F Steen and Song-Chun Zhu. 2014. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 216–223.
- Joo, Jungseock and Zachary C. Steinert-Threlkeld. 2018. “Image as Data: Automated Visual Content Analysis for Political Science.”
URL: <https://arxiv.org/abs/1810.01544>
- Koo, Terry K and Mae Y Li. 2016. “A guideline of selecting and reporting intraclass correlation coefficients for reliability research.” *Journal of chiropractic medicine* 15(2):155–163.
- Kovashka, Adriana, Devi Parikh and Kristen Grauman. 2012. Whittlesearch: Image search with relative attribute feedback. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE pp. 2973–2980.
- Liu, Ziwei, Ping Luo, Xiaogang Wang and Xiaoou Tang. 2015. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 3730–3738.
- Malik, Momin M, Hemank Lamba, Constantine Nakos and Jurgen Pfeffer. 2015. Population Bias in Geotagged Tweets. In *9th International AAAI Conference on Weblogs and Social Media*. pp. 18–27.
- Mellon, Jonathan and Christopher Prosser. 2017. “Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users.” *Research & Politics* 4(3):205316801772000.
- Moore, Will H. 2000. “The Repression of Dissent: A Substitution Model of Government Coercion.” *Journal of Conflict Resolution* 44(1):107–127.
- Parkhi, Omkar M, Andrea Vedaldi, Andrew Zisserman et al. 2015. Deep Face Recognition. In *BMVC*. Vol. 1 p. 6.
- Raleigh, Clionadh, Andrew Linke, Havard Hegre and Joakim Karlsen. 2010. “Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature.” *Journal of Peace Research* 47(5):651–660.
- Redmon, Joseph, Santosh Divvala, Ross Girshick and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 779–788.

- Ren, Shaoqing, Kaiming He, Ross Girshick and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. pp. 91–99.
- Rosenblatt, Frank. 1958. “The perceptron: a probabilistic model for information storage and organization in the brain.” *Psychological Review* 65(6):386.
- Salehyan, Idean, Cullen Hendrix, Jesse Hammer, Christina Case, Christopher Linebarger, Emily Stull and Jennifer Williams. 2012. “Social Conflict in Africa: A New Database.” *International Interactions* 38(4):503–511.
- Schroff, Florian, Dmitry Kalenichenko and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 815–823.
- Selvaraju, Ramprasaath R, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh and Dhruv Batra. 2016. “Grad-cam: Visual explanations from deep networks via gradient-based localization.” See <https://arxiv.org/abs/1610.02391> v3 7(8).
- Sloan, Luke and Jeffrey Morgan. 2015. “Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter.” *PLoS ONE* 10(11):1–15.
- Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. Cambridge University Press.
- Stephan, Maria J. and Erica Chenoweth. 2008. “Why Civil Resistance Works.” *International Security* 33(1):7–44.
- Sun, Yi, Yuheng Chen, Xiaogang Wang and Xiaoou Tang. 2014. Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*. pp. 1988–1996.
- Thomee, Bart, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth and Li-Jia Li. 2016. YFCC100M: The New Data in Multimedia Research. In *Communications of the ACM*. pp. 64–73.
- Urdal, Henrik and Kristian Hoelscher. 2012. “Explaining Urban Social Disorder and Violence: An Empirical Study of Event Data from Asian and Sub-Saharan African Cities.” *International Interactions* 38(4):512–528.
- Williams, Nora Webb, Andreu Casas and John D Wilkerson. 2020. *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification*. Cambridge University Press.
- Won, Donghyeon, Zachary C Steinert-Threlkeld and Jungseock Joo. 2017. Protest Activity Detection and Perceived Violence Estimation from Social Media Images. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM pp. 786–794.
- Xu, Huazhe, Yang Gao, Fisher Yu and Trevor Darrell. 2017. “End-to-end learning of driving models from large-scale video datasets.” *arXiv preprint* .