

ARTICLES

# Where to put bike counters? Stratifying bicycling patterns in the city using crowdsourced data

Vanessa Brum-Bastos\*, Colin J. Ferster†, Trisalyn Nelson‡, Meghan Winters\*\*

Keywords: machine learning, smart cities, urban planning, transportation, mobility, strava, bicycling

<https://doi.org/10.32866/10828>

---

## Transport Findings

When designing bicycle count programs, it can be difficult to know where to locate counters to generate a representative sample of bicycling ridership. Crowdsourced data on ridership has been shown to represent patterns of temporal ridership in dense urban areas. Here we use crowdsourced data and machine learning to categorize street segments into classes of temporal patterns of ridership. We used continuous signal processing to group 3,880 street segments in Ottawa, Ontario into six classes of temporal ridership that varied based on overall volume and daily patterns (commute vs non-commute). Transportation practitioners can use this data to strategically place counters across these strata to efficiently capture bicycling ridership counts that better represent the entire city.

---

## 1. RESEARCH QUESTION AND HYPOTHESIS

When a city aims to quantify bicycling volume, data are typically collected at a sample of locations via manual counts and automatic counters (Griffin et al. 2014; Nordback et al. 2013). The goal is to select locations for bicycle volume counting that are representative of volumes throughout the city. Systematic approaches that utilize fine spatial and temporal resolution crowdsourced data have the potential to improve sampling efficiency and representativeness for bicycle counts.

New bicycling ridership data, which are continuous across space and time, are available from crowdsourced tools (i.e., Strava) and provide an opportunity to develop methods to stratify sampling when locating count stations. Strava's data provide the number of app users on a street segment every minute and, despite a bias toward leisure bicycling, have been shown to moderately correlate with overall bicycling ridership (Jestico, Nelson, and Winters 2016; Boss et al. 2018). Strava can be used to differentiate high- and low-volume streets, as well as streets used during peak commute times. Streets with similar ridership patterns can be categorized, and these categories can be used as factors to stratify field- and volunteer-based bicycling volume counter programs to obtain more precise information (Nordback et al. 2019).

---

\* **Institution:** Arizona State University (AZ)

† **Institution:** University of Victoria **Department:** Department of Geography

‡ **Institution:** Arizona State University **Department:** School of Geographical Sciences and Urban Planning

\*\* **Institution:** Simon Fraser University **Department:** Faculty of Health Sciences

We used crowdsourced data on bike ridership and continuous signal processing data mining to classify street segments by the temporal patterns in ridership. These were demonstrated in a case study of bicycling patterns in Ottawa, Ontario, Canada.

## 2. METHODS AND DATA

Strava Metro provided bicyclist counts for 2016 at one-minute temporal resolution for a 20 km<sup>2</sup> area in downtown Ottawa, where correlation coefficients between daily total counts from Strava and automated counters ranged from 0.76–0.96 (Boss et al. 2018). In order to obtain temporal profiles of bicycling ridership, we computed the average hourly activity count for bicycling on weekdays during nonwinter months (March–September) for each hour of the day for the 3,880 non-zero-count street segments.

To quantify the differences between bicycling ridership profiles of all street segments, we computed dynamic time warping (DTW) distances. DTW finds the optimal global alignment between two time series by computing a pairwise distance matrix ( $M$ ) for all points in both time series (Equation 1) and by finding the least-cost path in  $M$  (Sakoe and Chiba 1978).

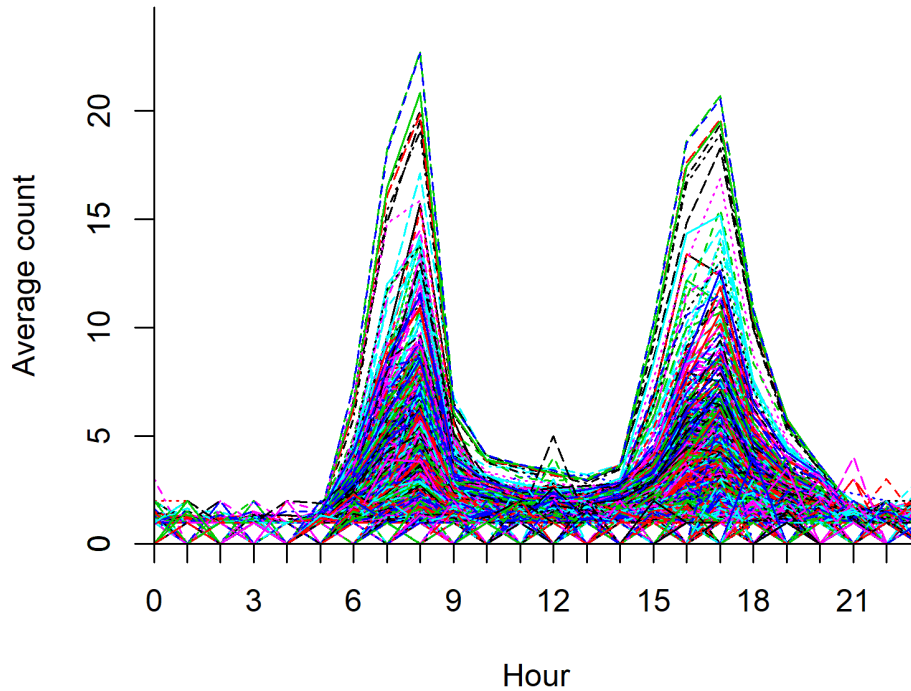
$$D = |A_i - B_j| + \min \left\{ \begin{array}{l} D [i - 1, j - 1] \\ D [i - 1, j] \\ D [i, j - 1] \end{array} \right\}$$

Equation 1

Here  $A_i$  is the mean cyclist count for street segment  $A$  at hour  $i$ ,  $B_j$  is the mean cyclist count for street segment  $B$  at hour  $j$ ,  $D [i - 1, j - 1]$  is the previously computed difference between mean cyclist counts in the previous hour for both time series,  $D [i - 1, j]$  is the previously computed difference between mean cyclist counts at the current hour for series  $B$  and the previous hour for series  $B$ , and  $D [i, j - 1]$  is the previously computed difference between mean cyclist counts at the current hour for series  $A$  and the previous hour for series  $B$ .

We used a Sakoe-Chiba band (Sakoe and Chiba 1978) to constrain warping to a three-hour interval centered at the hour being warped to avoid unrealistic least-cost paths (Zhang et al. 2017). Pairwise distances yielded by the least-cost path were used to generate a 3,880 by 3,880 matrix ( $W$ ) accounting for the differences among all bicycling ridership profiles.

In order to group street segments with similar bicycling ridership patterns, we applied Ward's hierarchical bottom-up clustering algorithm to the  $W$  matrix of time series data (no spatial relationships in the road network data were used in



**Figure 1:** Weekday Average Temporal Profiles of Bike Ridership for all Street Segments in the Study Area for Nonwinter Months (March–September)

the clustering). The algorithm starts with each street segment as its own group and successively merges them into clusters based on the minimum increase in the error sum of squares (Murtagh and Legendre 2014). We used the Calinski-Harabasz Index (CHI) (Calinski and Harabasz 1974) to select the optimal number of clusters. CHI considers the dispersion within and between groups; higher values indicate a better partition (Ahmed 2012).

We varied the number of clusters from 1–50 (more clusters than is practical to interpret) and selected the configuration with the highest CHI. We plotted the profiles within each cluster and analyzed their main characteristics to assign labels that are interpretable by city planners.

### 3. FINDINGS

Figure 1 shows the temporal profiles of ridership for all street segments in our study area. Similar profiles have been generated from single bicycling counters (Miranda-Moreno et al. 2013), but using Strava allows rich spatial resolution (every single street segment) in concert with the temporal richness. While Strava provides data based on only a sample of riders and there are demographic biases in the app users, research has shown the spatial patterns in this ridership data correlate with overall bicycle ridership volumes (Jestico, Nelson, and Winters 2016; Boss et al. 2018).

Figure 2 shows the temporal profiles for street segments within each bicycling ridership class identified by the clustering algorithm. We identify patterns of ridership and differences in volume (Figures 2a–d). The pattern we label as

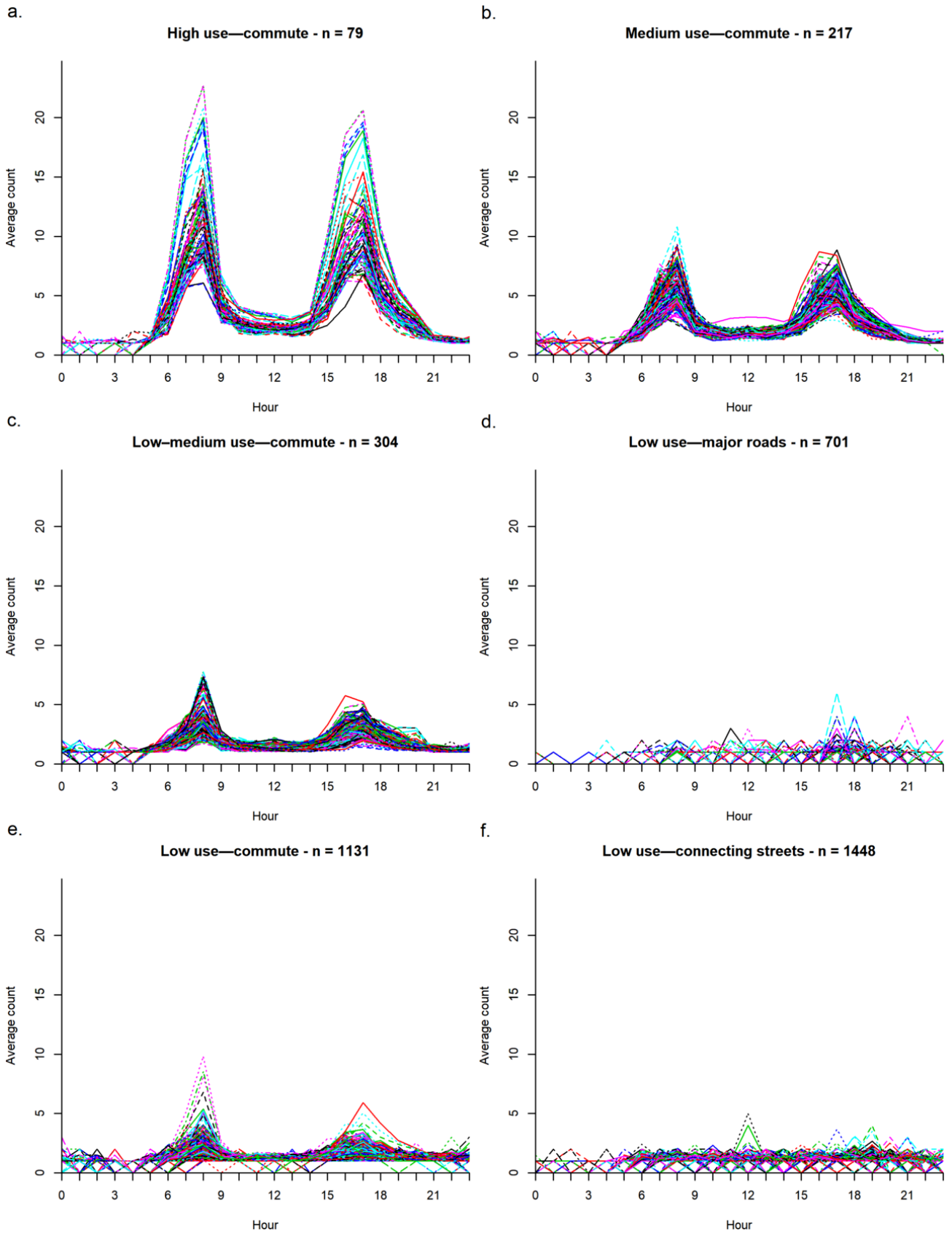
commute is characterized by two peaks (approximately 8:00 a.m. and 6:00 p.m.) with the maximum average volume of bicyclists varying from 25 (high use) to 5 (low use) at peak times. Figures 2e and 2f show the low-use classes, in which bike ridership is sporadic and the average volume of bicyclists is no higher than 5.

Figure 3 shows the mean temporal profile for each ridership class (a) and the spatial distribution of these classes (b). Street segments with high and medium commute volumes are along the Ottawa River, the Rideau Canal, the Trillium Pathway, and Laurier Avenue West. While the clusters had similar temporal patterns, differentiating classes by magnitude was useful for an end goal of stratified sampling.

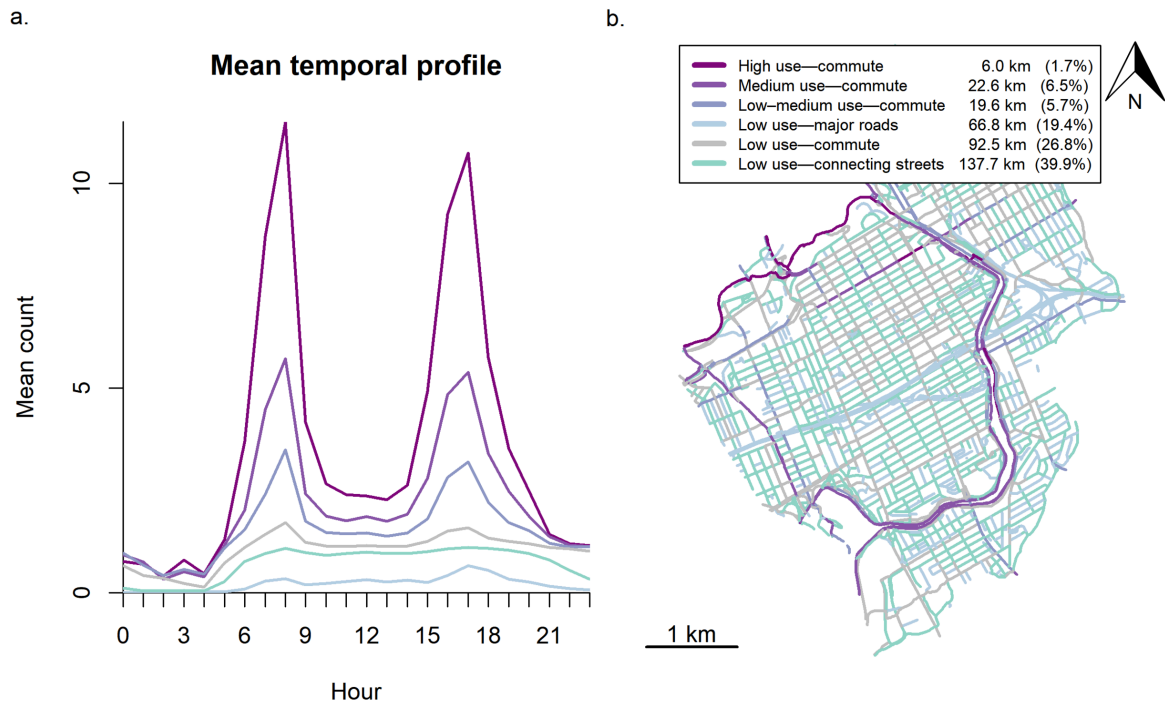
Using crowdsourced bicycling data and continuous time-series processing allowed us to map street segments into classes of temporal patterns of bicycle ridership that can be used for targeted interventions and stratifying count programs. Ridership classes can help planners determine the most appropriate locations and times to obtain bicycle volume counts that are representative of ridership across the city. We recommend placing four or more counters per bicycling ridership class to maximize the precision of bicycling volume estimates (Nordback et al. 2019) and at least 50 counters for a medium-sized city (Roy et al., Forthcoming). This method is a novel systematic approach based on spatially and temporally detailed crowdsourced data to determine where bicycle counts should be sampled to efficiently collect representative data.

#### **ACKNOWLEDGMENTS**

The authors would like to acknowledge Strava and the city of Ottawa for providing the data. This work was supported by a grant from the Public Health Agency of Canada to [BikeMaps.org](https://www.bikemaps.org).



**Figure 2:** Bike Ridership Temporal Profiles for Street Segments within Each Bike Ridership Class



**Figure 3:** a) Mean Temporal Profiles for the Classes of Bike Ridership; b) Map of Bike Ridership Classes for Study Area in Ottawa



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

## REFERENCES

- Ahmed, K.I. 2012. "Acoustic Data Optimization for Seabed Mapping with Visual and Computational Data Mining." (Unpublished master's thesis). Maynooth, Kildare, Ireland: National University of Ireland.
- 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>.
- Calinski, T., and J. Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics - Theory and Methods* 3 (1): 1–27. <https://doi.org/10.1080/03610927408827101>.
- Griffin, Greg, Krista Nordback, Thomas Götschi, Elizabeth Stolz, and Sirisha Kothuri. 2014. *Monitoring Bicyclist and Pedestrian Travel and Behavior*. Transportation Research Board. <https://doi.org/10.17226/22420>.
- Jestico, Ben, Trisalyn Nelson, and Meghan Winters. 2016. "Mapping Ridership Using Crowdsourced Cycling Data." *Journal of Transport Geography* 52 (April): 90–97. <https://doi.org/10.1016/j.jtrangeo.2016.03.006>.
- Miranda-Moreno, Luis F., Thomas Nosal, Robert J. Schneider, and Frank Proulx. 2013. "Classification of Bicycle Traffic Patterns in Five North American Cities." *Transportation Research Record: Journal of the Transportation Research Board* 2339 (1): 68–79. <https://doi.org/10.3141/2339-08>.
- Murtagh, Fionn, and Pierre Legendre. 2014. "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" *Journal of Classification* 31 (3): 274–95. <https://doi.org/10.1007/s00357-014-9161-z>.
- Nordback, Krista, Sirisha Kothuri, Dylan Johnstone, Greg Lindsey, Sherry Ryan, and Jeremy Raw. 2019. "Minimizing Annual Average Daily Nonmotorized Traffic Estimation Errors: How Many Counters Are Needed per Factor Group?" *Transportation Research Record: Journal of the Transportation Research Board*, May, 036119811984869. <https://doi.org/10.1177/0361198119848699>.
- Nordback, Krista, Wesley E. Marshall, Bruce N. Janson, and Elizabeth Stolz. 2013. "Estimating Annual Average Daily Bicyclists." *Transportation Research Record: Journal of the Transportation Research Board* 2339 (1): 90–97. <https://doi.org/10.3141/2339-10>.
- Roy, A. et al. Forthcoming. "Understanding Bicycle Ridership Patterns across North-American Cities Using Bias-Corrected Crowdsourced Data." *Journal of Transport Geography*.
- Sakoe, H., and S. Chiba. 1978. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26 (1): 43–49. <https://doi.org/10.1109/tassp.1978.1163055>.
- Zhang, Zheng, Romain Tavenard, Adeline Bailly, Xiaotong Tang, Ping Tang, and Thomas Corpetti. 2017. "Dynamic Time Warping under Limited Warping Path Length." *Information Sciences* 393 (July): 91–107. <https://doi.org/10.1016/j.ins.2017.02.018>.

## FIGURES, TABLES, AND SUPPLEMENTARY MATERIALS

### **r\_code.pdf**

Download: <http://app.scholasticahq.com/api/v1/attachments/26584/download>

---