Supplementary Materials for

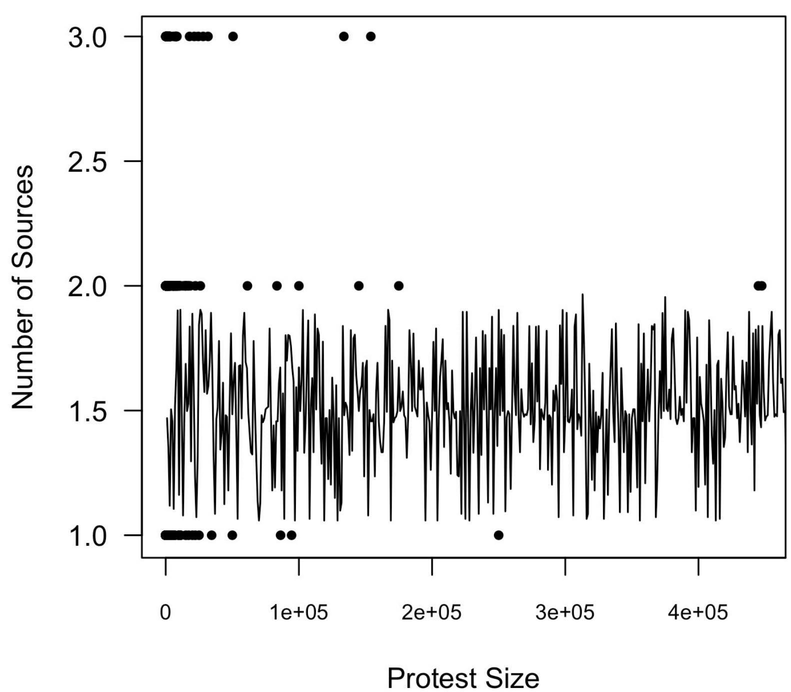
"News and Geolocated Social Media Accurately Measure Protest Size Variation"

S1 How Other Datasets Report Protest Size

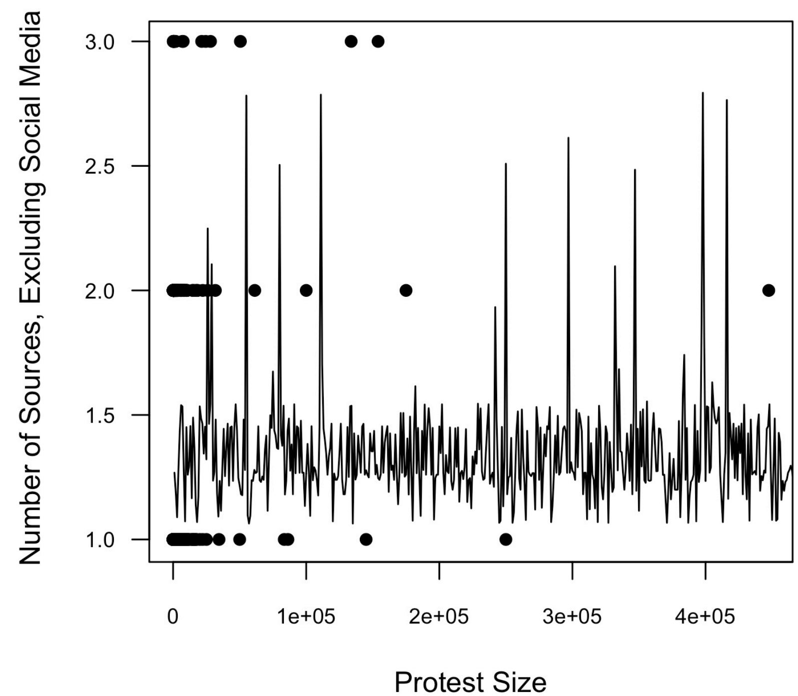
Many datasets report alternate measures of protest size. The Social Conflict Analysis Database (SCAD) transforms newspaper estimates into an ordinal variable on a logged scale (Salehyan et al., 2012). The Armed Conflict Location and Event Data project focuses on fatalities as a variable since those are believed to be easier than people in groups to count, and separate injury reports may be included in each event's "Notes" entry. Since February 2019, it provides reported size in the Notes column as a range, phrase ("scores", "dozens"), or "no report" (the most common response) (Jones, 2019). 72% of ACLED's recorded protests have no reported protest size, though that figure jumps to 98.27% when restricting to 2017. To investigate if a country's media environment affected these numbers, we added censorship and self-censorship scores from the Varieties of Democracy dataset (Coppedge et al., 2019) and kept those countries in the top 10% on both measures (Pakistan and Senegal). *None* of these reports contain protest size estimates. Other datasets, such as Mass Mobilization (Clark and Regan, 2018), Mass Mobilization in Autocracies Database (Weidmann and Rod, 2018), and Nonviolent and Violent Campaigns and Outcomes (Chenoweth, Pinckney and Lewis, 2018) do record the estimates provided in newspapers. Nonetheless, we are aware of no research that has shown that size data from newspapers can be used without introducing bias, though averaging estimates of event size from multiple reports may introduce bias (Cook and Weidmann, 2019). In addition to the event level version of MMAD, another version exists that records the raw reports that feed into the event dataset, allowing the interested researcher to check the veracity of media accounts. Work which takes protest size as its outcome instead has used the count of protests (Steinert-Threlkeld, 2017).

Figure A1 shows that the Crowd Counting Consortium records no relationship between the reported size of a protest and how many sources exist about the protest. This result suggests that multiple reports of protest size is a problem for all protests, not just large ones.

Figure A1: CCC Records No Relationship Between Protest Size and Number of Sources Reporting



(a) All Sources



(b) Excluding Social Media

S2 Cellphone Data

This section explains the cellphone data in greater detail, demonstrates how geohashes work, and validates its political representativeness.

SafeGraph, the cellphone data vendor, aggregates location pings from a collection of third-party smartphone applications ("apps") whose users have agreed to share their location data. Safegraph links these data across applications, resulting in a more complete time-series than relying on one app. (It does not disclose which apps provide it data.) A location ping is generated every time a smartphones operating system calls its location services. It enters SafeGraph's data, and therefore this paper, if a corresponding application is in active use or open in the background. Under normal circumstances, a phone pings approximately every ten minutes, which decreases to about every five seconds when greater location accuracy is needed, e.g. when driving. Low power modes can reduce the frequency of pinging. If a user closes an app from which SafeGraph buys location data, subsequent pings will not be in our data.

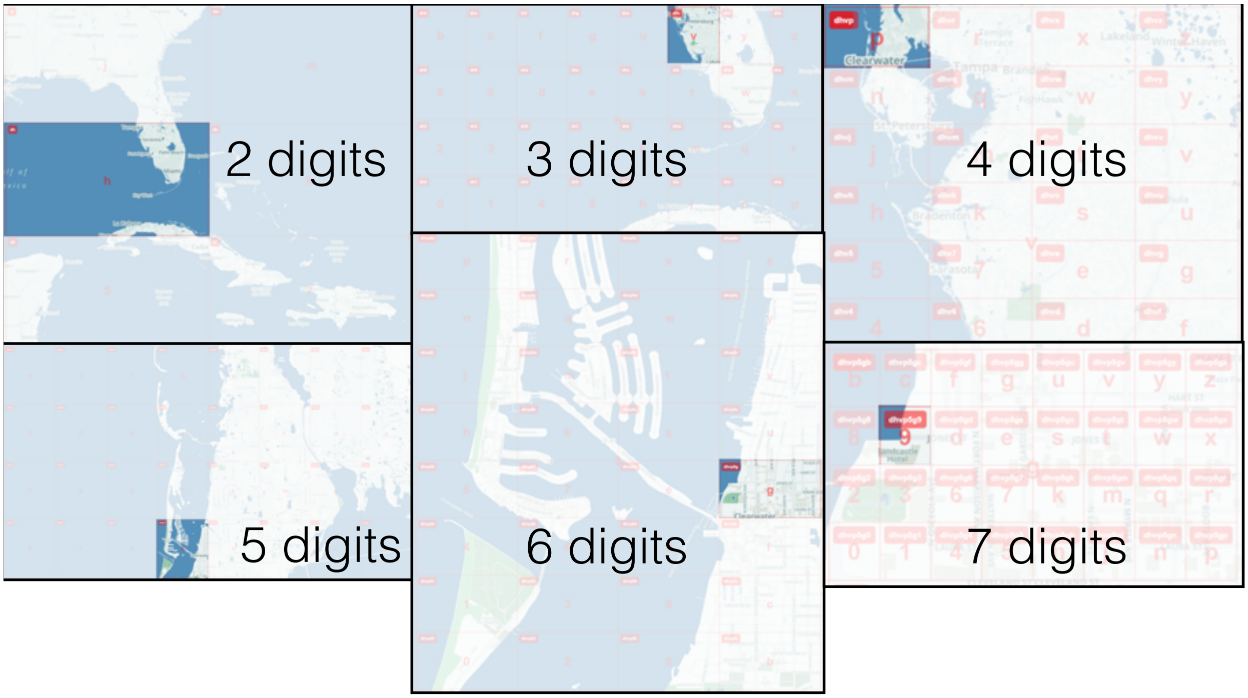
It anonymizes the data before providing them to clients or researchers.

Figure A2 uses Florida to show how location geohashes operate. The world is divided into successively smaller grids. As a geohash becomes longer, therefore, the area it represents shrinks. This figure shows the increasing resolution as a geohash shrinks from the eastern Gulf of Mexico (two digits) to the Sandcastle Hotel in Clearwater, Florida (seven digits).

Figure A3 shows the correlation between actual two-party vote share by county (x axis) in Texas and imputed vote share (y axis) using each cellphones home precinct. Each home location contributes its precinct's candidate vote share to the predicted vote share; those percentages are added by county to calculate the y-axis value. Counties are sized by population, with the largest labelled. The cellphone data are not political biased.

In our analysis, since we can differentiate individual smartphones from one another, the entire protest day is implicitly the window in which we compute counts of participating

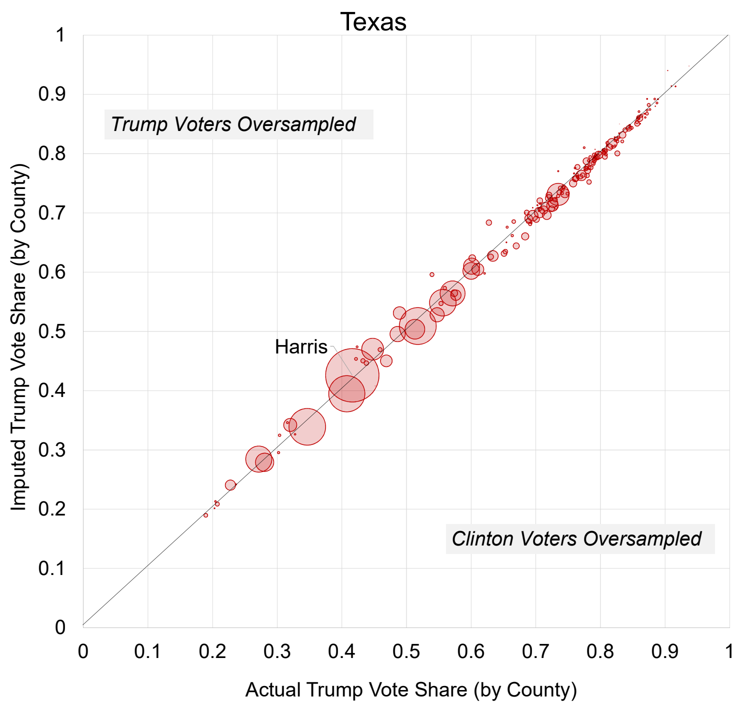
Figure A2: Florida Geohashes



Note: As more digits are added to the geohash, its resolution increases (covers a smaller area of land).

protesters. We then divide the protest day into sub-periods, facing two constraints. The first constraint is to minimize the number of sub-periods in which a cellphone does not report a location. The second constraint is to minimize the deviance from an individual's averaged location during a sub-period. Sub-periods of thirty minutes best satisfies these two constraints. Because we are interested only in attendance at the protest site, we have not attempted to study the means of transportation protesters used to come to and from these protests.

Figure A3: Ping to Geohash7 Location Recovers 2016 Presidential Vote Share by County in Texas

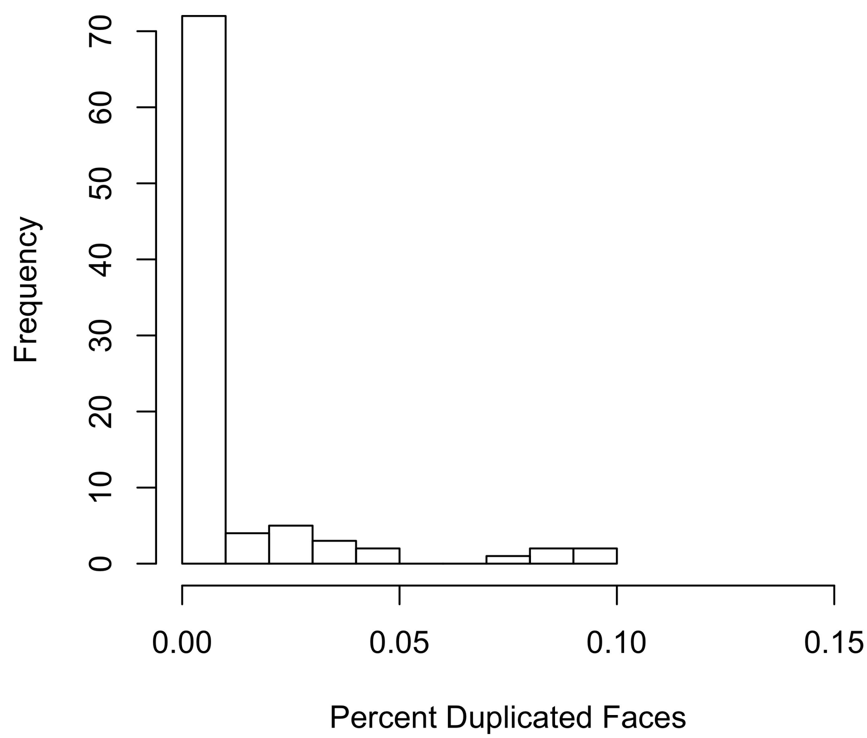


S3 Duplicate Faces

To confirm that the same person appearing multiple times across photos does not affect the size estimates, we calculate the percent of faces duplicated by city. For every city with more than 100 protest photos, we randomly sample 10% of those photos. This set contains 2,243 photos, from which 3,023 faces were detected by a face detector. Due to the excessive number of potential matches, we first filter out the majority of non-matching pairs of faces computationally using the face recognition API of dlib. The API applies a convolutional neural network (CNN) trained for face recognition to each face image and computes a feature descriptor of length 128 (He et al., 2016). In this feature space, the faces of the same individual should be closer to each other. We then calculate a Euclidean distance between every pair of faces in each city and disregard pairs whose distances are larger than 0.6. The threshold was recommended by dlib, but we found the threshold was liberal and yielded a large number of false positive matches. We therefore further manually verified every pair of faces classified as duplicates and obtained the final set of duplicates.

This procedure results in 2,993 unique faces and 30 duplicate faces, which makes the overall duplicate rate less than 1%. Of the 91 cities, 71 record no duplication; Nashville, Tennessee records a duplication rate of 9.5% (2 duplicates out of 21 faces), Cambridge, Massachusetts one of 9.1% (1 duplicate out of 11 faces). Since the duplication rate is very low to nonexistent, counting one face multiple times does not appear to drive our results. See Figure A4 for the distribution of duplication rate for these 91 cities.

Figure A4: Percent of Duplicate Faces



S4 Principal Component Analysis

We also construct two latent measures, using the first component from a principal component analysis, of protest size using *CCC: News*, *Twitter: Text Accounts*, *Twitter: Images Faces*, and *Twitter: Images Accounts*. PCA: All, the first measure, uses all four estimates. PCA: Images, the second, uses only *Twitter: Images Faces* and *Twitter: Images Accounts*. Only images are preferred for the second because this analysis will scale more easily than relying on newspapers or hand-constructed protest dictionaries.

The following sections include PCA estimates. Overall, PCA: All performs the best of all six measures, while PCA: Images performs almost identically to *Twitter: Images Faces*.

S5 Additional Model Fit Measurement

S5.1 Per Capita, Overlapping Observations

Table A1: Overlapping Observations, Include PCA

	CCC: News	PCA: Images	PCA: All	Twitter: Images Accounts	Twitter: Images Faces	Twitter: Text Accounts
Mean Error	0.06	0.09	0.00	0.10	0.08	0.07
Mean Error (Trimmed)	0.06	0.09	-0.07	0.11	0.14	0.04
Mean Absolute Error	0.58	0.75	0.84	0.82	0.71	0.60
Mean Absolute Error (Trimmed)	0.58	0.72	0.78	0.79	0.62	0.58
Best Predictor, %	0.22	0.10	0.20	0.16	0.11	0.21
Within 0.1 SDs	0.17	0.12	0.11	0.08	0.11	0.14
Within 0.5 SDs	0.52	0.50	0.50	0.38	0.52	0.54
Within 1 SDs	0.82	0.75	0.67	0.67	0.77	0.81

Note: "(Trimmed)" datasets drop observations where $|z| > 3$.

S5.2 Per Capita, All Observations

Table A2: All Observations, Include PCA

	CCC: News	PCA: Images	PCA: All	Twitter: Images Accounts	Twitter: Images Faces	Twitter: Text Accounts
Mean Error	0.03	0.11	0.00	0.06	0.11	0.13
Mean Error (Trimmed)	0.05	0.13	-0.07	0.10	0.16	0.12
Mean Absolute Error	0.60	0.73	0.84	0.83	0.71	0.61
Mean Absolute Error (Trimmed)	0.59	0.69	0.78	0.78	0.63	0.60
Best Predictor,	0.22	0.10	0.20	0.16	0.11	0.21
Within 0.1 SDs	0.17	0.11	0.11	0.08	0.09	0.13
Within 0.5 SDs	0.50	0.53	0.50	0.39	0.51	0.54
Within 1 SDs	0.81	0.76	0.67	0.68	0.78	0.80

Note: "(Trimmed)" datasets drop observations where $|z| > 3$.

S5.3 Not Per Capita, Overlapping Observations

Table A3: Not Per Capita, Include PCA, Overlapping Observations

	CCC: News	PCA: Images	PCA: All	Twitter: Images Accounts	Twitter: Images Faces	Twitter: Text Accounts
Mean Error	0.20	-0.22	-0.41	-0.14	-0.24	0.71
Mean Error (Trimmed)	0.20	-0.20	0.27	-0.14	-0.22	0.58
Mean Absolute Error	0.49	0.70	2.06	0.58	0.54	0.94
Mean Absolute Error (Trimmed)	0.49	0.68	1.42	0.58	0.52	0.83
Best Predictor, %	0.32	0.10	0.07	0.22	0.15	0.14
Within 0.1 SDs	0.19	0.06	0.03	0.12	0.08	0.07
Within 0.5 SDs	0.62	0.46	0.13	0.52	0.61	0.43
Within 1 SDs	0.88	0.79	0.33	0.82	0.88	0.66

Note: "(Trimmed)" datasets drop observations where $|z| > 3$.

S5.4 Not Per Capita, All Observations

Table A4: Not Per Capita, Include PCA, All Observations

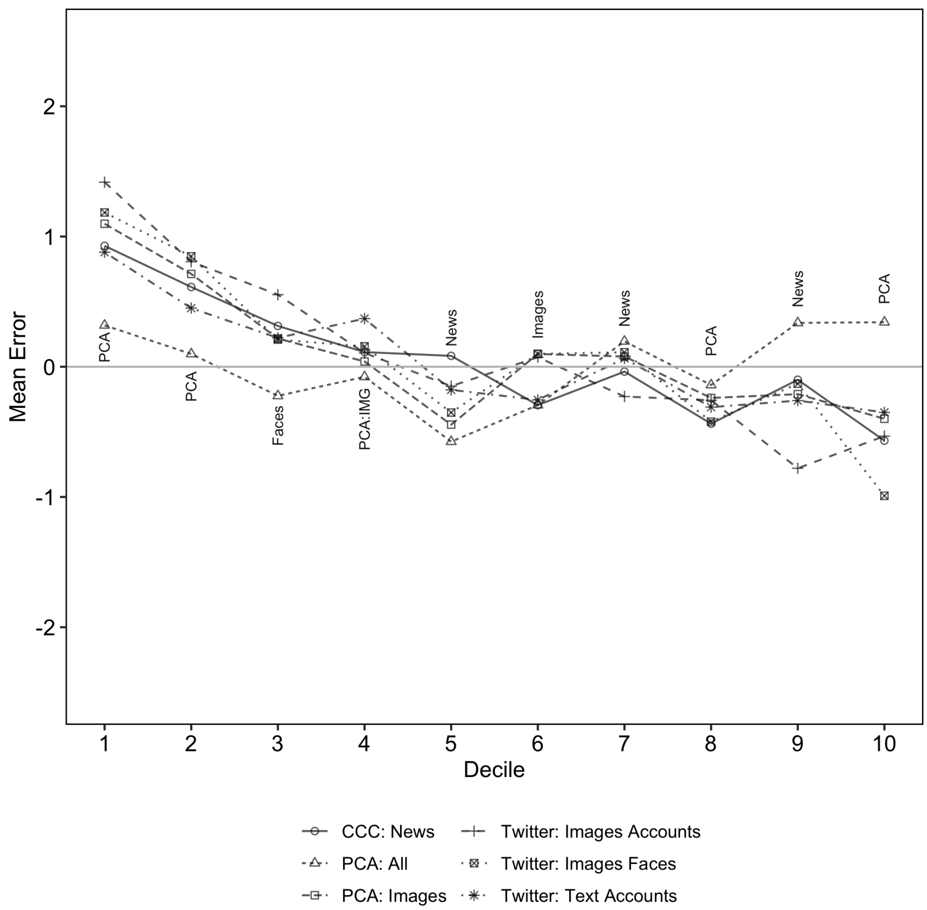
	CCC: News	PCA: Images	PCA: All	Twitter: Images Accounts	Twitter: Images Faces	Twitter: Text Accounts
Mean Error	0.03	0.11	0.00	0.06	0.11	0.13
Mean Error (Trimmed)	0.05	0.13	−0.07	0.10	0.16	0.12
Mean Absolute Error	0.60	0.73	0.84	0.83	0.71	0.61
Mean Absolute Error (Trimmed)	0.59	0.69	0.78	0.78	0.63	0.60
Best Predictor, %	0.22	0.10	0.20	0.16	0.11	0.21
Within 0.1 SDs	0.17	0.11	0.11	0.08	0.09	0.13
Within 0.5 SDs	0.50	0.53	0.50	0.39	0.51	0.54
Within 1 SDs	0.81	0.76	0.67	0.68	0.78	0.80

Note: "(Trimmed)" datasets drop observations where $|z| > 3$.

S6 Mean Error by Z-Score Decile

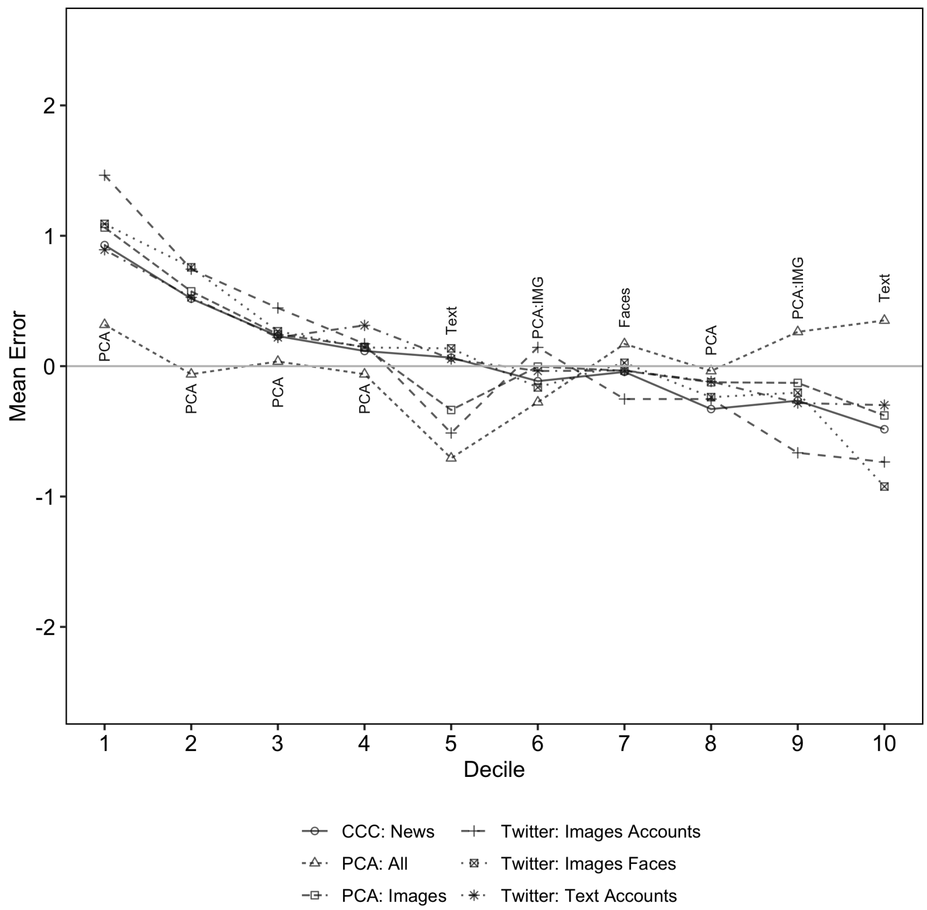
S6.1 Per Capita, Overlapping Observations

Figure A5: Mean Error by Decile, Overlapping Observations



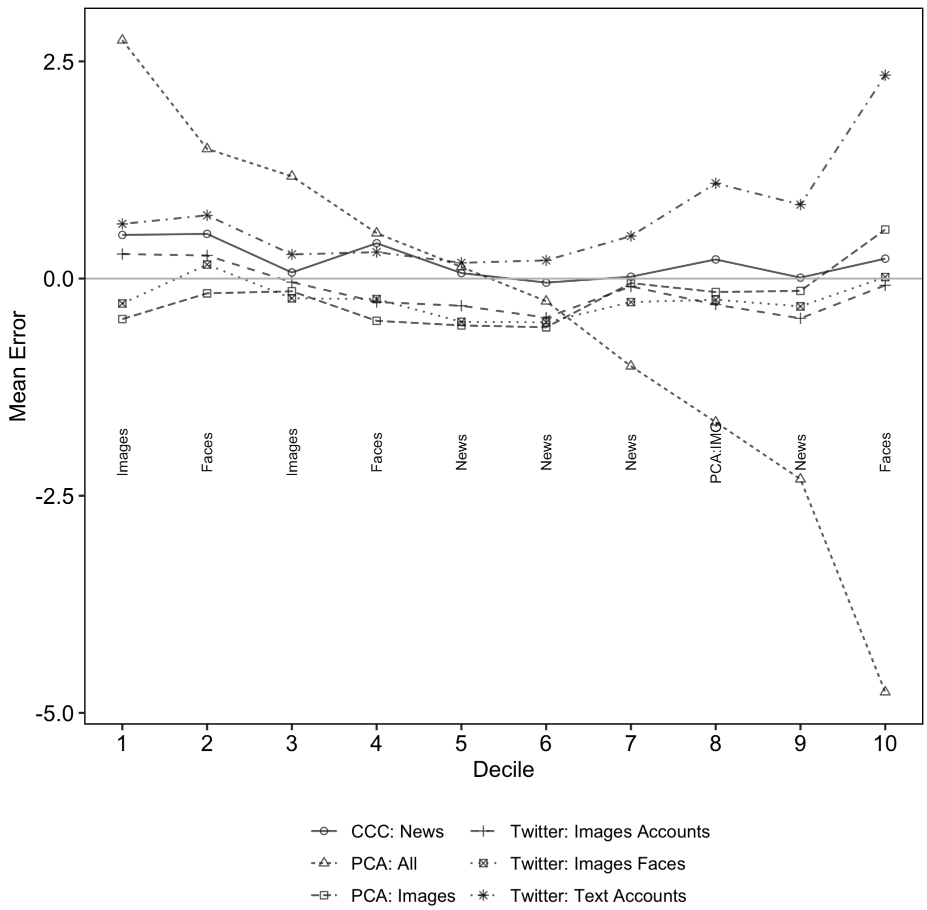
S6.2 Per Capita, All Observations

Figure A6: Mean Error by Decile, All Observations



S6.3 Not Per Capita, Overlapping Observations

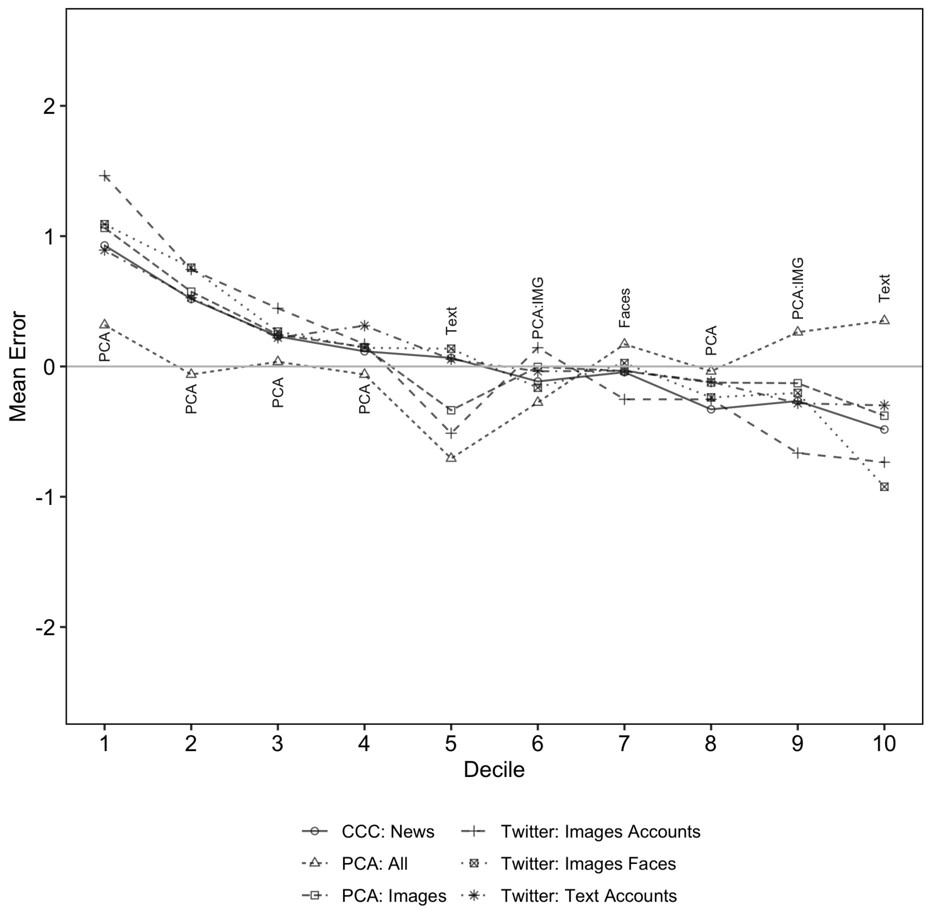
Figure A7: Mean Error by Decile, Overlapping Observations



PCA: All is calculated on the basis of the four other measures of protest size, and it systematically overestimates the size of small protests and underestimates large ones. Two factors drive this behavior. First, *Twitter: Images Accounts* has the largest variance of the four measures and is the most error prone estimate, as Figure 1 shows. Its variance therefore dominates the first component. Second, the other three measures are highly correlated but contribute primarily to the second principal component because of their much lower variance. As a result, the variation captured in the first principal component ignores information that *Twitter: Images Accounts* shares with others. Finally, *PCA: Images* outperforms *PCA: All* because it is calculated on the basis of only two variables, *Twitter: Images Accounts* and *Twitter: Images Faces*. In this case, by construction, the first principal component captures only information shared by both input variables.

S6.4 Not Per Capita, All Observations

Figure A8: Mean Error by Decile, All Observations



S7 Correlation Tables

Table A5 shows the pairwise correlation of the cellphone data estimates with the six protest measures and city population. Raw correlations are very high for all measures, with the news and social media measures appearing equally accurate. Because correlation is sensitive to outliers, our preferred measure is the correlation of the logged protest size values. Social media and news are almost equally accurate, though *Twitter: Images Accounts* is consistently the least accurate. Its size correlation is the lowest of the six measures, and the gap between its rank correlation and the next lowest (CCC, All Estimates (Low)) is large. Note as well that its size and rank correlation are low with the news and other Twitter estimates, suggesting that counting the number of accounts which share protest images captures a dynamic slightly different than protest attendance. While counting the number of accounts which share protest photos may appear less methodologically problematic than counting faces, it is clearly worse than the other two Twitter estimates and should be avoided.

Across almost all measurements of protest size, the placebo correlation is lower than the measures of protest size. The only times it is larger, shown in Tables A7b and A10b, is when the data are restricted to the observations common to all datasets. Even then, it is not much larger than *CCC: News* or *Twitter: Text Accounts*.

S7.1 Z-score of Log per Capita

This measure is the one used throughout the manuscript.

Table A5: Correlation of Z-score of Log per Capita

(a) All Observations per Pair

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	−0.585	0.655	0.655	0.652	0.681	0.424	0.395
Urbanized Area	0.585	1	0.327	0.327	0.338	0.428	0.059	0.216
CCC: News	0.655	0.327	1	1	0.988	0.807	0.609	0.391
CCC, All Estimates	0.655	0.327	1	1	0.988	0.807	0.609	0.391
CCC, All Estimates (Low)	0.652	0.338	0.988	0.988	1	0.780	0.600	0.366
Twitter: Text Accounts	0.681	0.428	0.807	0.807	0.780	1	0.574	0.563
Twitter: Images Faces	0.424	0.059	0.609	0.609	0.600	0.574	1	0.249
Twitter: Images Accounts	0.395	0.216	0.391	0.391	0.366	0.563	0.249	1

(b) Same Observations for All Pairs

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.409	0.609	0.619	0.612	0.642	0.432	0.337
Urbanized Area	0.409	1	0.359	0.312	0.317	0.501	0.109	0.089
CCC: News	0.609	0.359	1	1	0.991	0.802	0.606	0.305
CCC, All Estimates	0.619	0.312	1	1	0.991	0.752	0.592	0.209
CCC, All Estimates (Low)	0.612	0.317	0.991	0.991	1	0.736	0.585	0.191
Twitter: Text Accounts	0.642	0.501	0.802	0.752	0.736	1	0.611	0.319
Twitter: Images Faces	0.432	0.109	0.606	0.592	0.585	0.611	1	0.294
Twitter: Images Accounts	0.337	0.089	0.305	0.209	0.191	0.319	0.294	1

S7.2 Z-score of Log Size

This measure is the z-score of the logged protest size, not log per capita protest size.

Table A7: Correlation of Z-score of Log Measures

(a) All Observations per Pair

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.570	0.708	0.708	0.699	0.730	0.714	0.631
Urbanized Area	0.570	1	0.467	0.467	0.460	0.630	0.641	0.844
CCC: News	0.708	0.467	1	1	0.990	0.825	0.805	0.559
CCC, All Estimates	0.708	0.467	1	1	0.990	0.825	0.805	0.559
CCC, All Estimates (Low)	0.699	0.460	0.990	0.990	1	0.800	0.795	0.541
Twitter: Text Accounts	0.730	0.630	0.825	0.825	0.800	1	0.833	0.758
Twitter: Images Faces	0.714	0.641	0.805	0.805	0.795	0.833	1	0.687
Twitter: Images Accounts	0.631	0.844	0.559	0.559	0.541	0.758	0.687	1

(b) Same Observations for All Pairs

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.722	0.707	0.716	0.702	0.726	0.707	0.660
Urbanized Area	0.722	1	0.458	0.481	0.472	0.633	0.626	0.849
CCC: News	0.707	0.458	1	1	0.990	0.830	0.800	0.561
CCC, All Estimates	0.716	0.481	1	1	0.993	0.798	0.767	0.508
CCC, All Estimates (Low)	0.702	0.472	0.990	0.993	1	0.781	0.758	0.493
Twitter: Text Accounts	0.726	0.633	0.830	0.798	0.781	1	0.834	0.678
Twitter: Images Faces	0.707	0.626	0.800	0.767	0.758	0.834	1	0.681
Twitter: Images Accounts	0.660	0.849	0.561	0.508	0.493	0.678	0.681	1

S7.3 Log

Table A9: Correlation of Log(Size)

(a) All Observations per Pair

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.639	0.724	0.724	0.719	0.755	0.714	0.627
Urbanized Area	0.639	1	0.561	0.561	0.557	0.714	0.702	0.787
CCC: News	0.724	0.561	1	1	0.990	0.839	0.807	0.602
CCC, All Estimates	0.724	0.561	1	1	0.990	0.839	0.807	0.602
CCC, All Estimates (Low)	0.719	0.557	0.990	0.990	1	0.811	0.795	0.579
Twitter: Text Accounts	0.755	0.714	0.839	0.839	0.811	1	0.838	0.793
Twitter: Images Faces	0.714	0.702	0.807	0.807	0.795	0.838	1	0.710
Twitter: Images Accounts	0.627	0.787	0.602	0.602	0.579	0.793	0.710	1

(b) Same Observations for All Pairs

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.778	0.695	0.721	0.712	0.726	0.707	0.660
Urbanized Area	0.778	1	0.585	0.603	0.600	0.709	0.698	0.797
CCC: News	0.695	0.585	1	1	0.991	0.830	0.800	0.561
CCC, All Estimates	0.721	0.603	1	1	0.992	0.798	0.767	0.508
CCC, All Estimates (Low)	0.712	0.600	0.991	0.992	1	0.781	0.758	0.493
Twitter: Text Accounts	0.726	0.709	0.830	0.798	0.781	1	0.834	0.678
Twitter: Images Faces	0.707	0.698	0.800	0.767	0.758	0.834	1	0.681
Twitter: Images Accounts	0.660	0.797	0.561	0.508	0.493	0.678	0.681	1

S7.4 Raw

Table A10: Correlation of Size, no Transformation

(a) All Observations per Pair

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.587	0.798	0.798	0.841	0.794	0.829	0.640
Urbanized Area	0.587	1	0.674	0.674	0.648	0.678	0.583	0.870
CCC: News	0.798	0.674	1	1	0.972	0.978	0.966	0.819
CCC, All Estimates	0.798	0.674	1	1	0.972	0.978	0.966	0.819
CCC, All Estimates (Low)	0.841	0.648	0.972	0.972	1	0.931	0.943	0.760
Twitter: Text Accounts	0.794	0.678	0.978	0.978	0.931	1	0.971	0.849
Twitter: Images Faces	0.829	0.583	0.966	0.966	0.943	0.971	1	0.782
Twitter: Images Accounts	0.640	0.870	0.819	0.819	0.760	0.849	0.782	1

(b) Same Observations for All Pairs

	Cellphone Data	*Urbanized Area*	*CCC: News*	*CCC, All Estimates*	*CCC, All Estimates (Low)*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
Cellphone Data	1	0.587	0.798	0.798	0.841	0.794	0.829	0.640
Urbanized Area	0.587	1	0.674	0.674	0.648	0.678	0.583	0.870
CCC: News	0.798	0.674	1	1	0.972	0.978	0.966	0.819
CCC, All Estimates	0.798	0.674	1	1	0.972	0.978	0.966	0.819
CCC, All Estimates (Low)	0.841	0.648	0.972	0.972	1	0.931	0.943	0.760
Twitter: Text Accounts	0.794	0.678	0.978	0.978	0.931	1	0.971	0.849
Twitter: Images Accounts	0.640	0.870	0.819	0.819	0.760	0.849	0.782	1

S8 Results, Regression

This section regresses the cellphone data on a placebo model, the placebo model with the size estimate of the four measures, and a model pooling the four estimates. The placebo model includes variables for city population, median household income, income inequality, vote share for Candidate Hillary Clinton, and the number of unemployed. All models include state fixed effects. Ordinary least squares is the estimator, except for the models where the outcome is not transformed.

Each subsection uses a different transformation of protest size applied to the overlapping observations and all observations per measure. There are therefore eight tables, two per the four subsections.

The results mirror the main paper's. Across dependent variable transformations and subsets of data, *CCC: News* and *Twitter: Text Accounts* models explain the most variation of the cellphone protest size. In Tables A15 (z-score of log size) and A17 (log size), the placebo model outperforms the others models. In those tables, however, the model that pools the four measures' estimates with the placebo variables finds statistical significance for only *CCC: News* and *Twitter: Text Accounts*. Note as well that the placebo is worse than the other measures when using raw data, as seen in Tables A18 and A19.

S8.1 Z-score of Log per Capita

Table A12: Same Cases

	Z, Log(Per Capita)					
	(1)	(2)	(3)	(4)	(5)	(6)
Z, Log(*Twitter: Images Faces* per Capita)		0.273^{**}				0.050
		(0.105)				(0.088)
Z, Log(*Twitter: Text Accounts* per Capita)			0.791^{***}			0.545^{***}
			(0.104)			(0.171)
Z, Log(*Twitter: Images Accounts* per Capita)				0.363^{***}		0.020
				(0.124)		(0.119)
Z, Log(*CCC: News* per Capita)					0.694^{***}	0.297^{**}
					(0.106)	(0.147)
Median HHI	0.238	0.139	0.068	0.279	0.092	0.042
	(0.195)	(0.192)	(0.150)	(0.187)	(0.159)	(0.152)
Gini	0.178	0.212	0.085	0.015	0.101	0.078
	(0.151)	(0.146)	(0.115)	(0.154)	(0.122)	(0.123)
County Dem. Vote Share	−0.034	−0.160	−0.163	−0.060	-0.249^{*}	-0.240^{*}
	(0.179)	(0.180)	(0.137)	(0.171)	(0.148)	(0.140)
Unemployed Count	−0.300	−0.208	0.562	0.517	0.967	0.899
	(1.341)	(1.294)	(1.024)	(1.309)	(1.100)	(1.045)
Population	−0.014	−0.131	−0.990	−0.904	−1.336	−1.324
	(1.459)	(1.409)	(1.115)	(1.425)	(1.195)	(1.133)
Intercept	0.302	0.221	0.042	0.413	0.285	0.107
	(0.922)	(0.890)	(0.700)	(0.880)	(0.744)	(0.694)
Observations	127	127	127	127	127	127
Adjusted R^2	0.122	0.182	0.494	0.201	0.428	0.510
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A13: Maximum Cases per Measure

	Z, Log(Per Capita)					
	(1)	(2)	(3)	(4)	(5)	(6)
Z, Log(*Twitter: Images Faces* per Capita)		0.280***				0.050
		(0.091)				(0.088)
Z, Log(*Twitter: Text Accounts* per Capita)			0.683***			0.545***
			(0.076)			(0.171)
Z, Log(*Twitter: Images Accounts* per Capita)				0.281***		0.020
				(0.083)		(0.119)
Z, Log(*CCC: News* per Capita)					0.497***	0.297**
					(0.085)	(0.147)
Median HHI	0.213*	0.184	0.085	0.266*	0.238	0.042
	(0.120)	(0.158)	(0.113)	(0.143)	(0.148)	(0.152)
Gini	0.202*	0.262**	0.182*	0.058	0.164	0.078
	(0.103)	(0.126)	(0.095)	(0.129)	(0.115)	(0.123)
County Dem. Vote Share	−0.053	−0.166	−0.220**	−0.040	−0.216	−0.240*
	(0.116)	(0.153)	(0.108)	(0.138)	(0.144)	(0.140)
Unemployed Count	−0.566	−0.717	−0.456	−0.232	0.521	0.899
	(0.733)	(0.929)	(0.724)	(0.913)	(1.075)	(1.045)
Population	0.352	0.462	0.212	−0.031	−0.885	−1.324
	(0.757)	(0.968)	(0.753)	(0.949)	(1.169)	(1.133)
Intercept	−0.561	−1.274	−0.449	−0.844	−1.092*	0.107
	(0.590)	(0.853)	(0.514)	(0.651)	(0.600)	(0.694)
Observations	214	153	167	165	139	127
Adjusted R^2	0.365	0.244	0.492	0.206	0.357	0.510
Note:	*p<0.1; **p<0.05; ***p<0.01					

S8.2 Z-score of Log Size

Table A14: Same Cases

	Z, Log(Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Z, Log(*Twitter: Images Faces*)		0.334***				0.094
		(0.085)				(0.100)
Z, Log(*Twitter: Text Accounts*)			0.355***			0.199**
			(0.068)			(0.097)
Z, Log(*Twitter: Images Accounts*)				0.274**		0.056
				(0.124)		(0.130)
Z, Log(*CCC: News*)					0.436***	0.224*
					(0.092)	(0.121)
Median HHI	0.130	0.064	0.050	0.156	0.062	0.037
	(0.121)	(0.113)	(0.106)	(0.119)	(0.109)	(0.106)
Gini	−0.077	0.0004	−0.001	−0.085	−0.048	0.001
	(0.094)	(0.088)	(0.082)	(0.092)	(0.083)	(0.084)
County Dem. Vote Share	0.231**	0.045	0.011	0.146	0.021	−0.070
	(0.111)	(0.112)	(0.105)	(0.115)	(0.108)	(0.108)
Unemployed Count	−1.498*	−0.804	−0.110	−0.864	−0.307	0.213
	(0.821)	(0.775)	(0.759)	(0.851)	(0.770)	(0.772)
Population	2.014**	1.057	0.056	1.153	0.527	−0.289
	(0.893)	(0.856)	(0.858)	(0.954)	(0.851)	(0.876)
Intercept	0.953	0.744	0.183	0.846	0.578	0.249
	(0.573)	(0.530)	(0.518)	(0.562)	(0.514)	(0.508)
Observations	128	128	128	128	128	128
Adjusted R^2	0.362	0.461	0.522	0.392	0.499	0.542
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A15: Maximum Cases per Measure

	Z, Log(Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Z, Log(*Twitter: Images Faces*)		0.356***				0.094
		(0.065)				(0.100)
Z, Log(*Twitter: Text Accounts*)			0.344***			0.199**
			(0.050)			(0.097)
Z, Log(*Twitter: Images Accounts*)				0.338***		0.056
				(0.075)		(0.130)
Z, Log(*CCC: News*)					0.363***	0.224*
					(0.060)	(0.121)
Median HHI	0.380***	0.088	0.104	0.236***	0.197**	0.037
	(0.064)	(0.091)	(0.084)	(0.089)	(0.078)	(0.106)
Gini	0.052	0.026	0.070	−0.037	0.043	0.001
	(0.058)	(0.074)	(0.069)	(0.075)	(0.064)	(0.084)
County Dem. Vote Share	0.188***	0.022	−0.055	0.101	0.020	−0.070
	(0.063)	(0.093)	(0.085)	(0.088)	(0.080)	(0.108)
Unemployed Count	−0.453	−0.682	−0.617	−0.666	−0.576	0.213
	(0.496)	(0.539)	(0.529)	(0.571)	(0.674)	(0.772)
Population	0.725	0.858	0.628	0.837	0.865	−0.289
	(0.511)	(0.566)	(0.557)	(0.600)	(0.732)	(0.876)
Intercept	−1.275***	0.055	−0.672*	−0.779*	−0.985***	0.249
	(0.237)	(0.492)	(0.375)	(0.403)	(0.318)	(0.508)
Observations	295	154	170	166	187	128
Adjusted R^2	0.691	0.563	0.610	0.541	0.600	0.542
Note:	*p<0.1; **p<0.05; ***p<0.01					

S8.3 Log Size

Table A16: Same Cases

	Log(Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(*Twitter: Images Faces*)		0.297***				0.083
		(0.076)				(0.089)
Log(*Twitter: Text Accounts*)			0.697***			0.390**
			(0.133)			(0.190)
Log(*Twitter: Images Accounts*)				0.275**		0.056
				(0.124)		(0.130)
Log(*CCC: News*)					0.351***	0.180*
					(0.074)	(0.097)
Median HHI	0.091	0.044	0.035	0.109	0.043	0.026
	(0.085)	(0.079)	(0.074)	(0.083)	(0.076)	(0.074)
Gini	−0.054	0.0003	−0.0004	−0.059	−0.033	0.001
	(0.065)	(0.062)	(0.058)	(0.064)	(0.058)	(0.058)
County Dem. Vote Share	0.161**	0.032	0.007	0.102	0.014	−0.049
	(0.077)	(0.078)	(0.073)	(0.080)	(0.075)	(0.076)
Unemployed Count	−1.045*	−0.561	−0.077	−0.603	−0.215	0.149
	(0.573)	(0.541)	(0.530)	(0.594)	(0.537)	(0.539)
Population	1.406**	0.738	0.039	0.805	0.368	−0.202
	(0.623)	(0.597)	(0.599)	(0.666)	(0.594)	(0.612)
Intercept	2.395***	1.465***	1.400***	1.655***	1.075**	0.750
	(0.400)	(0.437)	(0.395)	(0.513)	(0.450)	(0.521)
Observations	128	128	128	128	128	128
Adjusted R^2	0.362	0.461	0.522	0.392	0.499	0.542
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A17: Maximum Cases per Measure

	Log(Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(*Twitter: Images Faces*)		0.317***				0.083
		(0.058)				(0.089)
Log(*Twitter: Text Accounts*)			0.675***			0.390**
			(0.099)			(0.190)
Log(*Twitter: Images Accounts*)				0.339***		0.056
				(0.075)		(0.130)
Log(*CCC: News*)					0.291***	0.180*
					(0.048)	(0.097)
Median HHI	0.265***	0.062	0.072	0.165***	0.137**	0.026
	(0.045)	(0.063)	(0.059)	(0.062)	(0.054)	(0.074)
Gini	0.036	0.018	0.049	−0.026	0.030	0.001
	(0.040)	(0.052)	(0.048)	(0.053)	(0.045)	(0.058)
County Dem. Vote Share	0.131***	0.016	−0.038	0.071	0.014	−0.049
	(0.044)	(0.065)	(0.059)	(0.061)	(0.056)	(0.076)
Unemployed Count	−0.317	−0.476	−0.431	−0.465	−0.402	0.149
	(0.346)	(0.377)	(0.369)	(0.399)	(0.471)	(0.539)
Population	0.506	0.599	0.438	0.585	0.604	−0.202
	(0.356)	(0.395)	(0.389)	(0.419)	(0.511)	(0.612)
Intercept	0.840***	0.933**	0.817***	0.367	0.164	0.750
	(0.166)	(0.364)	(0.265)	(0.332)	(0.276)	(0.521)
Observations	295	154	170	166	187	128
Adjusted R^2	0.691	0.563	0.610	0.541	0.600	0.542

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

S8.4 Size

Table A18: Same Cases

	Cellphone Data					
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter: Images Faces		0.00000				0.0001^{*}
		(0.00001)				(0.00003)
Twitter: Text Accounts			−0.0001			−0.0002
			(0.001)			(0.003)
Twitter: Images Accounts				−0.00004		−0.0001
				(0.0001)		(0.0001)
CCC: News					−0.00000	-0.00001^{*}
					(0.00000)	(0.00000)
Median HHI	0.132	0.129	0.134	0.137	0.138	0.111
	(0.110)	(0.111)	(0.111)	(0.110)	(0.111)	(0.110)
Gini	−0.036	−0.033	−0.037	−0.039	−0.041	−0.030
	(0.105)	(0.105)	(0.105)	(0.105)	(0.105)	(0.104)
County Dem. Vote Share	0.303^{***}	0.298^{***}	0.304^{***}	0.317^{***}	0.312^{***}	0.315^{***}
	(0.112)	(0.112)	(0.112)	(0.112)	(0.112)	(0.113)
Unemployed Count	−1.261	−1.178	−1.295	-1.589^{*}	−1.426	−1.435
	(0.854)	(0.910)	(0.883)	(0.949)	(0.888)	(1.045)
Population	2.057^{**}	1.957^{*}	2.098^{**}	2.509^{**}	2.259^{**}	2.580^{**}
	(0.919)	(1.005)	(0.977)	(1.124)	(0.984)	(1.213)
Intercept	4.958^{***}	4.953^{***}	4.960^{***}	5.005^{***}	4.967^{***}	5.060^{***}
	(0.080)	(0.084)	(0.084)	(0.109)	(0.084)	(0.116)
Observations	128	128	128	128	128	128
Log Likelihood	−768.393	−768.376	−768.390	−768.200	−768.326	−765.705
Akaike Inf. Crit.	1,548.787	1,550.751	1,550.781	1,550.401	1,550.652	1,551.410
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A19: Maximum Cases per Measure

	Cellphone Data					
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter: Images Faces		0.00000				0.0001^{*}
		(0.00001)				(0.00003)
Twitter: Text Accounts			−0.0001			−0.0002
			(0.001)			(0.003)
Twitter: Images Accounts				0.00001		−0.0001
				(0.0001)		(0.0001)
CCC: News					−0.00000	-0.00001^{*}
					(0.00000)	(0.00000)
Median HHI	0.669^{***}	0.194^{*}	0.297^{***}	0.264^{**}	0.316^{***}	0.111
	(0.086)	(0.111)	(0.106)	(0.107)	(0.097)	(0.110)
Gini	0.534^{***}	−0.029	0.111	0.127	0.326^{***}	−0.030
	(0.080)	(0.105)	(0.102)	(0.101)	(0.089)	(0.104)
County Dem. Vote Share	0.291^{***}	0.370^{***}	0.353^{***}	0.351^{***}	0.273^{***}	0.315^{***}
	(0.083)	(0.108)	(0.103)	(0.105)	(0.097)	(0.113)
Unemployed Count	−0.842	−0.471	−0.699	−0.613	-2.134^{**}	−1.435
	(0.607)	(0.740)	(0.761)	(0.757)	(0.949)	(1.045)
Population	1.187^{*}	1.041	1.334^{*}	1.190	3.205^{***}	2.580^{**}
	(0.623)	(0.777)	(0.799)	(0.803)	(1.045)	(1.213)
Intercept	4.487^{***}	4.832^{***}	4.741^{***}	4.737^{***}	4.672^{***}	5.060^{***}
	(0.058)	(0.079)	(0.078)	(0.091)	(0.077)	(0.116)
Observations	295	154	170	166	187	128
Log Likelihood	−1,709.422	−909.459	−991.326	−969.625	−1,050.807	−765.705
Akaike Inf. Crit.	3,430.844	1,832.918	1,996.653	1,953.250	2,115.614	1,551.410
Note:	*p<0.1; **p<0.05; ***p<0.01					

S8.5 Investigating Overlap

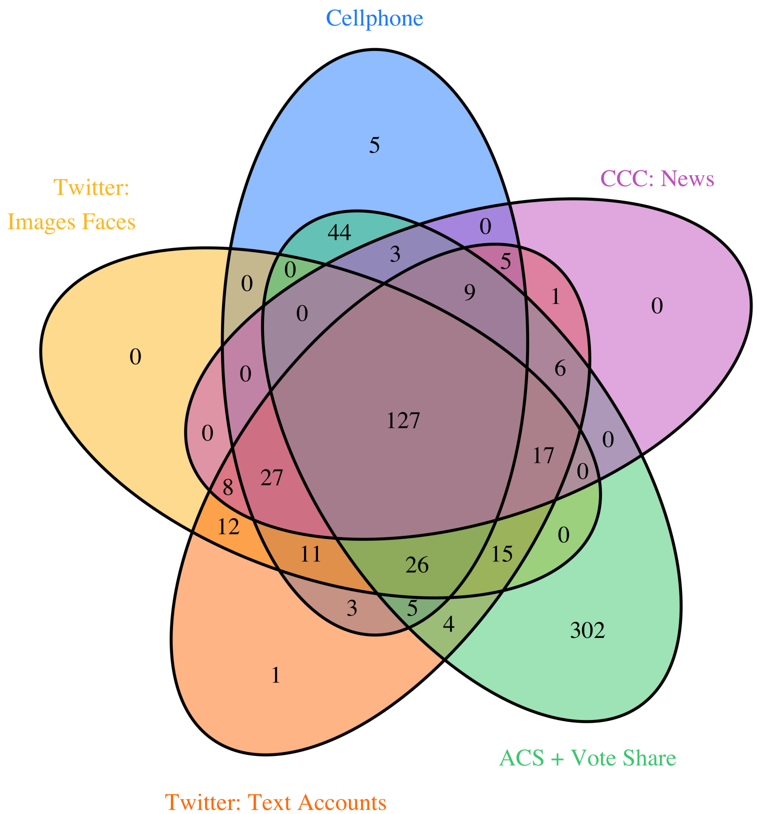
To better understand the overlap between the five measures, this subsection presents two sets of analysis. First, we construct two Venn Diagrams of the measures. The first includes the American Community Survey and voting data, and the second does not. Two diagrams are required because the package used to create them only allows five sets (Chen, 2018), so including the five measures and ACS in one diagram is not possible. Second, we generate a dependent variable that is a 1 if a measure recorded protest in a city and 0 otherwise. We then regress that on ACS measures of education, income, unemployment, and population.

Figure A9a shows the Venn-Diagram for the data used in the regressions in Table A13. Of the four size measures there, only the cellphone and *Twitter: Text Accounts* have cities exclusive to them, five and one respectively. Figure A9b shows a Venn-Diagram using the five size measures.[18] The clearest result is that only cellphone data captures protests that the other measures do not, if one is not also interested in demographic data. If one wants to model protest size and include demographic data, the size measures do not appear not distinct enough to induce bias. Of the 127 cities in each dataset, Los Angeles, California in the most populous and Aspen, Colorado the least.[19] The median city, based on population, is Madison, Wisconsin. Of the 27 ACS does not include that the five measures do, Montpelier, Vermont is the smallest and Santa Ana, California the largest. The complete list of cities is too large to convey easily on paper but is available via the replication material or directly from the corresponding author.

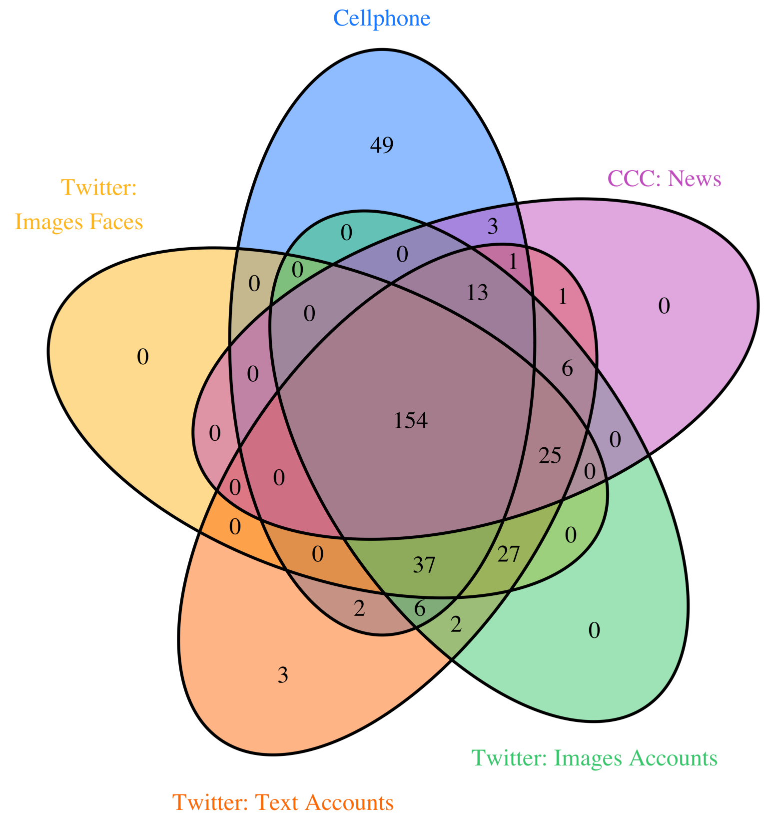
[18] The 27 not in these 127 are Ashland, OR; Augusta, GA; Bethlehem, PA; Boise, ID; Brighton, MI; Clemson, SC; Doylestown, PA; Jonesborough, TN; Montpelier, VT; Naples, FL; Newark, DE; Norfolk, VA; Northampton, MA; Ocean City, MD; Olympia, WA; Ontario, CA; Palm Springs, CA; Pequannock Township, NJ; Poughkeepsie, NY; Prescott, AZ; San Marcos, CA; Santa Ana, CA; Sarasota, FL; Seneca Falls, NY; Walnut Creek, CA; Westfield, NJ; and West Palm Beach, FL.

[19] New York City, New York is not part of this set because *CCC: News* records are from a Facebook group and a tweet.

Figure A9: Overlap of Cities



(a) Regression



(b) Excluding ACS + Voting Data

Table A20 presents the results of the second analysis, a series of logistic regressions of whether or not a size measure (column) records a protest (1=Yes) as a function of socioeconomic factors recorded by the ACS. More populous, higher income, and unequal cities are more likely to appear in all datasets, as are those with more inhabitants with a professional degree.

Table A20: Understanding Determinants of Recording Protest

	Dependent variable: 1 = In Size + ACS; 0 = No				
	Cellphone	*CCC: News*	*Twitter: Text Accounts*	*Twitter: Images Faces*	*Twitter: Images Accounts*
	(1)	(2)	(3)	(4)	(5)
Population	0.014^{**}	0.017^{***}	0.019^{***}	0.019^{***}	0.018^{***}
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Median HHI	0.043^{***}	0.045^{***}	0.039^{***}	0.042^{***}	0.042^{***}
	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)
Gini	0.142^{***}	0.117^{***}	0.120^{***}	0.131^{***}	0.123^{***}
	(0.037)	(0.036)	(0.034)	(0.036)	(0.035)
Unemployed Count	-0.033	-0.011	-0.058^{**}	-0.069^{***}	-0.056^{**}
	(0.029)	(0.023)	(0.023)	(0.023)	(0.022)
Education < HS	-0.00003^{***}	-0.00004^{***}	-0.00003^{***}	-0.00003^{***}	-0.00003^{***}
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Education = HS	-0.00003^{***}	-0.00003^{***}	-0.00003^{***}	-0.00003^{***}	-0.00003^{***}
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Educ. Some Coll.	0.00001	-0.00000	-0.00001	-0.00001	-0.00001
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Educ. = College	-0.0001^{***}	-0.0001^{***}	-0.0001^{***}	-0.0001^{***}	-0.0001^{***}
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Educ. = Master	-0.00003	-0.0001^{**}	-0.00003^{*}	-0.00003	-0.00003
	(0.00003)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Educ. = Professional	0.0004^{***}	0.0002^{**}	0.0002^{***}	0.0002^{***}	0.0002^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Educ. = PhD	-0.00004	0.0001	-0.00001	-0.00001	-0.00002
	(0.0001)	(0.0001)	(0.00004)	(0.00004)	(0.00004)
Constant	-11.040^{***}	-9.265^{***}	-8.932^{***}	-10.440^{***}	-9.272^{***}
	(2.123)	(1.996)	(1.903)	(2.081)	(1.930)
State FE	Y	Y	Y	Y	Y
Observations	569	569	569	569	569
Log Likelihood	-267.937	-268.626	-311.622	-287.502	-307.680

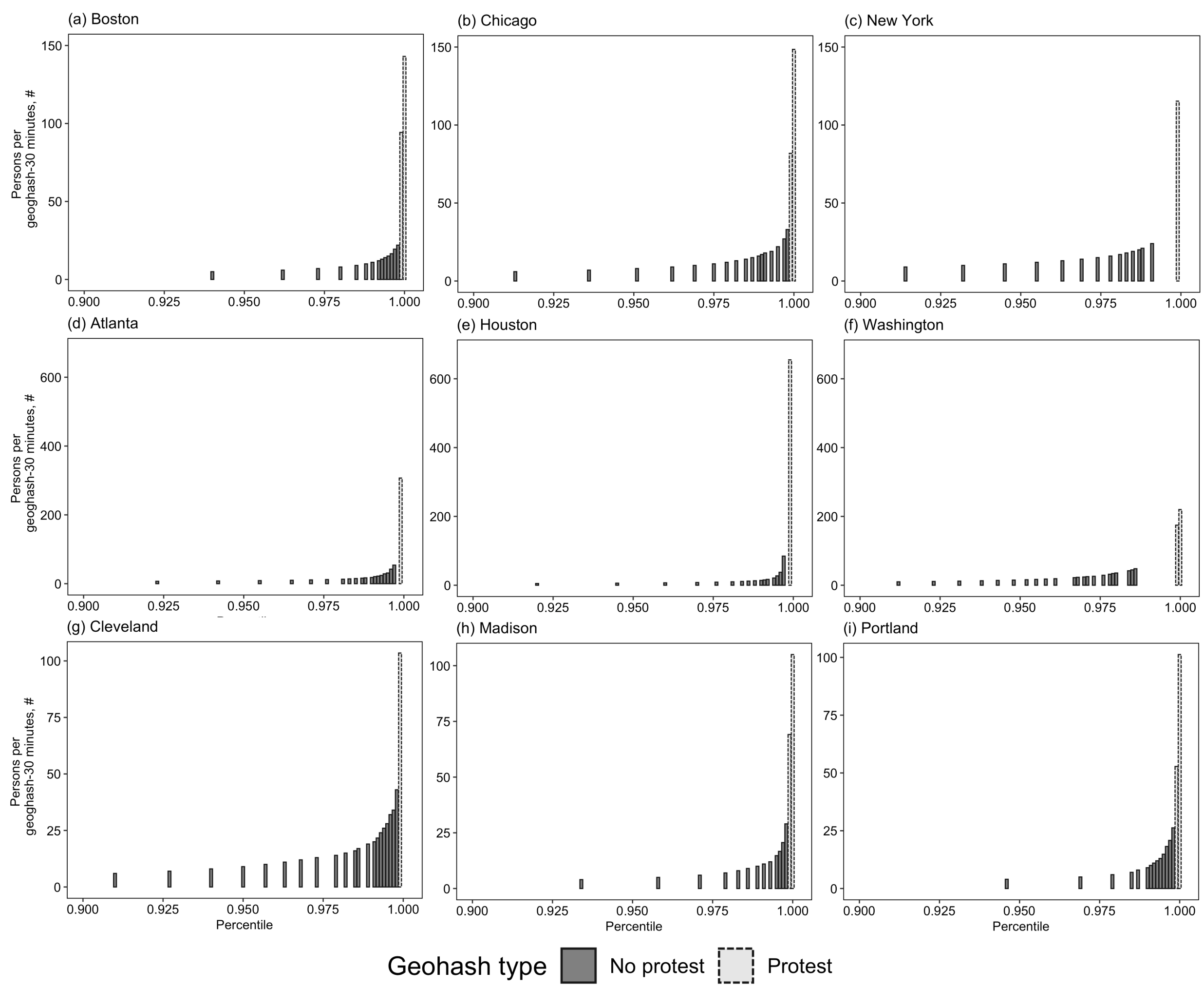
Notes: *p<0.1; **p<0.05; ***p<0.01
The education variables are for inhabitants over 25 years of age.
The number of observations per model is the same because of the ACS data.

S9 Automated Detection of Protest Geohashes

S9.1 Observations by percentile

Figure A10 shows the distribution of 30 minute geohash density in nine cities, with the protest geohashes in light grey with dashed borders. The cities were chosen to represent diverse urban forms and city sizes.

Figure A10: Persons per Geohash-30-Minutes



S9.2 Location of Protest Events: Declared Versus Detected

Figures A11-A19 shows maps of the nine cities from Figure A10 with two types of geohashes labeled. The red geohash corresponds to the reported location of the event as recorded in CCC. The green is those our detection method identified. The geohashes the automatic approach identifies are nearly identical to those reported in the media. The automatic approach identifies more locations, which makes sense because the Women's March protests were substantial enough to occupy a larger area than what one seven digit geohash encompasses.

Figure A11

Figure A12

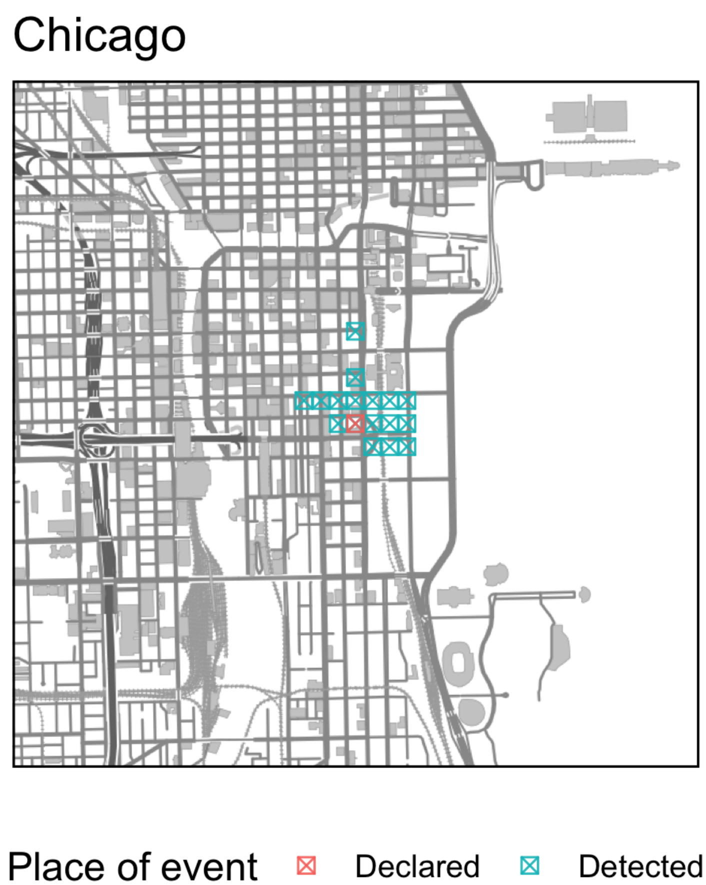


Figure A13

Figure A14

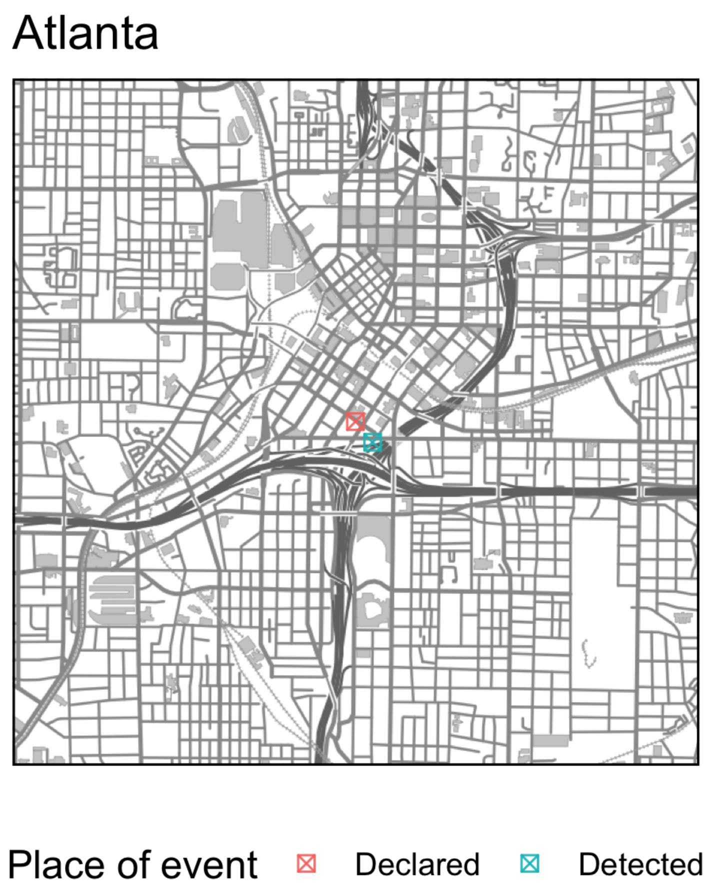


Figure A15

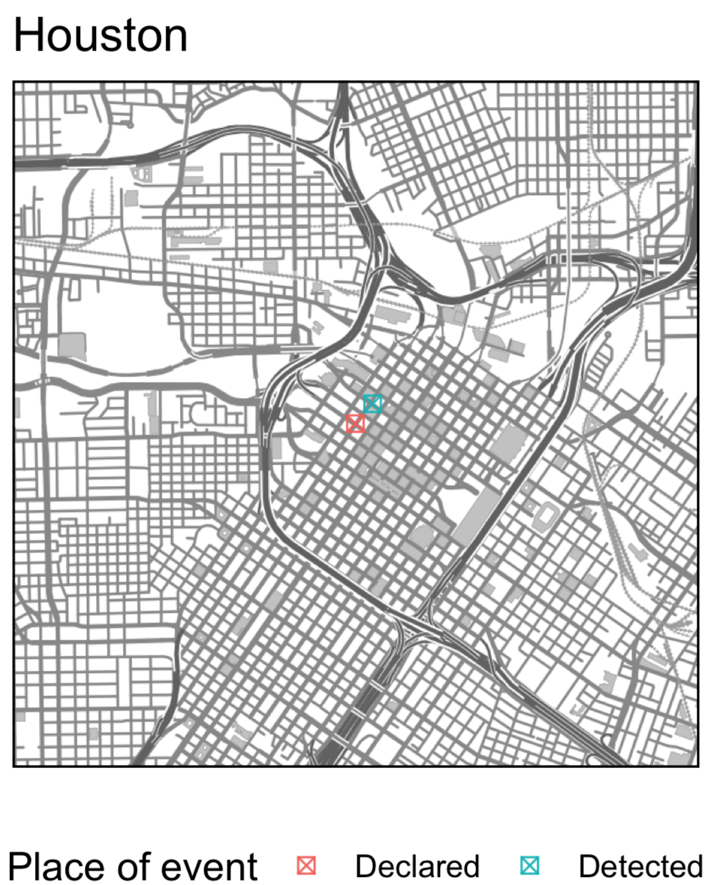


Figure A16

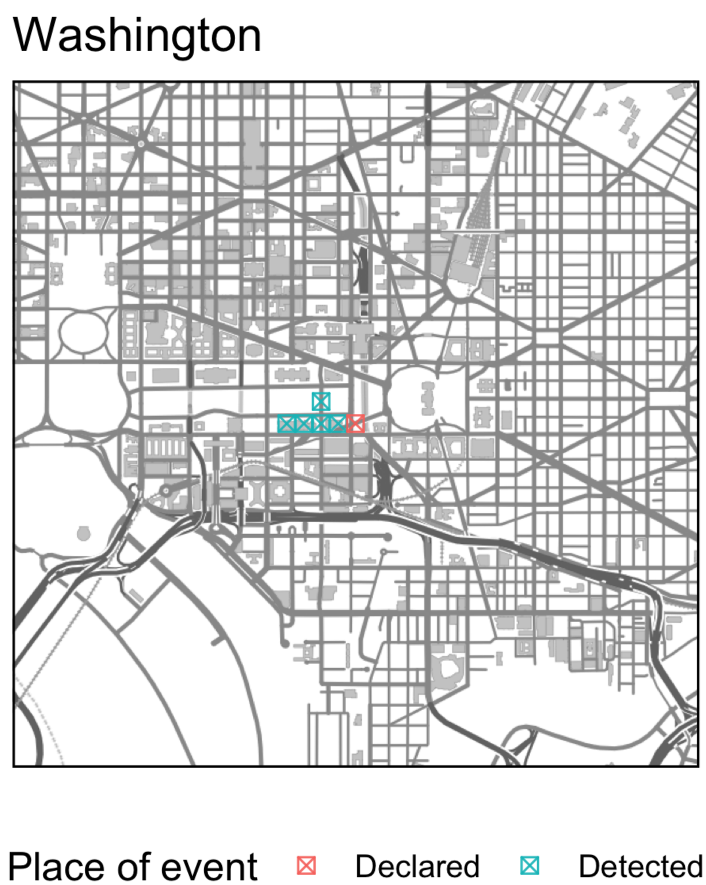
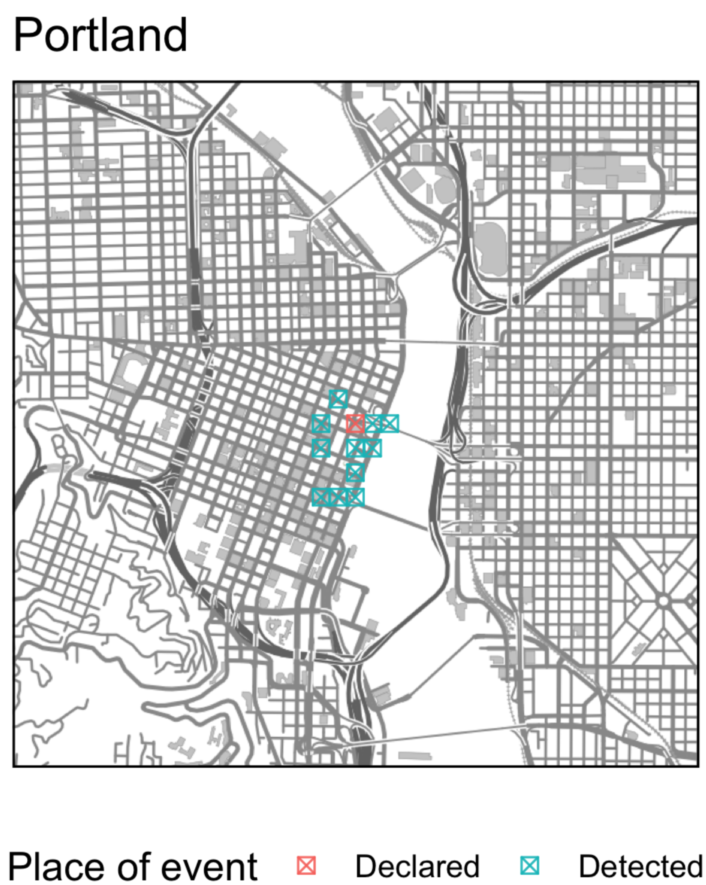


Figure A17

Figure A18

Figure A19



S10 Lower Geohash Density Threshold

This section replicates the papers' main result, using a lower geohash density threshold for inclusion as a protest geohash. In the initial setting, we defined the number of protesters as the number of people in a city's densest 7-digit geohashes (those whose density is in the 99^{th} percentile or greater). Here we change the threshold and define a city's densest 7-digit geohashes as those density is in the 95^{th} percentile or greater. Results do not change.

Table A21: Measuring Fit

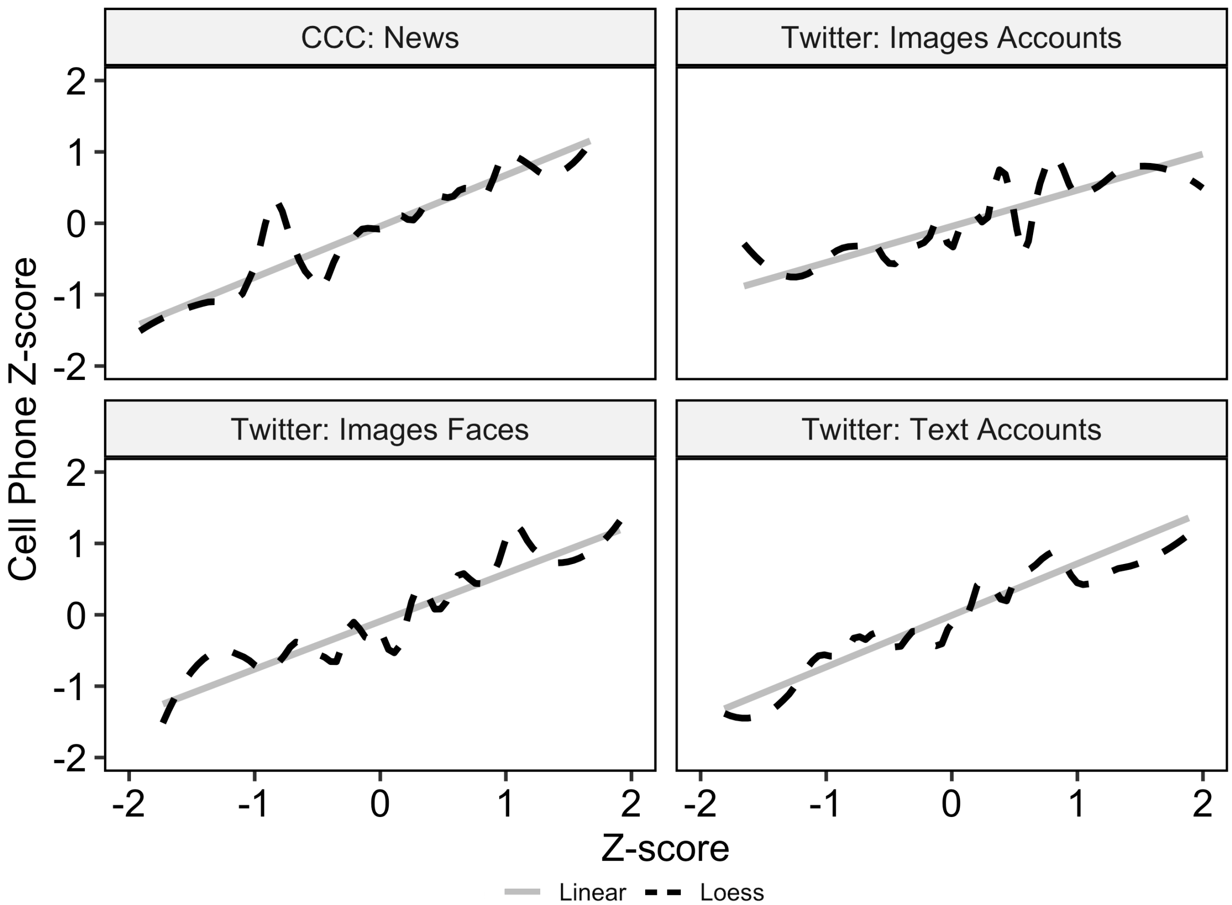
	CCC: News	*Twitter: Images Accounts*	*Twitter: Images Faces*	*Twitter: Text Accounts*
Mean Error	<u>-0.04</u>	0.02	-0.01	**0.00**
Mean Error (Trimmed)	-0.05	0.07	<u>0.08</u>	**-0.03**
Mean Absolute Error	**0.83**	<u>0.99</u>	0.97	0.86
Mean Absolute Error (Trimmed)	**0.81**	<u>0.92</u>	0.86	0.84
Closest Estimate %	**0.33**	0.31	0.19	<u>0.18</u>
Within 0.1 SD	0.08	<u>0.04</u>	**0.09**	0.08
Within 0.5 SD	**0.34**	<u>0.27</u>	0.30	**0.34**
Within 1 SD	**0.70**	<u>0.57</u>	0.66	0.68

Note: "(Trimmed)" datasets drop observations where $|z| > 3$. **Bold** is the best measure per row; <u>underline</u>, the worst.

S11 Non-Linearity

Figure A20 plots a loess curve (dashed line) against the linear fit (solid) from Figure 1. We use a window of twenty percent observations to calculate the loess smoother. There appears to be no non-linearity in the estimates produced in this study.

Figure A20: Correlation of Estimates: Linear VS Non-Linear



Supplementary Material References

Chen, Hanbo. 2018. "VennDiagram.".

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.

Clark, David H. and Patrick M. Regan. 2018. "Mass Mobilization Protest Data.". **URL:** *https://www.binghamton.edu/massmobilization/about.html*

Cook, Scott J. and Nils B Weidmann. 2019. "Lost in Aggregation: Improving Event Analysis with Report-Level Data." *American Journal of Political Science* 63(1):250–264.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Luhrmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daneil Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Steven Wilson, Agnes Cornell, Lisa Gastaldi, Haakon Gjerlow, Nina Ilchenko, Joshua Krusell, Laura Maxwell, Valeriya Mechkova, Juraj Medzihorsky, Josefine Pernes, Johannes von Romer, Natalia Stepanova, Aksel Sundstrom, Eitan Tzelgov, Yiting Wang, Tore Wig and Daniel Ziblatt. 2019. V-Dem [Country-Year/Country-Date] Dataset v9. Technical report Varieties of Democracy (V-Dem) Project.

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.

Jones, Sam. 2019. "E-mail correspondence.".

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.

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

Weidmann, Nils B. and Espen Geelmuyden Rod. 2018. *The Internet and Political Protest in Autocracies*. Oxford University Press.