MIT Connection Science
the technology of innovation

Effect of social distancing measures in the New York City metropolitan area

Michiel Bakker[1,2], Alex Berke[1], Matt Groh[1], Alex 'Sandy' Pentland[1,2], and Esteban Moro[2,3]

2020-03-29

[1] *MIT Media Lab*
[2] *MIT IDSS Connection Science*
[3] *GISC & Department of Mathematics, UC3M (Spain)*

Main findings

- Social distancing policies in the New York City metropolitan area have resulted in empirically verified, dramatic changes in where people spend their time and with how many people they interact. For example, here are three striking differences between the weekends of late February and last weekend (March 20th) that we discovered when analyzing anonymized geolocation data in New York City:
    - Distance travelled everyday dropped by 70 percent from a weekend average of 25 miles in February to 7 miles last weekend.
    - The number of social contacts in places decreased by 93% from 75 to 5.
    - The number of people staying home the whole day has increased from 20% to 60%.
- Supermarkets and grocery stores have become the most common place where social contact takes place.
- The national emergency declaration and school closure announcement on March 14th resulted in a huge surge of visits (up to 60% more) to many places. Most of this surge in activity happened at Grocery, Shopping, Food and Outdoor places. The reduction in distance travelled and daily social contacts became significant only after non-essential business closure measures were introduced on March 22nd.
- After the measures were introduced surges of activity appeared in places like the beaches and the Hamptons. A large fraction of people (5.5%) left the NYC area for other places across the US. For example, 0.37% of people left NYC for Florida, which is important to note because this kind of travel can bring the virus to new places.
- Normally, mobility and social contacts vary significantly by the demographic composition of a neighborhood. The social distancing policies have greatly reduced relative differences between different demographic groups as nearly everyone's mobility and social contacts has been dramatically reduced.

Intro

The World Health Organization has officially declared COVID-19 a pandemic. As of today (March 29, 2020), the number of new confirmed cases and deaths from coronavirus continues to grow exponentially in many regions of the world. Across the globe, nations are enacting extraordinary policies to reduce the spread of the coronavirus. In several countries, notably China, South Korea,

and Singapore, these policies have been extremely effective in reducing the growth rate of the virus [1].

In the US, social distancing has been encouraged and implemented with school closures and the strict "stay-in-place" policies. The intention of social distancing policies is to reduce the speed at which the virus spreads by reducing interpersonal contact [2,3,4,5,6]. By reducing the immediate burden on healthcare systems, social distancing is intended to save lives.

With respect to social distancing policy, there are two main empirical research questions. First, what is the effect of social distancing on the spread of coronavirus? Second, how well are people practicing social distancing? The first question can only be empirically answered retrospectively. The second question can be addressed right now. Specifically, we can examine how social distancing, as proxied by anonymized geolocation data from mobile phone apps, is taking place [7].

This report specifically focuses on the New York City metropolitan area. As of writing, New York City has 52,318 confirmed COVID-19 cases and 728 associated deaths. New York City is considered the epicenter of the pandemic in the US.

To understand our findings, we include a timeline of recent news that spread awareness of the severity of COVID-19 and policies that encouraged and enforced social distancing.

- *December 31st, 2019:* Chinese health officials inform the WHO about a cluster of 41 patients with a mysterious pneumonia
- *January 30th:* WHO declares a global public health emergency
- *February 29th*: US reports first death on American soil.
- *March 4th*: Governor of CA declares national emergency.
- *March 10-11th*: The need to "flatten the curve" with "social distancing" is published and goes viral. From this point onward, information about "flattening the curve" and encouraging others to practice "social distancing" spreads more quickly.
    - *March 10th*: Vox publishes "How canceled events and self-quarantines save lives, in one chart"
    - *March 10th*: "Coronavirus: Why You Must Act Now" is published on Medium.
    - *March 11th*: New York Times publishes "Social Distancing May Be Our Best Weapon to Fight the Coronavirus"
- *March 11th*: NBA announces season suspension.
- *March 11th (evening)*: Trump announces travel from Europe to the U.S. will be largely suspended for 30 days starting March 13.
- ***March 13th*: Trump declares national emergency.**
- ***March 15th:* NY City School System announces closure starting Monday March 16th.**
- *March 16th:* Casinos and gyms closed and restaurants and bars limited to serving take-out and delivery orders.
- *March 17th*: NYC Mayor Bill de Blasio says he may issue a "shelter in place" order for the city within the next two days in order to combat the new coronavirus.

- *March 21st:* Also closed until further notice: hair and nail salons, barbershops, tattoo parlors and other personal care businesses that can't comply with social distancing guidelines.
- *March 21st*: New Jersey Governor announces 'stay at home' order.
- ***March 22nd:* Non-essential businesses were ordered to close or end all in-person functions in New York, New Jersey and Connecticut.**
- *March 23rd*: New York City confirms 21,000 cases, making it the biggest epicenter of the outbreak in the US.

At the same time the number of cases of COVID-19 in the NYC area has increased exponentially. While at the beginning of March there were only 10 infected people and no deaths, by March 25th the number of cases had risen to 15,000 and more than 100 deaths.

The purpose of this first report is to examine how these recent events and policies have translated into social distancing in practice and related behavioral changes for people living in the New York City metropolitan area. In particular, we look at (1) how social distancing behavior has changed before and after social distancing policies, (2) how social distancing varies across the physical space of the city, and (3) how social distancing varies across demographic groups.

Data

Source and Privacy

The primary data source used is anonymized location data logged by a variety of applications on smartphone devices. This data is provided by Cuebiq, a location intelligence and measurement company. They supplied anonymized records of high-resolution timestamped GPS points from users who opted-in to share their data anonymously across the U.S. from January 1st 2020 to March 25th 2020. Data was shared through Cuebiq's Data for Good program where they provide access to de-identified and privacy-enhanced mobility data for academic research and humanitarian initiatives. Mobility data is derived from users who opted in to share their data anonymously through a GDPR and CCPA compliant framework.

We attribute the locations reported by devices in the dataset to individual anonymous users. Although the frequency at which a person's device reports a location varies, it typically covers most of the places they visited. In order to preserve privacy, the data provider aggregates residential and work areas to the Census Block Group level, thereby allowing for demographic analysis while obfuscating the true home location of anonymous users and prohibiting misuse of data.

The data we received is the sequence of pings reported by devices. We call each device's sequence a trajectory. From these trajectories, we extract "stays" as the places anonymous users stayed (stopped) for at least 5 minutes (see the Methods section). Some of those stays happen within places like restaurants, business, parks. We call these kinds of stays "visits." This results in a dataset of the places people stayed (with high spatial accuracy) including the points of interest that people visited and the most likely census tract of where the device owner lives and works.

Data Panel

We limit our analysis to stays that happen around the metropolitan area of New York City. We also limit our analysis to data from people selected for our data panel. We selected people who were active during the period February 17 to March 9 and for whom we have location data reporting that they stayed in their home Census Block Group more than 10 days. We specifically exclude all non-residents of the New York City metropolitan area. The panel is a sample of 567,000 people. We analyze the representativeness of this panel in the Methods sections.

Census Data

Our primary data source is location data from which we can infer the home Census Block Group of people in our data panel. We also use public census data[1] which reports estimates of demographics for people living in each of these census areas. We link this demographic information to the people in our data panel based on their home Census Block Groups to further analyze and compare changes across demographic groups.

Measuring Social Distance

We have developed metrics to measure the impact of social distancing policies in the New York City metropolitan area. The metrics are categorized by mobility and contacts:

- **Mobility** refers to how people move around a city. The core mobility metrics include:
  - Distance traveled: line distance traveled every day.
  - Radius of gyration: measures the typical size of the area covered by users [8].
  - Number of people staying home.
  - Number of stays in public places, which we call visits.
- **Contacts** refer to how many people each person comes into contact with. We estimate contacts by looking at instances where two people are within 25 meters distance from one another for at least 5 minutes. We present the estimated number of contacts by what we see in the data, which is a sample of an individual's true number of contacts.

Mobility and contacts are measuring something different by definition but they serve as proxies for social distance, and we should expect these measures to be correlated. For example, if businesses and workplaces close, then people stay more at home, and therefore both the distance travelled and the number of contacts decreases.

Mobility metrics

Figure 1 shows the results for the average daily mobility metrics in the New York City metropolitan area. We clearly see a change in the mobility metrics even before the declaration of national emergence. For example, the typical distance travelled dropped from ~40km per day before March 9th to only 15km from March 23rd, a reduction of 62%. A similar pattern is observed in the radius of gyration with a significant drop well before March 14th. Finally the fraction of people

[1] Our census demographic data is from the 2014-2018 American Community Survey estimates.

staying at home started increasing on the weekend before school closure, reaching around 60% of the people by March 23rd. The significant drop in these metrics show how profound is the change of mobility patterns in New York city after social distancing policies were enacted.

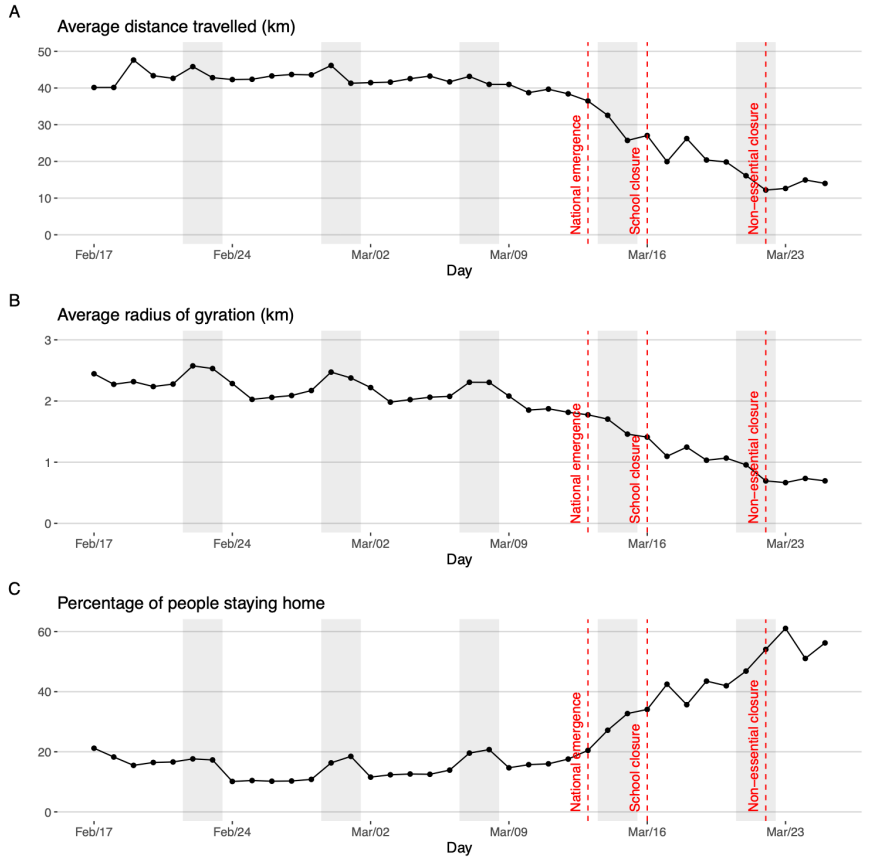

Figure 1: Change of average distance travelled (A), radius of gyration (B), and percentage of people staying home (C) in the NY area during February and March. Vertical red lines signal the declaration of national emergence, the starting date of school and non-essential business closure. Gray areas correspond to weekends.

We also see a decrease (around 83%) in the total number of visits to public places from March 12th onwards (see Figure 2A). However, the change in number of visits varies by type of place. As we can see in Figure 2B, visits to groceries surged (almost 60% more visits) from March 12th and March 13th when the national emergency was announced. There was another spike (25% more visits) on Monday March 16th when school closure happened, showing how these announcements directly result in panic-buying and grocery hoarding. The number of visits to

grocery stores continued to be relatively high until the weekend of March 23rd when it stabilized around 30% of what they had been before the pandemic. Other types of places experienced a more drastic and consistent decrease compared to grocery stores and outdoor places having seen less decrease. Note also that visits to health venues (including hospitals) have also diminished the same amount, probably because of cancellation of non-critical appointments and surgeries.

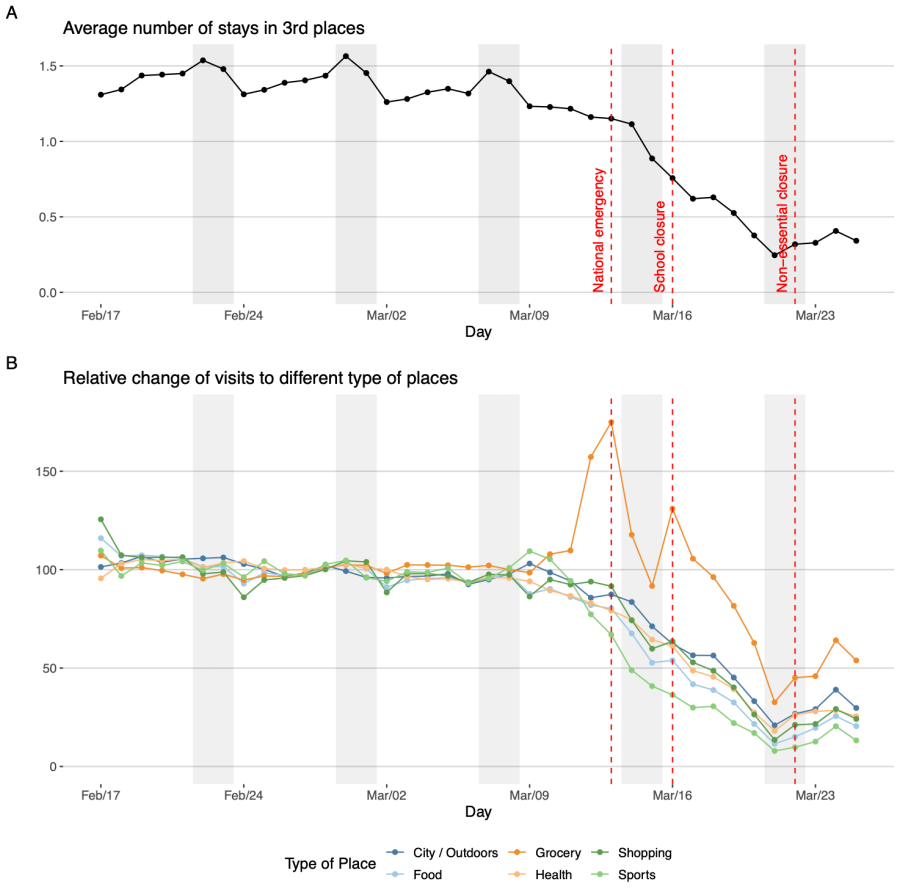

Figure 2: (A) Average number of visits to public places. (B) Relative change of average number of visits to different types of places compared to the average before March 9th (rescaled to 100).

The increase in social distancing metrics is happening homogeneously in the area NY. Figure 3 shows the average distance traveled by residents of each census tract during the weekends of before (2020-03-06) and after (2020-03-20) the lockdown measures. Nonetheless, there are still some tracts in which their residents are travelling large distances. Those areas are home to transportation hubs like airports.

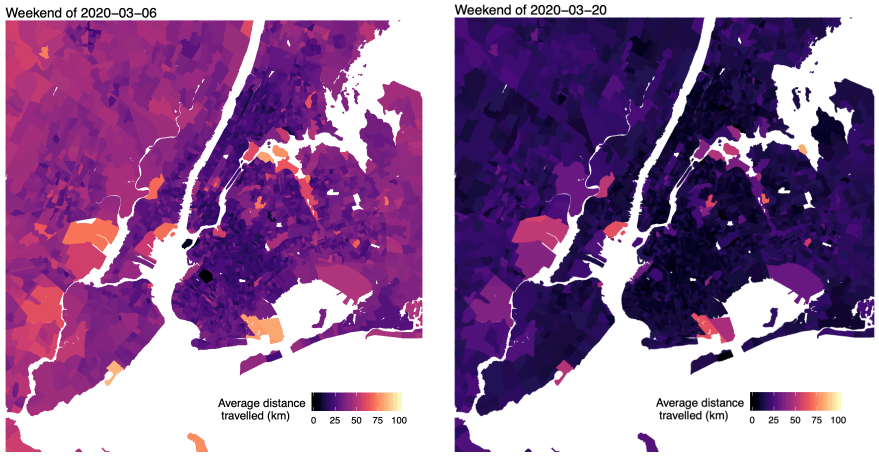

Figure 3: Average distance travel by residents in each census tract during the weekend of 2020-03-06 (left) and the one of 2020-03-20 (right). As we can see the average dropped significantly after the social distancing measures.

People that left the NYC area or the city

Another impact of the rapid spread of the virus in NYC and the subsequent social distance measures is that some people left the metro area for other places. This is a serious concern epidemiologically. Since NYC is the epicenter of the pandemic, they might have transmitted the infection to those places, accelerating the community transmission[2]. To investigate this, we have measured the relative change (in %) of the number of visits to public places by census tract in the NYC metro area during the last two weekends. As Figure 4 shows, in most census tracts that number of visits decreased significantly. However, there is a surge of visits to the beaches and the Hamptons with up to 100% more than the previous weekend. The influx of New Yorkers in this area is a concern for these communities and travel restrictions and recommendations have been issued not to travel to Long Island.[3]

[2] "Dr. Anthony S. Fauci, the director of the National Institute of Allergy and Infectious Diseases, said that New Yorkers who were "understandably" trying to leave for places like Florida needed to make sure they were not "seeding" the rest of the United States."
https://www.nytimes.com/2020/03/24/nyregion/coronavirus-new-york-update.html

[3] "Local leaders in the Hamptons are set to ask the governor to tell New Yorkers to stay away from the seaside destination, as tensions rise between residents and those fleeing New York City from the

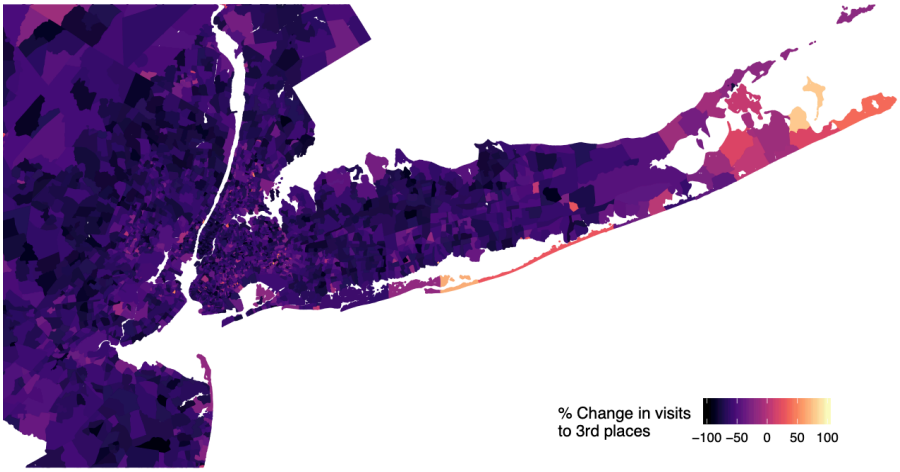

Figure 4. Percentage of change in the number of visits to public places by census tract between the weekends of 2020-03-14 and 2020-03-21. In most census tracts, the visits dropped significantly. However the beaches and the Hamptons showed a surge in the number of visits during the last week.

But fleeing the outbreak from NYC, people have gone to many other places in the US. During the weekend of 2020-02-20 we discovered that people in our panel are now showing up in other states. In aggregate, we see 5.5% of the users in our panel spending time in places well beyond the NY metro area. Figure 5 shows visits from those users that left the NYC area happened mostly in the NJ state (37% of the visits), NY upstate (23%), Pennsylvania (9.75%) and Florida (6.7%).

Commented [1]: In Figure 4 above we say "3rd places," can we change that?

Commented [2]: To what?

Commented [3]: "public places"

coronavirus outbreak." https://www.nbcnewyork.com/news/local/hamptons-leaders-want-cuomo-to-restrict-new-yorkers-from-visiting/2345567/

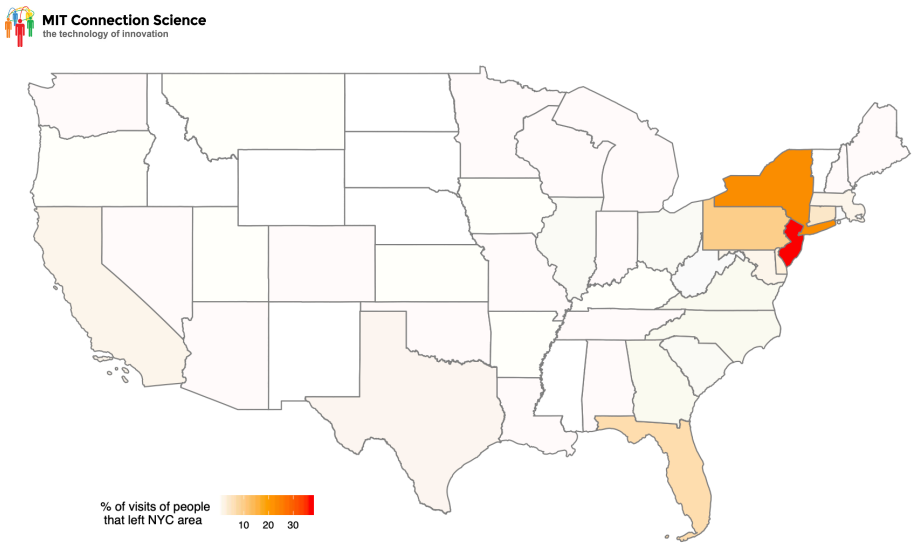

Figure 5: Percentage of visits to public places outside NYC metro area during the weekend of 2020-02-21 from people in our panel.

Impact of social distancing in contacts

We can directly evaluate the impact of social distancing by examining how many people each person comes into contact with. Based on anonymous geolocation data from mobile phones, we define a contact as any instance where two people are within 25 meters distance from one another for at least 5 minutes. It is important to keep in mind that the people in our panel represent about 3% of the New York City metropolitan area. This means we are significantly underestimating the true number of contacts. To approximate that number we have assumed a proportional relationship between the observed number of contacts and the true number of contacts given by the inverse of our sampling ratio. Note, that this is an estimation of the true number of contacts and since we don't have elevation in our data we are probably over-estimating them in places where a lot of people overlay on top of each other, like downtown Manhattan.

Figure 6 shows that social distancing metrics have severely impacted the number of contacts which decreased from 75 contacts to 5 daily contacts. This reduction in interpersonal contact is qualitatively similar to the changes in interpersonal contact in Wuhan and Shanghai during the COVID-19 social distancing period where personal surveys showed that daily contacts were reduced 7-9 fold [2]. Similarly in Italy, the average number of contacts (measured using similar high resolution mobility data) decreased around 34% [7]. Note however that the decrease in the NYC area (around 15-fold) is much larger, probably because our data does not include contacts around households.

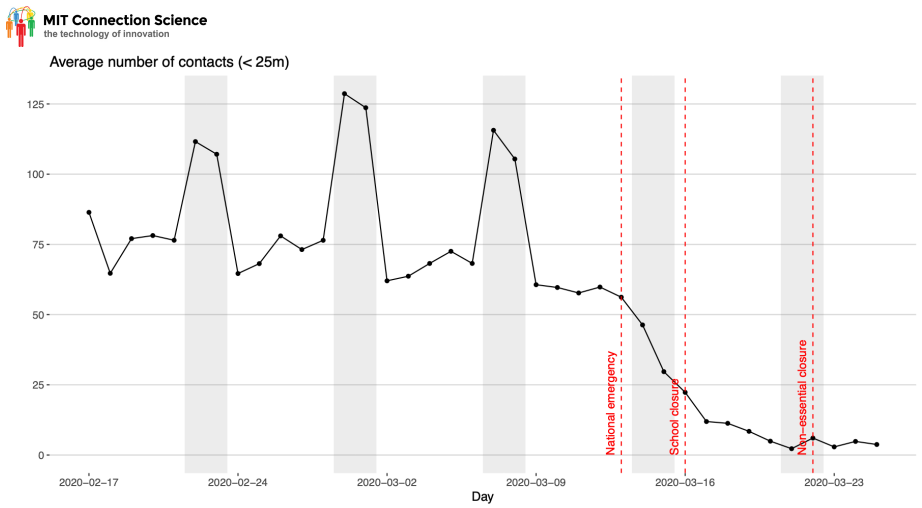

Figure 6. Estimated mean number of contacts between people in the NY area by day.

Social distancing is designed to decrease the number of interpersonal face-to-face contacts within a population and as a result, decrease the basic reproduction rate of the virus causing the epidemic. If the number of contacts is small, epidemics can propagate through interpersonal contact in public places. Using a database of points of interest (POIs), we are able to identify the place where those contacts are happening and their type (category). Figure 7 shows how the fraction of contacts by category have changed once the social distancing metrics were introduced in the NYC area. Before social distance, people came into contact with each other most frequently at work, school restaurants, and shopping. After the social distancing measures, supermarkets and grocery stores have become the main category where interpersonal contacts take place. Now, nearly 25% of the remaining contacts happen in grocery stores and about 15% in places categorized for food and shopping.

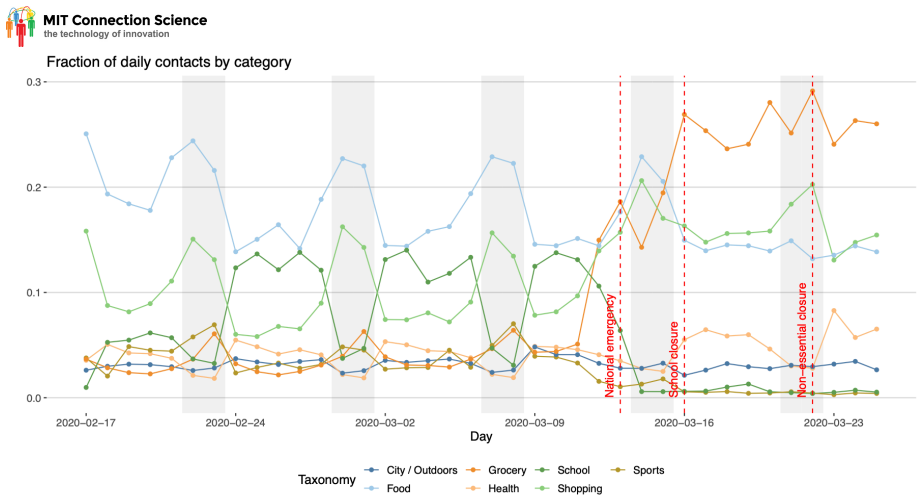

Figure 7. Fraction of total daily contacts by category.

How does social distancing vary across demographics?

Social distancing takes a different toll on everyone. Although some people can shelter-in-place and keep working from home, other people have to still travel large distances to work, get food and shop. In fact, for most of the world, social distancing is "an unimaginable luxury"[4]. Low and middle income neighborhoods often rely on the informal economy and work more in pay-per-hour jobs. For economically vulnerable people, staying home or even 6 feet away from other people is extremely burdensome.

If we had demographic data associated with individuals' mobility data, we could directly examine social distancing across demographics at the individual level. However, the anonymous location data does not contain demographic information. So, instead of examining demographic heterogeneity at the individual level, we examine demographic differences at the neighborhood level, which are defined by census tracts.

Figure 8 shows changes in the total number of contacts across different demographic groups. We specifically look at how income, insurance coverage, age, are related to social distancing. As Figure 8 below shows, neighborhoods with higher income, more health insurance and a lower proportion of people above the age of 65 had more interpersonal contacts before the beginning of the social distancing policies. After social distancing measures took effect, all groups appear to reduce their face-to-face contacts to the same low level. The relative change for each demographic group is different but now each group's level of interpersonal contacts is quite similar.

[4] https://qz.com/1822556/for-most-of-the-world-social-distancing-is-an-unimaginable-luxury/

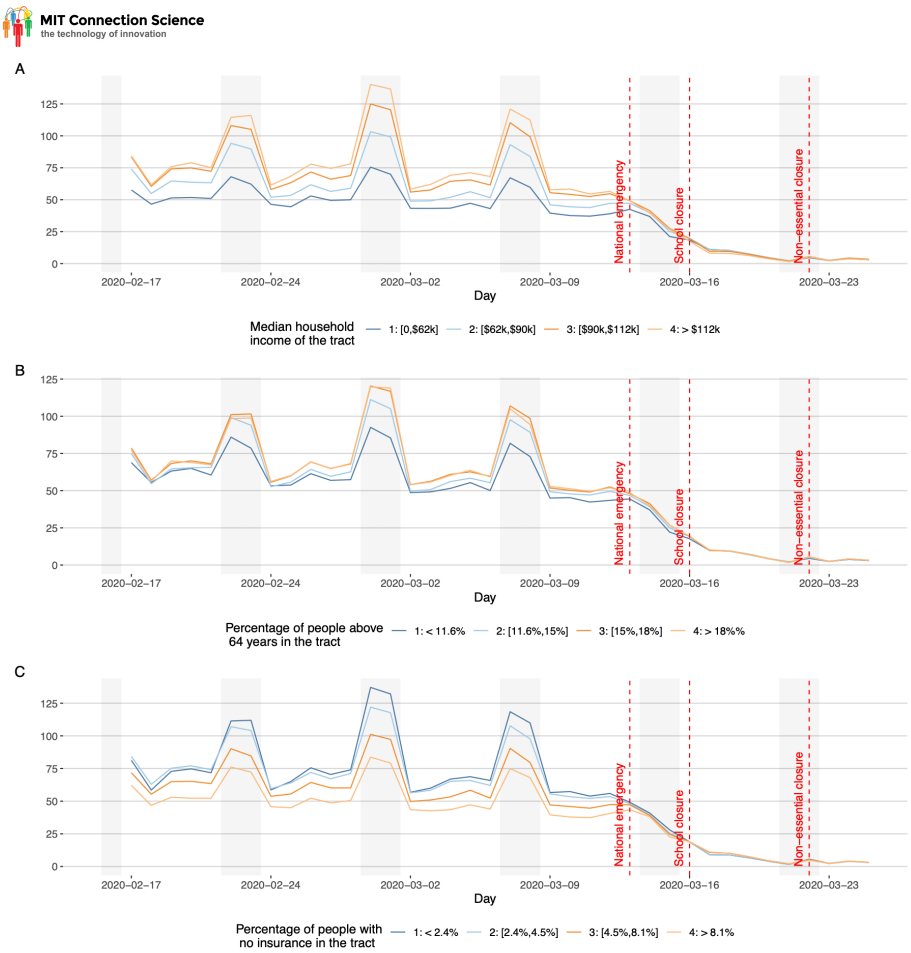

Figure 8. Average number of contacts by day by people in different groups of tracts. A) By median income, B) by percentage of population in the tract above 64 years old, and C) by percentage of people with no insurance.

Discussion

Social distancing is intended to slow the spread of the coronavirus [2,3,4,5,6]. Different countries have adopted different policies to encourage and enforce social distancing through travel or mobility restrictions, school closures or stopping non-essential business activities.

In the New York City metropolitan area, we see strong empirical evidence of social distancing behavior in response to the evolving pandemic and associated social distancing policies. People are travelling less, interacting with fewer other people, and generally staying at home. We estimate that social contacts have reduced by 93% from an average of 75 daily contacts to 5 daily

contacts. While mobility and social contacts vary across demographics during normal times, the social distancing policies have changed how people behave such that the data shows no difference in mobility or social contacts across demographic profiles like income, insurance coverage, and age. Part of the reduction in social contacts can be explained by what we estimate to be 5.5% of the New York City population leaving the city, which is important to note for how this might lead the virus to spread.

The next empirical question is how effective are these social distance policies at reducing the spread of the coronavirus? With high-resolution anonymous mobility data, we can study the effect of mobility-related policies on population-level behavioral responses and how the coronavirus spreads differently across places with different policies. Not only can high-resolution, anonymous mobility data monitor adherence to social distancing policies, but it can also inform epidemiological models based on real time contact matrices [9].

Methods

Stay calculation

For each device in the dataset, we observe a large number of “pings”, consisting of an anonymized device ID, the latitude and longitude, and the exact time and date when the device shared its location. When a person spends a longer time at a fixed location such as a store, we will observe a number of pings that are close to one another, scattered around a single location because of the uncertainty in location measurement and movement of a user. To make it easier to detect whether two users had contact, first map the spatiotemporal pings to “stays”, characterized by a person “spending some time in one place” [10].

The first part of computing the stays is finding the locations that best represent the location of pings we see clustered around point-of-interests like stores. For this mapping we use the infostop algorithm[5]. First, the algorithm clusters stationary points together by taking the median latitude and longitude of events that are consecutive in time and less than 25 meters apart. Subsequently, for each device, we connect different clusters of pings if they happen at different times but at locations that are less than 25 meters apart. Finally, the algorithm computes a single consistent location across times by computing the median latitude and longitude across all the pings that are associated with one of the connected clusters. If, for example, a device was located at the same store on multiple different days in our dataset, this step ensures that we still record this store as a single latitude and longitude pair for each user.

For each of the visits to a location, a stay is generated by taking the time a user arrived at a location (this could be multiple times across different days), the duration of the stay, and the median latitude and longitude computed for that location. To ensure that we are not matching people who walked or drove passed each other, the minimum duration of a stay is 5 min and each stay should consist of at least 2 pings.

[5] See infostop.readthedocs.io and github.com/ulfaslak/infosto

Data representativeness

Our panel comes from a sample of users in the dataset for which we identified the most probable home areas at the level of census tract. Figure 9A shows the scatter plot of the census population (obtained through the ACS 2018-2014) and the number of users in our panel by tract in % of the total. We observe a Pearson correlation of r = 0.68 (in log). Although this correlation is moderately high (especially at the spatial resolution of census tracts), we have also checked that a post-stratification using relative penetrations by census tracts does not significantly change the results. Figure 9B shows for example that the average distance travelled is only slightly modified in magnitude, but the relative change is the same in the panel and weighted panel results.

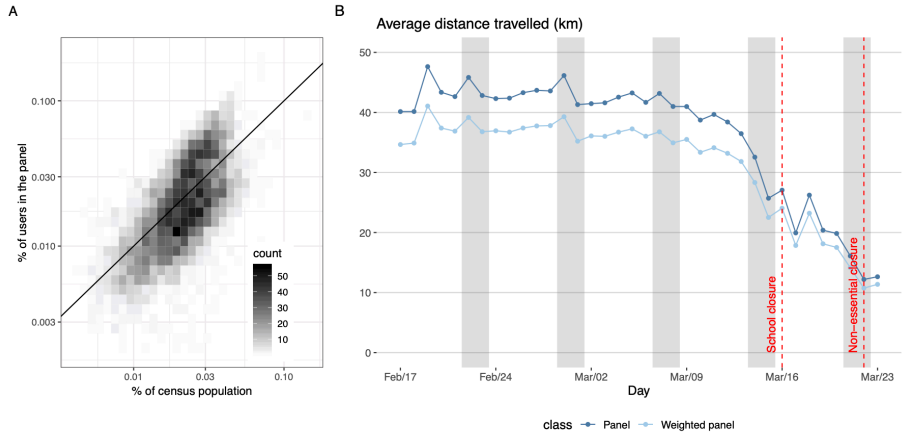

Figure 9: A) Scatter plot of the % of users in our panel in each census tract vs. the percentage of population from the census. B) Comparison between the average distance travelled by users in our panel with the one obtained weighting the % of users by census tract by the census population in that tract.

Glossary of Technical Terms

Contacts
The number of people that each person comes into contact with. We specifically define this as people who are co-located in the same 25 meter radius for at least 5 minutes.

Data panel
We limit our analysis to a sample of 567,000 individuals who appeared in the dataset consistently from February 17 to March 9.

Distance traveled
Line distance traveled every day.

Mobility
The empirical measurement of how people move around a city.

Public 3rd places
Anywhere that is neither a person's home or place of work.

Radius of gyration
The typical size of the area covered by users.

Stays
Any period of time where a user spends at least 5 minutes in the same 25 meter radius.

Trajectory
Timestamped sequence of GPS points reported by person's device.

References

[1] Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, eabb4218. http://doi.org/10.1126/science.abb4218

[2] Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., et al. (2020). Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China. medRxiv, 2020.03.19.20039107. http://doi.org/10.1101/2020.03.19.20039107

[3] Kissler, Stephen M., Christine Tedijanto, Marc Lipsitch, and Yonatan Grad. "Social distancing strategies for curbing the COVID-19 epidemic." *medRxiv* (2020).

[4] Jackson, Michael L., Gregory R. Hart, Denise J. McCulloch, Amanda Adler, Elisabeth Brandstetter, Kairsten Fay, Peter Han et al. "Effects of weather-related social distancing on city-scale transmission of respiratory viruses." *medRxiv* (2020).

[5] Glass, Robert J., Laura M. Glass, Walter E. Beyeler, and H. Jason Min. "Targeted social distancing designs for pandemic influenza." *Emerging infectious diseases* 12, no. 11 (2006): 1671.

[6] Caley, Peter, David J. Philp, and Kevin McCracken. "Quantifying social distancing arising from pandemic influenza." *Journal of the Royal Society Interface* 5, no. 23 (2008): 631-639.

[7] Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., & Tizzoni, M. (2020). COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown. medRxiv, 2020.03.22.20039933. http://doi.org/10.1101/2020.03.22.20039933

[8] Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. "Understanding individual human mobility patterns." *nature* 453.7196 (2008): 779-782.

[9] D. Martín-Corral, A. Aleta, A. Pentland, Y. Moreno, E. Moro, Effectiveness of social distancing strategies for protecting a community from a pandemic with a data- driven contact network based on census and real-world mobility data, COVID-19 Social Distancing Strategies (2020) https://covid-19-sds.github.io

[10] Hariharan, Ramaswamy, and Kentaro Toyama. "Project Lachesis: parsing and modeling location histories." International Conference on Geographic Information Science. Springer, Berlin, Heidelberg, 2004.