

Protest Event Data from Geolocated Social Media Content

Jungseock Joo¹ and Zachary C. Steinert-Threlkeld^{2,*}

¹University of California, Los Angeles

²University of California, Los Angeles.

*Corresponding author: zst@luskin.ucla.edu

September 3, 2020

Abstract

While it is understood that protester identity, violence, and emotions affect the size of protests, these concepts have proved difficult to measure at the protest-day level. Geolocated text and images from social media can improve these measurements. This advance is demonstrated on protests in Venezuela and Chile; it uncovers more protests in Venezuela and generates new measures in both countries. Moreover, the methodology generates daily city-day protest data in 107 countries containing 82.7% of the world's population and 97.15% of its GDP. These multimodal protest event data complement existing event datasets, though countries' population and income constrain the reach of any methodology.

Word count: 10,865 = 8,543 (body) + 2,312 (references).

1 Introduction

Geolocated text and images shared on social media can generate protest event data important to several strands of research on collective action. For example, many theories discuss how the demographics of protesters affect bystanders decision to protest (Oliver, Marwell and Teixeira, 1985, Lohmann, 1994), but quantitative research has not been able to measure such features cross-nationally, temporally, or at the level of the individual (Tufekci and Wilson, 2012, Asal et al., 2013). The size of protests and the severity of violence have also resisted precise measurement, so improved measurements may help resolve the repression-dissent puzzle (Davenport, 2007). Many scholars have also theorized about how emotions affect bystanders decision to protest (Jasper, 2011, Pearlman, 2013), but the difficulty of passively measuring emotions has forced researchers to rely on simulations or single-country experiments (Siegel, 2011, Young, 2019). This paper shows that the combination of large amounts of geolocated social media images and text ("content") and recent computer vision advances can generate these data. This combination will generate daily protest data at the city level in 107 countries that encompass 82.7% of the world's population and 97.15% of its GDP.

Creating event datasets, where the unit of observation is typically one actor performing an action directed towards another actor, has been an active area of research since at least the 1970s and continues to constitute a fruitful research program at the forefront of many methodological developments (Azar, 1980, Boschee et al., 2015, Croicu and Weidmann, 2015, Halterman et al., 2017). The use of text, often from newspapers, dominates these efforts. This paper shows that social media content, especially images, can contribute to these efforts. One strength of images is that visual communication is much closer to a universal language than text (Graber, 1996), so it is easier to identify protests from images than from text in countries with different written languages. Section 2 motivates the creation of multimodal event data in more detail.

Section 3 explains how to create multimodal event data, with an emphasis on generating data from images. Images could come from social media or newspapers, though Twitter is preferred for its availability. If text and image sources can be resolved to the same location and date, then

variables from images can be combined with measures only available via text. For example, images provide information about crowd size, violence, or protester demographics, while text sources will most likely record more types of actors and contemporaneous actions related to the protest. Non-newspaper text, such as social media posts, also provide measures of grievances before, during, and after a protest. These text also enable the measurement of emotions during protests, an important intervening variable otherwise unavailable to quantitative observational studies. Once primary sources are selected, one could measure these factors automatically, manually, or with a hybrid approach. This paper introduces an automatic methodology. Section 3.3 then explains the pipeline constructed to generate this paper’s event data, and other subsections discuss deduplication and bias.

Section 4 applies this pipeline to protests in Venezuela from 2014-2015 and Chile at the end of 2019. These countries were chosen because they experienced large waves of protest, are recorded in multiple text datasets, and contain enough people with enough income (see Section 4.1) to generate sufficient quantities of geolocated social media. Venezuela is used as a proof of concept for geolocated event data from images; we uncover more events than other datasets and new details about them (Boschee et al., 2015, Liang et al., 2018, Weidmann and Rod, 2018).

This comparison is not a competition, as the multimodal Chile example demonstrates. Using ACLED to identify protests (Raleigh et al., 2010), the pipeline adds information about magnitude, demographics, and emotions. These data form the basis of an expository statistical model investigating how violence, demographics, and emotions may affect protest size. The results are consistent with others’ findings, partially validating the methodology. Protester violence decreases protest size (Chenoweth and Stephan, 2011, Simpson, Willer and Feinberg, 2018), while state violence has the same curvilinear relationship for which others argue (Khawaja, 1993, Zhukov, 2018). The model finds a positive relationship between gender diversity and protest (Asal et al., 2013), though none for age (Nordås and Davenport, 2013). We are aware of no quantitative observational work that tests the relationship between emotions and protest, though the correlations these results exhibit match other’s theories (Pearlman, 2013), simulations (Siegel, 2011), and field experiments (Young,

2019). The results are meant to encourage the use of this paper’s methodology and not resolve ongoing theoretical debates.

This paper joins a growing body of work using computer vision to understand political processes. Social media have been used to study protests in China (Zhang and Pan, 2019), electoral violence (Muchlinski et al., Forthcoming, Senekal et al., 2020), and the effect of violence on support for ISIS (Mitts, Phillips and Walter, 2020). Computer vision also illuminates patterns of vote fraud (Cantu, 2019) and visual communication amongst elected officials (Xi et al., 2020). Like those, this paper does not explain computer vision methods but rather applies them; for an introduction to computer vision methods, see Williams, Casas and Wilkerson (2020).

Section 5 concludes with two observations. First, the results presented here represent a floor on the multimodal approach and several developments will improve these data going forward. Second, all event data projects face fundamental limits, as text and images are less likely to generate event data from low-income or sparsely population areas.

2 Multimodal Event Data

2.1 Images to Measure Violence, Size, and Protester Attributes

2.1.1 Violence

Generating violence data from geolocated images can contribute to the literature on repression and dissent by giving more precise estimates of violence’s severity than is available with existing approaches (Davenport, 2007). These continuously valued estimates contrast with the binary or ordinal measurements currently used in event data.

The repression-dissent puzzle refers to contradictory empirical findings about the effect of state violence on protest dynamics. Many studies find no correlation between repression and subsequent dissent (Ritter and Conrad, 2016). Others find a positive correlation (Francisco, 2004), a negative one (Ferrara, 2003), or argue that repression’s effects depend on the time period analyzed (Rasler, 1996). Part of the mixed results may be due to the difficulty of measuring violence from text sources, which is how current work measures violence. The coarseness of resulting measures means researchers may map violence of different intensities onto the same numeric values. New datasets

make violence a binary or ordinal variable with few values (Raleigh et al., 2010, Chenoweth, Pinckney and Lewis, 2018, Weidmann and Rod, 2018); Khawaja (1993) generates an ordinal variable with fourteen values, but this exception is not associated with a publicly available event dataset.¹

By asking annotators which image in a pair is more violent, the source of the violence, repeating this process across enough pairs, and scaling the output of the Bradley-Terry model (Bradley and Terry, 1952), continuous estimates of the magnitude of protester and state violence are feasible. This method has been used to measure accurately other subjective assessments such as favorability of media portrayals of politicians (Chen et al., 2016).

Images can also generate continuously valued estimates for protester violence as well. The literature on protester violence finds consistent negative correlations between protester violence and support for protest (Simpson, Willer and Feinberg, 2018), so this methodological may not be as impactful for that literature. Nonetheless, the methodology still permits more precise measurement of protester violence, and more complex relationships than are measurable from current estimates may reveal themselves.

2.1.2 Size

As more people protest, policy change becomes more likely. This relationship holds in democratizing countries (Chenoweth and Stephan, 2011), the United States (Gillion, 2012, McAdam and Su, 2002), Europe (Wouters and Walgrave, 2017), and non-Western contexts (Uba, 2005), as well as across branches of government (Walgrave and Vliegenthart, 2012). Understanding how large protests are and what causes them to grow or shrink in size is therefore essential to understanding when they succeed or repression “works”.

Crowd size estimation is its own field of study with a well-developed methodology (McPhail and McCarthy, 2004, Heller and Evans, 2010, Yip et al., 2010). These methodologies, because they require foreknowledge of a protest and are expensive to generate, rarely find their way into crowd size

¹Images could also generate categorical estimates of violence since they provide evidence of tear gas, fire, or property destruction, for example. We do not emphasize these possibilities since they more closely replicate existing measurements.

as reported in event datasets. Moreover, these numbers are often reported inexactly as "hundreds", "thousands", or as round numbers. Some datasets, such as SCAD, therefore report crowd size on an ordinal logarithmic scale; others, like the Crowd Counting Consortium ([Chenoweth and Pressman, 2017](#)), present low and high estimates; and others, like ACLED, report injuries and deaths, as those are easier to verify and so are more precisely reported.² None of the major machine-coded event datasets report crowd size, though individual projects have estimated crowd size from newspapers ([Acemoglu, Tahoun and Hassan, 2018](#)). Images therefore can generate direct estimates of a key variable in the study of collective action, and Section 4 shows that images from social media generate more estimates of size than existing approaches.

Social media posts can measure the size of protests. When an image contains faces, the number of faces are countable, and the number of faces across images at an event can be added together to generate an estimate of crowd size. Because counting faces raises ethical and measurement concerns, a coarser approach is to estimate when an image contains a small or large group. When using images derived from tweets, crowd size can be estimated by counting the number of tweets containing protest images. Crowd size can also be estimated counting the number of unique tweeters in a location at a given time, which works when applying a keyword filter. ([Sobolev et al., 2020](#)) uses cellphone location data to verify these image estimates in the United States.

2.1.3 Protester Demographics

A growing body of work focuses on how protester demographic attributes affect protests. Research on the youth bulge suggests that age matters because the young have their own grievances and have greater proclivity to protest ([Nordås and Davenport, 2013](#), [Hollander and Byun, 2015](#), [González, 2020](#)). Gender within movement leadership affects outcomes ([Robnett, 1996](#), [Asal et al., 2013](#)), and gender composition is a frequent object of study in surveys of protesters ([Kern, 2011](#), [Tufekci and Wilson, 2012](#)). Race is a common mobilizing feature across countries ([Scarritt, McMillan and Mozaffar, 2001](#)) and has featured prominently in social movements and protests in the United

²Starting in February 2019, the Notes column of each ACLED protest or riot event contains a size estimate. 63.71% of the time, no estimate is provided; most of the size estimates are actually descriptors ("hundreds", "dozens"), not a number.

States of America (Olzak, Shanahan and McEneaney, 1996, Bergesen and Herman, 1998, Wasow, 2020). These studies do not, however, measure temporal change in protester demographics, yet two leading theories about diversity and protest size require temporal variation in estimates. Information signaling models turn on the protesters coming from unexpected groups (Lohmann, 1994), while critical mass theories argue that initial protesters should come from homogenous groups (Oliver, Marwell and Teixeira, 1985).

To adjudicate between these theories therefore requires data about the changing composition of individual protesters, and we are aware of no event dataset which records attributes of the protesters themselves. Some, such as NAVCO 1.0 and 2.0, indicate whether a social movement includes individuals from various backgrounds (gender, racial, class, and education, for example), but it records that information as nominal variables at the movement, not protest, level. NAVCO 3.0, which provides daily data on events within movements, drops these attribute codings. SCAD codes whether an event is focused on female or LGBTQ issues but not the gender or sexual orientations of event participants. MMAD records actor attributes as a string variable *at the report level*, but the event level version of the dataset omits that variable; a researcher would therefore have to go to the MMAD report data and aggregate to the event level while keeping actor identifying information. Projects do code protester attributes (Heaney and Rojas, 2008, Kern, 2011, Tufekci and Wilson, 2012, Fisher, Dow and Ray, 2017). This type of work relies on survey data, sometimes *post hoc*, rarely include longitudinal observations, and are not part of larger event dataset efforts.

2.2 Text to Measure Emotions

How emotions affect individuals' decision to protest is a growing area of research (Jasper, 2011). They mobilize bystanders and mediate the effect of repression, depending on which are triggered (Pearlman, 2013). Because of the difficulty of measuring emotional states in real time, research in this area has been theoretical (Pearlman, 2013), experimental (Young, 2019), or uses simulations (Siegel, 2011). Estimating emotion from facial expressions is an active area of computer vision research (Soleymani et al., 2017), and some work has estimated the emotional content of protest images (Casas and Webb Williams, 2019).

Text is more appropriate for measuring emotions than images because many, such as pride or shame, are less apparent in images. Social media text is particularly useful because it is produced in large quantities. Emojis, digital faces used in conjunction with a post to express emotion, are transmitted as text, so their aliases (their representation as Unicode text) contribute to measuring emotions. To measure emotion, one can train new models as needed, rely on a large dictionary mapping words to anger, anticipation, disgust, fear, joy, sadness, surprise, trust (Mohammad and Turney, 2013), or use a Long Short-Term Memory trained on social media posts that treats emojis as labels (Felbo et al., 2017). Since this paper’s focus is on images, and natural language processing of social media data is a separate field of inquiry, we do not focus on methods for measuring emotions.

3 Generating Event Data from Images

This section explains the pipeline developed to extract magnitude and demographic information from geolocated social media images. To extract emotion from social media text, we use the R library `Syuzhet` to access Mohammad and Turney (2013)’s emotion dictionaries (Jockers, 2015).

3.1 Sources

Twitter, other social media platforms, and newspapers are the most common sources of images.

The easiest source of images is Twitter because its application programming interface (API) is much more open than competitors’. If downloading tweets via the streaming API, a random sample of tweets, most without images, will be provided. Based on our experience, approximately 10% of geolocated tweets contain images, and .45% of those are protest images. Since Twitter’s streaming API provides approximately 5 million geolocated tweets per day, approximately 2,250 geolocated protest images (.045% of tweets) are available per day. This estimate is a lower bound on the number of protest images from Twitter a researcher can obtain, and Section 5.1 discusses how to increase this number.

Instagram and Facebook contain many more images than Twitter, but their APIs are more restrictive. (Python libraries exist which will scrape each site. They violate each site’s terms of service, but the 9th Circuit Court of Appeals has ruled that platforms cannot prevent scraping of their public data (Berzon and Wallage, 2019).) Some work downloads images from Tumblr (Bourlai

and Herring, 2014, Xu et al., 2014), but its declining popularity means it is not likely to prove useful for future research. Messaging apps such as WeChat and WhatsApp are also increasingly popular platforms, but they are more difficult to access for research than Instagram and Facebook.

One could also retrieve images from newspaper articles. Though some newspapers maintain APIs, it is more common to acquire articles via scraping or social media. Each newspaper website will have its own structure, so building a pipeline to download images is time intensive and sensitive to changes. Instead, downloading articles via a Really Simple Syndication (RSS) feed or following news accounts on social media and downloading the shared article’s thumbnail image is easier. Newspapers also maintain archives of unpublished photos, the use of which would require directly partnering with an organization. Even if a research can access a large library of newspaper images, social media’s lower barriers to entry means such a corpora is likely to encompass fewer events than social media image datasets.

3.2 Three Methodologies

Event datasets fit into one of three categories depending on how they generate data. The earliest method, and perhaps still the most common, uses humans to extract data from text. The Conflict and Peace Databank (COPDAB) is the first large-scale project of which we are aware that codes events across countries (Azar, 1980). The World Event/Interaction Survey (WEIS) is the other large event dataset from this era (McClelland, 1984). Other prominent examples in this approach are the Nonviolent and Violent Campaigns and Outcomes (NAVCO) (Chenoweth and Schock, 2015), Social Conflict Analysis Database (SCAD) (Salehyan et al., 2012), and Dynamics of Collective Action in the United States (DCAUS) datasets (McAdam et al., 2009). The most comprehensive manual event dataset is the Armed Conflict Location and Event Dataset (ACLED) (Raleigh et al., 2010). Operating for over a decade with support from governments, research agencies, and non-profit institutions, it incorporates thousands of local news sources and experts plus several layers of coding review to produce daily updates.

A second method, using computers to algorithmically extract events, started in the early 1990s and remains an active area of methodological development. The Kansas Events Data System

(KEDS) pioneered this approach in the 1990s (Schrodt and Gerner, 1994), and the Global Database of Events, Location, and Tone (GDELT) as well the Integrated Conflict Early Warning System (ICEWS) are direct descendants (Leetaru and Schrodt, 2013, Boschee et al., 2015). New entrants in this area are the Machine-learning Protest Event Data System (MPEDS) (Hanna, 2017) and the Open Event Data Alliance (OEDA) (Halterman et al., 2017). The next generation of machine-coded event data uses advances in natural language processing, distributed computing, and machine translation to code actors and locations, including in non-English languages (Solaimani et al., 2016, Halterman, 2019). The most advanced datasets in this domain, as of this writing, are ICEWS and Temporally Extended, Regular, Reproducible International Event Records (TERRIER) (Liang et al., 2018). This paper uses this automated approach.

The third method is a hybrid: machines identify items, then humans extract event data. The Social, Political, and Economic Event Database (SPEED) is a prominent example of this approach (Hayes and Nardulli, 2011). The Uppsala Conflict Data Program has a variety of event data, and its Georeferenced Event Dataset uses a hybrid approach (Sundberg and Melander, 2013). Two new projects, Mass Mobilization and the Mass Mobilization in Autocracies Database, are more recent hybrid approaches focusing solely on collective action (Clark and Regan, 2018, Weidmann and Rod, 2018). The Crowd Counting Consortium manually inspects articles identified from the Count Love project's nightly scrapes of American news sites (Chenoweth and Pressman, 2017). For a comparison of machine-coded event data with manual and hybrid approaches, see Gleditsch, Metternich and Ruggeri (2014) and Wang et al. (2016).

3.3 Measuring Protest Magnitude and Protester Attributes from Images

3.3.1 Computer Vision for Automated Image Coding

Without recent advances in computer vision (LeCun, Bengio and Hinton, 2015), identifying images with protesters would require recruiting and training human coders who code content of each image according to a given schema. This approach is accurate but very expensive as the amount of data grows. It is also very slow, especially because most images do not contain protest.

Consider a hypothetical project to manually code 100,000 images. A human coder typically

codes about 100 images per hour. Since each image should receive attention from at least two individuals, this work requires 2,000 hours. Assuming 10 coders who can work for 20 hours per week, this project requires 10 weeks to complete and costs \$24,000, using California's minimum wage of \$12. The actual cost is certainly higher, as \$24,000 does not include training, team meetings, and error correction; many students also expect a wage higher than \$12/hour. Coders also have personal biases that can be difficult, if not impossible, to remove from their responses. If a study is based on longitudinal data coded by different sets of coders at different times, their responses may not be comparable. Overall, manual coding is not only expensive but also limited in objectivity and transferability, which has impeded quantitative analysis on large-scale visual data. For all this effort, only approximately 450 protest images will be found.

New computer vision approaches, fortunately, can automatically classify content of given images, such as whether one contains a protest event. Computer vision is a field of research, mostly studied in computer science and statistics, whose goal is to develop theories and algorithms to automatically understand content of visual data by recognizing objects, attributes, people and their interactions in a scene. Recognition occurs via statistical learning ("machine learning") from images or videos where an algorithm attempts to discover ("learn") patterns from given data. This algorithm often uses additional data to take advantage of human supervision, typically in the form of labels manually annotated on input images.

In particular, this paper uses deep learning (LeCun, Bengio and Hinton, 2015) and convolutional neural networks (CNN) (He et al., 2016). A CNN is a hierarchical and compositional model in which a complex visual concept, such as a human face, is broken down into smaller and less complex concepts, such as eyes or nose, repeatedly until it reaches a very low level representation of an input image, such as edges and pixels.

A CNN can classify images. It can take an image as input and compute its scores on a list of target labels. For example, a CNN trained with animal images and categories will compute how likely it is that a given input image contains a cat, a dog, or a horse, and the researcher may choose the label with the highest value as the output of the classifier. The visual features and knowledge

about their compositions necessary for this computation are automatically learned during training. The final output value is computed through a series of operations in multiple stages where the output of an earlier stage will be used as input for the next one. During the process, an input image will be first transformed into low level features (e.g. edge) and gradually turn into representations semantically more meaningful (e.g. objects and attributes). It is beyond the scope of this paper to elaborate its technical features in length. Interested readers can turn to [Joo and Steinert-Threlkeld \(2019\)](#) and [Williams, Casas and Wilkerson \(2020\)](#) for technical details.

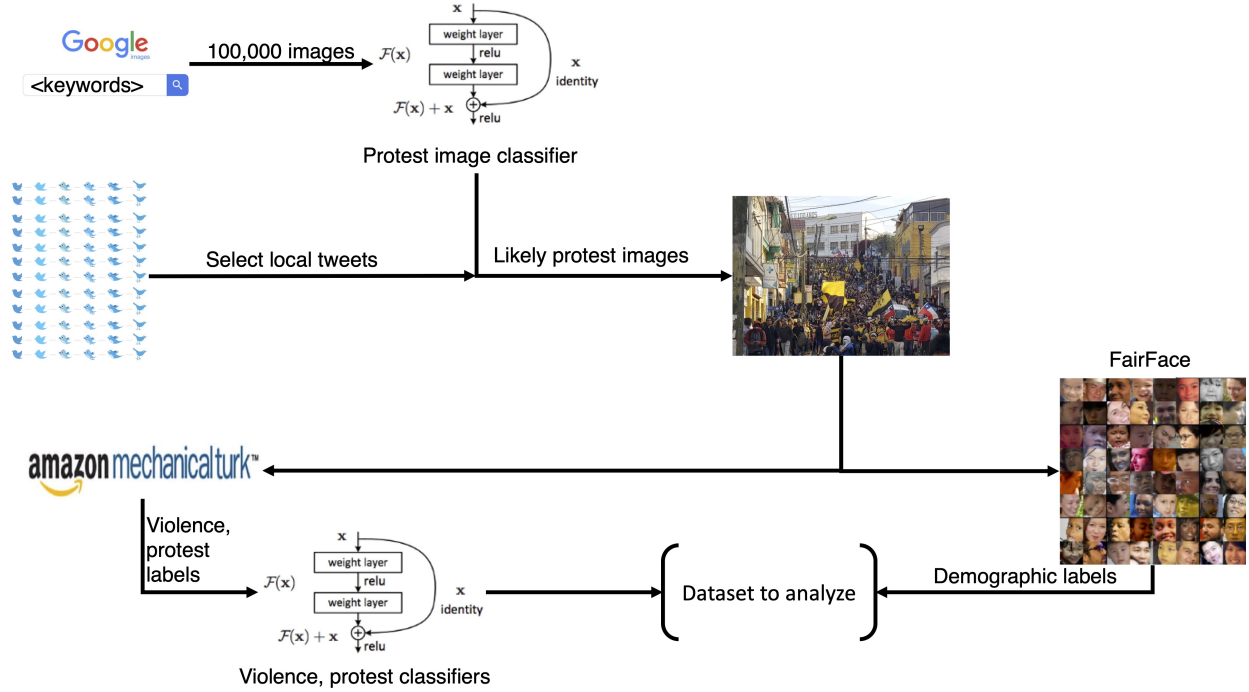
CNNs support feature sharing, which means the same visual feature can be shared for multiple tasks (this is called multi-task learning). This is an effective approach because features useful for protest detection, for example, signs, fire, and smoke, can be also used as indicators for violence estimation. These individual features – signs or fire – can be automatically learned during training. That is, the learning algorithm will automatically discover these features as discriminant patterns to identify protest images, and the CNN model will implicitly capture them. Researchers therefore do not need to define a list of visual features that constitute “protest” or “violence” because these will be learned automatically from the human annotations. This association is, however, not explicitly expressed in the model, and researchers will not know why the classifier classifies certain images as violent or not. This lack of explainability and interpretability has been pointed out as a weakness of deep learning and is an active area of research ([Zhang, Wu and Zhu, 2018](#)).

3.3.2 Training a CNN for Protest Image Analysis

For this paper, we train a CNN to handle two different tasks: (1) to classify whether an image describes a protest event and (2) estimate the level of protester and state violence depicted in the image. Figure 1 illustrates the steps in this paper’s pipeline, and the rest of this subsection describes them.

To identify protest images, we use weakly-supervised learning. Weakly-supervised learning combines the advantages of supervised learning (labeled data but a lesser amount) and unsupervised learning (more data but unlabeled). Training a CNN typically requires a large quantity of images describing the concepts of interest. It may be too time consuming to we randomly sample images

Fig. 1. Steps to Generate Magnitude and Demographic Estimates



from Twitter or other sources due to the low frequency of the concept, which is the case for protests.

To efficiently collect protest images in our train set, we first search Google Images using keywords related to protest such as “protest” or “BLM” and obtain about 10,000 positive images which may or may not capture a protest scene. We also used negative keywords such as “concert” to collect a large number of negative images that may look similar to a protest scene. Using such difficult examples in training can improve the robustness of the model. These labels (positive or negative) are noisy because we did not confirm their correct label through manual annotations, which is why this learning approach is called weakly supervised. Nevertheless, we trained our first rough model using these images and labels. The model is expected to perform reasonably well, although not perfect, which suffices the main purpose of this first model.

This model is then applied to about 4.46 million images from the countries and periods shown in Table A2. The protest image classifier runs on each of the images, and we discarded the images for which the classifier output for the protest concept was below a threshold of 0.6. The threshold was chosen empirically and we confirmed it was extremely unlikely that an actual protest image

receives a score below the threshold (a false negative). This filter results in 115,060 images that may be potential protest images. We further chose a subset of 40,764 images by stratified sampling.

We then used Amazon Mechanical Turk to manually verify ground-truth labels for the 40,746 images, of which 11,659 images were determined to contain protest. We assigned two workers to annotate each image and used a third annotator if their annotations differ. Figure S1.1 shows the interface along with sample images that were coded. These labels are then given to a ResNet-50 CNN (He et al., 2016); customizing a pre-trained model by using new training data is known as transfer learning and significantly reduces the time to construct a classifier. We then build a precision-recall curve to determine the threshold above which to conclude an image contains a protest. Figure A2 shows this curve. To be conservative, we chose the threshold providing the maximum recall while keeping precision at .85.

To train the second model to estimate protester and state violence, we used the same 11,659 protest images. While the concept of protest is binary, that of violence magnitude is continuous. Annotating the level of violence is also a more subjective and judgmental task than identifying protesters in a scene. Prior studies have used pairwise comparison (Chen et al., 2016) to obtain subjective annotations on image data as this typically produces a more reliable measure than using annotations made on individual images. Following this approach, we provide two images to an annotator and ask which image they think depicts a higher degree of violence. Table A1 shows intercoder reliability. These pairwise annotations are aggregated together and converted into a global ranking score using the Bradley-Terry model (Bradley and Terry, 1952). We then ask if the violence is from the state or protesters. These labeled images feed into another ResNet-50, generating a continuous violence score in the range of [0, 1]. A higher value indicates a more violent scene, and it is this classifier estimate that we mean when we discuss “continuously valued” estimates of violence.

Figure 2 shows sample images with their classifier estimate. The images for Protest and Protester Violence are from Chile; those for State Violence, from Venezuela. Each image is labeled with the city to which Twitter assigns it and the classifier estimate. In addition to the validation checks in

the appendix already mentioned, Figure A3 shows the receiver operating characteristic curve for the protest, protester violence, and state violence labels.

Fig. 2. Example Images and Classifier Annotation [0-1]

Protest



Protester Violence



State Violence



Finally, we used a publicly available model, FairFace, to recognize demographic attributes of people in images from their facial appearance. We chose the FairFace model (Karkkainen and Joo, 2019), which is based on a CNN. While there exist many different APIs and public models for facial attribute recognition, FairFace is the only model trained from a racially balanced dataset comprised of amateur images. FairFace’s training data substantially improves classification accuracy, especially for the non-White populations.

3.4 Deduplication

Like any automatically generated event data, duplication (one event recorded as several) is a concern. Social media image data facilitates duplicate detection in two ways.

First, it is simple to use convolutional neural networks to identify duplicate images. Once an image classifier is trained, its last layer will be a vector of length n corresponding to n learned features used to make the classifier's final estimate. Applying the classifier to each image generates as many feature vectors as there are images, and the distance between these vector pairs is calculated. A distance threshold is then set below which images are considered duplicates. (This process is similar to plagiarism detection for text.) For all duplicate images, only the first is kept.

Second, restricting images to those containing location and time metadata facilitates event deduplication, assuming multiple images from the same location at the same time document the same event. When aggregated to the desired location-time level, such as city-day, multiple images can be merged to create a single event. This deduplication is easier than deduplicating events from text because it is coarser: deduplicating events from text requires measuring meaning and action, which is much harder than measuring location and time.

This aggregation risks blending different events together if they occur in the same city on the same day. For example, a pro-regime counter protest occurring on the same day in the same city as an anti-regime protest will aggregate as the same protest. One could theoretically group images within cities, but in practice the vast majority of images a researcher can acquire are located only to the city, not to a point of interest. One could also group images by an intraday interval, but data are rarely prevalent enough to disaggregate days. Future work should investigate the prevalence of this event blending and methods to ameliorate it.

3.5 Bias

When generating event data, selection, location, and fatigue bias are common concerns. When generating event data from social media, one must also consider platform bias, a type of selection bias.

Perhaps the greatest concern when relying on social media data is that users on social media are often not representative samples from their country. In the United States, for example, Twitter skews young: 24% of Americans over 18 use Twitter (only Facebook and YouTube are used by a majority), but 45% of those 18-24 and 14% of those over 50 do (Smith, Anderson and Caiazza,

2018). There does not appear to be variation by gender or race, but wealthier and more educated individuals are more likely to use it (this pattern is true for all social media platforms).

It is not known if these biases affect the types of events that images record. Twitter users who geotag their tweets exhibit the same temporal patterns as non-geotagged users (Valkanias et al., 2014). There may also be variations by sex and occupation, though those effect sizes are quite small (Sloan and Morgan, 2015). There is variation of geotagging by language, with 8.8% of Turkish speaking users but only .9% of Arabic speaking users allowing longitude and latitude coordinates to be assigned to their tweets; the figure in that same study is 3.4% for the United States (Sloan and Morgan, 2015). We are not aware of other studies which compare behaviors of geotagging to non-geotagging users.

Regarding location bias, geotagged tweets from the United States are likelier to be wealthier, urban, and from Asian, Black, or Latinx individuals, compared to Census data (Malik et al., 2015). It is not clear if similar biases exist in other countries, though geotagged tweets come from every continent and most countries (Mocanu et al., 2013, Morstatter et al., 2013). When using geolocated tweets, geolocation is usually to the city; neighborhoods or points of interest require the user to input them and so are rare.³ The analysis of Venezuela and Chile shown in Section 4 shows that geolocated tweets can record events in at least as many cities as other event datasets, suggesting their location bias is no worse than the current state of the art.⁴

Fatigue bias means events appear to become less pervasive over time, even if their true intensity stays constant. This bias is especially concerning when event data derive from news sources because newspapers report events less often the longer they persist; this effect is sometimes referred to as novelty bias. This apparent decline in activity has been noted in coverage of the first Palestinian *intifada* (Gerner and Schrod, 1998), Syria's civil war (Masad, 2013), and Libya's 2011 protests

³Without metadata, location can be inferred from objects in the image, though assigning location based on object detection is computationally more difficult than relying on coordinates (Kaneko and Yanai, 2013).

⁴Improving geolocation in text is an active area of work in the event data community. Mordecai is a relatively new software library that uses neural networks and the Geonames gazetteer to assign a location from text (Haltermann et al., 2017), and it has been incorporated by the Open Event Data Alliance (Brandt et al., 2018). Lee, Liu and Ward (2018) improve ICEWS' geolocation by training a neural network on 1,000 articles with the true event location labeled (newspaper articles contain multiple locations, only one of which will refer to where the event occurred).

and subsequent civil war (Baum and Zhukov, 2015).

Data from social media should suffer less fatigue bias. Reporting fatigue derives from the need to keep readers’ attention with novel information, but most social media users do not face this pressure. That is, since most accounts on social media do not have to publish to satisfy an audience, their probability of posting about an event should not decline as a function of the event’s duration. A decline in the number of events recorded should therefore more likely reflect a true decline, not a fatigue-induced one.

Restrictions that social media companies impose on data availability may, however, bias against small events. To prevent malicious behavior, APIs usually limit the rate at which one can download data. Twitter, for example, only streams 1% of a day’s tweets for free, and the REST API also limits how much historical data is available. An event may therefore appear to no longer exist when in fact it does not exist strongly enough – not enough people tweet about it – to surmount sampling restrictions.

4 Demonstration

This section demonstrates the potential of generating multimodal protest data from geolocated social media content. It first identifies the countries in which this method will generate at least one protest event per week; these 107 countries contain 82.7% of the world’s population and 97.15% of its GDP. It then validates the ability of images to capture meaningful variation in protester attributes. Next, protest waves from Venezuela in 2014-2015 and Chile in 2019 are then analyzed and compared to records from existing event datasets. In Venezuela, the geocoded protest image records are compared to three datasets, two machine-coded text (ICEWS, TERRIER) and a hybrid text (MMAD). We are aware of no other datasets that provide daily protest and repression measures for Venezuela at that time. We then compare our approach in Chile to ACLED’s records. ACLED was chosen because MMAD and TERRIER do not record events through 2019, ICEWS performed poorly in Venezuela, and it is the most extensive hand-coded event dataset of which we are aware.

4.1 Country Income and Population Limit This Methodology

To estimate the geographic coverage geolocated images from Twitter provides, we downloaded extracted tweets from the seventeen countries and nineteen periods listed in Table A2. We then use the methodology described in Section 3.3 to identify those containing protest images. The bold rows are those for which at least one protest image per week is available at the city level.

Figure 3 plots each country against its population and gross domestic product per capita, revealing that those two variables bound the ability to generate event data from images. Red countries do not produce enough protest images – at least one per week in at least one city – using this methodology; green countries do. Size corresponds to the average number of protest images per day.

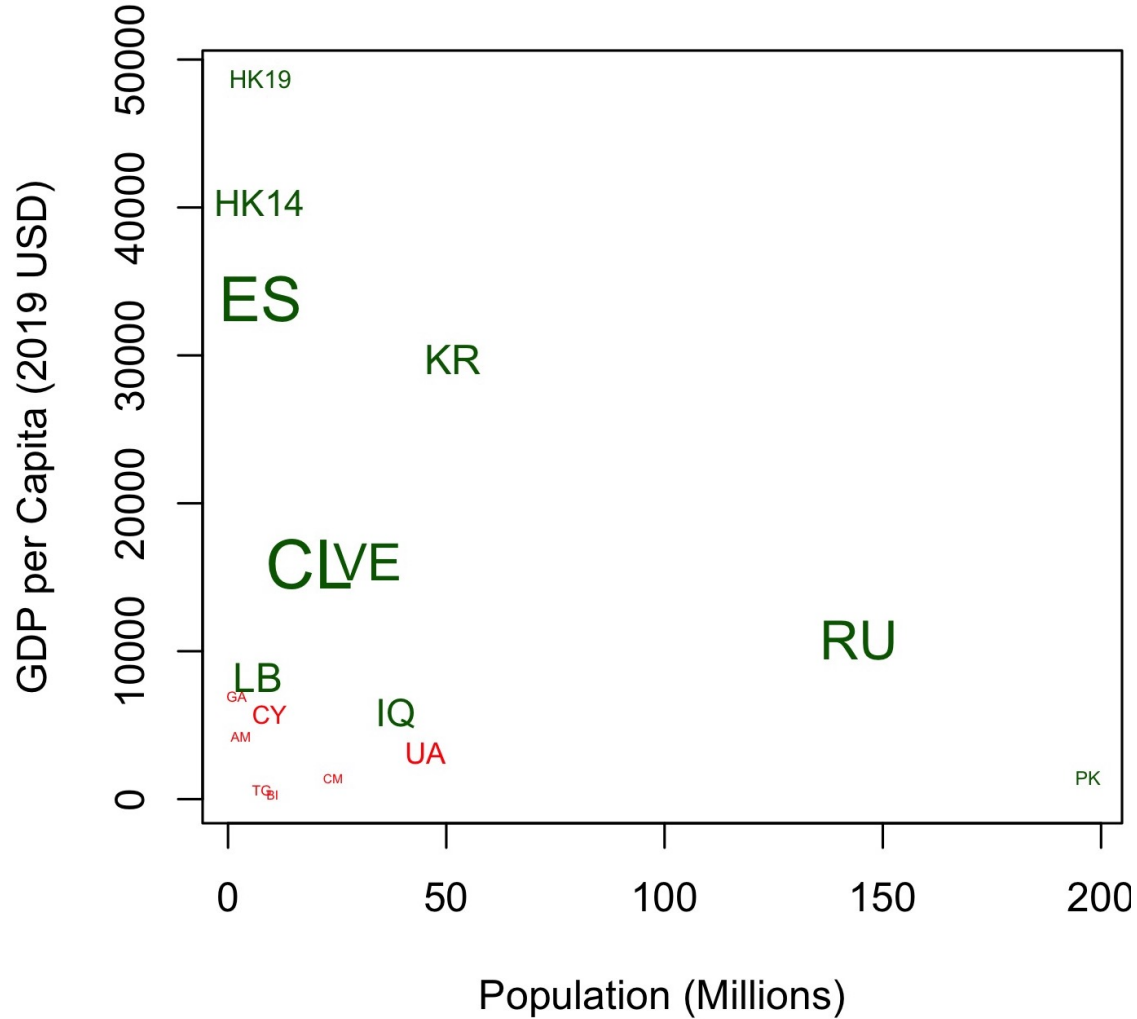
Since Figure 3 reveals a clear demarcation between countries that can produce city-day protest data and those that cannot, we fit a logistic regression with these nineteen events. The dependent variable is a 1 if a country contains at least one city producing at least one protest image per week (the green points in Figure 3, bold rows in Table A2). Using the World Bank’s World Development Indicators, we predict which countries will produce enough geolocated protest images to generate event data at the city-day level. Table 1 shows these results. Models 1-2 model if at least one city in a country contains one geolocated protest image once per week. Models 3-4 estimate the number of cities per country that meet this criterion. Model 1 is used to create Figure 4.

Based on these results, any country with a predicted probability greater than .38 is assumed to produce enough geolocated protest images to generate event data. Figure 4 shows these results.

Geolocated images shared on Twitter will produce protest event data in 107 countries, 53.5% of those for which population and gross domestic product per capita are available. These countries have an average population of 56.41 million and gross domestic product (GDP) per capita of \$29,425, versus 13.28 million and \$2,848 for those that do not produce enough geolocated social media protest images. They contain 82.7% of the world’s population and 97.14% of its GDP. Table A3 lists both sets of countries.⁵ These numbers will improve as future work incorporates more

⁵ACLED produces protest data in 142 countries; Mass Mobilization, 162; Mass Mobilization in Autocracies, 76; ICEWS, 199; and TERRIER, 220.

Fig. 3. Protest is Detected in Populous or Medium-Income Countries



Note: Red countries have fewer than one protest image per seven days, at a city level of resolution. Size corresponds to the number of protest images per day.

images, as Section 5.1 discusses.

4.2 Validation

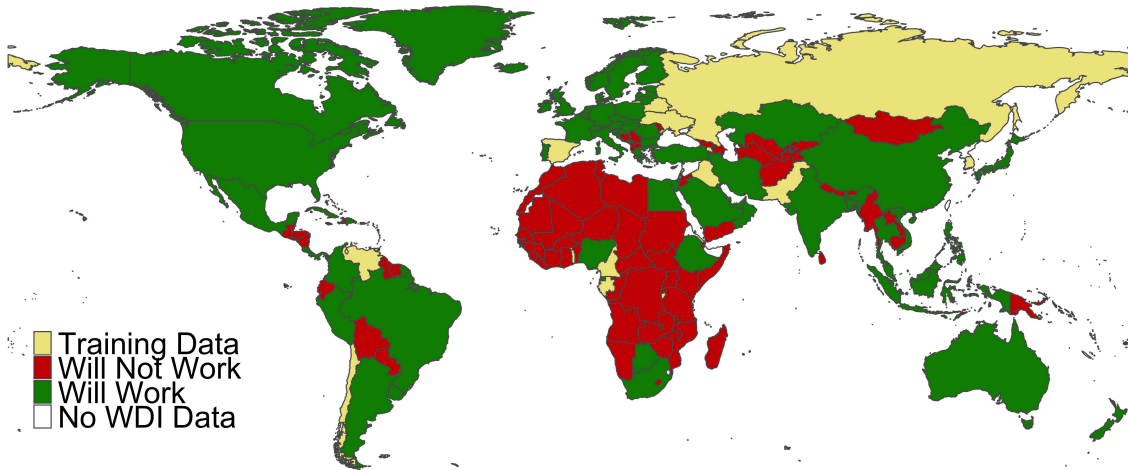
Images accurately recover demographic and magnitude information, based on descriptive patterns across four protest waves. The four are secessionist protests in Catalonia, Spain at the end of 2017; Hong Kong's Umbrella Movement protests in 2014; Pakistan's anti-blasphemy protests in November 2018; and anti-Maduro protests in Venezuela from the end of 2014 through the beginning of 2015.

These four were chosen to emphasize the differences images can record. Hong Kong images

Table 1. Determining Model Applicability

	Cities?		Cities per Country	
	<i>Logit</i>	<i>OLS</i>	<i>Poisson</i>	<i>OLS</i>
	(1)	(2)	(3)	(4)
Population (Millions)	0.021 (0.013)	0.005** (0.002)	0.001 (0.002)	0.006 (0.061)
GDP/1000 (USD)	(0.253)** (0.127)	0.024*** (0.006)	0.025*** (0.006)	0.163 (0.216)
Intercept	(-2.517)** (1.154)	0.095 (0.138)	1.253*** (0.195)	3.128 (5.102)
Observations	17	17	17	17
Adjusted R ²		0.536		-0.098
AIC	13.635		266.997	

Note: *p<0.1; **p<0.05; ***p<0.01

Fig. 4. Predicting the Geographic Coverage of Protest Event Data from Twitter Geolocated Images

Note: Using geolocated images shared on Twitter will produce protest event data in 107 countries. This number will increase if one aggregates by week or augments the image collection strategy.

should record a very high percentage of Asian faces, while the other three should record close to zero. Hong Kong’s protests were organized by student activists, so they should record the highest percentage of youth participants as well. All protests save Pakistan had reports of state violence. Since Venezuela’s protests aimed for President Maduro’s ouster, they may have the most protester violence. We also show the percent of protesters who are female but have no prior expectations about its values.

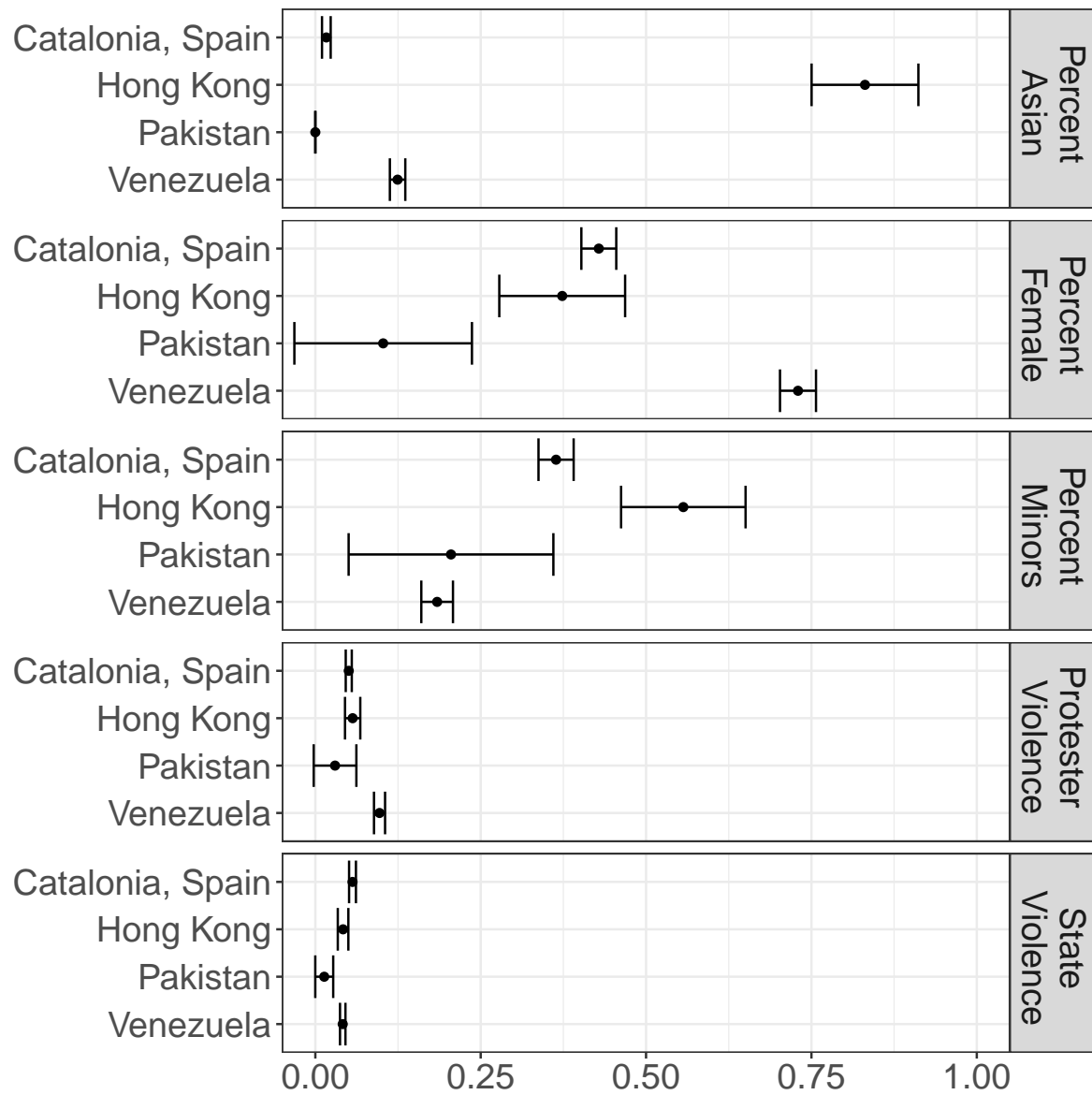
Figure 5 compares point estimates of an image classifier applied to protest images from these four protests. As expected, Hong Kong is the only protest which records a large percentage of faces as Asian. It also records a majority of its faces as being from those aged 0-19, which matches reports of youth leading the protests (Kaiman, 2014).⁶ Unexpectedly, almost 75% of protesters in photographs from Venezuela are female; Pakistan images suggest only 9.7% of protesters were female; the 2017 Global Gender Gap Report ranks it as the sixth most gender unequal country in the world, so this result is perhaps not surprising (Leopold, 2017). State violence is highest in Spain’s secession protests, where police frequently dispersed peaceful protests with violence (Beals, 2017). While there were reports from Spain of police beating up protesters, we were surprised to find it record more violence than Hong Kong or Venezuela. That Pakistan reports the lowest amount of state violence makes sense, as key cities were immobilized by protesters and leaders feared enraging them further. Finally, Venezuela records the highest amount of protester violence.

4.3 Venezuela Example

At the end of 2013, Nicolas Maduro, Hugh Chávez’s Vice-President, narrowly won Venezuela’s presidential election. Protests over security started in 2014 in response to the murder of former Miss Venezuela Mónica Spear, and they quickly expanded in scope and intensity; one non-profit organization recorded 9,286 that year (Patilla, 2015). Protests continued throughout 2015. Twitter is also a popular platform in Venezuela and is used by wide swaths of the population (Franceschi-Bicchierai, 2014, Forelle et al., 2015, Munger et al., 2019). We focus on November 1, 2014 - February 10, 2015 to keep the size of the data manageable, but this analysis could extend through

⁶Joshua Wong, one of its main organizers as well as a leader in the 2019-2020 protests, was 17 at the time.

Fig. 5. Demographic and Magnitude Variation Across Countries Confirms Image Methodology



2015, the end of MMAD and TERRIER’s data.

Since ICEWS and TERRIER use the same event ontology, Conflict and Mediation Event Observations (CAMEO), extracting the same type of events is straightforward. Any event with the root code of 14 is a protest. Repression can occur in various guises under root code 15 (“Exhibit Force Posture”), 17, (“Coerce”), or 18 (“Assault”). Event code 175 (“Use tactics of violent repression”) is the closest to protest policing, but the news datasets record few enough of those that the following results include all events we identify as repression.⁷ For MMAD, each row is a protest, and we use the `max_secengagement` variable, which is ordinal, to measure repression. Once protest and repression events are identified in these text datasets, they are merged with the image event data based on matching city days.

Results for magnitude, demographics, and bias follow. Section S3 analyzes which types of images are most likely to be duplicated.

4.3.1 Magnitude

We verify that images accurately measure protest size using data from Francisco Rodriguez; he has worked with a team in Caracas to estimate protest sizes by using carrying capacity estimates of public areas (Rodríguez, 2020).⁸ For the 18 protests they identify that our dataset contains, the correlation is .4101. Figure 6 shows that the residuals from regressing our estimate on theirs exhibit no bias, and results from a Shapiro-Wilks ($p = .9598$) and Kolmogorov-Smirnov ($p = .5975$) confirm this finding.

The image event data record different types of magnitude information for events that the text datasets also record, as Figures 7(a) and (b) show. ICEWS records a protest in Caracas on January 24, 2015, but it does not record protester violence (Figure 7a) or the presence of police (Figure 7b); TERRIER records no protest in Venezuela. MMAD records this protest, no protester violence, and the presence of police. The image, however, permits differentiation between normal and riot police as well as measuring the size of the police force.

⁷Event codes 151, 153, 170, 171, 1711, 1712, 1723, 1724, 173, 180, 182, 1823 are the other CAMEO events we include as repression.

⁸Several times, we contacted the Venezuelan Observatory of Social Conflict, which has been reported to have collected these data. Understandably, no response forthcame.

Fig. 6. Unbiased Estimates When Calculating Protest Size from Faces

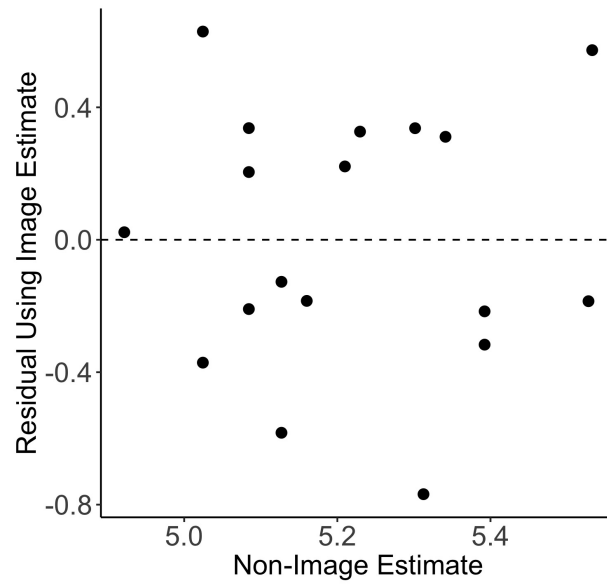


Fig. 7. Images Provide Additional Detail About Events in Other Datasets



(a) 01.24.2015: Protester Violence, Caracas

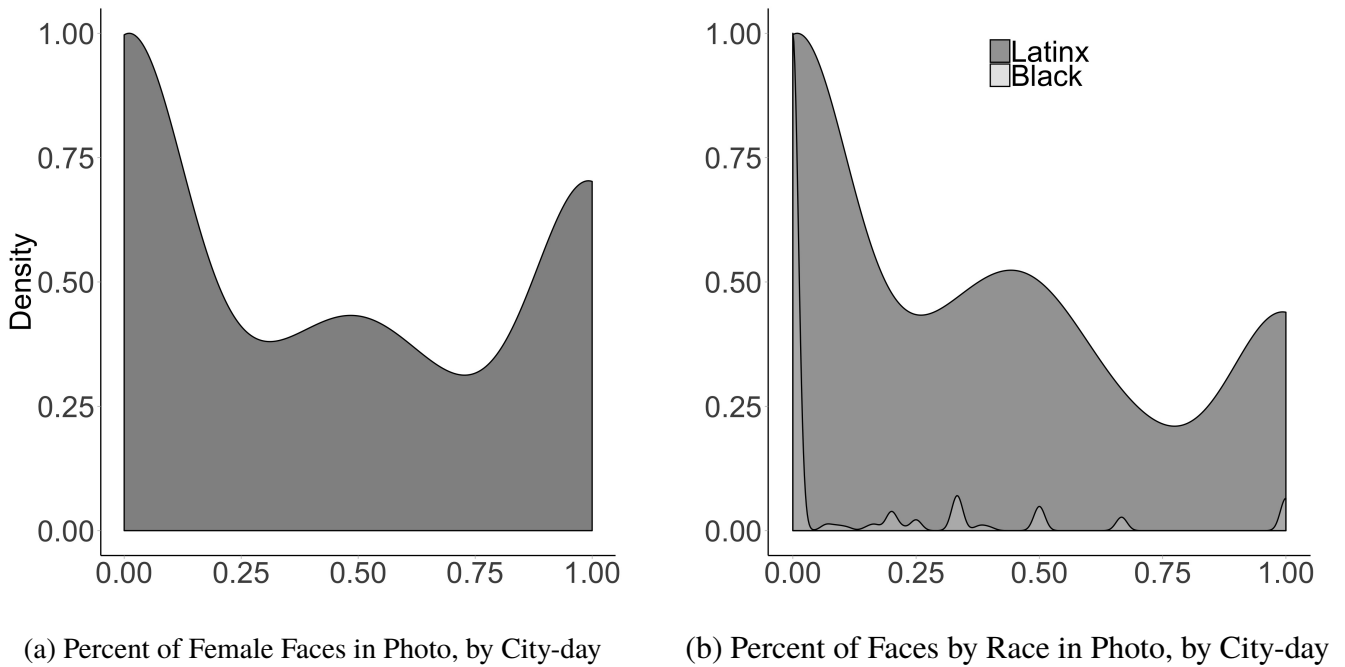


(b) 01.24.2015: State Response, Caracas

4.3.2 Protester Demographics

In addition to providing new information about the magnitude of an event, images permit reconstruction of demographics of protesters. Figure 8(a) shows the percent of protesters per city-day who are female, measured as the percentage of female faces out of all faces in protest photos. Figure 8(b) shows racial diversity. The FairFace classifier distinguishes between White, Middle Eastern, East Asian, Southeast Asian, Black, South Asian, and Latinx faces. While measuring race is contentious and not applicable in many countries, the legacy of colonialism in Venezuela, much like elsewhere, makes skin tone there a salient marker of politics (Page, 2015).

Fig. 8. Protester Demographics



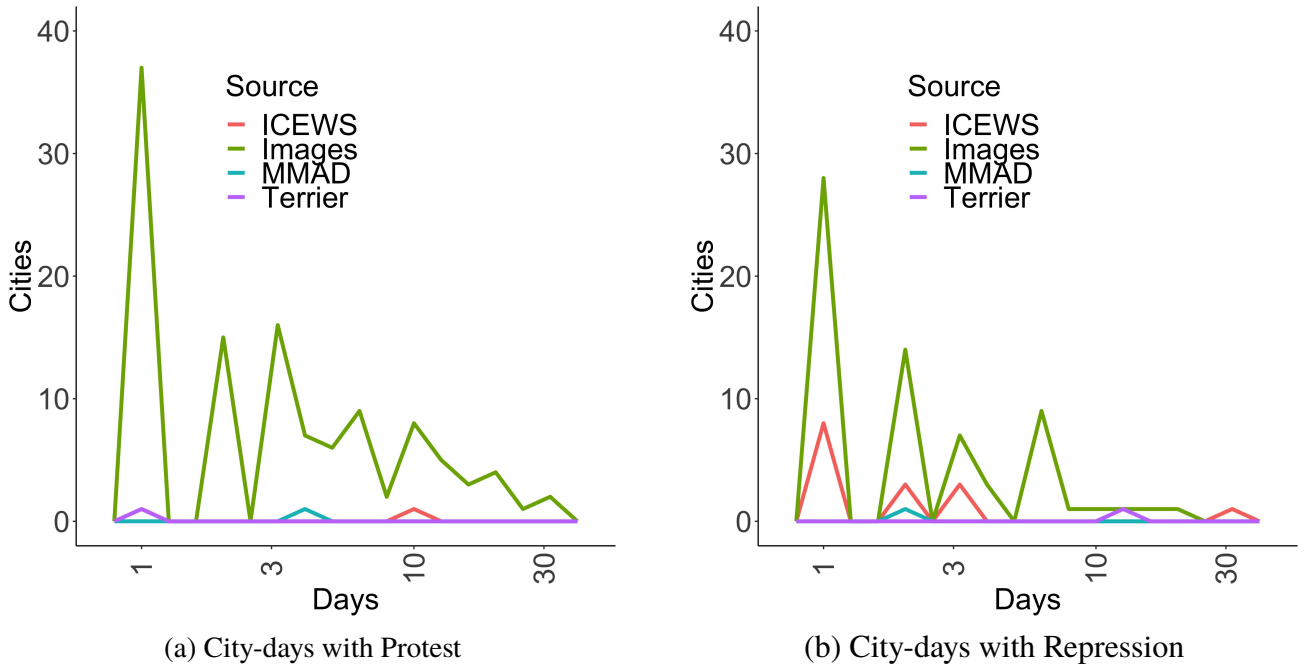
As far as we are aware, previous measures of demographics rely on surveys taken after (McAdam, 1986, Tufekci and Wilson, 2012) or during (Heaney and Rojas, 2008, Fisher, Dow and Ray, 2017) protests, not passive measures such as photos. No event dataset measures the demographics of protesters, though some are interested in ethnic identity and so measure group features of protesters (Vogt et al., 2015).

4.3.3 Bias

While there is no publicly available ground truth data for Venezuela protests, we attempt to measure bias by comparing the distribution of protests we identify to other datasets; we find more protests in more cities and confirm these are true positives.

Figure 9 shows the distribution of city-days for protest and repression. The image event data observes more protest and repression events in more cities than MMAD, TERRIER, or ICEWS, suggesting less location bias than those datasets. TERRIER only identifies one day of protest in this time period, Caracas on February 6, 2015. ICEWS identifies 9 protests in Caracas and 1 in San Cristobal on January 13, 2015. Twitter identifies 24 protests in San Cristobal across as many days. MMAD identifies 6 protests on 6 dates, all from Caracas.⁹

Fig. 9. Distribution of City-Days, Venezuela

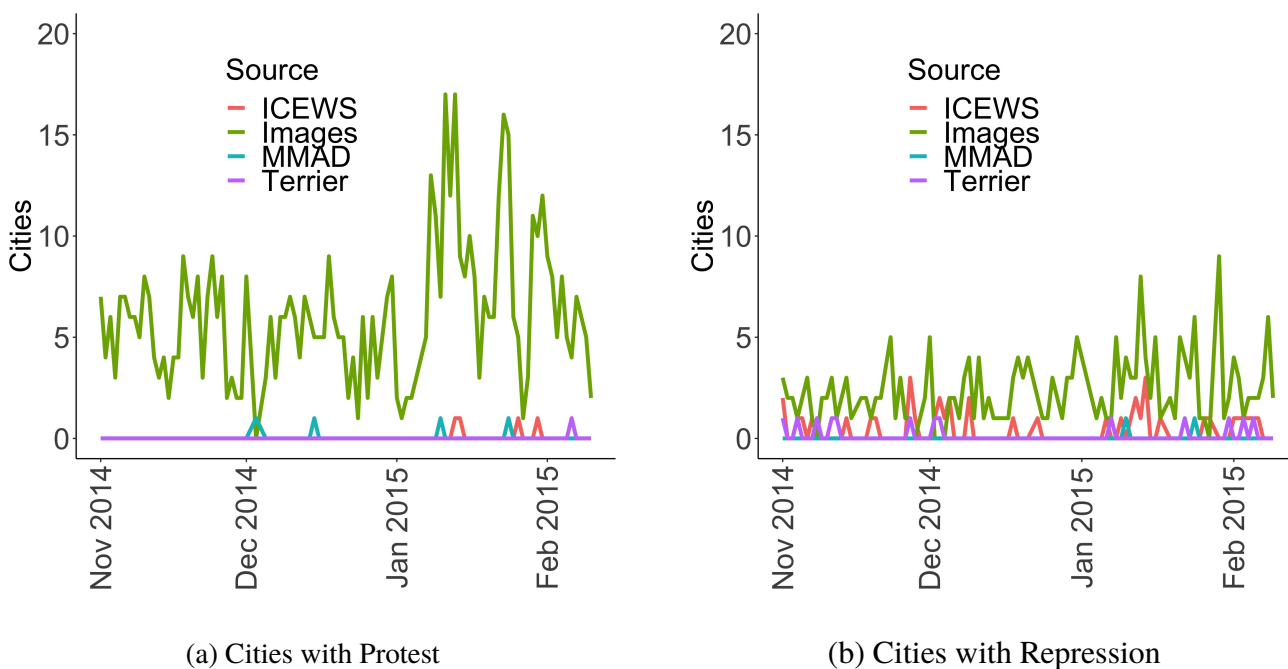


Protest event data from geolocated social media content appears to generate less fatigue bias than the comparison datasets, as the time series in Figure 10 show. Figure 10(a) shows that images provide a more continual measure of protests than the text datasets. The other datasets exhibit less fatigue bias for repressive events, as those datasets record more repression than they do protests,

⁹Twitter misses one of TERRIER's 15 events, 11 of ICEWS' 43, and none of MMAD's 4.

though only ICEWS records repression events consistently. Restricting analysis to cities included in both datasets shows they record similar numbers of repressive events, though “Use tactics of violent repression” (event code 175 and the most natural comparison to our measure) is recorded less frequently than in the image dataset.

Fig. 10. Trends in Protest and Repression in Venezuela



Figures 9 and 10 also hint at selection bias in the datasets. Matching previous research, the text datasets record more repressive events than protests (Myers and Caniglia, 2004, Hellmeier, Weidmann and Geelmuyden Rød, 2018). The image data, however, records more protests than repression events. While biases certainly exist in the image data, they do not appear to be towards violent events.

Importantly, the extra events the image data record are not false positives. Figures 11(a) and (b) show images from protests recorded in Merida and Caracas. Anti-Maduro protests occurred on February 4, 2015 in Merida (Patillia, 2015b); ICEWS records an arrest of citizens in Caracas on that day, and the other datasets record nothing. The image data also record a teacher protest in Caracas on November 6, 2014 that does not appear in the other datasets (Patilla, 2014).

Wikipedia provides the most comprehensive record of which we are aware of protests in

Fig. 11. Protest Events Recorded in Social Media but not Other Datasets



(a) 02.04.2015: Peaceful Protest, Merida



(b) 11.06.2014: Peaceful Protest, Caracas

Venezuela, with particular detail provided for 2015 and 2016. While checking every record there against protest images from Twitter is outside the scope of this paper, many protests from those pages appear in tweets in our corpus. For example, residents protested in Caracas on January 25th, 2015, and we have 12 tweets with protest images from there on that day; the other datasets record no protest then. The image data record 16 protests on January 23rd, compared to 0 in the other datasets. Wikipedia records protests on both of these days, further suggesting the image approach generates true positives.

4.4 Multimodal Event Data from Chile

This subsection combines image and text from Twitter with protests in Chile that ACLED identifies. Section S4 shows that the image approach and ACLED record similar amounts of protest and repression.

The resulting event data are multimodal in two ways. First, ACLED uses only text to identify protests. Second, part of the protest data from Twitter is emotions based on tweet text. Multimodal event data can be created exclusively with Twitter data, but this paper does not present that analysis in order to emphasize the power of adding social media data to existing event datasets.

To create the multimodal event data, we first identify protest events using the same image-based methodology as for Venezuela. In addition to generating magnitude and attribute variables from

images, each tweet receives an emotion type indicator; though *syuzhet* provides eight emotions, we analyze only anger, joy, fear, and sadness as they are what *Pearlman (2013)* theorizes should affect protests. We then kept all protests and riots ACLED recorded in Chile. Of these 1,364 city-days containing at least one protest or riot, 1,205 matching records exist from the Twitter images.

The outcome of interest is protest size, operationalized by summing the number of faces in Twitter protest photos. Though ACLED also provides size estimates, 63.71% of the protests do not have precise sizes, and almost all of the others are given as estimates (“approximately 300” or “at least 10”, for example). To measure violence variables, the average of the classifier output for protester violence and state violence is taken for all images from a city-day. We also keep ACLED’s *fatalities* variable as a measure of state violence. For protester attributes, gender is the percent of faces that are male and the percent that are female; age is the percent of individuals estimated to be between 20-29, to capture young adults purported proclivity to protest; and we do not use the race estimates. To measure emotion, we record the percent of tweets that contain keywords for anger, joy, sadness, or fear; categories are not mutually exclusive.

The regression model uses a log-transformed dependent variable and first-differences all variables. Standard errors are clustered by city. Table 2 presents these results.

While this paper does not develop theoretical expectations around violence, protester attributes, or emotions, the results are broadly consistent with others’ findings. Bystanders do not like protesters who behave violently (*Simpson, Willer and Feinberg, 2018*), and we find a negative, statistically significant correlation between protester violence and subsequent protest size. Regarding violence, the regression finds the same curvilinear relationship between state violence and subsequent protest size (*Zhukov, 2018*), though the result is not statistically significant. Fatalities, the most extreme result of state repression, should have a negative relationship with protest size; the regression’s sign is correct, but the result is not statistically significant. We are not aware of work that theorizes about how protester diversity should correlate with subsequent protest size, though these results match arguments about gender in movements and signalling theories or protest (*Lohmann, 1994*,

Asal et al., 2013). According to Pearlman (2013), anger and joy should positively correlate with increased protest, while fear and sadness should negatively correlate. The results in the full model match these expectations, though only sadness is statistically significant.¹⁰ Note as well that emotions provide essential no extra explanatory power for protest size.

Table 2. Violence, Demographics, Emotions, and Change in Protest Size

	Log ₁₀ (Total Faces) _{i,t}		
	(1) Image	(2) Text	(3) Multimodal
$\Delta\text{Protester Violence}_{i,t}$	-.0499*** (.0192)		-.0511*** (.0191)
$\Delta\text{State Violence}_{i,t}$.3216 (.2097)		.3342 (.2085)
$\Delta\text{State Violence}^2_{i,t}$	-.5352 (.3616)		-.5459 (.3596)
$\Delta\text{Fatalities}_{i,t}$	-.0039 (.0055)		-.0036 (.0054)
$\Delta\% \text{Faces Male}_{i,t}$.5350*** (.0616)		.5350*** (.0617)
$\Delta\% \text{Faces Female}_{i,t}$.4897*** (.0336)		.4911*** (.0339)
$\Delta\% \text{Faces Young Adult}_{i,t}$.0578 (.0565)		.0566 (.0565)
$\Delta\text{Tweets Angry}_{i,t}$		-.0722 (.0472)	.0039 (.0328)
$\Delta\text{Tweets Joy}_{i,t}$		-.0776*** (.0239)	.0071 (.0143)
$\Delta\text{Tweets Fear}_{i,t}$		-.0342 (.0410)	-.0169 (.0261)
$\Delta\text{Tweets Sadness}_{i,t}$		-.0032 (.0359)	-.0583* (.0342)
N	1,186	1,186	1,186
Adjusted R ²	.4530	.0008	.4518
Clustered SE	Y	Y	Y

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

All variables are first differenced. Standard errors are clustered by city.

¹⁰We also modeled protest size using only ACLED data. The result for $\Delta\text{Fatalities}_{i,t}$, the only independent variable from ACLED, are the same, though R² is much lower than Models 1 and 3.

5 Discussion

5.1 Future Work

The results presented in this paper constitute a floor on the methodology, for several reasons. First, it uses only 1% of tweets, the quantity Twitter provides for free. To gather more tweets, one could purchase geolocated tweets directly from Twitter or through enhanced access to the Search API. If working within the constraints of the Streaming API, which will provide one-third to one-half of all geolocated tweets (Steinert-Threlkeld, 2018), one can follow specific accounts, including news sources, that are likely to tweet about protests; if followed in real time via the Streaming API, all of an account’s tweets are provided. Improving the recall of the protest image classifier can theoretically almost quintuple the number of protest images available.

Future work could also improve geolocation. In Venezuela, for example, we discarded about half of the protest images because Twitter only provided their location at the country level. One can infer an account’s location, and assign it to all tweets from that account, using the user biography, self-reported location, or locations mentioned in the tweet itself (Sloan et al., 2013, McCormick et al., 2017). Supplementing the location Twitter provides with an inferred location will extend the applicability of this paper’s methodology to countries, such as Ukraine and Gabon, on the precipice of inclusion.

In addition, this paper only measures two kinds of events, protest and violent repression. Future work should focus on expanding the event types an image dataset records. The CAMEO coding scheme, the dominant event and actor ontology, records twenty categories of action (NAVCO 3.0 adds a 21st, defections) (Gerner et al., 2002, Chenoweth, Pinckney and Lewis, 2018). Many of those actions, especially the verbal ones such as offering assistance, are probably not available via images. It may be feasible, however, to generate events around speeches or military activity, as governments increasingly communicate on social media with text and image, and text from government accounts could be interpreted as actions falling into one of CAMEO’s categories.

This paper’s methodology still relies on general knowledge generated from mass media to know where and when to search for protests. This task is easier than that facing datasets such as MMAD

and ACLED which parse media to extract protest events. The equivalent approach in this paper’s methodology would be to download images from every tweet and detect if the image contains protest. We considered this approach but abandoned it because of hardware restrictions. We do not have enough bandwidth to download that many images, which would require approximately 44 gigabytes of storage per day.¹¹ *De novo* protest detection equivalent to others’ approaches therefore requires much greater financial support than we currently have, though it is technically feasible.

Future research should also explore creating multimodal protest event data from geolocated social media content using a hybrid approach. A protest image classifier can screen all images for those likely to contain a protest, and teams of trained human coders can manually annotate them. This approach will work best for the demographic and size variables. Depending on resources available, the hybrid approach may limit the creation of panel data.

Constructing a large-scale event dataset with this paper’s approach requires careful consideration of how to handle missing dates. Dates without records could represent dates without protest or dates with protest but without enough protest images to appear in a dataset. The coarser the geographic resolution the more likely are missing dates to be true zeroes. If image data record more true protests than text data, then a higher proportion of these missing dates should be treated as true zeroes, not as measurement error (Earl et al., 2004). Future research should work on identifying this boundary.

5.2 Fundamental Limits to Event Data

No approach to event data will circumvent a fundamental problem of any study whose units vary substantially on income, population, or both: there is less data from poorer, sparser parts of the world. This gradient means, for example, that protests in Namibia are less likely to receive coverage than in South Africa, riots in Paraguay less likely to enter a dataset than those in Brazil. Inferences made are therefore inferences about places of the world which produce legible data, and it is unclear if they have validity in less represented locations.

¹¹Roughly 10% of tweets contain images. Five million tweets per day are available for free, and each image requires about 88 kilobytes of storage. $5000000 * .1 * .000088 = 44$ gigabytes.

Nam (2006) shows that text based event datasets are substantially improved when they incorporate large amounts of local sources, a point also made in Herkenrath and Knoll (2011). Researchers do try to incorporate as many local news sources as possible, but an intrinsic tension exists between maximizing the number of local sources and maximizing geographic and temporal coverage. Machine-coded and hybrid approaches to text datasets are attempts to have both this depth and breadth, but they are still limited by their reliance on text. ACLED, as a result of a decade of partnerships and funding, is the best event dataset at incorporating local sources; it even occasionally sources events from Twitter and Facebook (Dowd et al., 2020).

One way to think of the potential of geocoded social media content as event data is that each item approximates a very local news record. Though text are easier to produce, when they do not come from major source it is harder to rigorously process them for inclusion in event datasets. Image processing, on the other hand, scales more easily because visual language is closer to a *lingua franca* than spoken language (Graber, 1996, Gazzaniga, 1998). Because the barrier to entry to publishing geocoded social media images is lower than for publishing a newspaper article, the barrier to entry for events in image datasets may be lower as well.

Event data from geolocated text and images is currently limited by the penetration of Twitter. But as the production and sharing of images by individuals proliferates, the viability of this paper's methodology increases. Since the preponderance of resources to date has been devoted to generating event data from text, future endeavors should also consider incorporating geolocated social media content.

References

- Acemoglu, Daron, Ahmed Tahoun and Tarek A. Hassan. 2018. "The Power of the Street: Evidence from Egypt's Arab Spring." *The Society for Financial Studies* 31(1):1–42.
- 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.
- Azar, Edward E. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24(1):143–152.
- Baum, Matthew A and Yuri M Zhukov. 2015. "Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war." *Journal of Peace Research* 52(3):384–400.
URL: <http://journals.sagepub.com/doi/10.1177/0022343314554791>
- Beals, Gregory. 2017. "Spanish Police Violence in Barcelona Inflames Catalonia and Shocks Europe."
- Bergesen, Albert and Max Herman. 1998. "Immigration, Race, and Riot: The 1992 Los Angeles Uprising." *American Sociological Review* 63(1):39–54.
- Berzon, Marsha S. and J. Clifford Wallage. 2019. "hiQ Labs, Inc. v. LinkedIn Corporation."
- 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>
- Bourlai, Elli and Susan C. Herring. 2014. Multimodal Communication on Tumblr: "I Have So Many Feels!". In *Proceedings of the 2014 ACM Conference on Web Science*. pp. 171–175.
- 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.

- Brandt, Patrick T., Muhammed Y. Idris, Maryam Imani, John Beielser, Christopher Boylan, Ben S. Fisher and Andy Halterman. 2018. “Open Event Data Alliance.”
URL: <https://github.com/openeventdata>
- Cantu, Francisco. 2019. “The Fingerprints of Fraud: Evidence From Mexico’s 1988 Presidential Election.” *American Political Science Review* 113(3):710–726.
- Casas, Andreu and Nora Webb Williams. 2019. “Images That Matter: Online Protests and the Mobilizing Role of Pictures.” *Political Research Quarterly* 72(2):360–375.
- Chen, Baiyu, Sergio Escalera, Isabelle Guyon, Víctor Ponce-López, Nihar Shah and Marc Oliu Simón. 2016. Overcoming calibration problems in pattern labeling with pairwise ratings: application to personality traits. In *European Conference on Computer Vision*. Springer pp. 419–432.
- Chenoweth, Erica and Jeremy Pressman. 2017. “Crowd Counting Consortium.”
URL: <https://sites.google.com/view/crowdcountingconsortium/home>
- Chenoweth, Erica, Jonathan Pinckney and Orion Lewis. 2018. “Days of rage : Introducing the NAVCO 3.0 Dataset.” *Journal of Peace Research* 55(4):524–534.
- Chenoweth, Erica and Kurt Schock. 2015. “Do Contemporaneous Armed Challenges Affect the Outcomes of Mass Nonviolent Campaigns?” *Mobilization: An International Quarterly* 20(4):427–451.
- Chenoweth, Erica and Maria J. Stephan. 2011. *Why Civil Resistance Works*. New York City: Columbia University Press.
- Clark, David H. and Patrick M. Regan. 2018. “Mass Mobilization Protest Data.”
URL: <https://www.binghamton.edu/massmobilization/about.html>
- Croicu, Mihai and Nils B Weidmann. 2015. “Improving the selection of news reports for event coding using ensemble classification.” *Research and Politics* 2(4):2053168015615596.

- Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10(1):1–23.
- Dowd, Caitriona, Patricia Justino, Roudabeh Kishi and Gauthier Marchais. 2020. "Comparing 'New' and 'Old' Media for Violence Monitoring and Crisis Response: Evidence from Kenya." *Research and Politics* 7(3).
- Earl, Jennifer, Andrew Martin, John D. McCarthy and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30:65–80.
- Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*. pp. 1615–1625.
- Ferrara, Federico. 2003. "Why Regimes Create Disorder: Hobbes's Dilemma During a Rangoon Summer." *Journal of Conflict Resolution* 47(3):302–325.
- Fisher, Dana R., Dawn M. Dow and Rashawn Ray. 2017. "Intersectionality takes it to the streets : Mobilizing across diverse interests for the Women' s March." *Science Advances* 3:1–8.
- Forelle, Michelle C, Philip N. Howard, Andres Monroy-Hernandez and Saiph Savage. 2015. "Political Bots and the Manipulation of Public Opinion in Venezuela."
URL: <https://arxiv.org/abs/1507.07109>
- Franceschi-Bicchierai, Lorenzo. 2014. "In Venezuela, the Only Free Media Is Twitter: How the Government Uses Twitter."
URL: <https://mashable.com/2014/02/28/venezuela-twitter/>
- Francisco, Ronald A. 2004. "After the Massacre: Mobilization in the Wake of Harsh Repression." *Mobilization: An International Journal* 9(2):107–126.
- Gazzaniga, Michael S. 1998. *The Mind's Past*. University of California Press.

- Gerner, Deborah J. and Philip A. Schrodt. 1998. The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East. In *Early Warning and Early Response*, ed. Susanne Schmeidl and Howard Adelman. New York City: Columbia University Press.
- Gerner, Deborah J., Philip A. Schrodt, Rajaa Abu-Jabr and Omur Yilmaz. 2002. Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions. In *Annual Meeting of the International Studies Association*.
- Gillion, Daniel Q. 2012. "Protest and congressional behavior: Assessing Racial and Ethnic Minority Protests in the District." *Journal of Politics* 74(4):950–962.
- Gleditsch, Kristian Skrede, Nils W Metternich and Andrea Ruggeri. 2014. "Data and progress in peace and conflict research." *Journal of Peace Research* 51(2):301–314.
- González, Felipe. 2020. "Collective action in networks: Evidence from the Chilean student movement." *Journal of Public Economics* 188:104220.
URL: <https://doi.org/10.1016/j.jpubeco.2020.104220>
- Graber, Doris A. 1996. "Say It with Pictures." *The ANNALS of the American Academy of Political and Social Science* 546(1):85–96.
- Halterman, Andrew. 2019. Geolocating Political Events in Text. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*. Minneapolis: pp. 29–39.
- Halterman, Andrew, Benjamin S. Fisher, Christopher Boylan, John Beielser, Muhammed Y. Idris and Philip A. Schrodt. 2017. "Open Event Data Alliance."
URL: <https://github.com/orgs/openeventdata/people>
- Hanna, Alex. 2017. "MPEDS : Automating the Generation of Protest Event Data." *Working Paper* pp. 1–40.
URL: <https://osf.io/preprints/socarxiv/xuqmv>

- Hayes, Matthew and Peter F. Nardulli. 2011. SPEED's Societal Stability Protocol and the Study of Civil Strife: an Overview and Comparison with Similar Event Data Projects. Technical report Cline Center for Democracy.
- 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.
- Heaney, Michael T. and Fabio Rojas. 2008. “Coalition Dissolution, Mobilization, and Network Dynamics in the U.S. Antiwar Movement.” *Research in Social Movements, Conflicts, and Change* 28(08):39–82.
- Heller, Patrick and Peter Evans. 2010. “Taking Tilly south: durable inequalities, democratic contestation, and citizenship in the Southern Metropolis.” *Theory and Society* 39(3-4):433–450.
- Hellmeier, Sebastian, Nils B. Weidmann and Espen Geelmuyden Rød. 2018. “In The Spotlight: Analyzing Sequential Attention Effects in Protest Reporting.” *Political Communication* 00(00):1–25.
- Herkenrath, Mark and Alex Knoll. 2011. “Protest events in international press coverage: An empirical critique of cross-national conflict databases.” *International Journal of Comparative Sociology* 52(3):163–180.
- Hollander, Ethan J and Chonghyun Christie Byun. 2015. “Explaining the Intensity of the Arab Spring.” *Digest of Middle East Studies* 24(1):26–46.
- Jasper, James M. 2011. “Emotions and Social Movements: Twenty Years of Theory and Research.” *Annual Review of Sociology* 37(1):285–303.
- Jockers, Matthew L. 2015. “Syuzhet: Extract Sentiment and Plot Arcs from Text.”
URL: <https://github.com/mjockers/syuzhet>

- Joo, Jungseock and Zachary C. Steinert-Threlkeld. 2019. "Image as Data: Automated Visual Content Analysis for Political Science." *Working paper* pp. 1–42.
URL: <https://arxiv.org/abs/1810.01544>
- Kaiman, Jonathan. 2014. "Who guides Hong Kong's 'Umbrella Revolution' pro-democracy movement?".
- Kaneko, Takamu and Keiji Yanai. 2013. "Visual Event Mining From Geo-Tweet Photos." *IEEE International Conference On Multimedia and Expo Workshops* .
- Karkkainen, Kimmo and Jungseock Joo. 2019. "FairFace: Face Attribute Dataset for Balanced Race.". **URL:** <https://arxiv.org/abs/1908.04913>
- Kern, Holger Lutz. 2011. "Foreign Media and Protest Diffusion in Authoritarian Regimes: The Case of the 1989 East German Revolution." *Comparative Political Studies* 44(9):1179–1205.
- Khawaja, Marwan. 1993. "Repression and Popular Collective Action: Evidence from the West Bank." *Sociological Forum* 8(1):47–71.
- LeCun, Yann, Yoshua Bengio and Geoffrey Hinton. 2015. "Deep learning." *nature* 521(7553):436–444.
- Lee, Sophie J., Howard Liu and Michael D. Ward. 2018. "Lost in Space: Geolocation in Event Data." *Political Science Research and Methods* pp. 1–18.
- Leetaru, Kalev and Philip Schrodt. 2013. GDELT: Global Data on Events, Language, and Tone, 1979-2012. In *International Studies Association*. International Studies Association San Diego, CA: .
- Leopold, Til Alexander. 2017. *The Global Gender Gap Report*. World Economic Forum.

- Liang, Yan, Khaled Jabr, Christan Grant, Jill A. Irvine and Andrew Halterman. 2018. New Techniques for Coding Political Events Across Languages. In *IEEE 18th International Conference on Reuse and Integration*. Salt Lake City: .
- Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91." *World Politics* 47(1):42–101.
- 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.
- Masad, David. 2013. "Studying the Syrian Civil War with GDELT." *Working paper* .
- McAdam, Doug. 1986. "Recruitment to High-Risk Activism: The Case of Freedom Summer." *American Journal of Sociology* 92(1):64–90.
- McAdam, Doug, John McCarthy, Susan Olzak and Sarah Soule. 2009. "Dynamics of Collective Action Dataset.".
- URL:** <https://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/>
- McAdam, Doug and Yang Su. 2002. "The War at Home: Antiwar Protests and Congressional Voting, 1965 to 1973." *American Sociological Review* 67(5):696–721.
- McClelland, Charles. 1984. "World Event/Interaction Survey (WEIS) Project, 1966-1978.".
- URL:** <https://doi.org/10.3886/ICPSR05211.v3>
- McCormick, T. H., H. Lee, N. Cesare, A. Shojaie and E. S. Spiro. 2017. "Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing." *Sociological Methods & Research* 46(3):390–421.
- McPhail, Clark and John McCarthy. 2004. "Who Counts and How: Estimating the Size of Protests." *Contexts* 3(3):12–18.

- Mitts, Tamar, Gregoire Phillips and Barbara F. Walter. 2020. "Studying the Impact of ISIS Propaganda Campaigns." *Working paper* .
- URL:** <https://www.dropbox.com/s/dxbliwko3jygu/propaganda.pdf?dl=0>
- Mocanu, Delia, Andrea Baronchelli, Nicola Perra, Alessandro Vespignani, Bruno Goncalves and Qian Zhang. 2013. "The Twitter of Babel: Mapping World Languages through Microblogging Platforms." *PLOS One* 8(4):e61981.
- Mohammad, Saif M. and Peter D. Turney. 2013. "Crowdsourcing a word-emotion association lexicon." *Computational Intelligence* 29(3):436–465.
- Morstatter, Fred, Jurgen Pfeffer, Kathleen M. Carley and Huan Liu. 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *Association for the Advancement of Artificial Intelligence*.
- Muchlinski, David, Xiao Yang, Sarah Birch, Craig MacDonald and Iadh Ounis. Forthcoming. "We Need to Go Deeper: Measuring Electoral Violence using Convolutional Neural Networks and Social Media." *Political Science Research and Methods* pp. 1–27.
- Munger, Kevin, Richard Bonneau, Jonathan Nagler and Joshua A Tucker. 2019. "Elites Tweet to Get Feet Off the Streets: Measuring Regime Social Media Strategies During Protest." *Political Science Research and Methods* 7(4):815–834.
- Myers, Daniel J. and Beth Schaefer Caniglia. 2004. "All the Rioting That's Fit to Print: Selection Effects in National Newspaper Coverage of Civil Disorders, 1968-1969." *American Sociological Review* 69:519–543.
- Nam, Taehyun. 2006. "What You Use Matters: Coding Protest Data." *PS: Political Science & Politics* 39(2):281–287.
- Nordås, Ragnhild and Christian Davenport. 2013. "Fight the Youth: Youth Bulges and State Repression." *American Journal of Political Science* 57(4):926–940.

- Oliver, Pamela, Gerald Marwell and Ruy Teixeira. 1985. "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91(3):522–556.
- Olzak, Susan, Suzanne Shanahan and Elizabeth H. McEneaney. 1996. "Poverty, Segregation, and Race Riots: 1960 to 1993." *American Sociological Review* 61(4):590–613.
- Page, Tiffany Linton. 2015. Race, Ethnicity, and Politics in Venezuela. In *The International Handbook of the Demography of Race and Ethnicity*, ed. Rogelio Saenz, David G. Emrick and Nestor P. Rodriguez. Vol. 4 chapter 6, pp. 111–128.
- Patilla, La. 2014. "https://www.lapatilla.com/2014/11/06/profesores-protestaron-cerquita-de-miraflores-fotos/".
- URL:** <https://www.lapatilla.com/2014/11/06/profesores-protestaron-cerquita-de-miraflores-fotos/>
- Patilla, La. 2015. "En 2014 se registraron 9.286 protestas, cifra inédita en Venezuela."
- URL:** <https://www.lapatilla.com/2015/01/19/en-2014-se-registraron-9-286-protestas-cifra-inedita-en-venezuela/>
- Patillia, La. 2015b. "#MaduroRenuncia: La consigna de la protesta merideña de este miércoles."
- URL:** <https://www.lapatilla.com/2015/02/04/madurorenuncia-la-consigna-de-la-protesta-meridena-de-este-miercoles-fotos/>
- Pearlman, Wendy. 2013. "Emotions and the Microfoundations of the Arab Uprisings." *Perspectives on Politics* 11(02):387–409.
- 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.

- Rasler, Karen. 1996. "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociological Review* 61(1):132–152.
- 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.
- Robnett, Belinda. 1996. "African-American Women in the Civil Rights Movement, 1954-1965: Gender, Leadership, and Micromobilization." *American Journal of Sociology* 101(6):1661–1693.
- 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.
- Scarritt, James R., Susan M. McMillan and Shaheen Mozaffar. 2001. "The Interaction Between Democracy and Ethnopolitical Protest and Rebellion in Africa." *Comparative Political Studies* 34(7):800–827.
- Schrodt, Philip A. and Deborah J. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-1992." *American Journal of Political Science* 38(3):825–854.
- Senekal, Burgert A, Walter Daelemans, Free State, South Africa, Sign Language, Deaf Studies and South Africa. 2020. "Automatic classification of social media reports on violent incidents in South Africa using machine learning." *South African Journal of Science* 116(3):4–11.
- 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.
- 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.
- Sloan, Luke, Jeffrey Morgan, William Housley, Matthew Williams, Adam Edwards, Pete Burnap and Omer Rana. 2013. “Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter.” *Sociological Research Online* 18(3):1–15.
- Smith, Aaron, Monkia Anderson and Tom Caiazza. 2018. Social Media Use in 2018. Technical report Pew Research Center.
- Sobolev, Anton, Jungseock Joo, Keith Chen and Zachary C. Steinert-Threlkeld. 2020. “News and Geolocated Social Media Accurately Measure Protest Size.” *American Political Science Review* pp. 1–9.
- Solaimani, Mohiuddin, Rajeevardhan Gopalan, Latifur Khan, Patrick T Brandt and Bhavani Thuraisingham. 2016. Spark-based Political Event Coding. In *IEEE Second International Conference on Big Data Computing Service and Applications*.
- Soleymani, Mohammad, David Garcia, Brendan Jou, Björn Schuller, Shih-fu Chang and Maja Pantic. 2017. “A survey of multimodal sentiment analysis.” *Image and Vision Computing* 65:3–14.
- Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. Cambridge University Press.
- Sundberg, Ralph and Erik Melander. 2013. “Introducing the UCDP Georeferenced Event Dataset.” *Journal of Peace Research* 50(4):523–532.
- Tufekci, Zeynep and Christopher Wilson. 2012. “Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square.” *Journal of Communication* 62(2):363–379.

- Uba, Katrin. 2005. "Political Protest and Policy Change: The Direct Impacts of Indian Anti-Privatization Mobilizations, 1990-2003." *Mobilization: An International Journal* 10(3):1990–2003.
- Valkanias, George, Ioannis Katakis, Dimitrios Gunopulos and Antony Stefanidis. 2014. Mining Twitter Data with Resource Constraints. In *2014 International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. Ieee pp. 157–164.
- Vogt, Manuel, Nils Christian Bormann, Seraina Rüegger, Lars Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59(7):1327–1342.
- Walgrave, Stefaan and Rens Vliegthart. 2012. "The Complex Agenda-Setting Power of Protest: Demonstrations, Media, Parliament, Government, and Legislations in Belgium, 1993-2000." *Mobilization: An International Journal* 17(2):129–156.
- Wang, Wei, Ryan Kennedy, David Lazer and Naren Ramakrishnan. 2016. "Growing pains for global monitoring of societal events: Automated event coding raises promise and concerns." *Science* 353(6307):1502–1504.
- 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. *The Internet and Political Protest in Autocracies*. Oxford University Press.
- 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*. Vol. 3794 Cambridge University Press.
- Wouters, Ruud and Stefaan Walgrave. 2017. "Demonstrating Power: How Protest Persuades Political Representatives." *American Sociological Review* 82(2):361–383.

- Xi, Nan, Di Ma, Marcus Liou, Zachary C. Steinert-Threlkeld, Jason Anastasopoulos and Jungseock Joo. 2020. Understanding the Political Ideology of Legislators from Social Media Images. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.
- Xu, Jiejun, Tsai-Ching Lu, Ryan Compton and David Allen. 2014. Civil Unrest Prediction: A Tumblr-Based Exploration. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, ed. William G. Kennedy, Nitin Agarwal and Shanchieh Jay Yang. Vol. 8393 LNCS Springer International Publishing pp. 403–411.
- Yip, P, R Watson, K Chan, E Lau, F Chen, Y Xu, L Xi, D Cheung, B Ip and D Liu. 2010. “Estimation of the Number of People in a Demonstration.” *Australian & New Zealand Journal of Statistics* 52(1):17–26.
- 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.
- Zhang, Quanshi, Ying Nian Wu and Song-Chun Zhu. 2018. Interpretable convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 8827–8836.
- Zhukov, Yuri M. 2018. “Repression Works (just not in moderation).” *Working paper*.

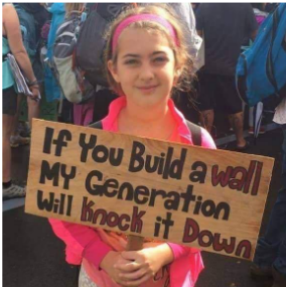




Appendix

S1 Methods Details

S1.1 Coding Interface

Figure A1 provides examples of our Amazon Mechanical Turk 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 were judged to contain a protest. In the second task, annotators label which of two images contain more violence; they then indicate if the violence is from protesters or the state.

Fig. A1. Examples of Our Annotation Interface (in Amazon Mechanical Turk)

 <p>Q1. Does this image contain a scene of protest?</p> <p><input checked="" type="radio"/> Protest <input type="radio"/> Uncertain <input type="radio"/> Not Protest</p>	 <p>Q2. Does this image contain a scene of protest?</p> <p><input type="radio"/> Protest <input type="radio"/> Uncertain <input checked="" type="radio"/> Not Protest</p>	 <p>Q3. Does this image contain a scene of protest?</p> <p><input type="radio"/> Protest <input type="radio"/> Uncertain <input checked="" type="radio"/> Not Protest</p>
<p>Q0. Choose the image that you feel is more violent.</p> <div><p><input checked="" type="radio"/> Image 1</p></div> <div><p><input type="radio"/> Similar</p></div> <div><p><input type="radio"/> Image 2</p></div>		

S1.2 Inter-coder Reliability

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 A1 shows the estimated reliability statistics.

Table A1. Inter-coder reliability

Label	Kappa
Perceived Protester Violence	.566
Perceived State Violence	.473

S1.3 Accuracy

Figure A2 shows the precision-recall curve for the protest classifier. We set the binary threshold at the label value that maximizes recall with 85% precision.

Figure A3 shows the receiver-operating characteristic curve for the protest, state violence, and protester violence labels.

Fig. A2. High Precision, Low Recall for Protest

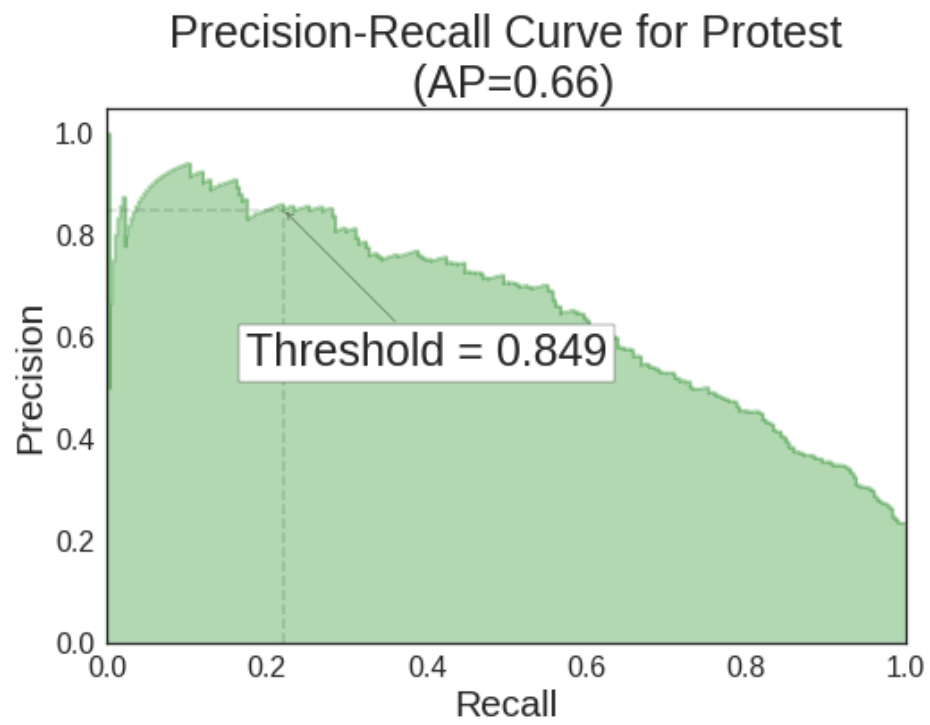
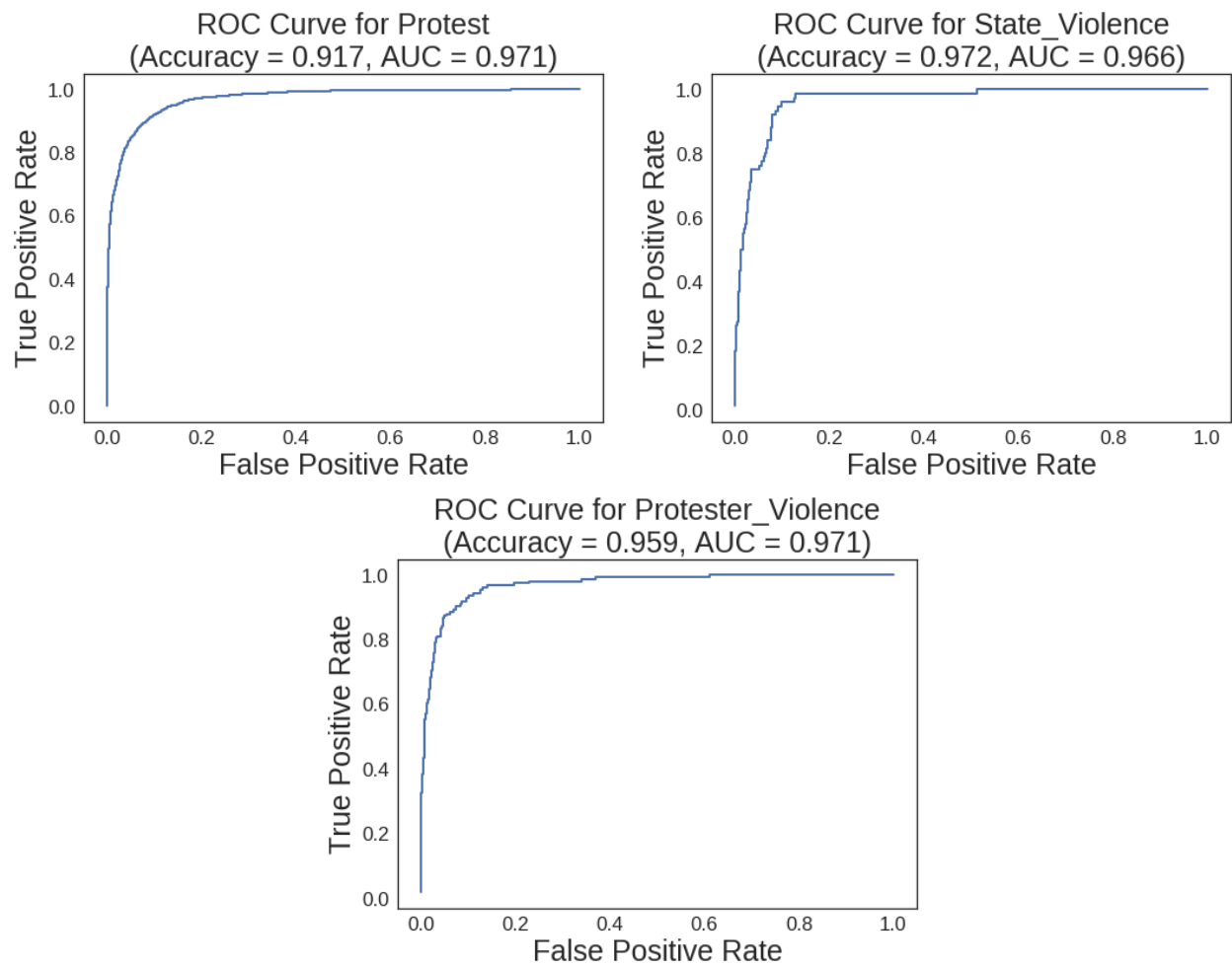


Fig. A3. Classifier Performance



S2 Protest Periods to Determine Limits

Table A2 shows the nineteen countries to which we applied the image methodology. **Bold rows** are those countries where at least one city contains a protest photo for at least $\frac{1}{7}$ of its period.

Table A2. Protest Periods

Country	Start	End	Issue	Protest Images/Day
Armenia	03.01.2018	05.31.2018	Anti-incumbency	.147
Belarus	02.18.2017	05.02.2017	Unemployment tax	2.18
Burundi	04.01.2015	12.01.2015	Elections	.06
Chile	10.01.2019	12.31.2019	Inequality	45.2
Cameroon	11.01.2016	12.01.2017	Bilingualism	.06
Egypt	06.01.2017	06.31.2017	Islands to Saudi Arabia	5.38
Gabon	08.20.2016	09.27.2016	Elections	.235
Hong Kong	2014.09.18	2014.12.23	China reforms	5.82
Hong Kong	03.01.2019	12.31.2019	China reforms	2.07
Iraq	10.01.2019	12.31.2019	Incompetence	6.63
Lebanon	10.01.2019	12.31.2019	Anti-tax	7.81
Pakistan	11.01.2017	11.30.2017	Blasphemy protests	.941
Russia	03.12.2017	04.26.2017	Corruption	19.3
Catalonia, Spain	2017.09.01	2017.12.31	Secession	31.8
South Korea	2016.10.20	2017.03.14	Anti-incumbency	8.04
Togo	08.01.2017	12.01.2017	Anti-incumbency	.23
Ukraine	11.21.2013	03.21.2014	European Integration	3.32
Venezuela	2014.11.01	2015.02.10	Grievances	31.20
Venezuela	12.29.2016	12.17.2017	Anti-Maduro	16.40

Bold rows are those countries where at least one city contains a protest photo for at least $\frac{1}{7}$ of its period.

Table A3 provides the list of countries in which this methodology should identify at least one protest per week.

Table A3. Countries Sorted by Applicability of Method

	Method Applicable	More Data Needed
1	American Samoa	Afghanistan
2	Andorra	Albania
3	Antigua and Barbuda	Algeria
4	Argentina	Angola
5	Aruba	Azerbaijan
6	Australia	Belize
7	Austria	Benin
8	Bahamas	Bhutan
9	Bahrain	Bolivia

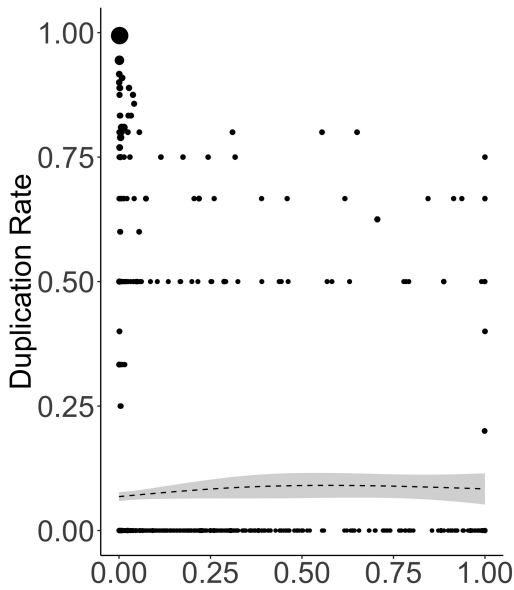
10	Bangladesh	Bosnia and Herzegovina
11	Barbados	Burkina Faso
12	Belgium	Cambodia
13	Bermuda	Cape Verde
14	Botswana	Central African Republic
15	Brazil	Chad
16	Brunei Darussalam	Comoros
17	Bulgaria	Côte d'Ivoire
18	Canada	Democratic Republic of the Congo
19	Cayman Islands	Djibouti
20	China	Dominica
21	Colombia	Ecuador
22	Costa Rica	El Salvador
23	Croatia	Federated States of Micronesia
24	Cuba	Fiji
25	Cyprus	Georgia
26	Czech Republic	Ghana
27	Denmark	Guatemala
28	Dominican Republic	Guinea
29	Egypt	Guinea-Bissau
30	Equatorial Guinea	Guyana
31	Estonia	Haiti
32	Ethiopia	Honduras
33	Finland	Jamaica
34	France	Jordan
35	French Polynesia	Kenya
36	Germany	Kiribati
37	Greece	Kosovo
38	Greenland	Kyrgyzstan
39	Grenada	Lao PDR
40	Guam	Lesotho
41	Hungary	Liberia
42	Iceland	Libya
43	India	Macedonia
44	Indonesia	Madagascar
45	Iran	Malawi
46	Ireland	Mali
47	Israel	Marshall Islands
48	Italy	Mauritania
49	Japan	Moldova
50	Kazakhstan	Mongolia
51	Kuwait	Morocco
52	Latvia	Mozambique
53	Lithuania	Myanmar
54	Luxembourg	Namibia
55	Macao	Nepal
56	Malaysia	Nicaragua
57	Maldives	Niger
58	Malta	Papua New Guinea
59	Mauritius	Paraguay
60	Mexico	Republic of Congo
61	Monaco	Rwanda
62	Montenegro	Samoa
63	Nauru	São Tomé and Príncipe
64	Netherlands	Senegal
65	New Zealand	Serbia
66	Nigeria	Sierra Leone

67	Northern Mariana Islands	Solomon Islands
68	Norway	Somalia
69	Oman	South Sudan
70	Palau	Sri Lanka
71	Panama	Sudan
72	Peru	Suriname
73	Philippines	Tajikistan
74	Poland	Tanzania
75	Portugal	The Gambia
76	Puerto Rico	Timor-Leste
77	Qatar	Tonga
78	Romania	Tunisia
79	Saint Kitts and Nevis	Turkmenistan
80	San Marino	Uganda
81	Saudi Arabia	Uzbekistan
82	Seychelles	Vanuatu
83	Singapore	Yemen
84	Slovakia	Zambia
85	Slovenia	Zimbabwe
86	South Africa	
87	Sweden	
88	Switzerland	
89	Thailand	
90	Trinidad and Tobago	
91	Turkey	
92	Turks and Caicos Islands	
93	United Arab Emirates	
94	United Kingdom	
95	United States	
96	United States Virgin Islands	
97	Uruguay	
98	Vietnam	
From training data		
99	Chile	Armenia
100	Hong Kong	Belarus
101	Iraq	Burundi
102	Lebanon	Cameroon
103	Pakistan	Egypt
104	Russia	Gabon
105	Spain	Togo
106	South Korea	Ukraine
107	Venezuela	

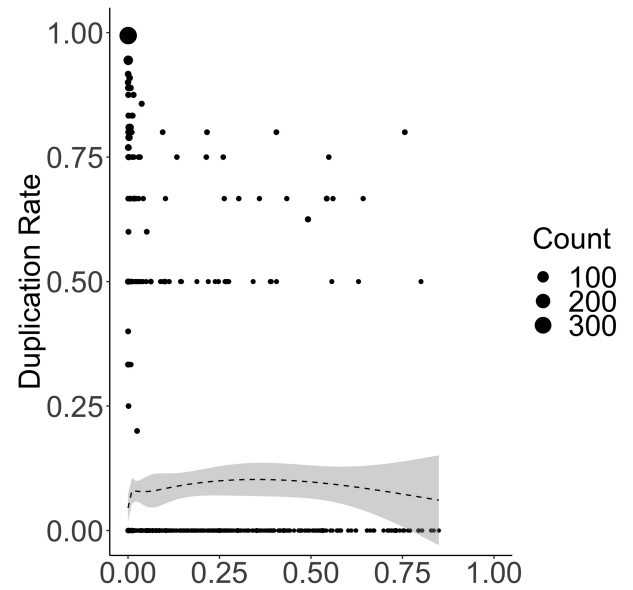
S3 Venezuela Duplication

Figure A4 shows that, in Venezuela, there is not a relationship between violence, from protesters or the state, and which images are shared more. In other words, while newspapers have a preference for covering violent events (Myers and Caniglia, 2004, Hellmeier, Weidmann and Geelmuyden Rød, 2018) and protesters may want to make a state seem to be behaving more violently than it is (Francisco, 2004, Young, 2019), that logic does not appear to transfer to Venezuela.

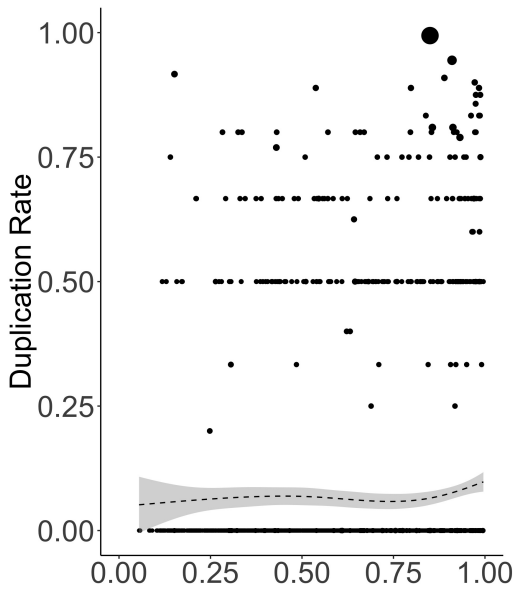
Fig. A4. No Correlation Between Violence and Duplication



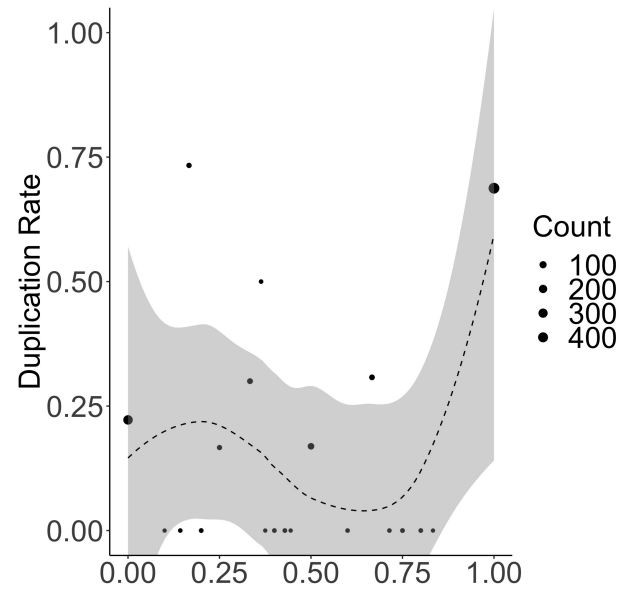
(a) Protester Violence



(b) State Violence



(c) Probability Photo Contains Small Group



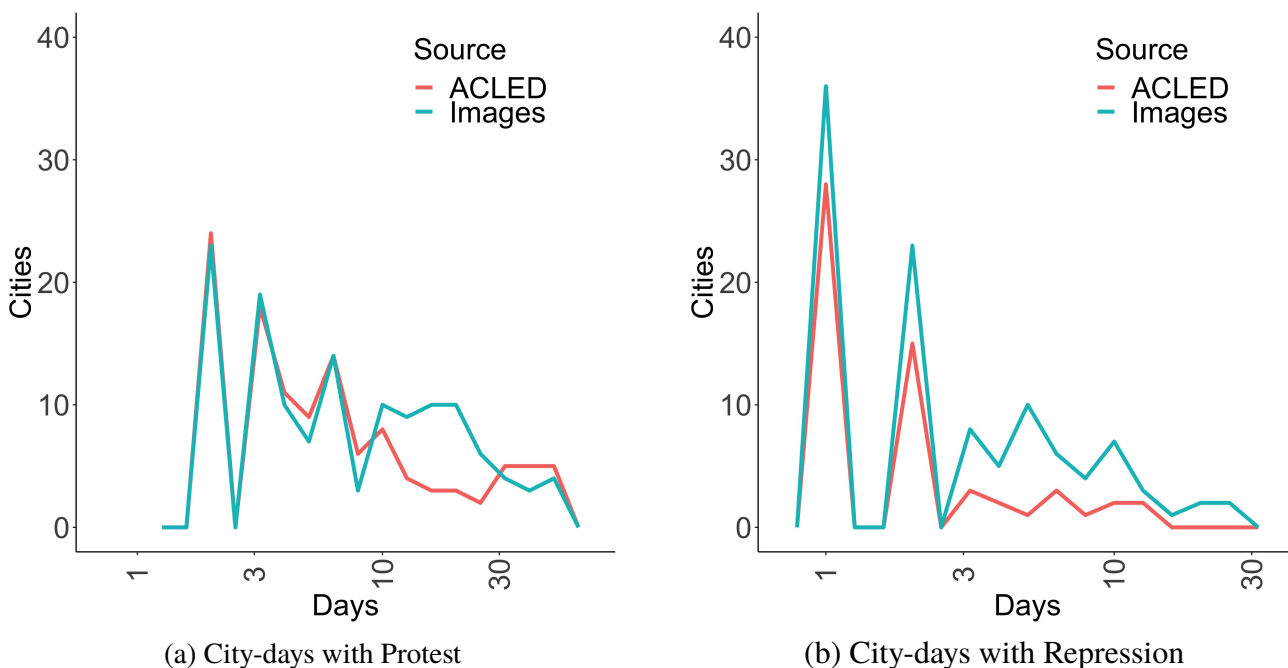
(d) Percent Protesters Female

In fact, we find duplication is driven by photos of small groups (Figure A4c) and women (Figure A4d). The most duplicated photo, 335 of the 1388 duplicates, is of a small group of women protesting deteriorating economic conditions. The second most duplicated photo is a small group of workers in a warehouse protesting wage arrears. Shared 54 times, these two account for 28% of the total number of duplicates.

S4 Images and ACLED in Chile

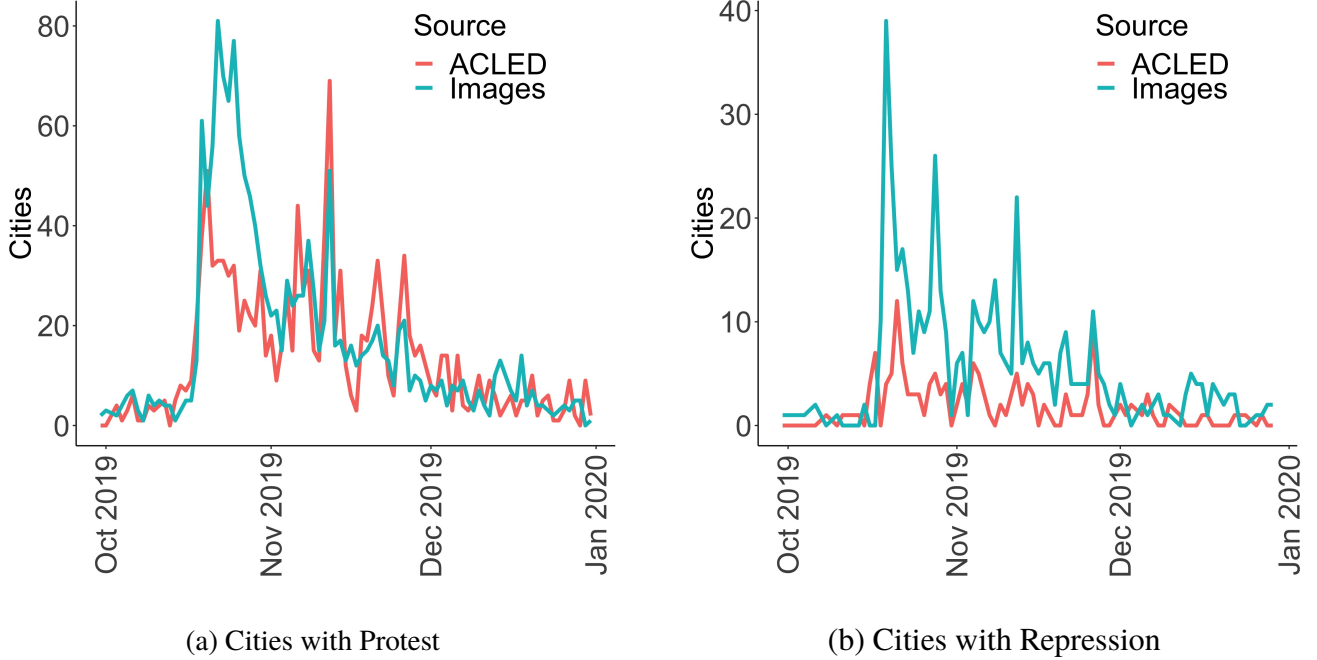
Images and ACLED record very similar trends in protest and repression in Chile, with images recording slightly more cities overall. Figure A5 compares the distribution of city days of protest and repression from ACLED and images. The distribution of each event type is much closer across the two sources than for Venezuela, though images record more repression events than ACLED. Note, however, that ACLED records more cities with 100 days of protest than images do. The time series each records, shown in Figure A6, also largely match, though images record more widespread protest and repression in mid-October.

Fig. A5. Distribution of City-Days, Chile



The two datasets complement each other on what they record. ACLED records 66 types of actors, 51 conducting action against 15. These actors, however, are broad labels, like "Labour

Fig. A6. Trends in Protest and Repression in Chile



Group" or "Civilians". The attribute characteristics that images can generate can then provide information about the gender, age, and race distribution of these actors. Regarding size, 63.65% of the events have sized recorded as "no report", 2.5% are "thousands", 2.24% are "hundreds", and the vast majority of the rest include phrases such as "around", "nearly", or "at least". Only 2.48% contain an exact number, reflecting the difficulty of measuring protest magnitude when relying on text. In contrast, 46.93% of events recognized from images contain countable faces and therefore a size estimate. Regarding violence, ACLED can infer protester violence from the "violent demonstration" or "mob violence" sub-event types and state violence from the "protest with intervention" and "excessive force against protesters" sub-event types or the fatalities variable. Except for fatalities, these variables are nominal, in contrast to the continuously valued estimates of protester and state violence images produce. For episodes in which text and image approaches generate significant record overlap, the strongest approach is to use both types of data, an argument to which Section 4.4 returns.

It is interesting to note that ACLED does not exhibit fatigue bias. The persistence of its records is in contrast to many datasets that rely on major international newspapers and reveals a major benefit

of incorporating local media and sources when possible. ACLED's persistence also suggests that the persistence observed in the Twitter data in Venezuela reflects true protests, not an artifact of the automatic process.