Methods for County-Wide Trail & Pathway Activity Estimates Using Strava Data

April 2018

HEADWATERS ECONOMICS

Overview

Headwaters Economics methods for estimating bike and pedestrian activities within Flathead County, Montana resulted from our exploration of "crowdsourced" data to help communities create economic opportunity and enhance quality of life for residents. Our goal was to evaluate how crowdsourced data could be used to engage community residents in projects that lead to better economic opportunity and health outcomes—for example, by encouraging participation in economic development projects, engaging residents in healthy outdoor recreation activities, and monitoring and demonstrating participation in ParksRx programs.

We investigated several platforms including Twitter, Flickr, Instagram, and Snapchat, and opted to develop this visualization around a popular activity and fitness tracking app: Strava. Strava is a website and mobile app used to track athletic activity via satellite navigation. Strava data has been used in urban areas to help city planners improve bike routes and understand commuting patterns. We believe this application of Strava for the community of Whitefish, Montana is the first time, to our knowledge, Strava has been used in a more rural setting to estimate recreational trail use and related local economic benefits.

We partnered with the Whitefish Legacy Partners in Whitefish, Montana to collect data using in-person surveys and infrared (IR) trail counters at key trailheads in the developing system of trails known as The Whitefish Trail. Additionally, we received support from the LOR Foundation and a grant from Strava Metro for Strava user data for the trail system for a concurrent time period and used it to compare to our IR counter and in-person survey data, as well as to extrapolate usage to parts of the trail system that were not surveyed directly.

Using Strava data calibrated to infrared counters, weestimated trail and pathway use in Flathead County by user type (local or tourist) and activity type (pedestrian vs. bike). These estimates are part of a larger project with local partners to understand the benefits of trail use on local businesses and residents' quality of life.

The Strava data enabled us to generate comprehensive estimates of trail use for the entire county (rather than only certain locations where we deployed infrared trail counters). The Strava-based data and visualization that we developed will be used by our local partners to support their trail network, which we demonstrated to be an economic asset to the community. Our approach to using Strava for estimating trail use is replicable in other communities and is substantially less expensive than estimating trail use and benefits using traditional analyses based entirely on local survey data.

Background

Crowdsourcing is an emerging concept. Jeff Howe first coined the term "crowdsourcing" in a 2006 Wired article,[1] but the concept of using the public to tackle a complicated problem or participate in a massive data-gathering effort (often voluntarily) dates to the early 18th century. In more recent times, crowdsourcing has been applied to such tasks as building photo collections (ShutterStock), searching interstellar noise for signs of extraterrestrial intelligence (SETI Project), and mapping the human response to earthquakes (USGS' "Did You Feel It?" Project).

Closely related to crowdsourcing, the term "social media" is applied to technologies that allow for the creation and sharing of user-generated content such as text, pictures, videos, etc., which can be distributed to other users via Internet-based applications. The reach of social media is increasing quickly. Facebook, by far the leader in the field, boasts that it has more than 1.94 billion monthly active users (MAU) as of March 2017. That usage represents a 17 percent increase from March 2016.[2]

Social media can be thought of as an emerging source of crowdsourced data as it capitalizes on the social graph, the interconnections between people, to propagate messages and feedback along the spines of the social graph.

Headwaters Economics explored the utility of many social media platforms included Twitter, Facebook, Flickr, Instagram, Snapchat and Strava. For each platform, we evaluated the type of data available, how that data could be gleaned from the application, what limitations there were on publicly available data, what technological know-how was required to get at the data, how information could be solicited from users, and, most importantly, how the gathered data might be used to address community development issues important to our local partners.

We eliminated several platforms due to either small or declining use or challenges accessing and customizing the application data. For example, we quickly eliminated Snapchat because it has a small market penetration and a limited number of users older than 30. And, we learned that while the number of Flickr photos taken of an area has been used as a proxy for visitation, Flickr use is declining overall.

We chose to pursue using Strava for measuring the use and economic impacts of recreational trails and pathways. We considered leveraging the geolocation capabilities built into many existing applications, but eventually settled on using Strava data as the most efficient way to leverage socially generated data to estimate trail and pathway use across a large geographic area.

[1] Howe, Jeff. "The Rise of Crowdsourcing." *Wired.* Condé Nast, 1 Jun. 2006. Web. 18 Jul. 2017.
[2] "Facebook Q1 2017 Earnings." *Investor Relations*. Facebook, Inc. 3 May. 2017. Web. 18 Jul. 2017.

Methods

Our objective was to see if a crowd-sourced dataset such as Strava data could be used to inform trail usage statistics given that it could be a much more cost-effective means of collecting such data "passively" over large geographic areas and timespans. We also wanted to evaluate the utility of Strava for estimating the economic impacts of trails. An important part of economic impact studies is quantifying the percentage of usage from local users vs. tourists. We wanted to know if crowdsourced data such as that generated by Strava users could serve as a reliable proxy for "origin of user."

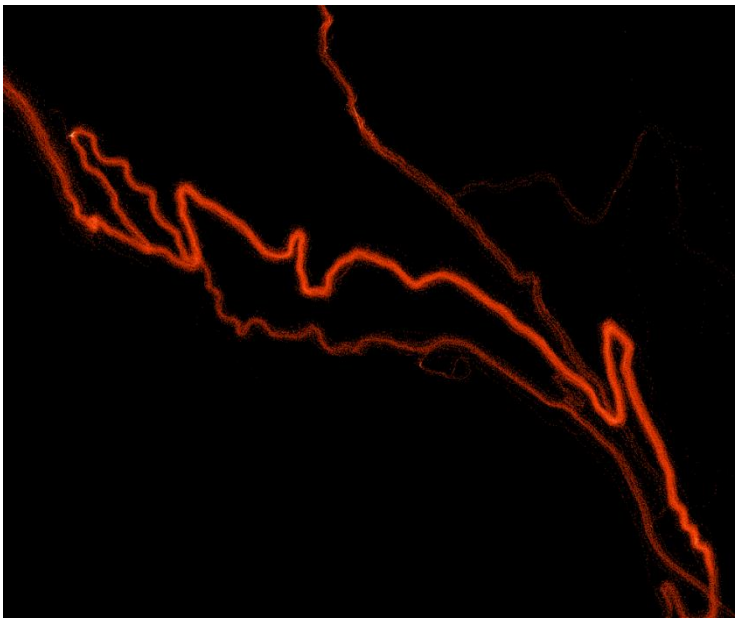
We partnered with the Whitefish Legacy Partners in Whitefish, Montana to collect data about trail users using both in-person surveys and infrared (IR) trail counters at key trailheads in the developing system of trails known as The Whitefish Trail[3]. Additionally, we acquired Strava data for the trail system for a concurrent time period and used it to compare to our IR counter and in-person survey data, as well as to extrapolate usage to parts of the trail system that were not surveyed directly.

We received a grant from StravaMetro for Strava user data for the period May 1, 2017 to September 26, 2017. The raw data collected by Strava are GPS point locations, a "heatmap," recorded by individual Strava users on mobile devices that are GPS-enabled (Figure 1). We used the heatmap data to decide what road/trail GIS network we would provide to StravaMetro so that they could transfer the point data to linear features (roads and trails) for us.

By default, StravaMetro will "snap" the raw data to the OpenStreetMap©[4] dataset. OpenStreetMap© is a free spatial dataset depicting roads, trails, railways, and many other features all over the world. Because OpenStreetMap© is an open-source product developed by voluntary contributors, it is not a comprehensive road/trail network. We found that some of the trails of most interest to ourselves and our partners, portions of The Whitefish Trail, were not represented in the OpenStreetMap© dataset.

[3] https://www.whitefishlegacy.org/projects/trail-construction/
[4] https://www.openstreetmap.org

Figure 1. Example raw trail usage data from StravaMetro.



We augmented the OpenStreetMap© dataset with additional trails that were needed for this project. We provided our augmented road/trail linear GIS dataset to StravaMetro and they returned to us four separate datasets: local pedestrian activity, tourist pedestrian activity, local bike (rider) activity, and tourist bike (rider) activity. For privacy reasons, the data provided by StravaMetro was anonymized, that is, information about the identity of individual users was stripped out of the data. Prior to anonymizing the data, however, StravaMetro grouped users into "locals" and "tourists." For the purposes of this study, "locals" are those that self-reported their home location to be in Flathead County, home to The Whitefish Trail. Those Strava users that self-identified their home location as outside of Flathead County were considered "tourists." Although, theoretically, each of the four datasets could have overlapping users (i.e., users are unique within but not across the four datasets), we felt it was safe to assume that a person was not walking and riding at the same time, nor could they be both a "local' and a "tourist." We limited our trail usage analysis to the number of activities rather than number of users, and that, in combination with our assumption about non-overlapping users, allowed us to add up total activities by trail segment.

Once the raw Strava data was imposed upon a linear GIS layer by StravaMetro, we had a GIS layer composed of many segments (portions of trails) from which we could glean the total number of activities at four levels (Table 1).

Table 1. Number of road/trail activities by category.

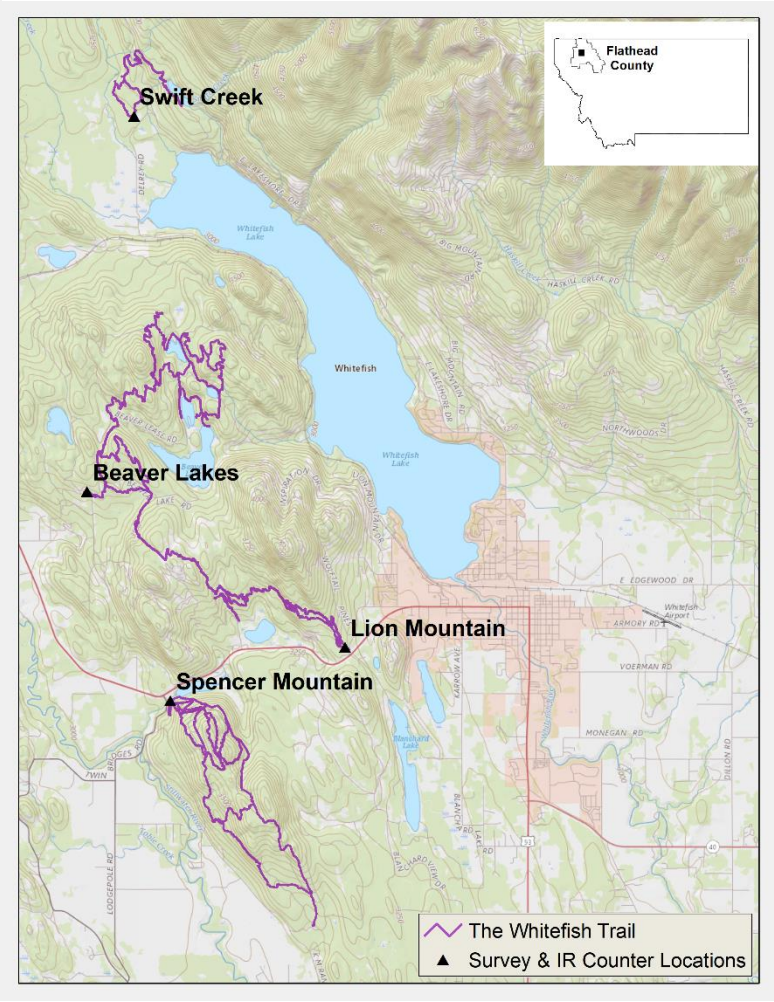
	Pedestrians	**Riders**
Locals	5,662	10,042
Tourists	7,862	6,236

The linear GIS layer to which the raw Strava data is snapped dictates the level of linear resolution at which the data can be summarized. For roads and trails that we adopted from the OpenSteetMaps© dataset, we were subject to the trail "segments" that have been defined by the contributors to that dataset. In many cases these segments were portions of trails that we were interested in. To compare Strava usage data to infrared counter data (calibration), we chose segments that were near the counters and that seemed to represent the four main sections of The Whitefish Trail system (Figure 2).

We treated the data collected by four infrared counters near the four trailheads as the actual count of trail users. Infrared counters may also count large animals such as domestic dogs; we made some corrections to the raw infrared counter data to account for dogs, based on a systematic comparison with

manual counts we conducted at the infrared counter sites. The infrared counters we deployed record trips in two directions, “in” and “out.” We used a single direction from the IR counters when attempting to calibrate the StravaMetro linear usage data. We also made sure to use the same “in” or “out” direction data in the StravaMetro dataset as those data also have a direction component.

Figure 2. Trails sections of The Whitefish Trail in Flathead County, MT.



Although the IR counters do not differentiate between walkers/hikers and bikers, we did collect information about what activity people were doing during the in-person surveys. Therefore, we had additional data, beyond Strava, by trailhead of the number of users that were walking/hiking vs. biking. We were able to confirm that the share of walkers/hikers and bikers in the Strava data were close to those observed in our in-person surveys. We applied the share of walkers/hikers and bikers in the Strava to inform our multipliers.

We assigned all trails to one of seven trail use areas, four of which corresponded to the separate sections of The Whitefish Trail, one of which corresponded to biking on Whitefish Mountain Resort and the adjacent Haskill Basin biking area, one of which corresponded to the popular Logan Pass paved ride in Glacier National Park, and one of which corresponded to all other roads and trails in Flathead County.

We chose representative trail segments from each of the four trails where we also conducted in-person surveys and recorded activity with the IR trail counters to do a comparison of the number of IR-counted users to Strava users. Comparing the number of users recorded by the IR counters to the number of activities recorded by Strava users gave us an idea of the percentage of total users that might be Strava users. The percentage of users that were Strava users varied from 1.5-6 percent (Table 2).

Table 2. Estimated percentage of total users that used Strava to record their activity by trail Whitefish Trail section.

Area	IR Counter Users[5]	Strava Trail Segment Users	Percent Using Strava
Lion Mountain	20681	451	2.2
Swift Creek	6257	92	1.5
Beaver Lakes	4531	114	2.5
Spencer Mountain	3183	191	6.0

We used the information about the percentage of trail users using Strava, information about the number of users that were participating in different activity types (walking hiking vs. biking), and expert opinion about geographic areas that likely would have similar usage patterns to come up with multipliers by activity type for each of the seven trail areas. We used those multipliers on the raw Strava usage data, by trail segment and trail area, to extrapolate usage for a broader geographic area—all of Flathead County.

Results

As we suspected, a small proportion of trail users utilized the Strava app to record their activities. We found this same result in the in-person interviews, although the reported percentages were higher than we found in the IR-counter/Strava data comparison. The total number of activities recorded by Strava users across the study was 29,802. Forty-seven percent of the total activities were conducted by "tourists" as opposed to 52 percent by "locals." Fifty-four percent of the Strava activities recorded were riding activities as opposed to 45 percent pedestrian activities. Local riders produced the largest share of the total activities (34%). Though the data was anonymized by StravaMetro prior to our receipt of it, we were provided with a demographic summary file of Strava users. In this study, the typical Strava user is a male between the ages of 35-45. Among females, that same age-class (35-45) as well as those in the next age-class down (25-34) were the most regular users of Strava.

We used the extrapolated usage to create a visualization of road and trail usage for our Whitefish partners. The data visualization allows users to zoom in on different parts of the Whitefish Trail (as well as other parts of Flathead County) and see the estimated usage in terms of different color-coded usage categories. Users can turn on and off different sections of trail so that their attention can be focused on one area alone or so that they can take in the big picture of usage across the entire county. Some basic charts accompany the usage map. One chart shows usage by day of the week, another shows usage by origin (local or tourist), and a third shows usage by activity type (pedestrian vs. bike).

Our partners in Whitefish intend to use the data we provided through the data visualization to inform a variety of decisions and to bolster support for their growing trail network. The information was presented briefly to the Whitefish Legacy Partners board in December 2017.

[5] After applying the "dog correction factor."

New sections of the Whitefish Trail are being developed currently, such as the Haskill Basin Area adjacent to Whitefish Mountain Resort. Our data visualization allows our Whitefish partners to ask such questions as:

- What routes are people using to get from Whitefish Mountain Resort to Haskill Basin?
- Are there any private property trespass issues that we need to be aware of (and that we can solve) based on the usage patterns we are seeing?
- Where are the bulk of tourist riders spending their time?
- What is the interaction between public trail usage and private entities such as the Whitefish Bike Retreat?

Conclusions

We know that Strava usage varies by gender, age, and even geographically within a study area. The percent of people that use Strava to record their outings is small (though likely increasing all the time).

Before one begins using the StravaMetro data, it is necessary to have a comprehensive road/trail dataset in hand. Using the OpenStreetMap© data, with its proliferation of small trail segments, was not ideal for this project, yet it was comprehensive and easy to come by. It is also important to consider though that any areas within the study area that do not have roads/trails on which to "snap" the raw data, will not show any usage. We probably should have used the OpenStreetMap© as a base but then replace all the Whitefish Trail portions with our own trail features. This approach would have resulted in linear usage patterns that made more sense to local partners.

In this study, we did not explicitly coordinate the location/timing of the IR-counters and the in-person interviews with the need to calibrate the Strava data. On the contrary, this experiment with Strava data was an add-on to a long-standing, existing project in Whitefish. With a better understanding of the Strava data, more thoughtful placement of IR-counters (or other counting devices), and more carefully targeted in-person survey questions (e.g., "Are you tracking today's workout with Strava?"), we think Strava data could be used to generate even more accurate area-wide usage estimates that would include information that could inform economic impact studies.

StravaMetro data has been used in urban areas to help city planners improve bike routes and understand commuting patterns. This project demonstrates the potential value of applying this data in a more rural setting to understand trail usage and try to relate it to local economic benefits.

Contact

Patty Gude, 406-599-7425, patty@headwaterseconomics.org

About Headwaters Economics

Headwaters Economics is an independent, nonprofit research group whose mission is to improve community development and land management decisions.
https://headwaterseconomics.org/