Twitter Detects Who is Social Distancing During COVID-19

Paiheng Xu,
Malone Center for Engineering in Healthcare
Center for Language and Speech Processing
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218
xpaihen1@jh.edu

David A. Broniatowski
Department of Engineering Management
and Systems Engineering
Institute for Data, Democracy, and Politics
The George Washington University
Washington, DC 20052
broniatowski@gwu.edu

Mark Dredze
Malone Center for Engineering in Healthcare
Center for Language and Speech Processing
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218
mdredze@cs.jhu.edu

Abstract

Adherence to social distancing practices are critical in stemming the spread of the COVID-19 pandemic, yet there remains mixed compliance in the United States with these recommendations. We utilize a recently released Twitter dataset on social mobility during COVID-19 [1] to study *who* may be social distancing and what factors correlate with this practice. We conduct an analysis of correlations between demographic and political affiliation with *reductions in mobility* as measured by public geolocation tweets to understand what factors may be at play. We find significant differences in mobility reduction between these groups, providing evidence to support public health policy-making.

1 Introduction

Without a vaccine to prevent new infections of the novel Coronavirus (COVID-19), adherence to social distancing practices are critical to efforts to stem the spread of the disease and control the scope of the pandemic [2, 3, 4]. Within the United States, public health officials have requested people to avoid large gatherings and limit contacts with others as part of a social distancing initiative [5]. However, mixed compliance with these recommendations limits their effect. Understanding who adheres to social distancing and what factors influence these practices may be critical to ensuring the effectiveness of these practices.

To understand social distancing practices public health officials have turned to online mobility data, a quantitative measure of travel patterns [6]. By measuring reductions and changes in movement, we can determine what kinds of messaging or policies are most effective. Survey data on mobility can be expensive to obtain and suffers from response bias [7]. Instead, GPS-based mobility data collected via mobile phones offers a massive, detailed indicator of mobility patterns, and has thus been widely used during the COVID-19 pandemic [8, 9, 10]. Despite its volume, analyses of these data are limited in that they cannot be correlated with attributes of individual users. Critically, we cannot answer *what type of person* has reduced their mobility.

Preprint. Under review.

Following a long line of research that uses social media data for public health [11, 12, 13, 14], work during the COVID-19 pandemic (and in previous epidemics [15]) has turned to Twitter as an alternative source of mobility data [1]. Twitter enables users to provide location information, and there are methods for automatic Twitter geolocation [16, 17, 18, 19], including work on patterns and trends in Twitter geotagged data [20].

An advantage to using publicly available Twitter location data is that it comes attached to content and user information. This information could enable studying how mobility changes correlate with individual characteristics that have been studied in the larger COVID-19 context, such as age [21, 22], income, race [23], and political affiliation [24, 25]. Prior work has shown how to infer these characteristics from Twitter data [26, 27, 28, 29], including gender [30, 31, 32], race [33, 34], age [35], and political affiliation [36, 37].

We use data from the Twitter Social Mobility Index [1] to study how demographic characteristics and political affiliation correlate with changes to mobility patterns, revealing insights on social distancing practices. We use demographic inference techniques, combined with user level mobility data, to examine how different groups have responded differently to the COVID-19 pandemic in the United States. Our findings can inform public health messaging and identify communities at higher risk from the virus.

Data and Privacy Our data are public tweets containing user provided geolocation information. To protect user privacy we remove all content and only use the information described in our analysis. We release these data to facilitate further research. The work conducted in this study falls under an exemption obtained from the Internal Review Board (IRB) at ANONYMIZED under 45 CFR 46 category 4.

2 User Level Mobility Data

Twitter Data We use data collected as part of the Twitter Social Mobility Index Project[1], which includes public geotagged data in the United States from January 1, 2019 to June 21, 2020. The index is computed by aggregating data for a user and measuring the standard deviation across locations within each week. Changes in mobility behavior are measured over time. See [1] for more details.

We select March 16, 2020 as the start of social distancing in the United States, since the national "Slow the Spread" guidelines announced on that date had the largest effect on mobility [1]. Furthermore, [10] showed mobility changes in many US counties before the state-level policies. We compare the time period before (January 1, 2019 - March 15, 2020) and after (March 16, 2020 - June 21, 2020) this date. We download the 3,200 most recent tweets for each of the 505,589 Twitter users in the collection who have mobility index in both before and after social distancing period. We exclude 51,447 users identified as organizations by either of the classifiers from [38, 35], leading to a total of 454,142 Twitter users.

Our user level dataset contains one entry for each user, including the mobility index, number of weekly geotagged tweets, and mean mobility index before and after the March 16, 2020. We augment each user in the dataset with demographic information.

Location The user's home city and state computed from the centroid of all of the geotagged tweets. In our analysis we use the home location to categorize a user as living in a high or low population density state, with a threshold set as the median US state population density. [2]

Age and Gender Age and gender are inferred using M3 [35], which uses both image (profile image) and text (name, username, user self description) features. We use the text-based model when the profile image is unavailable. For gender, the full model achieves 0.918 macro-F1 and the text-only model 0.907 when evaluated on heuristically-labeled self-report data. For age, the full model achieves 0.522 macro-F1 score. We use age categories of 18 and under, 19-29, 30-39 and over 39. In our ANOVA analysis we simplify this to be over/under 30.

[1] `http://socialmobility.covid19dataresources.org/`

[2] `https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density`

Race/Ethnicity We include categorical race/ethnicity based on the model of [34] which uses DistilBERT [39] to embed the latest 200 tweets of each user into a fixed-length representation, which is then passed through a logistic regression with l2 regularization. The model achieves 0.513 macro-F1 and 52.6% accuracy on a balanced dataset of self-reported race/ethnicity labels [33]. Both model and evaluation dataset provide the following race/ethnicity labels: white, black, Asian, Hispanic.

Political Affiliation We identify political affiliation (Democrat or Republican) using a method similar to [36]. We use three methods, and include each in the dataset. 1) A user is assigned a label if they follow a member of congressional leadership from *either* the Democrats – Nancy Pelosi (@SpeakerPelosi) or Chuck Schumer (@SenSchumer) – or the Republicans – Kevin McCarthy (@GOPLeader) or Mitch McConnell (@senatemajldr). Otherwise, they are assigned the label unknown. 2) The same approach but considering all current members of congress. [3] 3) Political affiliation assigned based on the home state's vote in the 2016 US presidential election. [4] We excluded President Trump's account from this method since it is very popular and widely followed.

Instead, we indicate if a user follows the US President as a separate field, which may be useful as recent work [40] suggests Trump supporters are less likely to accept a future COVID-19 vaccine.

We include two characteristics that reflect tweeted content.

COVID-19 Hashtags We indicate if this user tweeted or retweeted a COVID-19 hashtag in their most recent 3200 tweets. We collected hashtags usage from 81.1% of all the active users since 2020 March in [1], which is 1,103,749 users. We then manually identified COVID-19 hashtags by examining the 427 most popular hashtags whose total usage is above 30,000 tweets.

Social-distancing Hashtags We repeat the same process to identify social distancing hashtags. The hashtags for COVID-19 and social distancing are listed in Table 2 (Appendix A.)

3 Analysis and Result

Mobility Index Distribution Figure 1 shows the $\log(1 + \text{mean mobility})$ for before and after the start of social distancing for each characteristic in our dataset. In each subfigure, observe the reduction (blue vs. red) in mobility for each sub-group. For example, for age we observe a larger drop users who are older as compared to younger. People from high density states are more likely to reduce mobility, which may be because travel can be more easily limited in urban areas. For political affiliation, democrats have a much larger drop in mobility.

ANOVA Test While these figures give insights into individual characteristics, they do not say whether reductions are meaningful, and thus how we can compare mobility reductions across groups. Therefore, we run an analysis of variance (ANOVA) test, which tests whether two or more population means are different, typically employed when using categorical independent variables. Are the differences in mobility reduction among the demographics groups significant? We use difference in mobility (reduction) as the dependent variable. To gain more statistical power, we select age, gender, race, political affiliation considering all current members of congress, and state population density as independent variables. We binarize age into two brackets, i.e., people below or above 30, so that each intersection cell has sample size over 20. In this preliminary analysis we omit hashtags to ensure our group intersections have sufficient statistical power.

The results of the ANOVA test (Table 1) show that differences among all groups are significant, in large part due to our very large sample sizes for each group. These results show that male users, Asian and Hispanic users, older users, Democrats and people from higher population density states showed larger reductions in mobility. This suggests that different groups may be responding to social distancing recommendations differently. Additionally, the ANOVA test provides analyses of intersections of multiple variables. We show the significant intersections and mean mobility difference in Table 3 (Appendix B). These results are consistent with recent work analyzing other data sources. Residents in Democratic counties are more likely to completely stay at home after a state order has been implemented relative to those in Republican counties, based on geolocation

[3] `https://github.com/unitedstates/congress-legislators`
[4] `https://en.wikipedia.org/wiki/Political_party_strength_in_U.S._states`

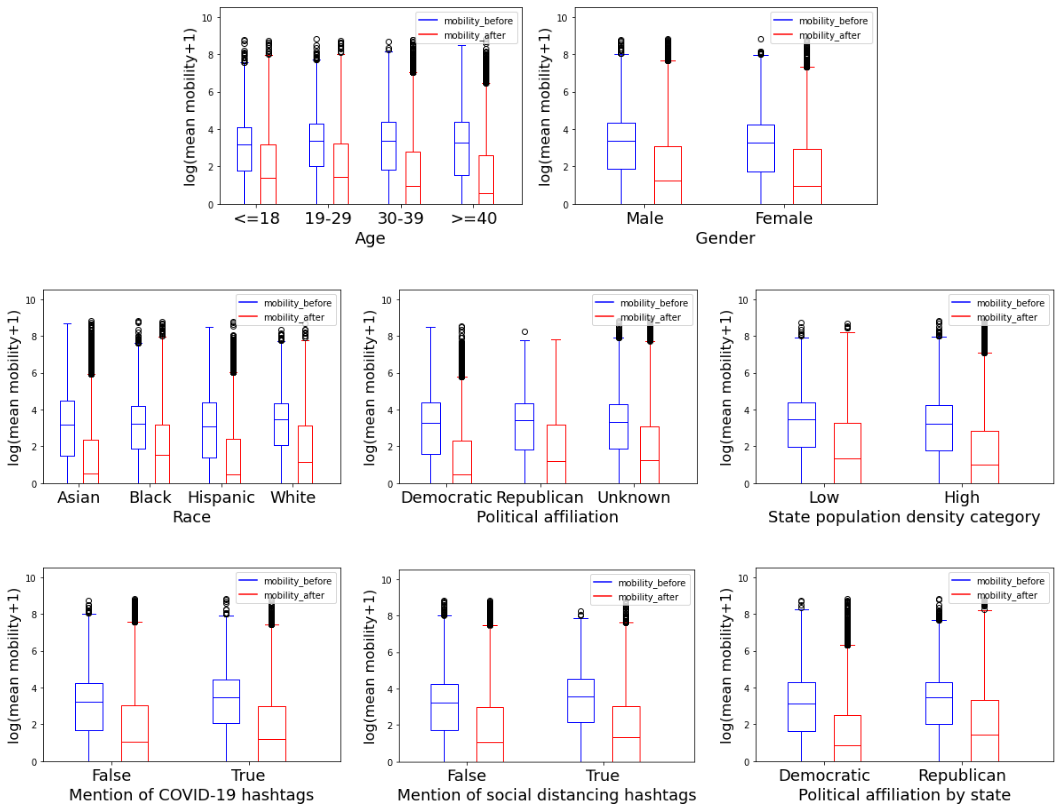

Figure 1: Mobility index distributions of each characteristic type for before and after the start of the COVID-19 pandemic.

Table 1: Summary statistics of ANOVA tests and mean mobility differences. p-values shown for each group.

Variable	Group	Mean mobility difference	Sample size
Gender (9.13×10^{-3})	male	36.40	258625
	female	32.48	194414
Race (7.60×10^{-16})	Asian	47.11	40193
	white	35.65	184654
	Hispanic	42.70	81509
	black	25.72	146683
Age (1.75×10^{-6})	< 30	28.40	245202
	≥ 30	42.18	207837
Political affiliation (3.18×10^{-4})	unknown	32.70	363927
	republican	36.44	25157
	democrat	45.53	63955
State population density category (1.19×10^{-3})	high	35.69	295719
	low	32.90	157320

data from SafeGraph [24]. Similar results were found using location data from smartphones [25]. Combining census data and COVID-19 test results shows that individuals from poor and immigrant neighborhoods and areas with predominantly black population in New York City are more likely to test positive [23]. A consistent picture emerging across multiple data sources and analyses indicates that some groups are practicing less social distancing, and thus may be at a higher risk of infection. Public health communication strategies should consider how to best reach these at risk groups.

Furthermore, we have demonstrated the utility of data from the Twitter Social Mobility Index, and thus open further avenues for research using this data. However, these preliminary results must be contextualized within the limitation of our data, which are discussed in the broader impact statement, and the limitations of our analysis. Critically, we consider individual characteristics rather than a holistic analysis that may illuminate other issues. For example, democrats may be more likely to reduce their mobility because they list in dense urban areas. Additionally, additional unavailable variables could be critical to understanding our conclusions. For example, we do not have access to socioeconomic information, but these factors may be confounding variables that explain why some groups have smaller reductions in mobility. Future work could extract county-level location from geotagged tweets and combine with census data to provide richer demographics variables.

Our future work will consider more detailed analyses to explore these issues.

Broader Impact

The current COVID-19 pandemic has clearly demonstrated the critical role of public health communications in ensuring an effective response to an infectious disease outbreak. With no vaccine and promising yet slow results with currently available therapies, communicating to the public best practices for avoiding an infection are the bulwark against the pandemic. Effective communication must consider how the message is understood and accepted by different communities, especially at-risk groups. This paper demonstrates how Twitter mobility data can be used to gain insights into adherence to public health guidelines, and our findings can inform public health messaging.

A responsible analysis must contextualize our results with the known, and potentially unknown, limitations of our data and methods. We enumerate some of these issues.

Twitter is a biased source of data on a population. It reflects a non-random sample of the underlying population, and users choose to share different types of information and use the platform in different ways. For example, different demographic groups do not use geotagging with the same prevalence [41]. [19] show that demographics like age and gender introduce bias that interacts with geographic inference and how geotagging may be used on Twitter. While Twitter has yielded numerous insights into population health [11] we must remain cautious about this source of bias as we explore each new issues.

Furthermore, our method for inferring demographic information, including gender, age and race/ethnicity are far from perfect. We report the accuracies of our selected systems in the body of the paper. Beyond raw accuracy, these systems all have biases in how they make demographic inference decisions. They mostly capture *perceived* demographics, which may not be consistent with an individuals self-identified demographics. Additionally, these models are limited in that they do not cover all groups within each demographic type. For example, 5 race/ethnicity exclude many individuals in the United States, and binary gender excludes gender minorities. These factors must be considered when drawing conclusions from our analysis.

We make an assumption shared by multiple other papers: changes in mobility tracks social distancing practices. While there is significant evidence to support this, the correlation is not perfect. Individuals may continue to travel but take careful precautions (6 foot distance, masks) while other remain close to home but maintain close, non-masked contact with others.

Despite the massive size of our dataset, it is still limited. We have only a few geotagged tweets each week for each user, and we do not have enough data to produce county level analyses for most locations in the United States. Therefore, these results should be compared to those from other data sources, and further work should more fully explore specific conclusions of the analysis.

Finally, human subjects research requires careful consideration of relevant ethical issues. In the context of social media anlaysis for health, attention to privacy is critical [42]. In addition to conducting this research under an IRB approved exemption, we anonymize the data such that we do not reveal actual user identifies. Our released user level data removes identifiable data and is generalized to preclude reverse identification [43].

References

[1] Paiheng Xu, Mark Dredze, and David A Broniatowski. The twitter social mobility index: Measuring social distancing practices from geolocated tweets. *arXiv preprint arXiv:2004.02397*, 2020.

[2] Charles Courtemanche, Joseph Garuccio, Anh Le, Joshua Pinkston, and Aaron Yelowitz. Strong social distancing measures in the united states reduced the covid-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed covid-19 cases across the united states. *Health Affairs*, pages 10–1377, 2020.

[3] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497, 2020.

[4] Fabiana Ganem, Fabio Macedo Mendes, Silvano Barbosa Oliveira, Victor Bertolo Gomes Porto, Wildo Araujo, Helder Nakaya, Fredi A Diaz-Quijano, and Julio Croda. The impact of early social distancing at covid-19 outbreak in the largest metropolitan area of brazil. *MedRxiv*, 2020.

[5] J Raifman, K Nocka, D Jones, J Bor, S Lipson, J Jay, and P Chan. Covid-19 us state policy database. *Boston, MA: Boston University*, 2020.

[6] Caroline O Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H Grad, Bryan Grenfell, M Elizabeth Halloran, Moritz UG Kraemer, et al. Aggregated mobility data could help fight covid-19. *Science (New York, NY)*, 368(6487):145, 2020.

[7] Alyce S Adams, Stephen B Soumerai, Jonathan Lomas, and Dennis Ross-Degnan. Evidence of self-report bias in assessing adherence to guidelines. *International Journal for Quality in Health Care*, 11(3):187–192, 1999.

[8] Arielle Lasry, Daniel Kidder, Marisa Hast, Jason Poovey, Gregory Sunshine, Nicole Zviedrite, Faruque Ahmed, and Kathleen A Ethier. Timing of cmmunity mitigation and changes in reported covid-19 and community mobility—four us metropolitan areas, february 26–april 1, 2020. 2020.

[9] Samuel Engle, John Stromme, and Anson Zhou. Staying at home: mobility effects of covid-19. *Available at SSRN*, 2020.

[10] Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020.

[11] Michael J Paul and Mark Dredze. Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 9(5):1–183, 2017.

[12] Michael J Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Fifth international AAAI conference on weblogs and social media*. Citeseer, 2011.

[13] Adrian Benton, Braden Hancock, Glen Coppersmith, John W Ayers, and Mark Dredze. After sandy hook elementary: A year in the gun control debate on twitter. *arXiv preprint arXiv:1610.02060*, 2016.

[14] Lauren Sinnenberg, Alison M Buttenheim, Kevin Padrez, Christina Mancheno, Lyle Ungar, and Raina M Merchant. Twitter as a tool for health research: a systematic review. *American journal of public health*, 107(1):e1–e8, 2017.

[15] Mark Dredze, Manuel García-Herranz, Alex Rutherford, and Gideon Mann. Twitter as a source of global mobility patterns for social good. *arXiv preprint arXiv:1606.06343*, 2016.

[16] Mark Dredze, Michael J Paul, Shane Bergsma, and Hieu Tran. Carmen: A twitter geolocation system with applications to public health. In *AAAI workshop on expanding the boundaries of health informatics using AI (HIAI)*, volume 23, page 45. Citeseer, 2013.

[17] Xin Zheng, Jialong Han, and Aixin Sun. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1652–1671, 2018.

[18] Bo Han, Paul Cook, and Timothy Baldwin. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500, 2014.

[19] Umashanthi Pavalanathan and Jacob Eisenstein. Confounds and consequences in geotagged twitter data. *arXiv preprint arXiv:1506.02275*, 2015.

[20] Mark Dredze, Miles Osborne, and Prabhanjan Kambadur. Geolocation for twitter: Timing matters. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1064–1069, 2016.

[21] Giuseppe Lippi, Camilla Mattiuzzi, Fabian Sanchis-Gomar, and Brandon M Henry. Clinical and demographic characteristics of patients dying from covid-19 in italy versus china. *Journal of medical virology*, 2020.

[22] Albert Esteve, Inaki Permanyer, Diederik Boertien, and James W Vaupel. National age and co-residence patterns shape covid-19 vulnerability. *medRxiv*, 2020.

[23] George J Borjas. Demographic determinants of testing incidence and covid-19 infections in new york city neighborhoods. Technical report, National Bureau of Economic Research, 2020.

[24] Marcus Painter and Tian Qiu. Political beliefs affect compliance with covid-19 social distancing orders. *Available at SSRN 3569098*, 2020.

[25] Hunt Allcott, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Y Yang. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *NBER Working Paper*, (w26946), 2020.

[26] Nina Cesare, Christan Grant, and Elaine O Nsoesie. Detection of user demographics on social media: A review of methods and recommendations for best practices. *arXiv preprint arXiv:1702.01807*, 2017.

[27] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist. Understanding the demographics of twitter users. *Icwsm*, 11(5th):25, 2011.

[28] Marco Pennacchiotti and Ana-Maria Popescu. A machine learning approach to twitter user classification. In *Fifth international AAAI conference on weblogs and social media*, 2011.

[29] Xiaolei Huang and Michael Paul. Neural user factor adaptation for text classification: Learning to generalize across author demographics. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (* SEM 2019)*, pages 136–146, 2019.

[30] John D Burger, John Henderson, George Kim, and Guido Zarrella. Discriminating gender on twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1301–1309, 2011.

[31] Zach Wood-Doughty, Nicholas Andrews, Rebecca Marvin, and Mark Dredze. Predicting twitter user demographics from names alone. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 105–111, 2018.

[32] Rebecca Knowles, Josh Carroll, and Mark Dredze. Demographer: Extremely simple name demographics. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 108–113, 2016.

[33] Daniel Preoţiuc-Pietro and Lyle Ungar. User-level race and ethnicity predictors from twitter text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1534–1545, 2018.

[34] Zach Wood-Doughty, Paiheng Xu, Xiao Liu, and Mark Dredze. Using noisy self-reports to predict twitter user demographics. *arXiv preprint arXiv:2005.00635*, 2020.

[35] Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *The World Wide Web Conference*, pages 2056–2067, 2019.

[36] Daniel Preoţiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar. Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 729–740, 2017.

[37] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 192–199. IEEE, 2011.

[38] Zach Wood-Doughty, Praateek Mahajan, and Mark Dredze. Johns hopkins or johnny-hopkins: Classifying individuals versus organizations on twitter. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 56–61, 2018.

[39] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[40] Matthew A Baum, Jennifer Lin, Katherine Ognyanova, Hanyu Chwe, Alexi Quintana, David Lazer, James Druckman, Roy H Perlis, Mauricio Santillana, John Della Volpe, Matthew Simonson, and Jon Green. The state of the nation: A 50-state covid-19 survey report# 13:public trust in institutionsand vaccine acceptance. 2020.

[41] Zach Wood-Doughty, Michael Smith, David Broniatowski, and Mark Dredze. How does twitter user behavior vary across demographic groups? In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 83–89, 2017.

[42] Adrian Benton, Glen Coppersmith, and Mark Dredze. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 94–102, 2017.

[43] John W Ayers, Theodore L Caputi, Camille Nebeker, and Mark Dredze. Don't quote me: reverse identification of research participants in social media studies. *NPJ digital medicine*, 1(1):1–2, 2018.

A COVID-19 and Social Distancing Related Hashtags

Table 2: Most popular hashtags identified that are related to COVID-19 and social distancing.

Topic	Hashtags
COVID-19	covid19, coronavirus, covid-19, covid_19, covid, coronaviruspandemic, covid2019.
Social distancing	stayhome, socialdistancing, quarantine, quarantinelife, stayathome

B ANOVA Test Result for Intersections

Table 3: Summary statistics of ANOVA tests and mean mobility differences of significant intersections. p-values shown for each group.

Variable	Group	Mean mobility difference	Sample size
Gender×Age (1.30×10^{-02})	male and < 30	28.34	123067
	male and ≥ 30	43.72	135558
	female and < 30	28.46	122135
	female and ≥ 30	39.27	72279
Race×Age (4.22×10^{-02})	Asian and < 30	36.50	20754
	Asian and ≥ 30	58.45	19439
	White and < 30	30.83	103108
	White and ≥ 30	41.74	81546
	Hispanic and < 30	34.57	24853
	Hispanic and ≥ 30	46.26	56656
	Black and < 30	22.47	96487
	Black and ≥ 30	31.97	50196
Gender×Political affiliation (2.68×10^{-02})	male and unknown	34.60	206562
	male and republican	38.21	17710
	male and democrat	46.30	34353
	female and unknown	30.21	157365
	female and republican	32.24	7447
	female and democrat	44.63	29602
Race×State population density category (4.23×10^{-03})	Asian and high	48.87	28630
	Asian and low	42.76	11563
	White and high	36.40	112551
	White and low	34.47	72103
	Hispanic and high	43.71	55738
	Hispanic and low	40.50	25771
	Black and high	26.53	98800
	Black and low	24.07	47883
Gender×Age×Political affiliation (2.58×10^{-02})	male and < 30 and unknown	27.15	109505
	male and < 30 and republican	32.19	3762
	male and < 30 and democrat	40.13	9800
	male and ≥ 30 and unknown	43.01	97057
	male and ≥ 30 and republican	39.84	13948
	male and ≥ 30 and democrat	48.76	24553
	female and < 30 and unknown	26.93	109157
	female and < 30 and republican	36.43	1934
	female and < 30 and democrat	42.24	11044
	female and ≥ 30 and unknown	37.64	48208
	female and ≥ 30 and republican	30.77	5513
	female and ≥ 30 and democrat	46.05	18558