How Social Media and Cellphone Data Can Be Used To Track Mobility During Pandemics

Jeff Chan

Assistant Professor, Department of Economics, Wilfrid Laurier University

University of Waterloo COVID-19 Data Privacy Conference

Abstract

The COVID-19 pandemic has prompted some technologies companies such as Facebook to publicly release mobility datasets, leveraging users' geolocation data to calculate how much people in different regions have moved around each day. Such resources can be a cost-effective way for researchers and policymakers to measure how Canadians' movement patterns have changed over time, particularly in response to policies like stay-at-home orders and lockdowns. In this paper, I first provide a brief explanation of mobility data and highlight potential uses, in part via an example from my own research on Canada's mobility changes during the pandemic. I argue that Canada could also benefit from increased availability of cellphone-based mobility data from marketing analytics companies, many of which have already released data for the United States. Finally, I discuss data privacy issues related to mobility data, such as lack of awareness amongst users about how their data will be used and provide some suggestions for how these issues might be mitigated.

Introduction

The COVID-19 pandemic has spurred a massive global effort to understand what types of interventions and policies are effective in stemming its spread. While progress on vaccination is being made, until mass rollout of vaccination programs can reach a sufficiently large number of people limiting contact with others will remain one of the most potent ways to limit the spread of the coronavirus. Staying at home has therefore been encouraged and even mandated through policies and messaging by governments around the world, including governments at all levels in Canada. Understanding the efficacy of such policies and messaging by measuring how movement has been reduced across Canada is thus important to our fight against COVID-19 and in understanding how Canadians' behaviour may be driving surges in cases.

To assess whether its citizens were changing their movement patterns and staying at home more, governments traditionally have had few options. First, they could implement surveys which ask whether people are staying home or reducing movements, but these surveys could potentially be costly to implement and logistically intensive. For example, if we want an evolving picture of how people are behaving, multiple waves of the survey must be administered; if we want a picture of how small, sub-provincial regions are doing across Canada, a large enough number of people must be surveyed each time to make that possible. Finally, people have also been known to respond untruthfully in surveys, particularly for contentious issues. Put simply, Canadians may not feel comfortable stating that they flouted COVID-related guidelines or bylaws in a government-administered survey.

In the absence of data, policy makers could also rely on piecemeal incidents, stories, and reports of people reducing their mobility. This approach, while valuable, can also give a skewed view of how Canadians as a whole, nationally or regionally, have changed their mobility in the

COVID era. It is challenging, without more information or data, to extrapolate from such incidents whether they represent a broader trend in the overall population or whether they are exceptional cases.

Fortunately, the current pandemic has also spurred unprecedented developments in the public availability of movement data. Tech giants such as Facebook and Google have released publicly available data based on their users that, for the first time, allows researchers to measure how much people are moving around or staying at home. Importantly, these data are freely available and are regularly updated. These datasets are also available for geographical regions below the provincial level and allow for comparisons for how mobility has changed across days. Researchers interested in comparing how residents of Waterloo have changed their movements outside of home, day-by-day, relative to those in Ottawa, can do so with data sets such as those from Facebook or Google.

Facebook Mobility Data

What exactly does this type of mobility data measure, and how is it constructed? In the case of Facebook, the data are created by using their users' location data. Facebook users that use the mobile app and consent to collection and use of location data will be part of the underlying database used to calculate their mobility measures. Facebook divides the entire surface of the Earth into squares, or grid cells, of roughly 600m by 600m. The mobility measure is then calculated as, for each region in Canada, the number of grid cells visited on average by eligible users for that region.

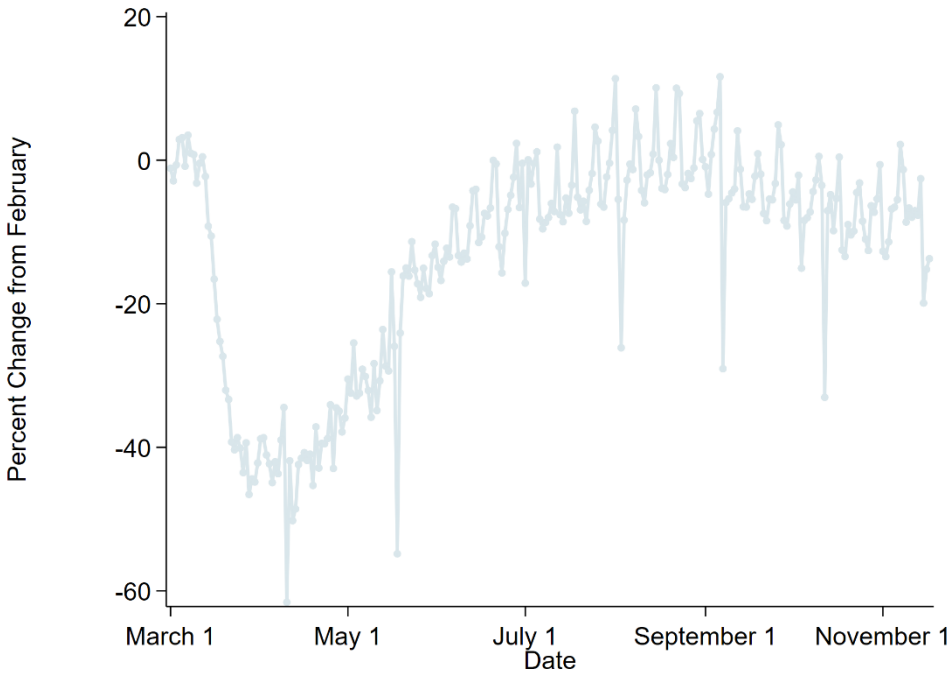

Figure 1: Movement Trend in Waterloo, March-Nov 2020

Figure 1 shows how the Facebook data can be used to trace movement trends across regions in Canada over time. Waterloo's average movement, compared to the baseline pre-COVID period in February, is plotted and shows that there was a huge drop in movement at the start of the pandemic but that movement started crawling upwards by summer. This figure illustrates the helpfulness of the Facebook data, which is regularly updated. If one were interested in finding out whether people in Ottawa were moving around more at the end of November compared to the beginning of November, the Facebook data would provide such an answer. Having measures of how much people move around is key to understanding the causes and consequences of movement limitations in the COVID era. For example, with many regions in Canada and around the world faced with second and even later waves of COVID-19, a critical question is whether these waves have been brought about by people becoming more lax in their precautions and increasing their activities and movements. The Facebook data, and other similar

movement data, can form a crucial part of answering this question. Perhaps most helpfully, the Facebook data are also available at no cost to researchers and can be easily obtained in a few clicks by anyone with a computer and an internet connection, dramatically lowering the barrier to accessing high-quality movement data.

My Research

While the trend shown in Figure 1 shows that people in Waterloo have been moving around more over time, what is not clear is whether this has also been the experience of other regions in Canada. In peer-reviewed research (see Chan (2020a)), I used the Facebook data to address this knowledge gap and describe whether the demographic makeup of different Canadian regions can explain the extent of their movement decline from February to April.[1]

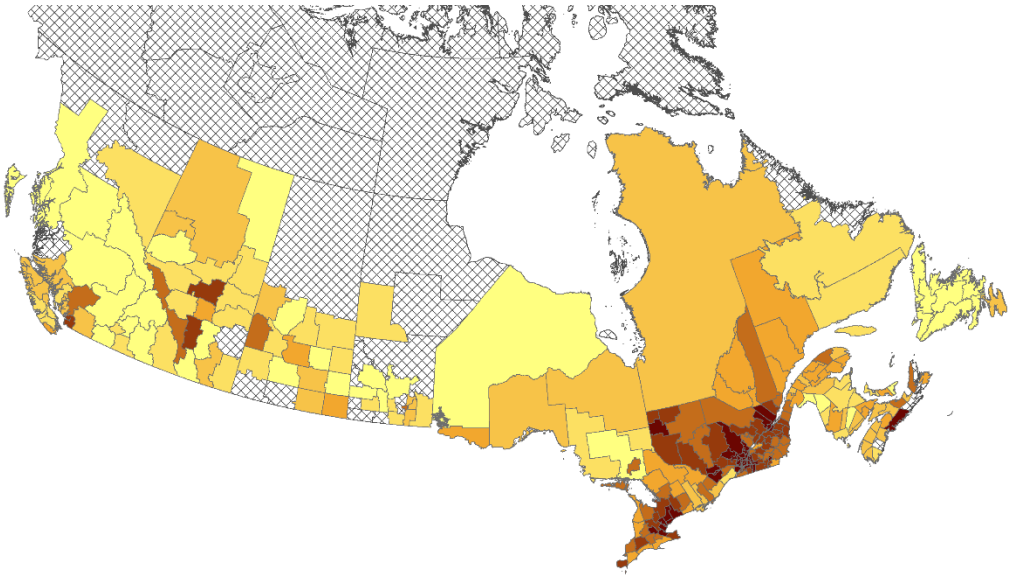

Figure 2: Movement Changes Across Canada, Feb-Apr 2020

[1] In a companion unpublished working paper (Chan (2020b)), I use Google mobility data at the province-day level to perform an analysis of mobility trends across provinces and their relation to COVID-19 cases and deaths as well as other factors that might influence mobility patterns.

Figure 2 helps motivate my analysis by illustrating that there are differences across regions in Canada in how much residents reduced their movements from February to April, 2020. The map, where darker colours represent larger movement reductions, clearly shows that regions in Southern Ontario and Quebec reduced their movements by the most. What is absent from this map, however, is an understanding of why those regions reduced their movements by more than others.

To answer this question, I paired the Facebook data with data from the 2016 Census to obtain information on the characteristics of regions such as the size of the population, population density (population divided by land area), average age, average income, reliance on government transfers, unemployment rate, share of employment in healthcare, share of employment in services, the share of immigrants in the population, the share of people that are married, the share of married couples with children, the share of the population without postsecondary education, and the share of the population who live in apartments. This combination of Facebook and StatCan data allows me to use econometric analysis to examine what types of regions are more likely to reduce their mobility than others, even within the same province. This type of work was especially important at the time during the early stages of the pandemic, because much speculation was floating around on what types of people were more likely to flout social distancing based on specific incidents and anecdotal stories.

I find that highly populated and denser parts of Canada saw larger mobility reductions in April. This is consistent with the graphical evidence from the map in Figure 2. Conversely, regions with a higher share of its population living in apartments reduced their mobility by less. I do not find strong or compelling evidence that the age, education, immigrant or income composition of a region have any association with how much it reduces its mobility. This

suggests that, at least during the early stages of the pandemic, much of the reporting around younger people, immigrants, or those with lower incomes ignoring stay-at-home orders and guidelines may have been driven by isolated incidents and overly active speculation.

I also tested whether different regions differed in how much they were able to maintain their February-April movement reductions into May. More specifically, I took the change in movements by region from February to April and compared these changes to the change in movements from April to May. A lack of difference in movement changes between these two periods would suggest that a Canadian region did not experience much so-called lockdown fatigue, maintaining the reduction in movement that they experienced from February-April into May. I did not find much evidence that lockdown fatigue was driven by most things, like age or income, that were being speculated on at the time. While this conclusion may have changed as the pandemic continues, the continued availability of the Facebook data makes follow-up studies easy to conduct for interested researchers.

The work I did has limitations. The results were suggestive that some groups of Canadians may be reducing their movements at different rates than the rest of the country but was not strong causal evidence based on a randomized or natural experiment. My paper also highlighted relationships between regional characteristics and movement reductions during COVID, but did not provide further evidence on why these relationships exist. For example, although more apartment dwelling in a region was found to be related to smaller movement reductions, my paper was silent on whether this was because apartment dwellers had less space and thus compensated by leaving their homes more often or whether there was another mechanism at play. Nonetheless, my work was useful, particularly during the early stages of the pandemic, because simply knowing which Canadians may be reducing mobility by less can help

policymakers understand when and why policies and messaging that encourage staying at home are falling short. For example, if apartment dwellers reduce their movements by less in the COVID era, as my results suggest, is it because they have less space in their homes and need additional supports? Descriptive research such as mine provides sign posts to dig deeper into issues like these and follow up with further investigation. As many data-minded researchers often echo, correlation is not causation but can be a good first step.

My research was made possible thanks to the easy accessibility of the Facebook data. Available at no cost and with no administrative hurdles to clear beyond clicking a link to download, researchers interested in using this data can do so rapidly.[2] This greatly increases the number of people that can work on these issues and aids in the democratization of COVID research, particularly for more junior researchers and those with limited research funding. Combined with Facebook's regular updating of its data, this will allow us to continue adding to what we know about movement during the COVID era.

Cellphone Data

In the United States and other countries, additional publicly available mobility data exists which differ from the Facebook data by not coming from one particular source but multiple unnamed applications which collect their users' data. Like the Facebook data, though, location information on users is generated by collecting mobile device locations through the GPS "pinging" their cellphones throughout the day. These cellphone-based datasets have been produced and publicly released by marketing analytics companies such as Safegraph, Descartes

[2] The data are available, as of January 7, 2021, from the following link in a zipped folder: https://data.humdata.org/dataset/movement-range-maps

Labs, and PlaceIQ. Like the Facebook data, these datasets have been regularly updated, increasing their potential for time-sensitive research or statistical modelling.

These cellphone-based datasets do also provide additional advantages over mobility data from Facebook, Google, or Apple. Data from Safegraph and other similar companies are often drawn from a non-disclosed list of numerous applications that then provide their data to these companies; this means that Safegraph and other similar data are more likely to fairly represent the movement patterns of Canadians with mobile devices, compared to data sources that rely on the usage of one app, like Facebook. In addition, some, like the data from Safegraph, are available at significantly more granular levels of geography, allowing researchers to examine neighbourhood level mobility patterns within a given city; in contrast, currently available mobility data for Canada can drill down to the city level at best. This finer-grain detail can allow researchers to examine whether certain neighbourhoods are staying at home more, and whether these behaviours are translating into differences in COVID infections. The data released by PlaceIQ in partnership with academics, in contrast, contains measures of exposure that not only measures movement but also exposure to other people, as measured by other devices that are encountered while away from home.

Unfortunately, few analytics companies (as of mid-February 2021) have publicly released extensive mobility metrics for Canada at the same level of quality and detail as in the US. Safegraph, as one example, only recently released a beta version of their data for Canada with a limited sample that does not yet extend back to the early stages of the pandemic, as of the time of writing. Environics Analytics, for another example, have used cellphone GPS data to construct postal-code level movement measures to produce reports on mobility in Canada. Unfortunately,

the Environics data is not currently publicly available and can only be obtained at a cost that would be significant for researchers, especially those without generous research funding.

Privacy Issues

Facebook has implemented safeguards to protect the privacy of users when releasing mobility data. Facebook does not report data for regions where the number of people that the calculations are based on is too small; such small numbers might make it possible for users of the data to identify individual movement patterns. In addition, Facebook introduces statistical noise into the mobility data, meaning that each subnational unit's actual mobility cannot be precisely obtained from the public data. This is similar to what Statistics Canada does in its publicly released Census data on the demographics of small regions in Canada. This added noise does not prevent researchers from using the data for statistical analyses but does make it more difficult to identify any particular Facebook user.

One final privacy safeguard is that only Facebook mobile app users that allow Facebook to access their location history and collect this information in the background are included in the mobility calculations. Any user can therefore prevent themselves from being included in the mobility dataset via opting out in the app's settings. Despite this ability to opt out, users that do initially consent may not fully understand the ramifications of their actions. Put more simply, users may not know that their personal data and movements may be tracked and included in Facebook's publicly released mobility data; if they did, their decision to opt out might be different.

Cellphone-based data present additional privacy challenges since measures are often based on data from unnamed third-party apps, meaning that Canadians have no easy way of

knowing if they are included in the data. Cellphone users may unwittingly give consent for their data to be sold or forwarded to others when agreeing to terms and policies associated with apps on their phones if they do not read agreements in full. Once given, it is not clear if or how users can remove themselves from such mobility datasets, especially since there is often no way to know what third party apps these companies are creating their data sets from.

To further safeguard privacy and ensure that Canadians are making informed choices about their data privacy, more awareness in how companies are using individuals' data should be made a key priority; in particular, Canadians should be informed and equipped with tools to find out if their data has been sold to third parties by applications or businesses that they have provided their data to. There should also be a way for Canadians to find out what databases their data have been included in, especially for those datasets with no obvious singular source. Finally, easier and more transparent ways to withdraw personal data from inclusion in data sets, mobility or otherwise, would also be an excellent step towards providing Canadian users with more control over their own personal information in the Internet age.

Conclusion

Mobility data provides researchers with a powerful tool to observe movement trends across regions in Canada during the COVID-19 pandemic. Even in its aggregated and de-identified format, these data sets are a very cost and easy-to-access tool. Provided that these companies continue to regularly release updates to the mobility measures, such data would also be suitable for use in forecasting and statistical modelling, especially for regions or researchers that have limited resources. Researchers are also not the only beneficiaries. Public health officials have also utilized mobility data to help understand how lockdown policies have

impacted movement, and to help with COVID-related projections.[3] Finally, journalists have used mobility data to provide timely analyses to the public; for example, a January 10, 2021 Toronto Star article by Jennifer Yang referenced data analysis using Environics Analytics cellphone mobility data to try to connect post-Christmas COVID outbreaks with mobility rate spikes during the holidays across Ontario.

On the other hand, the unprecedented public release of these data sets highlights privacy concerns with the collection and use of personal information. One key issue is lack of awareness among users, which makes it challenging to know whether those included would truly be comfortable with their inclusion in public releases of mobility datasets, even if privacy safeguards have been enacted. Empowering Canadians to better understand how their data are being used by companies and providing more ways to remove personal data from inclusion in such databases would be useful in ensuring that Canadians have the level of data privacy that they want.

References

Chan, Jeff. 2020a. "The Geography of Social Distancing in Canada: Evidence from Facebook." *Canadian Public Policy*, Vol. 46(S1): S19-S28.

Chan, Jeff. 2020b. "Using Google Data to Understand Canadian Movement Reductions During the COVID-19 Pandemic." SSRN Working Paper 3599227.

[3] In a December 10, 2020 briefing on COVID projections, Ontario public health officials presented mobility trends across several Ontario regions. Slides are available, as of January 10, 2021, here: https://covid19-sciencetable.ca/wp-content/uploads/2020/12/Update-on-COVID-19-Projections_December-10-2020.pdf

Yang, Jennifer. 2021. “Why did Ontario COVID-19 rates surge after Christmas? New cellphone mobility data offers some clues.” Toronto Star, 10 January. Accessed 10 January at https://www.thestar.com/news/gta/2021/01/10/why-did-ontario-covid-19-rates-surge-after-christmas-new-cellphone-mobility-data-offers-some-clues.html.