

TRANSPORT FINDINGS

Changes in the Representativeness of Strava Bicycling Data during COVID-19

Jaimy Fischer¹ , Trisalyn Nelson² , Meghan Winters¹ ¹ Faculty of Health Sciences, Simon Fraser University, ² Department of Geography, University of California Santa Barbara

Keywords: active transportation, bicycling, big data, crowdsourced, COVID-19, Strava

<https://doi.org/10.32866/001c.33280>

Findings

COVID-19 prompted a large global increase in Strava app use that could have implications for the representativeness of Strava bicycling data. We evaluated change in Strava representation of the general bicycling population from 2019 to 2020 by correlating Strava to counts of all bicyclists in Vancouver and Victoria, Canada. Strava data became more representative during COVID-19 lockdowns, likely because more bicycling activities were logged on the app. Age and gender bias in Strava data decreased and demographics of app users became more representative of the general population. For cities looking to increase Strava data reliability we suggest a benchmark Strava usage rate of ~7% of the total bicycle volumes.

1. Questions

Strava is a social fitness app that people use to track their bicycling activities. Strava data, generated from crowdsourced activity tracking on the Strava fitness app, provide bicycle ridership data with unprecedented spatial and temporal resolution and the data are used in a growing number of bicycling research, planning, and advocacy projects (Lee and Sener 2021; Nelson, Ferster, et al. 2021). Strava data have shown moderate to strong correlations with the general bicycling population (Lee and Sener 2021). Bias related to who uses Strava to track their rides is well documented (Lee and Sener 2021), with the data tending to underrepresent women and older adults (Griffin and Jiao 2019). Research on sociodemographic representation outside of age and gender (e.g., by race or socioeconomic status) is sparse because Strava does not systematically collect such characteristics, but the limited work available suggests that fitness app users are predominantly White and more affluent (Garber, Watkins, and Kramer 2019)—a trend that mirrors the sociodemographic characteristics of who bikes in Canada (Firth et al. 2021). Yet, through the pandemic time many more people have started using Strava, with app usership rising 33% to 73 million, and collectively users uploading 82% more activities (1.1 billion) from 2019 to 2020 (Strava 2020). These huge increases provoke further questions about changes in data representativeness. Our goal is 1) to evaluate the impacts

^a Corresponding author:
Jaimy Fischer
jaimyf@sfu.ca
778-676-7773

of increased Strava app use on city-level correlation with counts of all bicycling and 2) to identify changes in correlation in subsets of Strava data representing trip purpose and demographics.

2. Methods

Study area

The study area is in the west coast of Canada and includes the city of Vancouver and Greater Victoria (including Victoria, Oak Bay, Saanich, and Esquimalt municipalities), BC, Canada (Statistics Canada 2016). We chose these cities because both had official bike counts, measures of all bicycle ridership that are needed as a benchmark to assess change in representativeness of bicycling. Both cities also have similarly high bicycling to work mode shares relative to provincial and national averages and have substantial city investment in all ages and abilities bicycling infrastructure but differ in size and the extent of Strava usage ([Table 1](#)).

Bicycle count data

We created a bicycle count database for the period April 1-October 31 (2019 and 2020). City staff provided counts from 19 permanent bicycle counter locations in Vancouver, and 29 temporary and permanent locations in Victoria. City count data were provided at hourly aggregations. We used the Strava dashboard to access city data profiles on the total number of people and public bicycling activities recorded on Strava and to download the Strava bicycle count data used in analysis (Strava Metro 2022). These data were provided as aggregate hourly-level counts of bicycling activities across all links in the street network. We matched counts from both sources by location, date, and hour of the day, removing observations with missing data. Then we grouped observations by year for analysis.

Table 1. Study area characteristics.

| Location | Population ^a | | Temperature and precipitation ^b | | | Bicycle facilities ^c | | Bicycle mode share ^a (%) | | |
|-------------------|-------------------------|------------------------------------|--|--------------------------|------------------------------|---------------------------------|---------------------------|-------------------------------------|------|--------|
| | People | Density (ppl/ km ²) | Winter avg. temp (°C) | Summer avg. temp (°C) | Annual precipitation (mm) | Total distance (km) | Density (km/ 1000 ppl) | City | B.C. | Canada |
| City of Vancouver | 631,486 | 5,492.0 | 4.1 | 18.0 | 1189.0 | 495.6 | 0.8 | 6.1 | 2.5 | 1.4 |
| Greater Victoria | 235,689 | 1,673.2 | 4.6 | 16.9 | 882.9 | 205.4 | 0.9 | 8.7 | 2.5 | 1.4 |

^a Population and mode share data were obtained from the 2016 Canadian census (Statistics Canada 2016)

^b Climate data were obtained from Environment Canada (Environment and Climate Change Canada 2021)

^c Bicycle infrastructure was obtained from open data for Vancouver (City of Vancouver 2022) and staff from the Capital Regional District for Greater Victoria (including Victoria, Oak Bay, Saanich, and Esquimalt).

Table 2. Change in Strava usage rates and city-level correlation.

| City | Year | Strava app use | Trips logged to Strava | City-level Strava usage rate | <i>R</i> , city-level (pooled sample) |
|-----------|------|----------------|------------------------|------------------------------|---------------------------------------|
| Vancouver | 2019 | Reference | Reference | 5.7% | 0.42 |
| | 2020 | +48 % | +41 % | 12.1% | 0.73 |
| Victoria | 2019 | Reference | Reference | 7.0% | 0.85 |
| | 2020 | +26 % | +27 % | 15.8% | 0.87 |

Analysis

To compare trends between cities, we calculated the proportional change in the number of app users and bicycling activities logged to Strava using data profiles from the Strava dashboard (Strava Metro 2022). To evaluate how Strava ridership represented counts of all bicyclists, we measured Strava usage rates and assessed correlation between Strava and counts of all bicyclists. The Strava usage rate is the proportion of all trips counted at city bike counters that were captured in the Strava data (Camacho-Torregrosa et al. 2021), and the correlation indicates whether Strava bicyclists provide a good approximation of all bicyclists (Lee and Sener 2021). Subsets of Strava data may be more representative of the general population and drive trends observed in the pooled sample (all observations combined) (Fischer, Nelson, and Winters 2020; Nelson, Roy, et al. 2021); so, we calculated usage rates and correlation measures for the pooled sample of Strava ridership and for subsets of Strava data by trip purpose and demographics. Specifically, we used labels in the Strava data to extract data subsets for recreation and commute ridership, and for the sample of ridership generated by women and by older adults (55+).

3. Findings

The number of people using Strava to track bicycling activities and the total volume of bicycling activities recorded on Strava increased in both cities, with the greatest increases in Vancouver (Table 2). In Vancouver app usership nearly doubled (48% increase) and Strava bicycling activities increased by 41%, whereas in Victoria app usage and activities increased by 26 and 27%, respectively.

City-level Strava usage rates and correlations between Strava and counts of all bicyclists appear to be linked, with usage rates ~7-10% possibly corresponding to strong city-level correlation ($R \geq 0.7$). To illustrate, the Strava usage rate rose from 5.7 to 12.1% in Vancouver and city-level correlation increased substantially, jumping from 0.42 in 2019 to 0.73 in 2020. Pre-COVID correlation was already strong in Victoria when the Strava usage rate was 7% (2019) and increased marginally in 2020 (0.85 to 0.87) even as the Strava usage rate exceeded 15%.

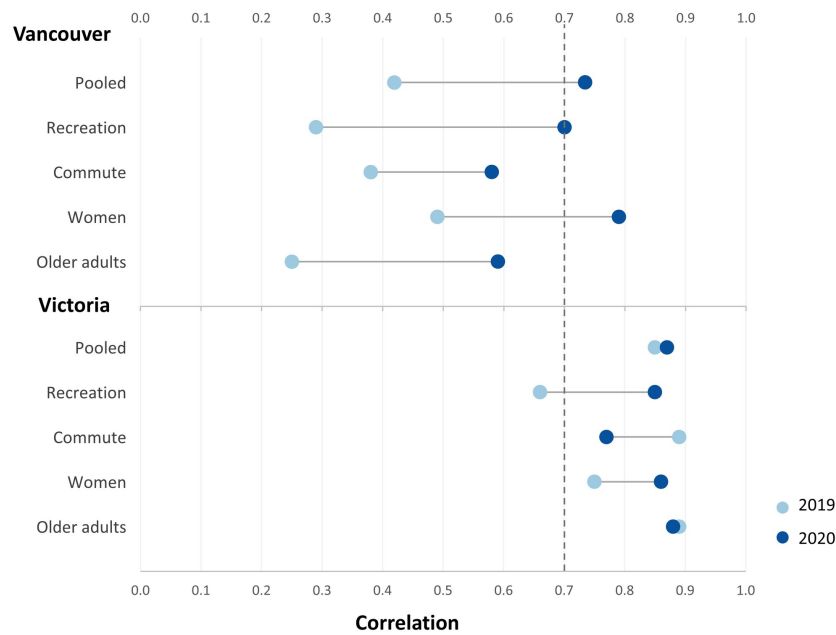


Figure 1. Change in correlation (2019-2020) between Strava and counts of all bicyclists. The dashed black line marks the value for strong correlation (0.7 or higher). Year-on-year correlation increased in both cities, but Vancouver had the greatest gains, with correlation jumping from 0.42 in 2019 to 0.73 in 2020. Pre-COVID correlation was already strong in Victoria in 2019 and increased marginally in 2020. The subset of Strava data generated by women demonstrated strong correlations in both cities in 2020, and correlation in the subset generated by older adults increased substantially in Vancouver.

Differences in correlation by trip purpose and demographics imply that subsets of Strava ridership may be more representative of the general bicycling population. In Vancouver, all Strava data subsets showed year-on-year increase in correlation with counts of all bicycling, and the strongest correlation overall was for the sample of Strava ridership generated by women (0.79). In Victoria the strongest correlation occurred in the sample of ridership generated by older adults (0.88). In both cities, large increases in correlation (28-140%) in the recreational ridership sample substantiate research demonstrating that bicycling for recreation increased dramatically when cities went into lockdown during the first wave of the pandemic (Buehler and Pucher 2021).

With higher city-level Strava usage rates, we also see stronger correlations in the data subsets for women and older adults, suggesting these higher usage rates may be indicative of better representation of women and older adults in the data. Bias in Strava data related to who is using the app impacts data reliability (Lee and Sener 2021). Our findings imply that age and gender bias in Strava app use may have decreased in these cities and that Strava data became more representative when app use surged in the first months of the pandemic. These findings also reflect how bicycling, especially for recreation, increased as cities went into pandemic lockdowns in 2020. Virtual connection, work from home, and policy that supports active transportation are likely to continue (van Wee and Witlox 2021), but the long-term effects of COVID-19 on bicycling

and Strava use are still unknown. We expect at least some of the demographic shifts in Strava data to persist, but future studies would help clarify these uncertainties.

Through this analysis we show that Strava representativeness of the general bicycling population is dynamic and influenced by level of app use, which is influenced by external individual (Garber, Watkins, and Kramer 2019), social (Couture 2021), and environmental forces (e.g., pandemic lockdown measures and more recently, extreme weather events (Strava 2020, 2021)). With better representation of women and older adults in Strava data, planners and researchers might use these subsets more purposefully to advance active transportation equity goals as these are priority populations for increasing bicycling in Canada. As well, when enough women and older adults are using Strava they may be a subsample of Strava data that best represent population-level bicycling patterns. In both cities, increases in Strava app use corresponded with better representation (i.e., higher correlation) of the general bicycling population, but further gains may be marginal when the proportion of all bicycling trips logged by Strava reaches some threshold. For cities looking to increase Strava data representativeness we suggest a benchmark Strava usage rate of $\geq \sim 7\%$. This recommendation could be bolstered by additional studies in medium, small, and low bicycling cities. For cities looking to increase Strava usage rates, campaigns that encourage bicyclists to map their rides on Strava can increase app uptake (Boss et al. 2018).

Acknowledgements

The authors would like to acknowledge Strava for providing the data and timely review of our work and staff at the Capital Regional District and the City of Vancouver for providing us with official bicycle count data.

Submitted: February 10, 2022 AEDT, Accepted: March 09, 2022 AEDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Boss, Darren, Trisalyn Nelson, Meghan Winters, and Colin J. Ferster. 2018. "Using Crowdsourced Data to Monitor Change in Spatial Patterns of Bicycle Ridership." *Journal of Transport & Health* 9 (June): 226–33. <https://doi.org/10.1016/j.jth.2018.02.008>.
- Buehler, Ralph, and John Pucher. 2021. "COVID-19 Impacts on Cycling, 2019–2020." *Transport Reviews* 41 (4): 393–400. <https://doi.org/10.1080/01441647.2021.1914900>.
- Camacho-Torregrosa, Francisco Javier, David Llopis-Castelló, Griselda López-Maldonado, and Alfredo García. 2021. "An Examination of the Strava Usage Rate—a Parameter to Estimate Average Annual Daily Bicycle Volumes on Rural Roadways." *Safety* 7 (1): 8. <https://doi.org/10.3390/safety7010008>.
- City of Vancouver. 2022. "Bikeways." https://opendata.vancouver.ca/explore/dataset/bikeways/information/?disjunctive.bike_route_name&disjunctive.bikeway_type&disjunctive.year_of_construction.
- Couture, Jesse. 2021. "Reflections from the 'Strava-Sphere': Kudos, Community, and (Self-)Surveillance on a Social Network for Athletes." *Qualitative Research in Sport, Exercise and Health* 13 (1): 184–200. <https://doi.org/10.1080/2159676x.2020.1836514>.
- Environment and Climate Change Canada. 2021. "Canadian Climate Normals." Government of Canada. https://climate.weather.gc.ca/climate_normals/index_e.html.
- Firth, Caislin L., Michael Branion-Calles, Meghan Winters, and M. Anne Harris. 2021. "Who Bikes? An Assessment of Leisure and Commuting Bicycling from the Canadian Community Health Survey." *Findings*, May. <https://doi.org/10.32866/001c.22163>.
- Fischer, Jaimy, Trisalyn Nelson, and Meghan Winters. 2020. "Comparing Spatial Associations of Commuting versus Recreational Ridership Captured by the Strava Fitness App." *Findings* 770 (September): 1–9. <https://doi.org/10.32866/001c.16710>.
- Garber, Michael D., Kari E. Watkins, and Michael R. Kramer. 2019. "Comparing Bicyclists Who Use Smartphone Apps to Record Rides with Those Who Do Not: Implications for Representativeness and Selection Bias." *Journal of Transport & Health* 15 (December): 100661. <https://doi.org/10.1016/j.jth.2019.100661>.
- Griffin, Greg, and Junfeng Jiao. 2019. "Crowdsourcing Bicycle Volumes: Exploring the Role of Volunteered Geographic Information and Established Monitoring Methods." *SocArXiv*, January. <https://doi.org/10.31235/osf.io/e3hbc>.
- Lee, Kyuhyun, and Ipek Nese Sener. 2021. "Strava Metro Data for Bicycle Monitoring: A Literature Review." *Transport Reviews* 41 (1): 27–47. <https://doi.org/10.1080/01441647.2020.1798558>.
- Nelson, Trisalyn, Colin Ferster, Karen Laberee, Daniel Fuller, and Meghan Winters. 2021. "Crowdsourced Data for Bicycling Research and Practice." *Transport Reviews* 41 (1): 97–114. <https://doi.org/10.1080/01441647.2020.1806943>.
- Nelson, Trisalyn, Avipsa Roy, Colin Ferster, Jaimy Fischer, Vanessa Brum-Bastos, Karen Laberee, Hanchen Yu, and Meghan Winters. 2021. "Generalized Model for Mapping Bicycle Ridership with Crowdsourced Data." *Transportation Research Part C: Emerging Technologies* 125 (April): 102981. <https://doi.org/10.1016/j.trc.2021.102981>.
- Statistics Canada. 2016. "Census Profile, 2016 Census." <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E>.
- Strava. 2020. "Year in Sport." https://1n4rcn88bk4ziht713dla5ub-wpengine.netdna-ssl.com/wp-content/uploads/2020/12/USA_YIS_2020.pdf.
- . 2021. "Year in Sport." <https://1n4rcn88bk4ziht713dla5ub-wpengine.netdna-ssl.com/wp-content/uploads/2021/12/YIS-2021-Press-Book-United-States.pdf>.

Strava Metro. 2022. “Home.” 2022. <https://metro.strava.com/>.

van Wee, Bert, and Frank Witlox. 2021. “COVID-19 and Its Long-Term Effects on Activity Participation and Travel Behaviour: A Multiperspective View.” *Journal of Transport Geography* 95 (2021): 103144. <https://doi.org/10.1016/j.jtrangeo.2021.103144>.