The Twitter Social Mobility Index:
Measuring Social Distancing Practices from Geolocated Tweets

Paiheng Xu, Mark Dredze
Malone Center for Engineering in Healthcare
Center for Language and Speech Processing
Department of Computer Science
Johns Hopkins University
`paiheng,mdredze@jhu.edu`

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

Abstract

Social distancing is an important component of the response to the novel Coronavirus (COVID-19) pandemic. Minimizing social interactions and travel reduces the rate at which the infection spreads, and "flattens the curve" such that the medical system can better treat infected individuals. However, it remains unclear how the public will respond to these policies. This paper presents the Twitter Social Mobility Index, a measure of social distancing and travel derived from Twitter data. We use public geolocated Twitter data to measure how much a user travels in a given week. We find a large reduction in travel in the United States after the implementation of social distancing policies, with larger reductions in states that were early adopters and smaller changes in states without policies. Our findings are presented on `http://socialmobility.covid19dataresources.org` and we will continue to update our analysis during the pandemic.

1 Introduction

The outbreak of the SARS-CoV-2 virus, a Coronavirus that causes the disease COVID-19, has caused a pandemic on a scale unseen in a generation. Without an available vaccine to reduce transmission of the virus, public health and elected officials have called on the public to practice social distancing. Social distancing is a set of practices in which individuals maintain a physical distance so as to reduce the number of physical contacts they encounter (Maharaj and Kleczkowski, 2012; Kelso et al., 2009). These practices include maintaining a distance of at least six feet and avoiding large gatherings (Glass et al., 2006). At the time of this writing, in the United States nearly every state has implemented state-wide "stay-at-home" orders to enforce social distancing practices (Zeleny, 2020).

While an important tool in the fight against COVID-19, the implementation of social distancing by the general public can vary widely. While a state governor may issue an order for the practice, individuals in different states may respond in different ways. Understanding actual reductions in travel and social contacts is critical to measuring the effectiveness of the policy. These policies may remain in effect for an extended period of time. Thus, the public may begin to relax their practices, making additional policies necessary. Additionally, epidemiologists already model the impact of social distancing policies on the course of an outbreak (Prem et al., 2020; Fenichel et al., 2011; Caley et al., 2008). These models may be more effective when incorporating actual measures of social distancing, rather than assuming official policies are implemented in practice.

It can be challenging to obtain data on the efficacy of social distancing practices, especially during an ongoing pandemic. A recent Gallup poll surveyed Americans to find that many adults are taking precautions to keep their distance from others (Saad, 2020). However, while polling can provide insights, it cannot provide a solution. Polling is relatively expensive, making it a poor choice for ongoing population surveillance practices and providing data on specific geographic locales, i.e. US States and major cities (Dredze et al., 2016a). Additionally, polling around public health issues suffers from response bias, as individuals may overstate their compliance with established public health recommendations (Adams et al., 1999).

Over the past decade, analyses of social media and web data have been widely adopted to support public health objectives (Paul and Dredze, 2017). In this vein, several efforts have emerged over the past few weeks to track social distancing practices using these data sources. Google has released "COVID-19 Community Mobility Reports"

which use Google data to "chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential" (Google, 2020). The Unacast "Social Distancing Scoreboard" uses data collected from 127 million monthly active users to measure the implementation of social distancing practices (Unacast, 2020). Researchers at the Institute for Disease Modeling have used data from Facebook's "Data for Good" program to model the decline in mobility in the Greater Seattle area and its effect on the spread of COVID-19 (Burstein et al., 2020). Using cell phone data, the New York Times completed an analysis that showed that stay-at-home orders dramatically reduced travel, but that states that have waited to enact such orders have continued to travel widely (Glanz et al., 2020). These efforts provide new and important opportunities to study social distancing in real-time.

We present the Twitter Social Mobility Index, a measure of social distancing and travel patterns derived from public Twitter data. We use public geolocated Twitter data to measure how much a user travels in a given week. We compute a metric based on the standard deviation of a user's geolocated tweets each week, and aggregate these data over an entire population to produce a metric for the United States as a whole, for individual states and for some US cities. We find that, taking the US as a whole, there has been a dramatic drop in travel in recent weeks, with travel between March 16 and March 29, 2020 showing the lowest amount of travel since January 1, 2019, the start of our dataset. Additionally, we find that travel reductions are not uniform across the United States, but vary from state to state. A key advantage of our approach is that, unlike other travel and social distancing analyses referenced above, we rely on entirely public data, enabling others to replicate our findings and explore different aspects of these data. Additionally, since Twitter contains user generated content in addition to location information, future analyses can correlate attitudes, beliefs, and behaviors with changes in social mobility.

Our findings are presented on `http://socialmobility.covid19dataresources.org` and we will continue to update our analysis during the pandemic.

2 Data

Twitter offers several ways in which a user can indicate their location. If a user is tweeting from a GPS enabled device, they can attach their exact coordinate to that tweet. Twitter may then display to the user, and provide in their API, the specific place that corresponds to these coordinates. Alternatively, a user can explicitly select a location, which can be a point of interest (coffee shop), a neighborhood, a city, state, or country. If the tweet is public, this geolocation information is supplied with the tweet.

We used the Twitter streaming API[1] to download tweets based on location. We used a bounding box that covered the entire United States, including territories. We used data from this collection starting on January 1, 2019 and ending on March 29, 2020. In total, this included 4,600,287 Twitter users and 544,546,651 tweets.

3 Location Data

We process the two types of geolocation information described in the previous section.

Coordinates The exact coordinates (latitude/longitude) provided by the user ("coordinates" field in the Twitter JSON object). About 8.02% of our data included "coordinates".

Place The "place" field in the Twitter json object indicates a known location in which the tweet was authored. A place can be a point of interest (a specific hotel), a neighborhood ("Downtown Jacksonville"), a city ("Kokomo, IN"), a state ("Arizona") or a country ("United States"). The place object contains a unique ID, a bounding box, the country and a name. More information about the location is available from the Twitter GEO API. A place is available with a tweet in either of two conditions. First, Twitter identifies the coordinates provided by the user as occurring in a known place. Second, if the user manually selects the place when authoring the tweet.

Since coordinates give a more precise location, we use them instead of place when available. If we only have a place, we assume that the user is in the center of the place, as given by the place's bounding box.

For points of interest and neighborhoods, Twitter only provides the country in the associated metadata. While in some cases the city can be parsed

---

[1] https://developer.twitter.com/en/docs/tweets/filter-realtime/overview/statuses-filter

from the name, and the state inferred, we opted to exclude these places from our analysis. The full location details can be obtained from querying the Twitter API, but the magnitude of data in our analysis made this too time consuming. This excluded about 1.8% of our data.

We include an analysis of New York City. For this analysis, we include places that corresponded to each of the five New York City boroughs (Brooklyn, Manhattan, Queens, Staten Island, The Bronx). We also included points of interest that had in the name the words "New York City", e.g. "New York City Center".

In summary, for each geolocated tweet we have an associated latitude and longitude.

4 Computing Mobility

We define the Twitter Social Mobility Index as follows. For each user, we collect all locations (coordinates) in a one week period, where a week starts on Monday and ends the following Sunday. We compute the centroid of all of the the coordinates and consider this the "home" location for the user for that week. We then measure the distance between each location and the centroid for that week. For distance, we measure the geodesic distance in kilometers between two adjacent records using geopy[2]. After collecting the distances we measure the standard deviation of these distances. In summary, this measure reflects the area of travel for a user, rather than the raw distance traveled. Therefore, a user who takes a long trip with a small number of checkins would have a larger social mobility measure than a user with many checkins who traveled in a small area.

We aggregate the results by week by taking the mean measure of all users in a given geographic area. We also record the variance of these measures to study the travel variance in the population, which will indicate if some users are not reducing travel.

We produce aggregate scores by geopgraphic area for the United State as a whole, for each US state and territory, and for New York City. We determine the geographic area of a user based on their centroid location for that week. This allows users to move geographic areas by week.

We compute the social mobility index for each week between January 1, 2019 and March 29, 2020. We select the date of March 16, 2020 as the start of social distancing on the national level, though

individual states have implemented practices at different times. Therefore, we divide the data into two time periods: before social distancing (January 1, 2019 - March 15, 2020) and after social distancing (March 16th, 2020 - March 29, 2020).

We then compute the reduction in social mobility as follows:

$$\text{Mobility Reduction} = 1 - \frac{\text{mobility after social distancing}}{\text{mobility before social distancing}}.$$
(1)

To handle sparse data issues in our dataset, we exclude (1) users with less than 3 geolocated tweets overall, and (2) a weekly record for a user if that user has less than 2 geolocated tweets in that week. Additionally, due to data loss in our data collection process we remove two weeks with far less data than other weeks: January 20 - January 27, 2020 and February 17 - February 24, 2020.

5 Results

Table 1 shows the Twitter Social Mobility Index measured in kilometers for every state and territory in United States, New York City, and United States as a whole. We also include the rank of each state or territory.

A few observations. The overall drop in mobility across the United States was large: 48.77%. Second, every US state and territory saw a drop in mobility, ranging from 21.54% to 63.13% travel compared to numbers before March 16, 2020. However, the variance by state was high. States that were early adopters of social distancing practices are ranked highly on the reduction in travel: e.g. Washington (3) and Maryland (5). In contrast, the eight states that do not currently have state wide orders (Zeleny, 2020) rank poorly: Arkansas (37), Iowa (50), Nebraska (29), North Dakota (13), South Carolina (34), South Dakota (39), Oklahoma (52), Utah (17), Wyoming (53). We measured two cities: New York City (51.16%) and Washington, DC (63.13%). The fact that Washington, DC, ranks first in Table 1 is consistent with the result in Unacast (2020), where DC has the highest grade (A-) as of date March 29, 2020. More work is needed to understand how to compare US states with cities.

Figure 1 shows the weekly social mobility index for the United States for the entire time period of our dataset. The figure reflects a massive drop in mobility starting in March, with the two most recent weeks the lowest on record in our dataset.

---

[2]https://github.com/geopy/geopy

Finally, Figure 2 shows a box-plot of the mobility variance across all user in a given time period. The distribution is long-tailed and with a lot zeros, so we take the log of 1 plus each mobility index. While mobility is reduced in general, some users are still showing a lot of movement, suggesting that social distancing is not being uniformly practiced. These results clearly demonstrate that our metric can track drops in travel, suggesting that it can be used as part of ongoing pandemic response planning.

6 Related Work

There is a long line of work on geolocation prediction for Twitter, which requires inferring a location for a specific tweet or user (Dredze et al., 2013; Zheng et al., 2018; Han et al., 2014; Pavalanathan and Eisenstein, 2015). This includes work on patterns and trends in Twitter geotagged data (Dredze et al., 2016c). While most of this work focused on a user, and thus not suitable for tracking a user's movements, there may be opportunities to combine these methods with our approach.

There have been many studies that have analyzed Twitter geolocation data to study population movements. Hawelka et al. (2014) demonstrated a method for computing global travel patterns from Twitter, and Dredze et al. (2016b) adapted this method to support efforts in combating the Zika epidemic.

Several studies have used human mobility patterns from Twitter data (Jurdak et al., 2015; Huang and Wong, 2015; Birkin et al., 2014; Hasan et al., 2013). These studies have included analyses of urban mobility patterns (Luo et al., 2016; Soliman et al., 2017; Kurkcu et al., 2016). Finally, some of these analyses have considered mobility patterns around mass events (Steiger et al., 2015).

7 Conclusion

We presented the Twitter Social Mobility Index, a measure of social mobility based on public Twitter geolocated tweets. Our analysis shows that overall in the United States there has been a large drop in mobility. However, the drop is inconsistent and varies significantly by state. Anecdotally, states that were early adopters of social distancing practices have more significant drops than states that have not yet implemented these practices.

Our work on this data is ongoing, and there are several directions that warrant further study. First,

our analysis does not incorporate state by state variations in when social distancing was implemented. Second, we only include two cities in our current analysis, but our approach can be extended to dozens of US metro areas. Third, we focused on the United States, but Twitter data provides sufficient coverage for many countries to replicate our analysis. Finally, for each user in the dataset there exists tweet content, that can reflect a user's attitudes, beliefs and behaviors. Studying these together with their mobility reduction could yield further insights.

Our findings are presented on `http://socialmobility.covid19dataresources.org` and we will continue to update our analysis during the pandemic.

References

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

Mark Birkin, Kirk Harland, Nicolas Malleson, Philip Cross, and Martin Clarke. 2014. An examination of personal mobility patterns in space and time using twitter. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, 5(3):55–72.

Roy Burstein, Hao Hu, Niket Thakkar, Andrew Schroeder, Mike Famulare, and Daniel Klein. 2020. Understanding the impact of covid-19 policy change in the greater seattle area using mobility data. `https://covid.idmod.org/data/Understanding_impact_of_COVID_policy_change_Seattle.pdf`.

Peter Caley, David J Philp, and Kevin McCracken. 2008. Quantifying social distancing arising from pandemic influenza. *Journal of the Royal Society Interface*, 5(23):631–639.

Mark Dredze, David A Broniatowski, Michael C Smith, and Karen M Hilyard. 2016a. Understanding vaccine refusal: why we need social media now. *American journal of preventive medicine*, 50(4):550–552.

Mark Dredze, Manuel García-Herranz, Alex Rutherford, and Gideon Mann. 2016b. Twitter as a source of global mobility patterns for social good. In *ICML Workshop on Data4Good: Machine Learning in Social Good Applications*.

Mark Dredze, Miles Osborne, and Prabhanjan Kambadur. 2016c. Geolocation for twitter: Timing matters. In *Proceedings of the 2016 Conference of the*
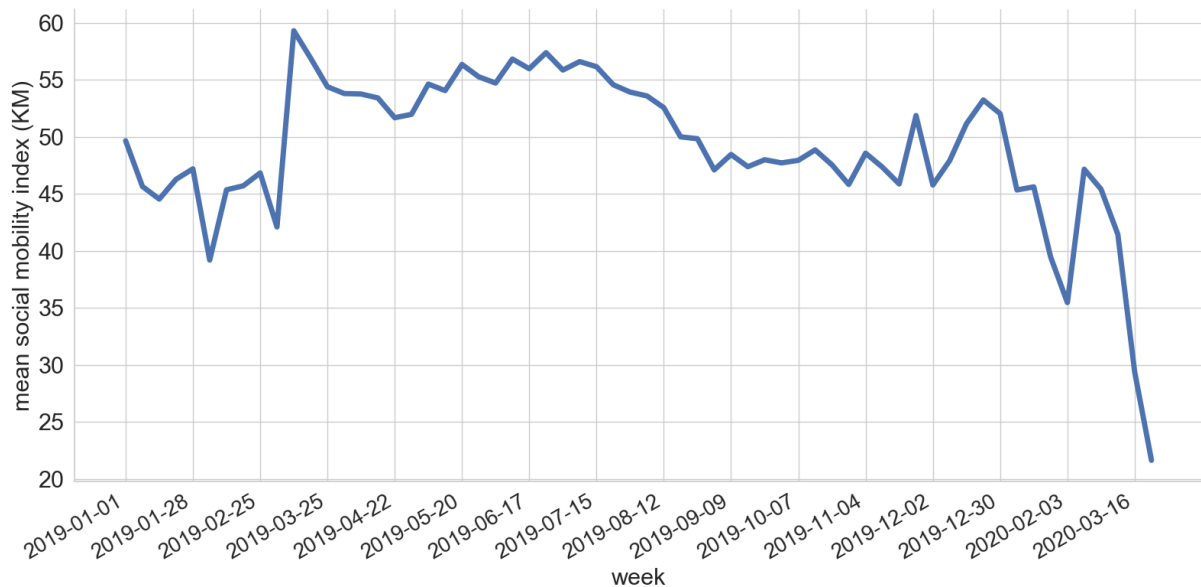
Figure 1: Mean social mobility index (KM) in United States from January 1, 2019 to March 29, 2020. Weeks with missing data are excluded from the figure.
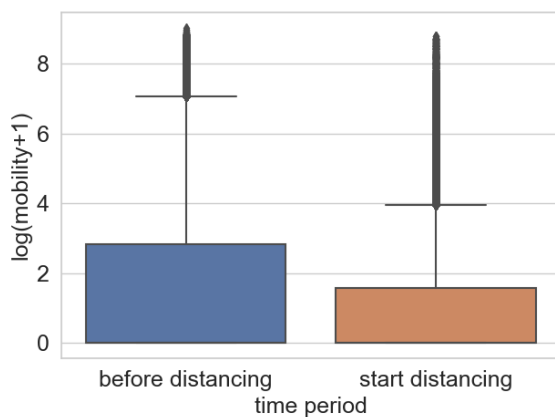

Figure 2: User distribution of mean social mobility index (KM) before/after social distancing in United States.

North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1064–1069.

Mark Dredze, Michael J Paul, Shane Bergsma, and Hieu Tran. 2013. Carmen: A twitter geolocation system with applications to public health. In *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Eli P Fenichel, Carlos Castillo-Chavez, M Graziano Ceddia, Gerardo Chowell, Paula A Gonzalez Parra, Graham J Hickling, Garth Holloway, Richard Horan, Benjamin Morin, Charles Perrings, et al. 2011. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311.

James Glanz, Benedict Carey, Josh Holder, Derek Watkins, Jennifer Valentino-DeVries, Rick Rojas, and Lauren Leather. 2020. *Where America Didn't Stay Home Even as the Virus Spread*. https://www.nytimes.com/interactive/2020/04/02/us/coronavirus-social-distancing.html.

Robert J Glass, Laura M Glass, Walter E Beyeler, and H Jason Min. 2006. Targeted social distancing designs for pandemic influenza. *Emerging infectious diseases*, 12(11):1671.

Google. 2020. COVID-19 community mobility reports. https://www.google.com/covid19/mobility/.

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

Samiul Hasan, Xianyuan Zhan, and Satish V Ukkusuri. 2013. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pages 1–8.

Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271.

Qunying Huang and David WS Wong. 2015. Modeling and visualizing regular human mobility patterns with uncertainty: An example using twitter data. *Annals of the Association of American Geographers*, 105(6):1179–1197.

Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice Abou-Jaoude, Mark Cameron, and David Newth. 2015. Understanding human mobility from twitter. *PloS one*, 10(7).

Joel K Kelso, George J Milne, and Heath Kelly. 2009. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC public health*, 9(1):117.

Abdullah Kurkcu, K Ozbay, and EF Morgul. 2016. Evaluating the usability of geo-located twitter as a tool for human activity and mobility patterns: A case study for nyc. In *Transportation Research Board's 95th Annual Meeting*, pages 1–20.

Feixiong Luo, Guofeng Cao, Kevin Mulligan, and Xiang Li. 2016. Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of chicago. *Applied Geography*, 70:11–25.

Savi Maharaj and Adam Kleczkowski. 2012. Controlling epidemic spread by social distancing: Do it well or not at all. *BMC Public Health*, 12(1):679.

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

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

Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl AB Pearson, James D Munday, et al. 2020. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*.

Lydia Saad. 2020. Americans step up their social distancing even further. https://news.gallup.com/opinion/gallup/298310/americans-step-social-distancing-even-further.aspx.

Aiman Soliman, Kiumars Soltani, Junjun Yin, Anand Padmanabhan, and Shaowen Wang. 2017. Social sensing of urban land use based on analysis of twitter users' mobility patterns. *PloS one*, 12(7):e0181657.

Enrico Steiger, Timothy Ellersiek, Bernd Resch, and Alexander Zipf. 2015. Uncovering latent mobility patterns from twitter during mass events. *GI_Forum*, 1:525–534.

Unacast. 2020. Social distancing scoreboard. https://www.unacast.com/covid19/social-distancing-scoreboard.

Jeff Zeleny. 2020. Why these 8 Republican governors are holding out on statewide stay-at-home orders. https://www.cnn.com/2020/04/04/politics/republican-governors-stay-at-home-orders-coronavirus/index.html.

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

Location	Mobility January 1, 2019 to March 15, 2020	March 16 - March 29, 2020	Mobility Reduction	Rank
AK	84.75	31.49	62.84%	2
AL	37.46	23.63	36.93%	47
AR	39.19	22.34	43.00%	37
AZ	47.30	25.37	46.37%	32
CA	59.47	28.64	51.84%	18
CO	54.21	23.82	56.05%	9
CT	34.16	15.35	55.05%	11
DC	58.45	21.55	63.13%	1
DE	33.38	17.39	47.90%	28
FL	59.29	34.61	41.62%	40
GA	51.43	29.46	42.73%	38
HI	112.71	70.14	37.77%	46
IA	38.23	24.49	35.94%	50
ID	53.61	34.07	36.44%	49
IL	41.95	18.97	54.79%	12
IN	34.69	19.30	44.35%	35
KS	50.44	23.67	53.07%	15
KY	33.99	16.48	51.51%	20
LA	36.53	21.65	40.74%	43
MA	43.50	18.59	57.27%	7
MD	34.94	14.81	57.61%	5
ME	46.42	27.19	41.42%	42
MI	42.79	22.12	48.31%	26
MN	47.75	23.23	51.35%	21
MO	39.89	20.49	48.64%	24
MS	39.63	25.59	35.42%	51
MT	52.46	32.23	38.56%	45
NC	39.89	20.32	49.06%	23
ND	48.94	22.50	54.02%	13
NE	40.66	21.26	47.73%	29
NH	41.97	21.66	48.38%	25
NJ	37.23	15.54	58.28%	4
NM	44.87	25.54	43.08%	36
NV	62.72	34.33	45.26%	33
NY	53.32	25.73	51.75%	19
OH	34.22	16.70	51.19%	22
OK	40.44	27.96	30.87%	52
OR	53.96	24.10	55.34%	10
PA	42.14	19.54	53.62%	14
PR	35.30	20.63	41.55%	41
RI	34.36	14.69	57.24%	8
SC	36.78	20.16	45.18%	34
SD	49.80	28.82	42.12%	39
TN	42.64	22.42	47.42%	30
TX	56.00	29.09	48.05%	27
UT	50.62	24.29	52.01%	17
VA	44.22	20.83	52.89%	16
VI	100.91	63.86	36.72%	48
VT	41.19	17.51	57.48%	6
WA	55.74	22.99	58.76%	3
WI	42.78	22.83	46.63%	31
WV	36.25	22.18	38.82%	44
WY	53.10	41.66	21.54%	53
United States	49.81	25.52	48.77%	-
New York City	63.98	31.25	51.16%	-

Table 1: Reduction of mobility for all states and territories in United States, New York City and United States.