Journal of Park and Recreation Administration
https://doi.org/10.18666/JPRA-2024-12457
Volume 42, Issue 4, Winter 2024, pp. 86-96

Research Note

Chasing Informed Decisions: A Research Note on the Potential for Strava to Support Rural Recreation Planning

Emily Reinhardt,[a] Kimberly J. Coleman,[b] Mark Baran,[a] Elizabeth E. Perry,[c] and Mary Alldred[a]

[a] Center for Earth and Environmental Science, SUNY Plattsburgh, NY
[b] Rubenstein School of Environmental and Natural Resources, The University of Vermont, Burlington, VT
[c] Department of Community Sustainability, Michigan State University, East Lansing, MI

Please send correspondence to Mary Alldred, malld001@plattsburgh.edu

Abstract

With user-generated data from social media platforms becoming more widely available to researchers, these datasets may offer a valuable opportunity for recreation and tourism managers to enhance decision-making processes in their communities. The activity-tracking app Strava is one potentially underutilized data source in rural recreation planning. Strava data offer a useful supplementary tool that can be used in combination with traditional methods of data collection. We tested the usability of data provided by Strava Metro (a non-profit data service) to assess its capabilities with a case study of Kingdom Trails. We used trail counter data to assess the validity of Strava's bike ride-counts on five trails. We also integrated traffic count data to determine peak times for both bike riders and vehicles where safety concerns may arise. The results of our assessment indicate that Strava data may provide additional insights into trail usage patterns, intensity variations, and visitor-use hotspots for a more holistic understanding of recreation at different spatial and temporal scales.

Keywords

Bicycle infrastructure, cycling volumes, mountain biking, user-generated data, tourism management

Introduction

Exploration of user-generated data from social media sources such as Flickr, Twitter (now X), Instagram, and Facebook, can assist recreation and tourism managers in making informed management decisions (Pickering et al., 2023). For example, a large dataset of geotagged tweets from Louisville, Kentucky, allowed planners to explore problems of socio-spatial inequality (Shelton et al., 2015); scholars have analyzed geotagged data from Flickr to understand the public values and behaviors associated with a given landscape (e.g., Dunkel, 2015; Pickering et al., 2020); and Johnson et al. (2019)

used Twitter to map cultural ecosystem services in urban parks. Because people worldwide use various social media platforms to comment on everyday life, share photos, and document activities (Cope, 2015), user-generated data provide a way to explore larger volumes of visitor information than traditional means, including on-site visitor surveys, gate monitoring, entrance fees, trail counters, and focus groups (e.g., Leggett et al., 2017; Pickering et al., 2023).

Although these novel sources have great potential to increase the scope of our data collection (Cope, 2015), each platform comes with its own set of challenges, including bias, quality, and usability of the raw data (Heikinheimo et al., 2020). Scholars have also warned that user-generated data could be misapplied, particularly if researchers do not assess whether social media users are representative of the intended study population (Wilkins et al., 2021). Thus, individual platforms should be vetted for the appropriateness of their use for decision-making generally but also to answer specific questions. The goal of this research note is to provide an introductory look at user-generated data from the platform Strava and to test the suitability of potential applications for this and similar data sources. We provide an overview of Strava, how it works, and a brief review of its use as a data source for scholarship and decision-making. We include a test case of how Strava data can be used to inform outdoor recreation management decisions by comparing it to traditional data collection methods and provide recommendations for future work.

Background

Strava is a social media and fitness tracking app that is growing in popularity amongst fitness enthusiasts, recreationists, and athletes of all levels. Through Strava, users record and upload the GPS-track of the route they took, which can be self-reported as commuting or a leisure activity. In addition to recording GPS tracks, known as "segments," Strava also allows users to collect performance data, such as personal best times, heart rate, speed, and power output, if supported by the athlete's equipment (e.g., a cyclist must have a power meter to accurately record power). The social media element of Strava encourages users to interact online, by giving each other "kudos" (Strava's version of "likes"), by creating "challenges" in which groups of Strava users all work towards a shared goal (e.g., completing the "Festive 500" by cycling 500 km between Christmas and New Years), and by awarding "King of the Mountain" (KOM) or "Queen of the Mountain" (QOM) to the user with the fastest recorded time on a given segment. Additionally, Strava allows users to create routes through its mapping features, and provide suggested routes for users. These opportunities for engagement have made Strava increasingly popular, with over 100 million users in 195 countries, including athletes, recreationists, commuters, and utility users (Strava Business, 2024).

Given Strava's growing popularity, it may prove a useful source of data that can help researchers map where recreationists go in space and time (Wilkins et al., 2021). Strava provides user-generated data through Strava Metro, a nonprofit data service that partners with organizations that plan, own, or maintain active transportation infrastructure or seek to positively influence planning processes (Strava Metro, 2024). Strava Metro provides free data access to "public agencies" including departments of transportation, planning organizations, trail organizations, and cities, to support projects that seek to improve infrastructure for bicyclists and pedestrians (Lee & Sener, 2021).

Strava's applicability to the field of rural recreation planning has not been fully explored. Indeed, a recent literature review by Wilkins et al. (2021) on using social media to inform management in parks and protected areas revealed that only two of the papers in their corpus utilized Strava data. Strava data have been predominantly used in the context of urban planning. Researchers have explored topics including mapping volume and patterns of bike usage in metropolitan areas (Lin & Fan, 2020; Roy et al., 2019), tracking mobility trends following catastrophic events like COVID-19 and natural disasters (Fischer et al., 2022; Liu et al., 2022), exploring incidence of bike-car collisions (Ferster et al., 2021; Garber et al., 2023; Lee & Sener, 2021), and assessing exposure to air pollution for commuters (Sun et al., 2017).

The goal of this research was to explore the applicability of Strava data in a rural outdoor recreation planning context. The need to ground truth these data is imperative for accurate insights into local usage patterns. Due to the competitive nature of the Strava platform, patterns of usage on specific segments may not accurately represent recreation behavior in the typical sense because of recreationists' commitment to obtaining a KOM or QOM designation for example. Specifically, we sought to test the validity of Strava data with historical trail counter data, identify potential challenges to analyzing Strava data, and explore the feasibility for rural recreation managers to utilize Strava data to answer questions to inform better decision-making.

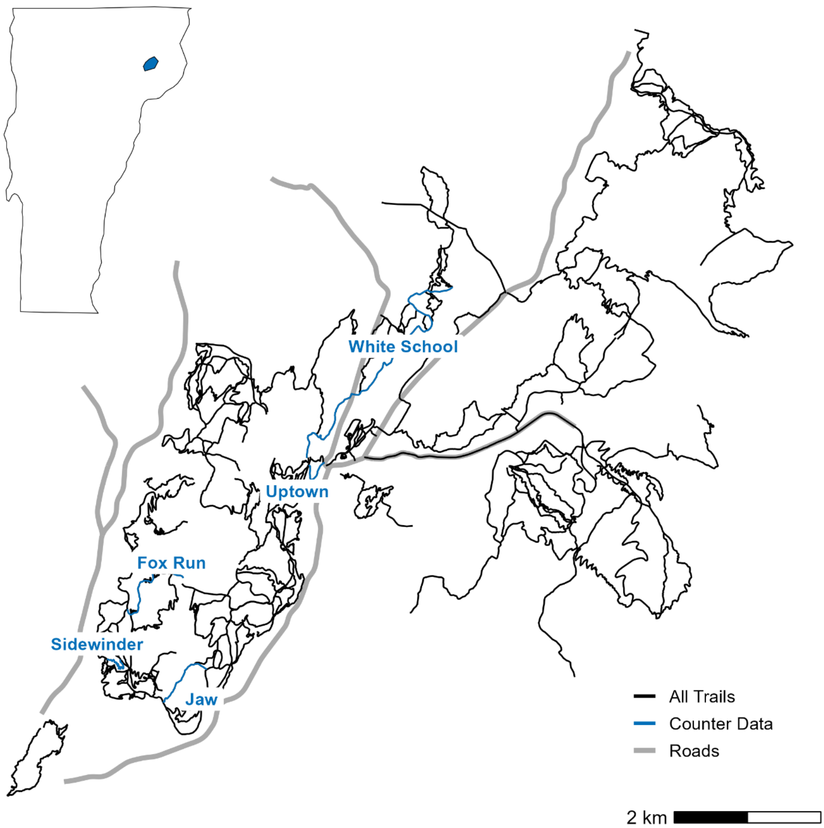
Methods

To explore the usability of Strava Metro data in rural recreation planning, we partnered with Kingdom Trails, a popular non-profit trail network. Kingdom Trails is located in Burke, VT, a rural community in the northeastern corner of the state, colloquially referred to as "The Northeast Kingdom" (Figure 1). The rural, mountainous environment of Kingdom Trails, as well as a previous network capacity study on the site that generated trail-counter data, provided an ideal test case. Our project partners at Kingdom Trails identified several central concerns and questions about their trail network. These included understanding the volume of mountain bikers using the trail network, spatiotemporal patterns in that volume, and areas where high volumes of mountain bikers might conflict with traffic on busy roads that bisect the trail network.

Kingdom Trails and the Burke County Chamber of Commerce hired the independent consultant SE Group in 2020 to deploy trail counters to explore questions of network capacity (SE Group, 2020). Counters were placed near primary access points on trails and calibrated using best practices outlined in Pettebone et al. (2010). While several trail counters were repositioned during the trail capacity study, all counters remained at a fixed position near a primary access point for the duration of July 2020. Final calibrated daily counts for July 2020 were shared via our project partners at Kingdoms Trails and provided an opportunity to validate Strava data using a traditional sampling method (SE Group, 2020).

Data analysis and visualization was conducted using R version 4.2.2 and is documented in the supplementary material. We requested data for Caledonia County, Vermont, from the Strava Metro portal and were able to access daily user-reported bike ride counts for segments from 2019 to 2022. Strava ride counts are aggregated into bins of five to protect individual user privacy; therefore, a segment with between one and five bike rides per day would appear as a count of five in the shared dataset. Importantly, activity types (pedestrian or bike) are user-reported and not auto-generated by the app. We used trail counter data collected by SE Group for Kingdom Trails' Network

Figure 1
Map of trails within the Kingdom Trail network (black), roads within the network boundary (gray), and trails where trail counters were deployed in 2020 (blue). Map inset indicates the location of the Kingdom Trails network in the state of Vermont.

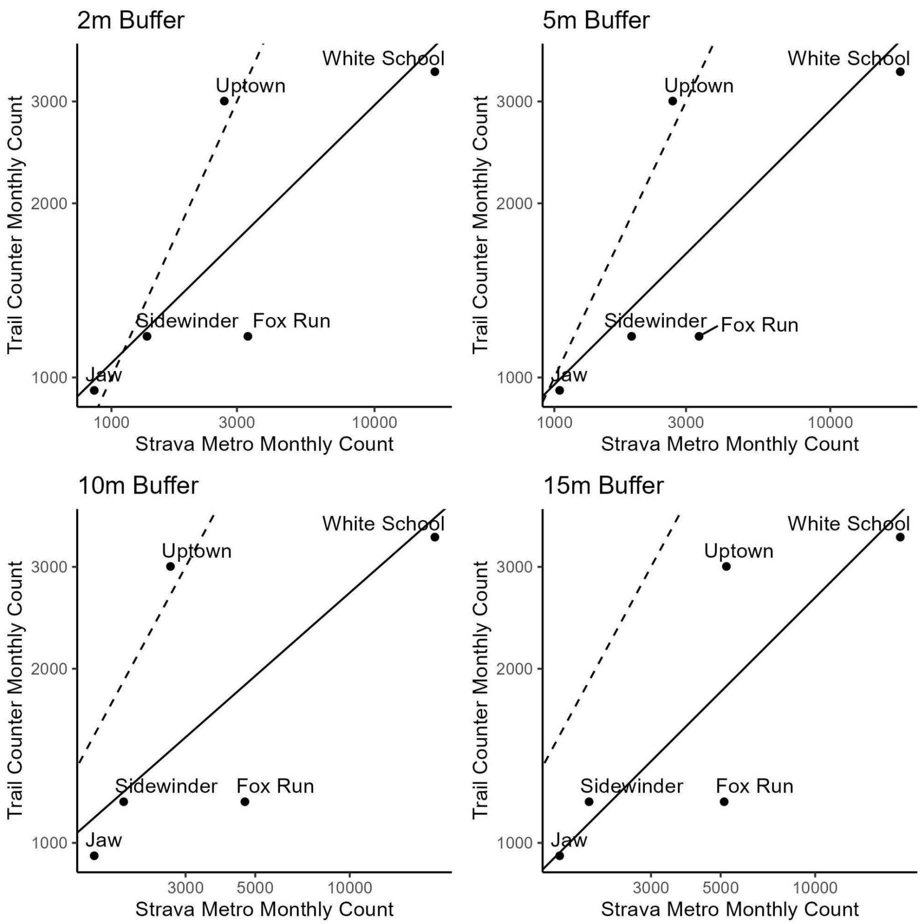


Capacity Study from July 2020, for five trails within the trail network as the basis to test the validity of ride-count data from Strava Metro, and annual average daily traffic (AADT) data collected by the Vermont Agency of Transportation to examine road traffic volume. We created buffer areas of 2m, 5m, 10m, and 15m around the five trails with associated trail counter data and extracted all Strava counts from within those buffer areas. We used model II linear regression to test for an association between trail-counter and Strava count. To demonstrate how Strava Metro data can be summarized and visualized for trail planning, we created maps showing hotspots in monthly bike ride counts, as well as spatial overlap between average daily bike traffic and average daily road traffic during 2020.

Results

Our results showed that ride-count estimates on trails containing only one or two Strava segments (Jaw, Sidewinder, Uptown) aligned very closely to the trail counter data when using the smallest, 2m, buffer area (Figure 2). Trails containing many Strava segments (Fox Run [5], White School [19]) tended to overestimate ride counts relative

Figure 2
Comparison of bike-ride data from deployed trail counters, and Strava ride counts extracted from within a 2m, 5m, 10m, or 15m buffer area around the trail. Solid line shows model II (also "major-axis") regression and dashed line shows 1:1 relationship.



to trail-counter estimates. The relationship between trail counter and Strava data was statistically significant, and had the greatest explanatory power, when using smallest, 2m ($m = 0.45$, $p = 0.03$, $R^2 = 0.61$), and largest, 15m ($m = 0.54$, $p = 0.03$, $R^2 = 0.69$) buffer areas around the trail; intermediate-size buffers of 5m ($m = 0.47$, $p = 0.09$, $R^2 = 0.57$) and 10m ($m = 0.48$, $p = 0.20$, $R^2 = 0.46$) performed less well (Figure 2). As we increased the size of the buffer area around the mapped trail, we found that Strava ride counts were more likely to provide an overestimate of trail use.

Strava data revealed landscape level patterns in the volume of bike traffic, both temporally and spatially, with peak use occurring in the southwest portion of the trail network during the summer months (Figure 3). We also observed considerable spatial variation in the parts of the trail network that were accessed over the course of the year (Figure 3). Looking at the network as a whole (i.e., at the landscape level) did allow for integration with traffic data, and the identification of potential risky areas where high volumes of cyclists and motorists overlap (Figure 4).

Figure 3
Heat map of monthly bike ride counts at Kingdom Trails based on daily Strava Metro data

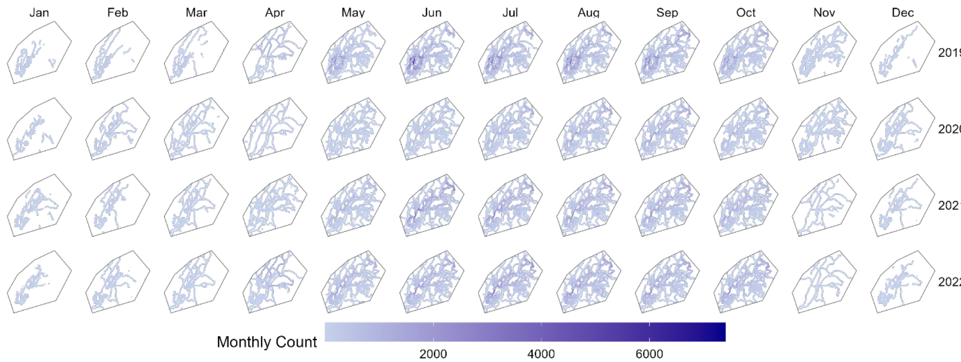
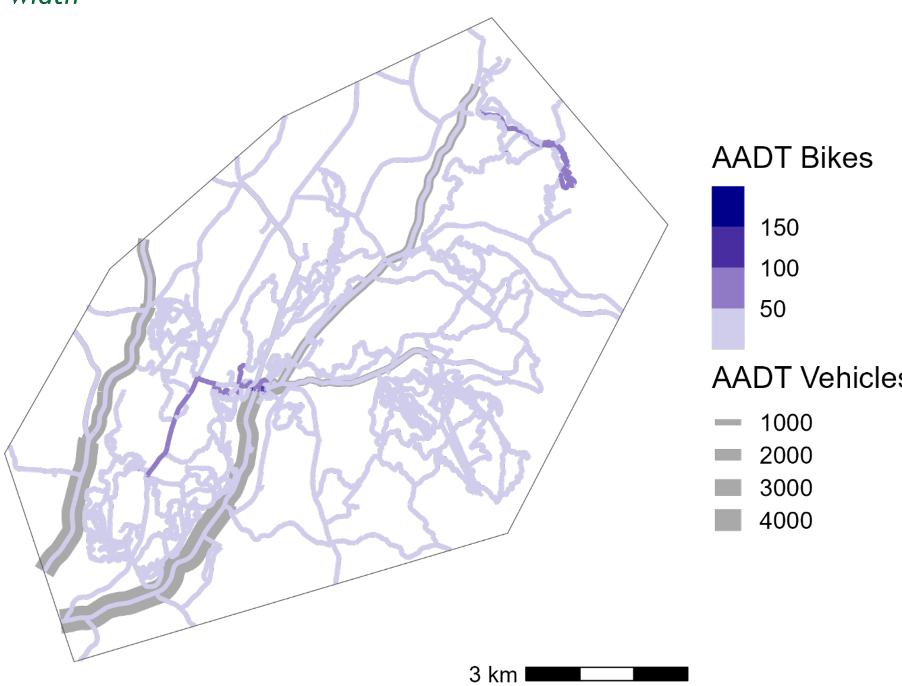


Figure 4
Kingdom Trails network with annual average daily traffic (AADT) for bikes indicated in color ramp and AADT for vehicles indicated with relative line width



Discussion

Our analysis suggests that Strava data can provide precise estimates of rural trail use in cases where Strava segments align well with mapped trails (i.e., Jaw, Sidewinder, and Uptown). However, for two mapped trails, Fox Run and White School, Strava data overestimated actual use relative to trail-counter estimates. Notably, both of these trails contained multiple Strava segments (5 and 19, respectively) and were in close geographic proximity to other mapped trails in the network where trail counters were not deployed. We suspect that in many cases, weak satellite signals or cell coverage common to rural, especially mountainous, areas may result in a mismatch between reported Strava segments and actual trails. In such cases, a planner may not be able to use Strava to quantify ride volumes for a specific trail, but could determine usage for that region of the trail network. We expect that these mismatches are more likely to occur in areas like Kingdom Trails, which consist of an intricate network of interconnected trails in close geographic proximity, where some trails can only be accessed by first traversing a segment (or multiple segments) of another. These sources of error are less likely to become an issue for trail networks with designated trails spaced further apart with separate access points. If trails are located near roads, counts from road biking may also be misassigned to trails.

Other potential sources of error commonly identified in Strava use cases were not apparent in our analysis. Strava aggregates ride counts into five-count bins to protect individual user privacy prior to sharing data (Lee & Sener, 2021). Users must also have consented to sharing their information with others and with the Strava Metro server for their data to be included in the five-count aggregates. In urban settings, the worry is that these privacy precautions could lead to underreporting of total ride counts; whereas, on sparsely used rural trails, segments with only one or two bikers per day would be reported as a five-ride count, potentially leading to overestimates of use. We found little evidence for this source of error in our data, as Strava ride counts from the most sparsely used trails, Jaw and Sidewinder, correlated well with trail-counter data. It is also possible that Strava could capture more users than traditional methods. Trail counters are deployed in specific locations along trails and may not capture all users particularly for trails with multiple entry points. In some cases bikers may also avoid or "hop" trail counters. However, these sources of error are unlikely to result in significant counting errors, and we did not observe any evidence of systematic overcounting or undercounting across the trail network. Indeed, Strava counts within 2m for three out of five of the mapped trails corresponded nearly 1:1 with trail-counter estimates.

As the number of social media users increases globally, the amount of available data grows. Researchers are continuously trying to harness the capabilities of the diverse user-generated data available to help answer pressing questions resource managers and decision-makers face relating to outdoor recreation and planning (Pickering et al., 2023). Being able to quantify the ways in which users are interacting with trails allows for informed on-the-ground management decisions. The type and magnitude of trail use can influence demand for infrastructure, timing and types of resourcing, minimizing environmental impacts, enhancing visitor experiences, supporting conservation of natural areas and strategic destination marketing (Kajala et al., 2007). Analyzing user-generated data from Strava Metro, with an appropriate understanding of their limitations, is one way in which recreation and tourism managers can broaden insights into recreation participation and spatial patterns on the landscape. Data from social media and activity-tracking apps like Strava may be well suited for rural recreation planning objectives as budget constraints often hinder more in-depth studies and implementation of technology (e.g., trail counters) to assist in extrapolating dispersed recreational trail use patterns is often expensive (Lawson, 2021).

The need to respond efficiently and effectively often leads managers of parks and protected areas, such as Kingdom Trails, to use data that represent a small snapshot of overall use patterns. Our results revealed considerable spatial-temporal variation in trail use, even within a single season (e.g., summer or winter months). Compared to trail-counter data collected at a limited number of points on the landscape, Strava data are collected with a much higher spatial-temporal resolution, and thus expand managers' ability to consider variation in usage, as well as intersections with other datasets such as traffic counts. Though these datasets should be considered carefully, and validated using traditional methods whenever possible, they do provide an ability to be more responsive and nimble in management planning than with traditional visitor use pattern count methods alone.

The management challenges faced by Kingdoms Trails, including the need to track visitor use and ensure visitor safety around rural roads with higher speed limits and areas of poor visibility, are common among managers of rural parks and protected areas (Wilkins et al., 2021). As particular recreation activities such as mountain biking continue to gain in popularity, and necessitate discussions about safety between riders and vehicles, user-generated data may provide valuable spatial information that would be logistically difficult or prohibitively expensive to obtain using traditional methods. This information is critical in safety considerations and can assist managers in retroactive inquiries that promote more proactive planning.

Limitations

Inferences based on Strava Metro data are not without limitations, many of which are related to the unique experience of the Strava app and its online community. Because Strava relies on cell phone GPS to assign user tracks to segments, spatial inaccuracies may arise in areas with poor cell and satellite reception. Rural and mountainous areas, such as northeastern Vermont, are typically less favorable for accurate GPS locations, so we would expect more segment mismatches in such environments. Another area of concern is whether Strava data are truly representative of the full user base of a specific trail. With its focus on competitiveness, Strava users may skew towards those seeking more difficult or technical rides, leading to undercounts of users on more casual trail segments. Similar limitations were found in a study where Strava commuter data and the U.S. Census Bureau's American Community Survey data were compared to determine whether Strava could provide a representative sample of commuters in four cities. This study reported inaccuracies in GPS-tracked data along with acknowledgment that Strava users represent a subset of people who are particularly active as compared to the general population (Whitfield et al., 2016). Both of these limitations were present in our analysis as well and should remain a consideration in the usage of user-generated data overall. Nonetheless, an online dataset like Strava can be a useful tool for rural communities with recreation and tourism planning and management questions, particularly when used together with other measures for ground truthing trail usage.

Conclusion

Our analysis suggests that Strava Metro can provide valuable data for describing and visualizing patterns of trail usage in rural recreation planning. Though commonly used in urban planning, Strava Metro remains underutilized as a data source in rural recreation planning. Strava can reveal landscape-level patterns in use, which are likely useful for planners and managers and would be difficult to obtain using traditional sampling methods. Although this paper reports on data from a small, private trail network, the methods discussed here are applicable to public land management. Ultimately, making planning and management decisions based on information from larger datasets may prove to be more equitable than traditional means of collecting public comment (Cope, 2015), and Strava Metro may provide such an opportunity for public land managers. In the context of rural recreation planning, Strava can provide insight into trail use patterns and intensity over time and space. These patterns may be important for informing decisions on where resources are best suited. Land managers can supplement traditional means of data collection with Strava's user-generated data as a way to answer questions of usage patterns across an entire trail network (e.g., Where are usage hotspots? What time of year is the busiest? Are some areas more popular at certain times of the year vs others?). Future work should aim to incorporate user-generated data in combination with other datasets like traffic camera counts, trail counters, visitor logs or other mobility analytics to holistically understand recreation at various spatial scales in many different contexts.

Acknowledgments: We would like to thank Taygin Jump and Meghan Bargabos for contributions to the literature review, and Jessica Leahy and three anonymous reviewers for helpful comments on an earlier draft of this manuscript. We would also like to thank Strava Metro for partnering in this analysis and Abby Long, Executive Director of Kingdom Trails for access to trail-counter and geospatial data and many helpful conversations regarding ongoing planning challenges in rural areas.

Disclosure Statement: The authors have no conflicts of interest to declare.

Funding: This project was supported by the Northeastern States Research Cooperative through funding made available by the USDA Forest Service. The conclusions and opinions in this paper are those of the authors and not of the NSRC, the Forest Service, or the USDA.

References

Cope, M. (2015). Commentary: Geographies of digital lives: Trajectories in the production of knowledge with user-generated content. *Special Issue: Critical Approaches to Landscape Visualization, 142*, 212–214. https://doi.org/10.1016/j.landurbplan.2015.08.009

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Special Issue: Critical Approaches to Landscape Visualization, 142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

Ferster, C., Nelson, T., Laberee, K., & Winters, M. (2021). Mapping bicycling exposure and safety risk using Strava Metro. *Applied Geography, 127*, 102388. https://doi.org/10.1016/j.apgeog.2021.102388

Fischer, J., Nelson, T., & Winters, M. (2022). Riding through the pandemic: Using Strava data to monitor the impacts of COVID-19 on spatial patterns of bicycling. *Transportation Research Interdisciplinary Perspectives, 15*, 100667. https://doi.org/10.1016/j.trip.2022.100667

Garber, M. D., Watkins, K. E., Flanders, W. D., Kramer, M. R., Lobelo, R. L. F., Mooney, S. J., Ederer, D. J., & McCullough, L. E. (2023). Bicycle infrastructure and the incidence rate of crashes with cars: A case-control study with Strava data in Atlanta. *Journal of Transport & Health, 32*, 101669. https://doi.org/10.1016/j.jth.2023.101669

Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T., & Toivonen, T. (2020). Understanding the use of urban green spaces from user-generated geographic information. *Landscape and Urban Planning, 201*, 103845. https://doi.org/10.1016/j.landurbplan.2020.103845

Johnson, M. L., Campbell, L. K., Svendsen, E. S., & McMillen, H. L. (2019). Mapping urban park cultural ecosystem services: A comparison of Twitter and semi-structured interview methods. *Sustainability, 11*(21). https://doi.org/10.3390/su11216137

Kajala, L., Almik, A., Dahl, R., Dikšaitė, L., Erkkonen, J., Fredman, P., Jensen, F. S., Karoles, K., Sievänen, T., Skov-Petersen, H., Vistad, O. I., & Wallsten, P. (2007). *Visitor monitoring in nature areas: A manual based on experiences from the Nordic and Baltic countries.* Nordic Council of Ministers. http://www.naturvardsverket.se/Documents/bokhandeln/bokhandeln.htm

Lawson, M. (2021). *Innovative new ways to count outdoor recreation: Using data from cell phones, fitness trackers, social media, and other novel data sources* (p. 38). Headwaters Economics. https://headwaterseconomics.org/outdoor-recreation/counting-outdoor-recreation/

Lee, K., & Sener, I. N. (2021). Strava Metro data for bicycle monitoring: A literature review. *Transport Reviews, 41*(1), 27–47. https://doi.org/10.1080/01441647.2020.1798558

Leggett, C., Horsch, E., Smith, C., & Unsworth, R. (2017). *Estimating recreational visitation to federally managed lands.* Report prepared for National Park Service, Environmental Quality Division. https://rsginc.com/wp-content/uploads/2018/05/Leggett-2015-NPS-NRDA-Report.pdf

Lin, Z., & Fan, W. (David). (2020). Modeling bicycle volume using crowdsourced data from Strava smartphone application. *International Journal of Transportation Science and Technology, 9*(4), 334–343. https://doi.org/10.1016/j.ijtst.2020.03.003

Liu, Z., Maneekul, P., Pendergrast, C., Doubleday, A., Miles, S. B., Errett, N. A., & Choe, Y. (2022). Physical activity monitoring data following disasters. *Sustainable Cities and Society, 81*, 103814. https://doi.org/10.1016/j.scs.2022.103814

Pettebone, D., Newman, P., & Lawson, S. R. (2010). Estimating visitor use at attraction sites and trailheads in Yosemite National Park using automated visitor counters. *Landscape and Urban Planning, 97*(4), 229–238. https://doi.org/10.1016/j.landurbplan.2010.06.006

Pickering, C., Olafsson, A. S., & Hansen, A. S. (2023). Editorial for Special Issue of the Journal of Outdoor Tourism and Recreation on social media and other user created content for outdoor recreation and nature-based tourism research. *Social Media and Other User Created Content for Outdoor Recreation and Nature-Based Tourism Research, 44,* 100727. https://doi.org/10.1016/j.jort.2023.100727

Pickering, C., Walden-Schreiner, C., Barros, A., & Rossi, S. D. (2020). Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism, 29*, 100252. https://doi.org/10.1016/j.jort.2019.100252

Roy, A., Nelson, T. A., Fotheringham, A. S., & Winters, M. (2019). Correcting bias in crowdsourced data to map bicycle ridership of all bicyclists. *Urban Science, 3*(2). https://doi.org/10.3390/urbansci3020062

SE Group. (2020). *Kingdom Trails Network Capacity Study.* ArcGIS StoryMaps. https://arcg.is/9mTiD0

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Special Issue: Critical Approaches to Landscape Visualization, 142*, 198–211. https://doi.org/10.1016/j.landurbplan.2015.02.020

Strava Business. (2024). *What makes a typical Strava user?* https://business.strava.com/blog/what-makes-typical-strava-user

Strava Metro. (2024). *Strava Metro FAQ*. https://metro.strava.com/faq

Sun, Y., Moshfeghi, Y., & Liu, Z. (2017). Exploiting crowdsourced geographic information and GIS for assessment of air pollution exposure during active travel. *Journal of Transport & Health, 6*, 93–104. https://doi.org/10.1016/j.jth.2017.06.004

Whitfield, G. P., Ussery, E. N., Riordan, B., & Wendel, A. M. (2016). Association between user-generated commuting data and population-representative active commuting surveillance data—Four cities, 2014–2015. *Morbidity and Mortality Weekly Report, 65*(36), 959–962.

Wilkins, E. J., Wood, S. A., & Smith, J. W. (2021). Uses and limitations of social media to inform visitor use management in parks and protected areas: A systematic review. *Environmental Management, 67*(1), 120–132. https://doi.org/10.1007/s00267-020-01373-7