Mask Mandates Work, Especially in Republican Counties

Xiaofeng Lin[a], Georgia Kernell[a,b,1], Tim Groeling[a], Jungseock Joo[a], Jun Luo[a], Zachary C. Steinert-Threlkeld[b,c]

[a] Department of Communication, University of California - Los Angeles, Los Angeles, CA 90024;
[b] Department of Political Science, University of California - Los Angeles, Los Angeles, CA 90024;
[c] Department of Public Policy, Luskin School of Public Affairs, University of California - Los Angeles, Los Angeles, CA 90024

[1] To whom correspondence should be addressed. Email: gkernell@ucla.edu.

Abstract

Wearing masks reduces the spread of COVID-19, but compliance with mask mandates varies across individuals, time, and space. Accurate and continuous measures of mask wearing, as well as other health-related behaviors, are important for public health policies. This article presents a novel approach to estimate mask-wearing using geotagged Twitter image data from March through September, 2020 in the United States. We validate our measure using individual survey data and extend the analysis to investigate county-level differences in mask wearing. We find strong evidence that mask mandates work – on average they increase mask wearing by 20% - and moreover that their effectiveness is greatest in Republican-leaning counties. The findings have important implications for understanding how governmental policies shape citizen responses to public health crises.

Significance Statement

We use Twitter images to measure when and where people wear masks in the United States. Our data includes every publicly available geocoded image posted on Twitter from March 1 through September 11, 2020 (8 million images). We find that Twitter images closely track polling data on aggregate mask wearing, but they have the added advantage of being free, available over a long period of time, and accessible retrospectively. Using multilevel regression, we find that mask mandates work, but they are especially effective for individuals living in Republican-leaning counties.

Introduction

Widespread mask wearing greatly reduces COVID-19 transmission [1–6]. Accurate and continuous measures of mask use are therefore necessary for public health agencies to understand and predict outbreaks, identify susceptible populations, and formulate timely policy responses. Throughout the COVID-19 pandemic, health and government officials, as well as the general public, received real-time access to important information such as hospitalizations, deaths, and vaccination rates. Yet, data on preventative behavior is largely retrospective or unknown.

Our study addresses this shortcoming by presenting a novel way to measure individual-level behaviors in real-time using geolocated social media images. The contribution is threefold. First, we develop an automated image classifier using a convolutional neural network (CNN) to detect images of people wearing masks and apply this classifier to geotagged Twitter data from March 1 through September 11, 2020. Twitter data were collected in real time and represent all publicly available (and approximately one-third of actual) geolocated Tweets from the United States at this time. Second, we demonstrate that social media behavior closely tracks survey data using YouGov's COVID-19 Public Monitor and UCLA's Nationscape. In doing so, we also find that several individual-level correlates of mask wearing documented in observational research – age, gender, and partisanship – are also predictive of mask wearing on Twitter. Third, we investigate county-level factors that may impact widespread mask wearing. With 170,014,835 Tweets from 2,603,654 users, Twitter data allow us to examine less populated areas commonly underrepresented in surveys. We combine this information on local mask wearing with county-level data on mask mandate policies, COVID-19 death rates, voting behavior, and individual mobility, as well as national media attention. While we find that a county's 2016 GOP presidential vote share is negatively associated with mask wearing overall, the introduction of mask mandates in Republican-leaning counties leads to larger increases in mask-wearing images than in counties that lean Democratic. National media attention and urbanization are also predictive of mask wearing, though local measures of population density and mobility are not.

Results

Validating Twitter Image Data Using National Surveys

To assess how well Twitter image data captures mask-wearing trends across the nation, we compare aggregate levels of mask wearing in our sample with self-reports from two nationally-representative surveys run during this period: YouGov's COVID-19 Public Monitor [7] and UCLA Nationscape [8]. These online surveys asked individuals whether or not they wore a mask in public during the past week.[1] The Twitter measure is the number of users who posted at least one mask image in the span of a week divided by the number of users who posted at least one image at all during the same time period.

Fig 1 reveals a strong association between the estimates using Twitter images and both surveys.[2] Both Twitter and YouGov reveal similar increases in public mask

[1] YouGov's survey stated: "Thinking about the last 7 days, how often have you: worn a face mask outside your home to protect yourself or others from coronavirus (COVID-19)?" Response options include "always," "frequently," "sometimes," "rarely," and "not at all." The UCLA Nationscape survey asks respondents "Have you done any of the following in the past week?" with "Worn a mask when going out in public" as one of the possible categories and a binary response option of "Yes" or "No."

[2] Fig 1 uses dual y-axes to accommodate different magnitudes of surveys and Twitter data. Past research [9] has illustrated that trends estimated using geolocated Twitter data can be excellent predictors of real world events, despite sometimes drastically different levels of measurement.

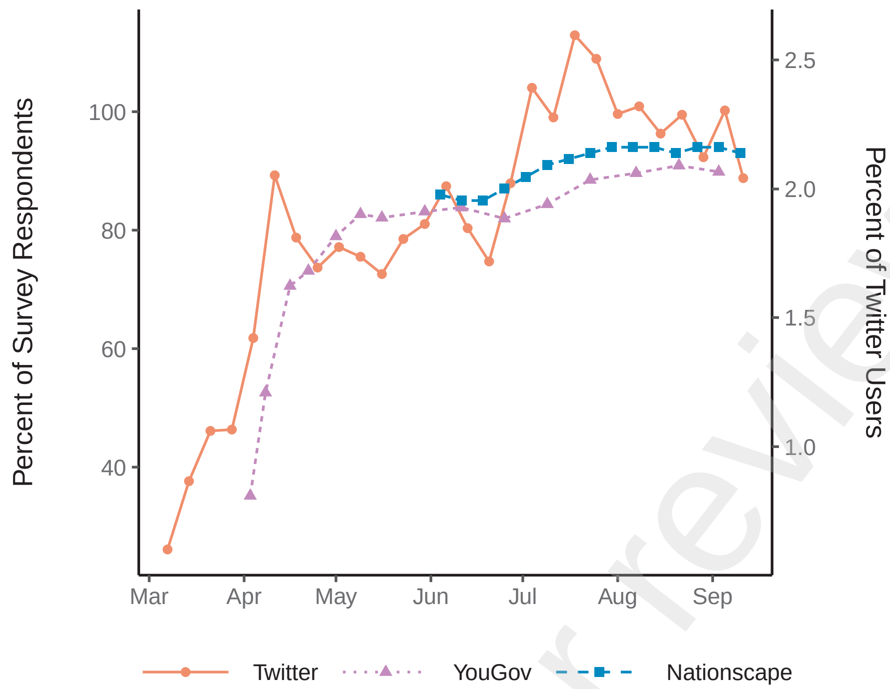

Fig 1. Mask wearing from March through September, 2020. The left vertical axis shows percentages of survey respondents who reported wearing a mask in public. The right vertical axis shows the percentage of Twitter users who posted a mask image (among those who posted at least one image in a geotagged Tweet).

wearing from April to May. (Nationscape did not ask about mask wearing for the first half of the data.) All estimates increased at a similar level from June to September. High correlations between mask usage on Twitter and self reports from YouGov ($r = 0.60$) and Nationscape ($r = 0.79$) suggest that social media images are highly predictive of similar behaviors offline.[3]

Individual Level Determinants

Previous research examining individual-level factors finds that older individuals are more likely to wear masks than their younger counterparts [8, 11–13].[4] Females are also more likely than males to don a mask [12, 16–18], and Democrats are consistently more likely than Republicans to not only to wear a mask, but to support mask mandates [16–18]. To further validate our data, we compare mask wearing across age, gender, and political ideology for Twitter users. A user's age and gender are estimated by applying a deep neural network facial classifier trained by FairFace [19] on the user's profile image. *Conservative Leaning* measures a user's ideology based on the number of influential partisan-leaning Twitter accounts that individual follows. Users with a *Conservative Leaning* score greater than 0 are generally considered Republicans [20].

Table 1 reports odds ratios for each demographic group in relation to the specified reference group. In line with previous research, we find that seniors, females, and

[3]Moreover, the Twitter data correlate with a different YouGov study running from March through September at $r = 0.84$. We do not include this study in Figure 1 because the individual-level data is not publicly available. [10]

[4]Although see [14, 15].

Democrats demonstrate significantly higher likelihoods of mask wearing.[5] See the Materials and Methodology section for details on sampling individual users.

Variable	Odds Ratio	95% CI
Age: 30-49	1.57	(1.49, 1.66)
Age: 50 plus	2.14	(1.99, 2.30)
Female	1.35	(1.28, 1.43)
Democrat	1.65	(1.60, 1.70)
N (Ages and Genders)		35321
N (Partisanship)		68076

Table 1. Odds Ratios. Odds ratios for each demographic group against the corresponding reference group (age 20-29. male, Republican leaning). The numbers of missing values in profiles and ideological leaning are different, resulting in different sample sizes.

Contextual Determinants: Mask Mandates, Mobility, Death Rates, and County Partisanship

To examine contextual factors that shape mask wearing, we calculate the percent of geotagged image-posting Twitter users who tweet at least one mask image for each county-week in the data. The denominator is users posting any image in order to control for unobserved heterogeneity that may reflect differences between users who share images (of anything) and those who do not.

Mask mandate data come from [21]. This database includes the earliest effective dates of local and state mask mandates for every U.S county and runs through August 4, 2020. Figure 2 shows the share and distribution of counties, as well as the corresponding share of the U.S. population, under an active county or state mandate (or both) from May 1 through August 1, 2020. As of August 1, 2020, 66.3% of U.S counties, and 87.4% of the U.S. population were under active mask-wearing mandates.

Table 2 presents summary statistics for each of the continuous explanatory variables of interest. Subscript i denotes different counties, t denotes different weeks, and $t-1$ indicates a lag of one week. *Deaths per 10k* is the number of reported COVID-19 deaths in the county during the prior week per 10,000 residents [22]. *GOP Vote 2016* is the share of the two-party vote that went to Donald Trump in the 2016 presidential election. *Urban Population Percentage* [23] is the percent of a county's population living in a Census-defined urban area. *Population Density* [24] is the county's population (using 2010 Census data) divided by its area in square miles. Media coverage of the pandemic may drive public attention to COVID-19 and increase mask wearing. *COVID News* measures the average number of times the word "COVID" is used per news show on ABC, CBS, and NBC [25]. Last, images of mask wearing may decrease as people venture outside their homes. We use Safegraph's Places of Interest dataset to measure changes in activity [26]. Specifically, *Retail Visit per Hundred* measures the number of visits to a grocery store, as defined by those with NAICS business codes 44 or 45, in a county during the current week per 100 residents. In addition, the following models include a *Week Counter*, which ranges from 1 to 22 and is not shown in Table 2, to control for unobserved time trends [27].

Like population, Tweets are imbalanced across counties, making sample sizes too small to use in the majority of the country's 3,006 counties. As a result, the model is

[5] FairFace is also capable of classifying race, but the accuracy on Twitter profiles is unsatisfactory (64% on a random sample of 1000 profiles). The odds ratios between race groups classified by FairFace are not significant in our data.

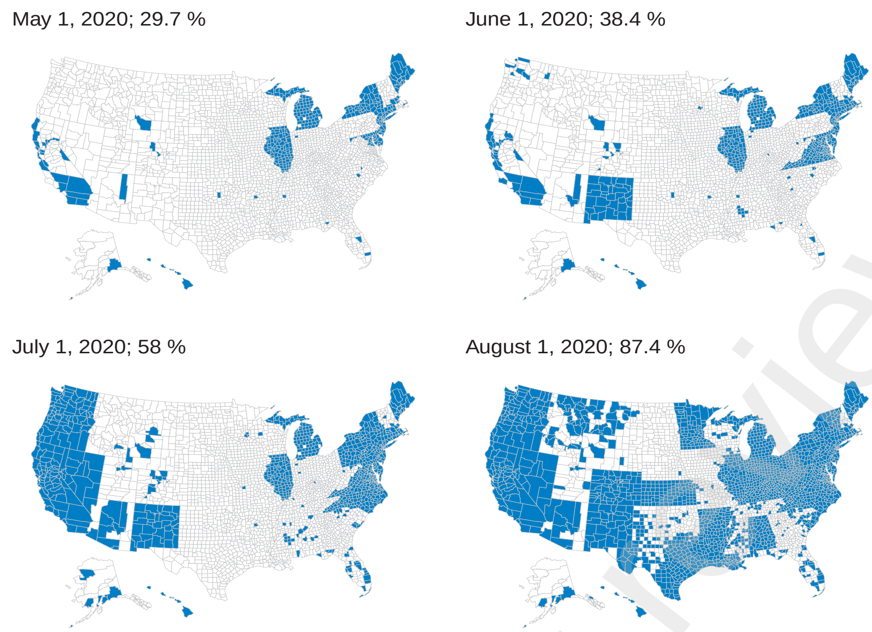

Fig 2. Counties under a mask mandate. Percentages indicate the share of the U.S. population under a mask mandate on the corresponding date.

	Top 100 Counties	Top 300 Counties	Top 500 Counties
Deaths per 10k$_{i,t-1}$	0.2.2	0,1.6	0,1.2
GOP Vote 2016$_i$	9.8, 60.5	15.3, 70.1	16.7, 77.4
Urban Population Percentage$_i$	89.2, 100	70, 100	54, 100
Population Density$_i$	106.7, 32903.3	81.7, 11379.5	50.9, 7671.5
COVID News$_t$	0.9, 11	0.9, 11	0.9, 11
Retail Visit per Hundred$_{i,t}$	574.7, 3444.1	674.2, 4240.2	739.9, 4664.7

Table 2. Explanatory Variables 95% Intervals for the 100, 300, and 500 most populous counties.

run for the 100, 300, and 500 most populous counties, which together account for approximately 43, 64, and 76% of the United States' population. The Materials and Methods section provides the power calculation informing this decision. There are no substantive differences found when we include all counties, as shown in table S1.

Table 3 presents results from a multilevel generalized linear model in the binomial family with random effects for week, county, and state.[6] Let us start by examining the full effect of mask-wearing mandates and county vote. (Because we include an interaction between *Mask Mandate* and *GOP Vote 2016*, we must interpret the effect of the two variables jointly. For example, the negative coefficient on *Mask Mandate* in the third model can only be interpreted for a county with zero percent GOP vote, which is never the case.) Under no mask mandate, county level Republican leaning shows a negative, significant association with mask wearing: for a one standard deviation increase (12.8 percentage points) in GOP vote share in the 2016 presidential election, the expected level of mask wearing decreases between 6.2 and 9.0 percent across the three models. Not surprisingly, mask wearing is lower in Republican strongholds, all else

[6] Our dependent variable is a proportion. A generalized linear model with a binomial family is ideal for modeling proportions that are otherwise prone to heteroskedasticity while accommodating lower and upper bounds at 0 and 1.

	Top 100 Counties	Top 300 Counties	Top 500 Counties
Intercept	$\mathbf{-5.508}^{***}$	$\mathbf{-5.066}^{***}$	$\mathbf{-5.025}^{***}$
	(0.684)	(0.256)	(0.183)
Mask Mandate	0.005	−0.052	$\mathbf{-0.070}^{**}$
	(0.032)	(0.027)	(0.026)
GOP Vote 2016	$\mathbf{-0.005}^{*}$	$\mathbf{-0.006}^{***}$	$\mathbf{-0.006}^{***}$
	(0.002)	(0.001)	(0.001)
Mask Mandate * GOP Vote 2016	$\mathbf{0.002}^{**}$	$\mathbf{0.003}^{***}$	$\mathbf{0.003}^{***}$
	(0.001)	(0.001)	(0.001)
Deaths per 10k	0.010	0.011	0.013
	(0.008)	(0.007)	(0.007)
Week Counter	$\mathbf{0.025}^{***}$	$\mathbf{0.029}^{***}$	$\mathbf{0.031}^{***}$
	(0.007)	(0.006)	(0.006)
Urban Population Percentage	0.008	$\mathbf{0.004}^{*}$	$\mathbf{0.005}^{***}$
	(0.007)	(0.002)	(0.001)
Population Density	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
COVID News	$\mathbf{0.068}^{***}$	$\mathbf{0.064}^{***}$	$\mathbf{0.060}^{***}$
	(0.015)	(0.014)	(0.014)
Retail Visit per Hundred	$\mathbf{0.057}^{*}$	0.008	−0.021
	(0.025)	(0.016)	(0.013)
AIC	14879.610	32527.223	43410.279
BIC	14953.661	32615.556	43505.252
Log Likelihood	−7426.805	−16250.612	−21692.139
County-Weeks	2200	6600	11000
States (Including D.C)	31	45	50
County Intercept	0.024	0.037	0.043
State Intercept	0.008	0.015	0.018
Week Intercept	0.030	0.028	0.025

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3. Contextual effects of mask-wearing images. The dependent variable is the proportion of Twitter users who posted mask wearing images, among users that posted any Tweets with images.

equal.

The introduction of a mandate, however, is strongly associated with an increase in mask wearing. Moreover, this effect is especially large in counties that favored Donald Trump in 2016. To illustrate these interactive effects, Figure 3 shows the marginal effect of a mandate for varying levels of GOP vote share. Moving from the county with the lowest percentage vote for the GOP to the county with the highest vote increases the effect of mandates by 25 percentage points. The marginal effect is stronger when less populous counties are included, as adding more counties increases the range of GOP vote share. Note that although mask mandates have a greater effect in Republican counties, the overall level of mask wearing remains higher in Democratic counties. Indeed, Republican counties with mask mandates display similar levels of mask wearing to Democratic counties without mandates. These results suggest that in majority-Democratic counties individuals are more likely to wear a mask regardless of whether there is a legal mandate requiring them to do so or not, whereas in majority-Republican counties mandates actually cause people to put on a mask, thus narrowing the gap caused by partisanship. [7]

News coverage has the second strongest effect: when the average occurrence of the word "COVID" increases by one standard deviation (2.9 times), mask wearing increases

[7] Using a binned estimator [45] to evaluate the marginal effects produces similar results.

by 19% to 21.8% across the three models.[8] To a large extent, the news variable mirrors what people are Tweeting about. In particular, COVID-19 news decreased by approximately 60% from week 13 to week 16 during the protests following George Floyd's death, and mask wearing images decreased during this time as well.

Mask wearing also increases with urbanization. A one standard deviation increase in urbanization (25 percentage points) corresponds to between a 10.4% and 13.3% increase across models. Surprisingly, the remaining population variables - deaths and density - are not statistically significant in most models. Mask-wearing images also do not decrease with higher rates of grocery store visits. Not surprisingly, the positive effect of the linear week counter demonstrates an overall increasing trend in public mask wearing. As time goes by, mask wearing becomes a habit for many individuals. Table S3 shows the estimation with a logged week counter and demonstrates the same effect of time.

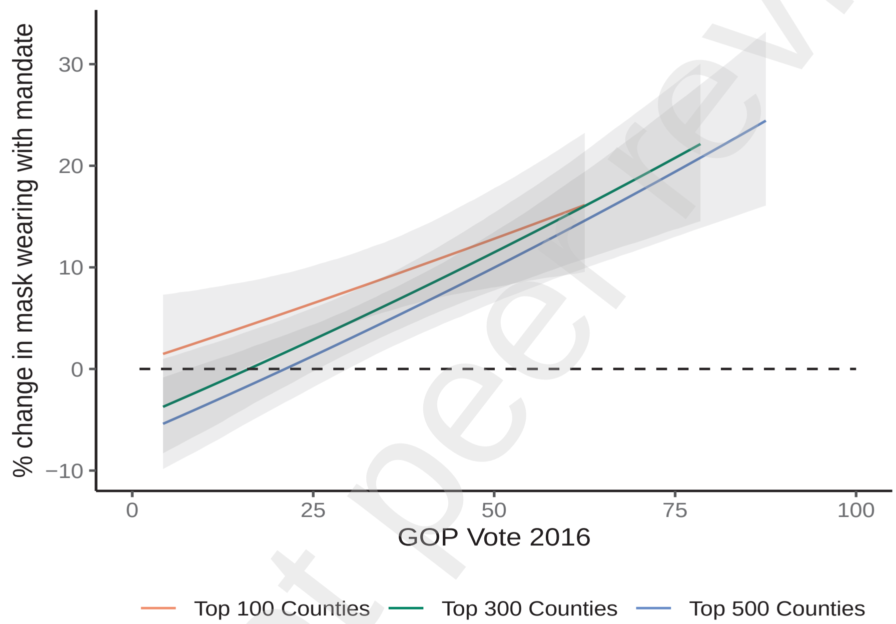

Fig 3. Marginal Effect of Mask Mandates for Varying GOP County Vote. Shaded areas represent 95% confidence intervals.

Discussion

Using Twitter, and digital trace data more broadly, confers a number of benefits. In contrast with surveys, Twitter data do not rely on retrospective responses, which often suffer due to memory loss and social desirability biases. Twitter data provide a measure that is free and can be collected and stored on demand in real time. The new Academic Search Product also provides any published Tweet that is still public, facilitating post hoc studies.

In addition, the sheer volume of Twitter data allows researchers to estimate effects across groups that would otherwise be prohibitively costly. In this study, we were able to investigate the effects of ideology, mobility, mandates, and death rates in counties that might normally have too few respondents for standard survey analysis. Though

[8] Table S2 shows the same regression models using cable rather than network news coverage and finds the same positive and significant association with mask wearing.

Twitter is not a representative sample of Americans [28], biases do not appear to affect its appropriateness as a sensor for large-scale social behaviors [29]. Moreover, social media image data provide an invaluable and unobtrusive way to measure public health behaviors [30]. To our knowledge, this is among the earliest work to collect COVID-19 Twitter image data and apply deep learning based image classification [31], and the first to measure public health behaviors related to COVID-19 with such methods. Future research may employ social media images to identify other behaviors or attributes, such as smoking, alcohol or drug use, obesity, and seat-belt compliance.

The principal finding that mask mandates matter most in areas with more Republican voters demonstrates that governmental policies have the ability to equalize health behaviors, and potentially health outcomes. We know that Conservative voters are less likely to vaccinate against COVID-19 [32], and early research suggests that vaccine mandates are effective at increasing compliance [33, 34]. Although systematic county-level analyses have yet to be conducted, our study suggests that vaccine mandates will likely have the greatest effect in GOP strongholds.

Last, it is worth noting that Twitter postings are not simply a random slice of life, but might also reflect individuals' strategic choices aimed at influencing friends or public opinion. These decisions likely change over time in response to the prevalence and politicization of mask wearing, and future research would benefit from exploring a dynamic relationship between mask images, mandates, and the political climate.

Materials and Methods

Geolocated Twitter Database. We compiled a dataset for all geocoded publicly-available Tweets from March 1 through September 11, 2020, collected in real time using Twitter's POST statuses/filter streaming endpoint. Twitter returns all Tweets up to a 1% ceiling *set by the number of all Tweets* and approximately 3% of Tweets in English are geocoded [35, 36]. This process results in 170,014,835 Tweets from 2,603,654 U.S. users. Of these, 18,968,029 Tweets contained an image, and 1,451,591 users posted at least one image.

Each geotagged Tweet contains a global bounding box of coordinates identifying the Tweet's location. However, the size of the bounding box varies because users can specify one of five levels to geolocate their Tweets: country, state, city, neighborhood, or point of interest. We dropped all Tweets with boxes above the city level (7% of Tweets), because they were too large to provide county-level location. For a city that spans multiple counties (14% of Tweets that specify the city level), we assign the Tweet to the county that covers most of the city's population.

Mask Detection. To automatically detect mask-wearing images, we developed an image classifier using a convolutional neural network (CNN) with the ResNet-50 architecture [37]. CNNs have been widely used for automated visual content analysis including facial mask detection [38–40]. We took a supervised learning approach to collect and annotate images with mask-wearing labels and train and evaluate our model. The quality of training data is critical to ensure the model's performance. Therefore, we used an iterative approach to collect diverse and challenging data to make the final model robust.

To this end, we first collected $8,000$ images using Google's image search API for three mask-related keyword phrases ('wearing masks,' 'face covering,' and 'mask selfie') and three keyword phrases not related to mask wearing ('selfie,' 'hangout,' and 'stay at home'). Approximately $1,300$ images were collected for each phrase. Figure 4 shows sample images selected in this initial step. The first two rows of images contain a mask-wearing face. Others do not, but these negative images likely contain similar visuals such as human faces or masks. Such "hard" negatives help train a more robust classifier (see [41], for example).

We then manually annotated each of the 8,000 images, dividing them into positive and negative samples. Images in the training set are annotated with a binary, non-exclusive attribute: Mask. To receive this label, an image must show at least one human face looking at the camera or occupying at least 5% of its area and at least one of those faces must wear a covering over the mouth and nose. An initial ResNet-50 classifier was trained on these 8,000 images with a binary cross entropy loss. Optimization was reached with an SGD optimizer, 500 training epochs, learning rate = 0.005, weight decay = 0.0004, and momentum = 0.9. We used a model pre-trained on Imagenet data [42]. This classifier was then applied to all geotagged Tweet images from March $1-5$, April $1-5$, and May $20-25$. Out of these 1,272,334 total images, 9,391 were classified as containing masks and were again manually verified, annotated, and added to the training set as true positive (5,032 images) and false positive (4,359 images) examples, resulting in the final set of 17,391 images. We did not include images predicted as negative because there were too many of them and the recall was very low. Finally, we randomly divided the annotated images in a training set (80%, 13,913 images) and a validation set (20%, 3,478 images). These labeled images further fine-tune the initial classifier with the same training parameters. One coder completed the labeling mentioned above, and two coders verified this work by annotating a 1000-image random subset of the training set. The Krippendorff's alpha estimated on this random subset was 0.852, showing strong inter-rater reliability.

Last, we chose an optimal decision threshold in the following way because naively choosing 0.5 as the threshold may yield an undesirable result (e.g., a too high false positive rate). Table 4 displays the classifier performance when using different decision thresholds on the same set. Since thresholds 0.7, 0.8, and 0.9 produce similarly high F1 scores (0.85, 0.86, and 0.85, respectively), we use a threshold of 0.9 in order to maximize precision (0.93).[9] With this threshold, the model has high classification accuracy, as the Receiver Operating Characteristic graph in Figure 5 shows.

Threshold	0.50	0.60	0.70	0.80	0.90
Precision	0.77	0.81	0.86	0.92	0.93
Recall	0.87	0.85	0.83	0.81	0.77
F1 Score	0.82	0.83	0.85	0.86	0.85

Table 4. Test accuracy using different thresholds. Precision is calculated as proportion of predicted mask images that actually shows a mask. Recall is the proportion of actual mask images also predicted as a mask image. F1 score = 2 × (Precision × Recall) / (Precision + Recall).

Mask Mandate Coding. For the county level analysis, the unit of observation is the county-week dyad, which covers a Monday-Sunday week. Each observation is coded as having a mask mandate in place if a mandate started on or before the Thursday of that week.

Case-control Sampling. For the individual-level analyses, we perform a case-controlled sampling design by collecting all Tweets with mask-wearing images, as well as a random sample of 1000 Tweets per day from the set of non-mask images [43]. Then we extract demographic features of users who posted these Tweets to calculate odds ratios.

[9] Among all Twitter images, the proportion showing a mask is low. We manually verified a random subset of 1,000 geotagged Twitter images posted on Aug 1, 2020 and found only 11 positive cases. To maximize our model precision in the face of potentially high false positives, we used the 0.9 threshold.

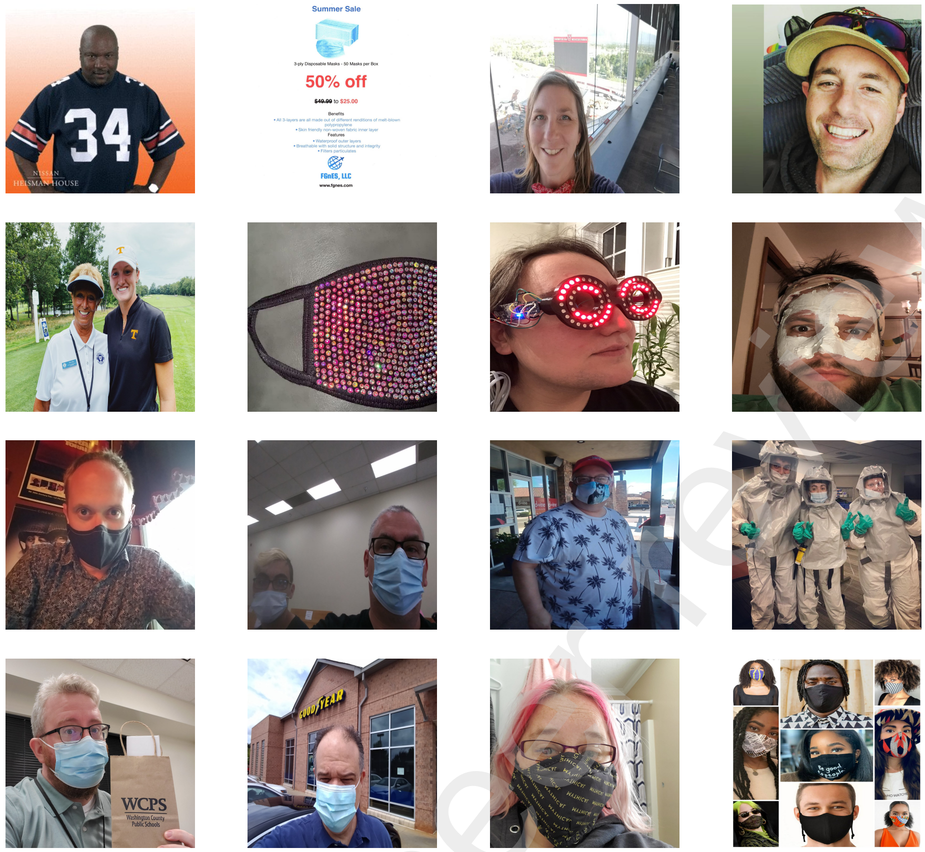

Fig 4. Examples of training set images. The first and second row show negative samples (no mask wearing); the third and forth row show positive samples (mask wearing).

Power calculation. To determine the number of Tweets with images needed per county, the following calculation is performed. The null and alternative hypothesis proportions are $p_\mu = 0.005$ and $p_\alpha = 0.02$, respectively, which are equal to the approximate minimum (p_μ) and maximum (p_α) values of the nationally aggregated mask wearing on Twitter. Let the level of significance $\alpha = 0.05$ and the desired power $= 1 - \beta = 0.9$, where $\beta = 0.1$ is the probability of committing a Type II error. Using Cohen's arcsine transformation [44], the effect size is $h = 2 * (\text{arcsin}\sqrt{p_\alpha} - \text{arcsin}\sqrt{p_\mu}) = 0.142$. Under the null hypothesis, μ_0, the decision boundary $\hat{p}$ should have z-score Z_0, where $P(Z < Z_0) = 1 - \alpha \div 2 = 0.975$ for $Z \sim N(0, 1)$, so $Z_0 = 1.959$. Under the alternative hypothesis, μ_a, $\bar{X}$ should have a z-score Z_a where $P(Z < Z_a) = \beta = 0.1$, so $Z_a =$ -1.281. Thus the required sample size $n = (\frac{Z_0 - Z_a}{h})^2 = 519$. In our sample, 488 of 3,006 counties posted at least 519 Tweets with images.

Acknowledgement

Funding: This work was supported by National Science Foundation SBE/SMA #1831848.
Competing interests: The authors declare no competing interests.
Author contributions: Conceptualization: X.L., G.K., T.G., J.J, J.L., and Z.C.S.T. Research and Writing: X.L. and G.K. Analyses and Figures: X.L. Image Classifier: J.J.

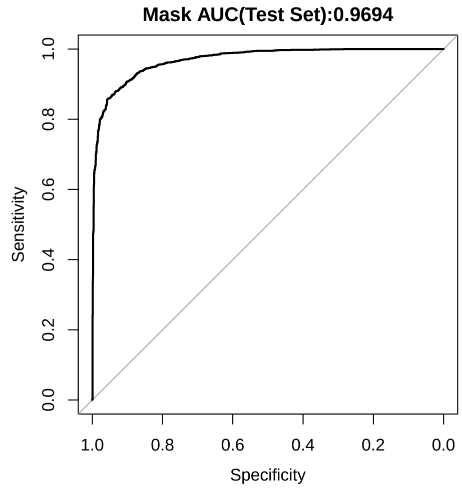

Fig 5. ROC curve for mask classifier on the validation set. Sensitivity is the fraction of actual positive data points that the model correctly classifies as positive. Specificity is the fraction of actual negative data points that the model correctly classifies as negative. Higher AUC values are associated with a greater ability to distinguish positive from negative examples.

and X.L. Twitter Data Acquisition: Z.C.S.T. Review and Editing: X.L., G.K., T.G., J.J., J.L., and Z.C.S.T. Funding Acquisition: J.J.
Data and materials availability: All data and replication files will be made publicly available on the Dryad website.

References

1. J. Howard, A. Huang, Z. Li, Z. Tufekci, V. Zdimal , H. M. van der Westhuizen , A. von Delft , A. Price , L. Fridman , L. H. Tang, V. Tang , G. L. Watson , C. E. Bax , R. Shaikh, F. Questier, D. Hernandez, L. F. Chu, C. M. Ramirez , A. W. Rimoin, An evidence review of face masks against COVID-19. *Proceedings of the National Academy of Sciences* **118** (4) (2021).

2. J. T. Brooks, J. C. Butler, Effectiveness of mask wearing to control community spread of SARS-CoV-2. *JAMA* **325**, 998-999 (2021).

3. National Center for Immunization and Respiratory Diseases, Division of Viral Diseases, "Scientific Brief: Community Use of Cloth Masks to Control the Spread of SARS-CoV-2." https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/masking-science-sars-cov2.html (2021).

4. V. Costantino, C. Raina MacIntyre, The impact of universal mask use on SARS-COV-2 in Victoria, Australia on the epidemic trajectory of COVID-19. *Frontiers in Public Health* **9**, 307 (2021).

5. W. Lyu, G. L. Wehby, Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US. *Health Affairs* **39**, 1419-1425 (2020).

6. J. Abaluck, L. H. Kwong, A. Styczynski, A. Haque, M. A. Kabir, E. Bates-Jeffries, E. Crawford, J. Benjamin-Chung, S. Raihan, S. Rahman, S. Benhachmi,

N. Zaman, P. J. Winch, M. M. Hossain, H. M. Reza, A. A. Jaber, S. G. Momen, F. L. Bani, A. Rahman, T. S. Huq, S. P. Luby, A. M. Mobarak, The impact of community masking on COVID-19: A cluster randomized trial in Bangladesh. https://elischolar.library.yale.edu/cowles-discussion-paper-series/2642 (2021).

7. S. P. Jones, Imperial College London YouGov COVID Data Hub, https://github.com/YouGov-Data/covid-19-tracker, Imperial College London Big Data Analytical Unit and YouGov Plc (2020).

8. C. Tausanovitch, L. Vavreck, Democracy Fund + UCLA Nationscape. https://www.voterstudygroup.org/covid-19-updates# (2020).

9. A. Sobolev, M. K. Chen, J. Joo, Z. C. Steinert-Threlkeld, News and geolocated social media accurately measure protest size variation. *American Political Science Review* **114**, 1343–1351 (2020).

10. YouGov, "Personal Measures Taken to Avoid COVID-19." https://today.yougov.com/topics/international/articles-reports/2020/03/17/personal-measures-taken-avoid-covid-19 (2021).

11. H. J. Hutchins, B. Wolff, R. Leeb, J. Y. Ko, E. Odom, J. Willey, A. Friedman, R. H. Bitsko, COVID-19 mitigation behaviors by age group — United States, April–June 2020. *Morbidity and Mortality Weekly Report* **69**, 1584–1590 (2020).

12. M. H. Haischer, R. Beilfuss, M. R. Hart, L. Opielinski, D. Wrucke, G. Zirgaitis, T. D. Uhrich, S. K. Hunter, Who is wearing a mask? gender-, age-, and location-related differences during the COVID-19 pandemic. *PLOS ONE* **15**, e0240785 (2020).

13. E. S. Knotek II, R. Schoenle, A. Dietrich, G. Müller, K. O. R. Myrseth, M. Weber, Consumers and COVID-19: Survey results on mask-wearing behaviors and beliefs. *Economic Commentary (Federal Reserve Bank of Cleveland)* pp. 1–7 (2020).

14. C. R. MacIntyre, P. Y. Nguyen, A. A. Chughtai, M. Trent, B. Gerber, K. Steinhofel, H. Seale, Mask use, risk-mitigation behaviours and pandemic fatigue during the COVID-19 pandemic in five cities in Australia, the UK and USA: A cross-sectional survey. *International Journal of Infectious Diseases* **106**, 199-207 (2021).

15. M. C. Howard, The relations between age, face mask perceptions and face mask wearing. *Journal of Public Health (Oxford, England)* **fdab018** (2021).

16. A. Naeim, R. Baxter-King, N. Wenger, A. L. Stanton, K. Sepucha, L. Vavreck, Effects of age, gender, health status, and political party on COVID-19–related concerns and prevention behaviors: Results of a large, longitudinal cross-sectional survey. *JMIR Public Health and Surveillance* **7**, e24277 (2021).

17. M. Brenan, "Americans' Face Mask Usage Varies Greatly by Demographics." https://news.gallup.com/poll/315590/americans-face-mask-usage-varies-greatly-demographics.aspx, Gallup (2020).

18. S. Kramer, "More Americans Say They Are Regularly Wearing Masks in Stores and Other Businesses." https://pewrsr.ch/32ttrRi, Pew Research Center (2020).

19. K. Karkkainen, J. Joo, Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* pp. 1548–1558 (2021).

20. P. Barberá, Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* **23**, 76–91 (2015).

21. A. L. Wright, G. Chawla, L. Chen, A. Farmer, Tracking mask mandates during the COVID-19 pandemic. *University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-104* (2020). Available at SSRN:https://ssrn.com/abstract=3667149.

22. M. Ponce, A. Sandhel, covid19.analytics: An R package to obtain, analyze and visualize data from the Coronavirus disease pandemic. https://arxiv.org/abs/2009.01091 (2020).

23. U.S. Census Bureau, 2010 Census Urban and Rural Classification and Urban Area Criteria. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html (2010).

24. E. Magdel, U.S. Population Density. https://github.com/camillol/cs424p3 (2011).

25. J. Joo, F. F. Steen, M. Turner, Red Hen Lab: Dataset and tools for multimodal human communication research. *KI-Künstliche Intelligenz* **31**, 357-361 (2017).

26. Safegraph, Weekly Patterns. https://docs.safegraph.com/v4.0/docs/weekly-patterns (2020).

27. T. McGovern, S. Larson, B. Morris, J. Ro, M. Hodges, US County-level Presidential Election Results. https://github.com/tonmcg/US_County_Level_Election_Results_08-20 (2020).

28. M. M. Malik, H. Lamba, C. Nakos, J. Pfeffer, Population Bias in Geotagged Tweets. *9th International AAAI Conference on Weblogs and Social Media* pp. 18–27 (2015).

29. N. A. Christakis, J. H. Fowler, Social network sensors for early detection of contagious outbreaks. *PLOS ONE* **5**, e12948 (2010).

30. L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, R. M. Merchant, Twitter as a tool for health research: A systematic review. *American Journal of Public Health* **107**, e1 (2017).

31. S. Khurana, R. Chopra, B. Khurana, Automated processing of social media content for radiologists: Applied deep learning to radiological content on twitter during COVID-19 pandemic. *Emergency Radiology* **28**, 477-483 (2021).

32. S. G. Stolberg, "G.O.P. Seethes at Biden Mandate, Even in States Requiring Other Vaccines." https://www.nytimes.com/2021/09/12/us/politics/vaccine-mandates-republicans.html, The New York Times (2021).

33. A. Blake, "The Evidence is Building: Vaccine Mandates Work — and Well." https://www.washingtonpost.com/politics/2021/09/29/evidence-is-building-vaccine-mandates-work-well/, The Washington Post (2021).

34. A. Hsu, "Faced with Losing Their Jobs, Even the Most Hesitant are Getting Vaccinated." https://www.npr.org/2021/10/07/1043332198/employer-vaccine-mandates-success-workers-get-shots-to-keep-jobs, NPR (2021).

35. Z. C. Steinert-Threlkeld, *Twitter as Data (Elements in Quantitative and Computational Methods for the Social Sciences)* (Cambridge University Press, Cambridge, United Kingdom, 2018).

36. B. Huang, K. M. Carley, A large-scale empirical study of geotagging behavior on Twitter. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* pp. 365–373 (2019).

37. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition* pp. 770–778 (2016).

38. M. R. Bhuiyan, S. A. Khushbu, M. S. Islam, A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. *2020 11th International Conference on Computing, Communication and Networking Technologies* pp. 1–5 (2020).

39. S. Sethi, M. Kathuria, T. Kaushik, Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *Journal of Biomedical Informatics* **120**, 103848 (2021).

40. M. Loey, G. Manogaran, M. H. N. Taha, N. E. M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities Society* **65**, 102600 (2021).

41. A. Shrivastava, A. Gupta, R. Girshick, Training region-based object detectors with online hard example mining. *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 761–769 (2016).

42. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* pp. 248–255 (2009).

43. W. W. LaMorte, "Case-Control Studies." https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_analyticoverview/EP713_AnalyticOverview5.html#headingtaglink_1 (2017).

44. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, New York, NY, United States, 1988).

45. J. Hainmueller, J. Mummolo, Y. Xu, How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis* **27**, pp. 163–192 (2019).

Supporting Information

1 County Level Regression Using All Counties

	All US counties
Variable	Estimate(SE)
Intercept	$-\mathbf{4.809}^{***}$
	(0.129)
Mask Mandate	$-\mathbf{0.085}^{***}$
	(0.024)
GOP Vote 2016	$-\mathbf{0.005}^{***}$
	(0.001)
Mask Mandate * GOP Vote 2016	$\mathbf{0.003}^{***}$
	(0.001)
Deaths per 10k	$\mathbf{0.015}^{*}$
	(0.007)
Week Counter	$\mathbf{0.032}^{***}$
	(0.006)
Urban Population Percentage	$\mathbf{0.001}^{*}$
	(0.001)
Population Density	$\mathbf{0.000}^{*}$
	(0.000)
COVID News	$\mathbf{0.091}^{***}$
	(0.020)
Retail Visit per Hundred	$-\mathbf{0.032}^{***}$
	(0.008)
AIC	68033.740
BIC	68144.507
Log Likelihood	−34003.870
County-Weeks	37070
Counties	1685
States (Including D.C)	51
Weeks	22
County Intercept	0.052
State Intercept	0.011
Week Intercept	0.023

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table S1. Contextual effects of mask-wearing images. All counties with at least one user posted mask images are included.

2 County Level Regression with COVID News from FOX/CNN/MSNBC

	Top 100 Counties	Top 300 Counties	Top 500 Counties
Variable	Estimate(SE)	Estimate(SE)	Estimate(SE)
Intercept	$-\mathbf{5.512}^{***}$	$-\mathbf{5.075}^{***}$	$-\mathbf{5.036}^{***}$
	(0.683)	(0.256)	(0.182)
Mask Mandate	0.005	−0.052	$-\mathbf{0.069}^{**}$
	(0.032)	(0.027)	(0.026)
GOP Vote 2016	$-\mathbf{0.005}^{*}$	$-\mathbf{0.006}^{***}$	$-\mathbf{0.006}^{***}$
	(0.002)	(0.001)	(0.001)
Mask Mandate * GOP Vote 2016	$\mathbf{0.002}^{**}$	$\mathbf{0.003}^{***}$	$\mathbf{0.003}^{***}$
	(0.001)	(0.001)	(0.001)
Deaths per 10k	0.010	0.011	0.013
	(0.008)	(0.007)	(0.007)
Week Counter	$\mathbf{0.026}^{***}$	$\mathbf{0.029}^{***}$	$\mathbf{0.031}^{***}$
	(0.007)	(0.006)	(0.006)
Urban Population Percentage	0.008	$\mathbf{0.004}^{*}$	$\mathbf{0.005}^{***}$
	(0.007)	(0.002)	(0.001)
Population Density	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
COVID News	$\mathbf{0.105}^{***}$	$\mathbf{0.100}^{***}$	$\mathbf{0.093}^{***}$
	(0.023)	(0.022)	(0.021)
Retail Visit per Hundred	$\mathbf{0.057}^{*}$	0.008	−0.021
	(0.025)	(0.016)	(0.013)
AIC	14879.494	32526.891	43409.854
BIC	14953.545	32615.224	43504.828
Log Likelihood	−7426.747	−16250.446	−21691.927
County-Weeks	2200	6600	11000
States (Including D.C)	31	45	50
County Intercept	0.024	0.037	0.043
State Intercept	0.008	0.015	0.018
Week Intercept	0.030	0.027	0.025

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table S2. Contextual effects of mask-wearing images. *COVID News* in this model only data from FOX/CNN/MSNBC.

3 County Level Regression Using Logged Time

	Top 100 Counties	Top 300 Counties	Top 500 Counties
Variable	Estimate(SE)	Estimate(SE)	Estimate(SE)
Intercept	$\mathbf{-5.652}^{***}$	$\mathbf{-5.221}^{***}$	$\mathbf{-5.193}^{***}$
	(0.679)	(0.248)	(0.171)
Mask Mandate	0.000	$\mathbf{-0.054}^{*}$	$\mathbf{-0.072}^{**}$
	(0.032)	(0.027)	(0.026)
GOP Vote 2016	$\mathbf{-0.004}^{*}$	$\mathbf{-0.006}^{***}$	$\mathbf{-0.006}^{***}$
	(0.002)	(0.001)	(0.001)
Mask Mandate * GOP Vote 2016	$\mathbf{0.002}^{**}$	$\mathbf{0.003}^{***}$	$\mathbf{0.003}^{***}$
	(0.001)	(0.001)	(0.001)
Deaths per 10k	0.010	0.011	0.013
	(0.008)	(0.007)	(0.007)
Week Counter Log	$\mathbf{0.292}^{***}$	$\mathbf{0.318}^{***}$	$\mathbf{0.329}^{***}$
	(0.045)	(0.042)	(0.039)
Urban Population Percentage	0.008	$\mathbf{0.004}^{*}$	$\mathbf{0.005}^{***}$
	(0.007)	(0.002)	(0.001)
Population Density	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
COVID News	$\mathbf{0.042}^{**}$	$\mathbf{0.038}^{**}$	$\mathbf{0.034}^{**}$
	(0.013)	(0.012)	(0.011)
Retail Visit per Hundred	$\mathbf{0.053}^{*}$	0.006	−0.022
	(0.025)	(0.016)	(0.013)
AIC	14822.656	32422.481	43279.280
BIC	14896.707	32510.814	43374.253
Log Likelihood	−7398.328	−16198.241	−21626.640
County-Weeks	2200	6600	11000
States (Including D.C)	31	45	50
County Intercept	0.024	0.038	0.043
State Intercept	0.008	0.015	0.017
Week Intercept	0.017	0.014	0.013

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table S3. Contextual effects of mask-wearing images, with log week counter.