# CDOT COLORADO
Department of Transportation

# Strava Metro
# Data Analysis Summary

June 2018

**STRAVA | METRO**

# ACKNOWLEDGEMENTS

The Strava Metro Data Analysis Summary was prepared for The Colorado Department of Transportation by Atkins, a member of the SNC-Lavalin Group. The completion of this Summary would not have been possible without the support of the following CDOT and Atkins staff:

*Prepared for:*

*Prepared by:*

# TABLE OF CONTENTS

# APPENDICES

# FIGURES

# TABLES

# ACRONYMS

| | |
|---|---|
| **CDOT** | Colorado Department of Transportation |
| **C-470** | Colorado Highway 470 |
| **GB** | gigabyte |
| **GIS** | Geographic Information System |
| **GPS** | Global Positioning System |
| **I-##** | Interstate ## |
| **IQR** | Interquartile Range |
| **NMM** | Non-Motorized Monitoring |
| **OSM** | Open Street Map |
| **OTIS** | Online Transportation Information System |
| **$R^2$** | **coefficient of determination** |
| **SH ##** | State Highway ## |
| **US ##** | U.S. Highway ## |

# 1. INTRODUCTION

## Background

In 2012, the Colorado Department of Transportation (CDOT) adopted its first *Statewide Bicycle and Pedestrian Plan*, which laid out a policy foundation for state and local planning agencies to prioritize investments in bicycle and pedestrian infrastructure. The plan called for an expansion of CDOT's Non-Motorized Monitoring (NMM) Program to better understand the current level of bicycle and pedestrian activity around the state. In 2016, CDOT completed the NMM Strategic Plan. The goals and objectives established in the strategic plan help to ensure that data collected are managed in an efficient manner, one that meets the needs of state and local agencies as well as private and non-profit organizations. The NMM Strategic Plan recognized the limitations of CDOT's count program and recommended that CDOT consider emerging datasets, such as Strava and other examples of "big data," to develop a comprehensive understanding of bicycle travel in the state.

CDOT has partnered with Strava, a mobile fitness application that allows bicyclists and other recreational users to track their activities using Global Positioning System (GPS)-enabled mobile devices. Strava anonymously compiles the activity information and aggregates the data into a product called the "Strava Metro." CDOT purchased a 24-month Strava Metro dataset that consists of bicycle trips recorded by Strava users across the state of Colorado. CDOT is interested in using this new type of bicycle data to better understand non-motorized travel in Colorado.

The purpose of this study is to use Strava data to gain insights about bicyclist activity patterns in the state of Colorado. This study focuses on three primary objectives through the analysis of Strava data:

- Develop best practices for database management and quality control of crowd-sourced bicycling data as it relates specifically to big data and mobile application data sources.

- Correlate permanent continuous bicycle counter data with Strava data to develop parameters for extrapolating actual activity from Strava trip counts, and to estimate bicycle activity across Colorado.

- Identify and classify bicycle corridors within the CDOT-managed system into high-, medium-, and low-use categories based on Strava bicycle trip data.

## Relevant Studies

As an emerging source of bicycle data, Strava has been assessed to examine its level of representativeness of actual bicycle ridership in different studies all over the world. As the use of the application has become more prevalent, interest in the use of the data has also increased. Depending on the location and types of users, there is a wide range of results. Griffin and Jiao (2013) demonstrated that Strava counts represented 2.8 percent to 8.8 percent of the total number of bicyclists on trails in downtown Austin, Texas. In 2014, the Oregon Department of Transportation compared Strava counts with monthly counter data on the Hawthorne Bridge in Portland and found that Strava can represent 1.4 percent of the total bicyclists over a year at that location. Jestico et al. (2016) correlated Strava data with manual bicycling counts in Victoria, British Columbia by hourly, AM and PM peak, and peak period totals separated by season. The results indicated a linear relationship between the two types of data in which an increase of one Strava bicyclist would correspond to 51 more bicyclists. Boss et al. (2018) explored using spatial analysis methods to analyze crowd-sourced bicycling data to monitor changes in ridership patterns at a city level. The research compared Strava data with ground counts in Ottawa-Gatineau, Canada, and found that Strava samples of bicyclists were correlated with automated counts at 11 locations with the coefficient of determination ($R^2$) ranging from 0.76 to 0.96.

An increasing number of studies have used Strava data to investigate bicycling behavior for planning purposes. Moore et al. (2015) used Strava data along with other factors to model bicyclist choice of routes in Auburn, Alabama. The Vermont Agency of Transportation used Strava data as its key data input for statewide bicycle infrastructure planning (2016). Hochmair et al. (2017) used Strava data to model bicycling ridership at different temporal levels for Miami-Dade County, Florida. The results indicated Strava data are useful to estimate bicycling volumes for large areas. Proulx et al. (2017) studied estimating bicycle ridership by fusing various demand datasets. The findings illustrated that Strava is the most predictive dataset against the observed volumes compared to other bicycle travel demand datasets managed by state and local planning agencies in San Francisco, California. These studies indicate that, despite the range of results, Strava continues to grow in its utility for understanding bicyclist behavior, and to plan for future bicycle infrastructure.

This page intentionally left blank.

# 2.  DATA SOURCES

To carry out the analysis for this study, three categories of data were acquired, including: counter data, Strava Metro data, and roadway spatial data. This section discusses each type of data in detail and presents the data in different views that help visualize what is included in various datasets.

## CDOT Non-Motorized Monitoring Counts

CDOT's NMM counts were correlated with Strava data to determine the extent to which Strava trips represent the total number of bicycle trips in Colorado. Counts were collected from 16 permanent continuous monitoring locations on an hourly basis per day over the course of 2017. Counters generally are located in the Front Range, on Interstate 70 (I-70) in the west, and positioned along major shared use trail and sidewalk segments. The counter locations are illustrated in Figure 1 and Table 1. CDOT conducted data quality checks to ensure that common reporting errors were not included in permanent count records. With erroneous data removed from the counts through CDOT's quality control process, available days of data ranged from 193 days to 365 days among these locations.

## Strava Metro

The Strava Metro dataset contains three sub-sets in three spatial formats: Edges (streets), Origin-Destination, and Nodes (intersections). For each type, the data product provides minute-to-minute data, rolled-up summary data, **geometry files, and demographic files.** The *Strava Metro Comprehensive User Guide* (Version 5.01, 2017) (found in Appendix A) provides users with detailed step-by-step instructions on how to understand and use the data product. In addition, Metro DataView is a web-based interactive tool that displays the activities recorded by Strava users in aggregate at different spatial levels (see Figure 2). It is a useful interface that gives users visual presentations of the dataset at both state and local levels.

*Figure 1. CDOT Counter Locations (2017)*
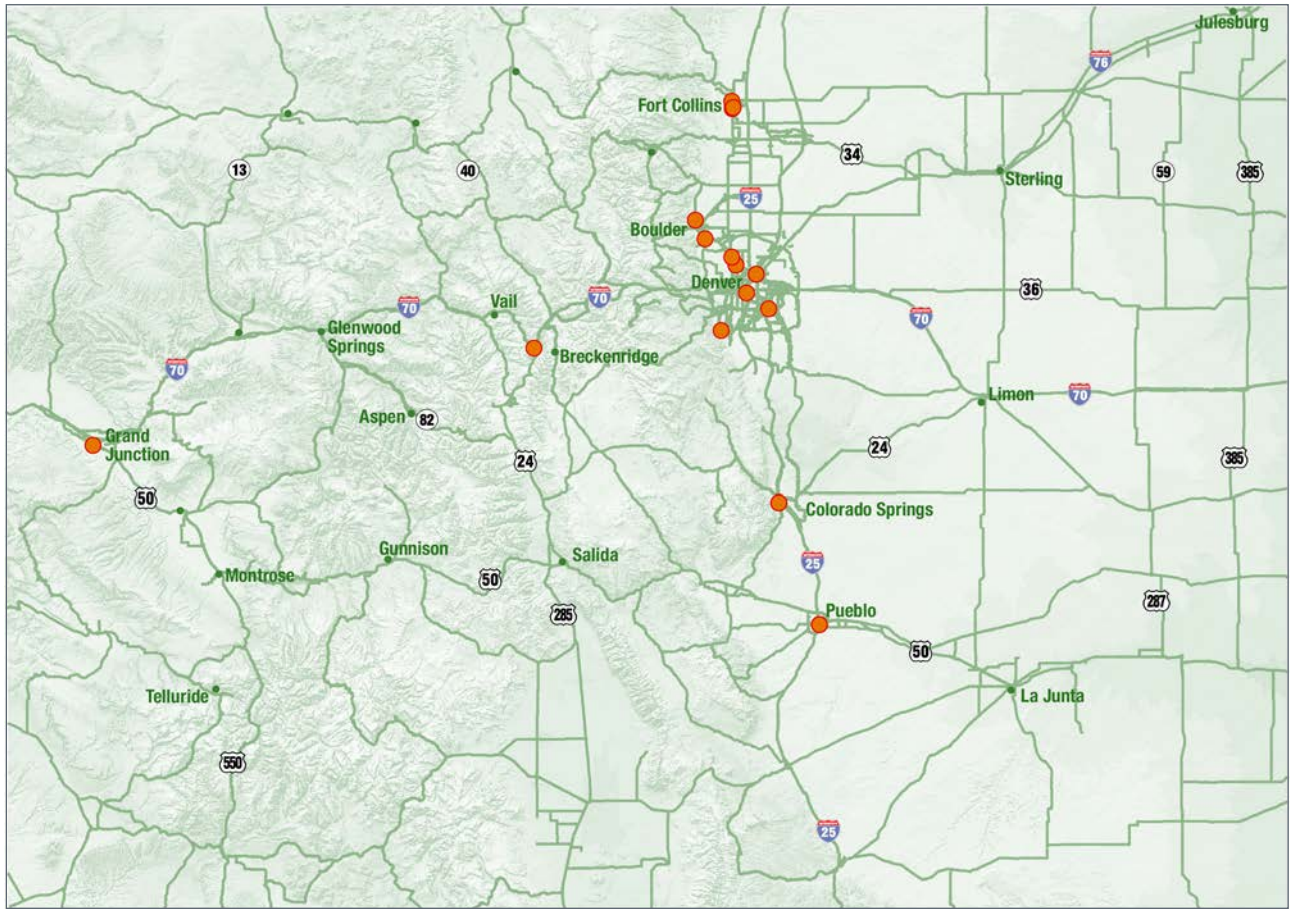


● Counter

*Table 1.  Continuous Count Location Summary*

| Counter ID | Lat/Long | Location | City / County | Analysis Period [1] | Total Counts |
|---|---|---|---|---|---|
| B60001 | 40.58332, -105.07886 | Mason Street north of Magnolia Street | Fort Collins | 2017 Jul-Dec | 81,660 |
| B60002 | 40.59681, -105.08242 | Poudre River Trail—Lee Martinez Park | Fort Collins | 2017 Jul-Dec | 93,117 |
| B60003 | 40.56887, -105.07533 | Remington Street north of East Lake Street | Fort Collins | 2017 Jan-Dec | 162,325 |
| B90004 | 40.068823, -105.283131 | US 36 | Boulder | 2017 Jan-Dec | 99,817 |
| B90009 | 39.674289, -104.88008 | Highline Canal Trail | Arapahoe County | 2017 Jan-Dec | 92,596 |
| B90010/ B90011 | 38.274513, -104.601797 / 38.274359, -104.602658 | East 8th Street WB/EB | Pueblo | 2017 Jan-Dec | 46,547 |
| B90013 | 39.827076, -104.94971 | South Platte River Trail | Denver | 2017 Jan-July | 34,088 |
| B90015 | 39.500175, -106.16472 | Tenmile Canyon Trail | Copper Mountain | 2017 Jul-Dec | 52,881 |

| Counter ID | Lat/Long | Location | City / County | Analysis Period [1] | Total Counts |
|---|---|---|---|---|---|
| B90018 | 39.068108, -108.5797 | Broadway Avenue Separated Path | Grand Junction | 2017 Jul-Dec | 14,709 |
| B90020 | 40.562972, -105.079855 | Mason Trail | Fort Collins | 2017 Jan-Dec | 355,161 |
| B90022 | 38.81481, -104.82444 / 38.81491, - 104.82468 | Tejon Street NB/SB | Colorado Springs | 2017 Jan-Dec | 36,821 |
| B90023 | 39.86781, -105.05895 | US 36 Bikeway (1)—1,800 feet south of Westminster Boulevard | Westminster | 2017 Jan-Dec | 33,826 |
| B90024 | 39.903054, -105.083024 | US 36 Bikeway (2)—550 feet north of Uptown Avenue | Broomfield | 2017 Jan-Dec | 72,682 |
| B90029 | 39.983483, -105.229389 | US 36 Bikeway (3)—parallel to US 36 on-ramp from Foothills Parkway | Boulder | 2017 Jan-Dec | 66,685 |
| B90030 | 39.74594, -105.00057 | Cherry Creek Trail at Lawrence | Denver | 2017 Jul-Dec | 405,259 |
| C90026 | 39.57841, - 105.140599 | C-470 Trail south of Ken Caryl Avenue | Jefferson County | 2017 Jan-Dec | 62,054 |

[1]  Through this quality control process, some erroneous data are removed; this can result in some of the datasets not containing the entire year of bicycle trip counts.

*Figure 2.  Strava Metro DataView Snapshot*
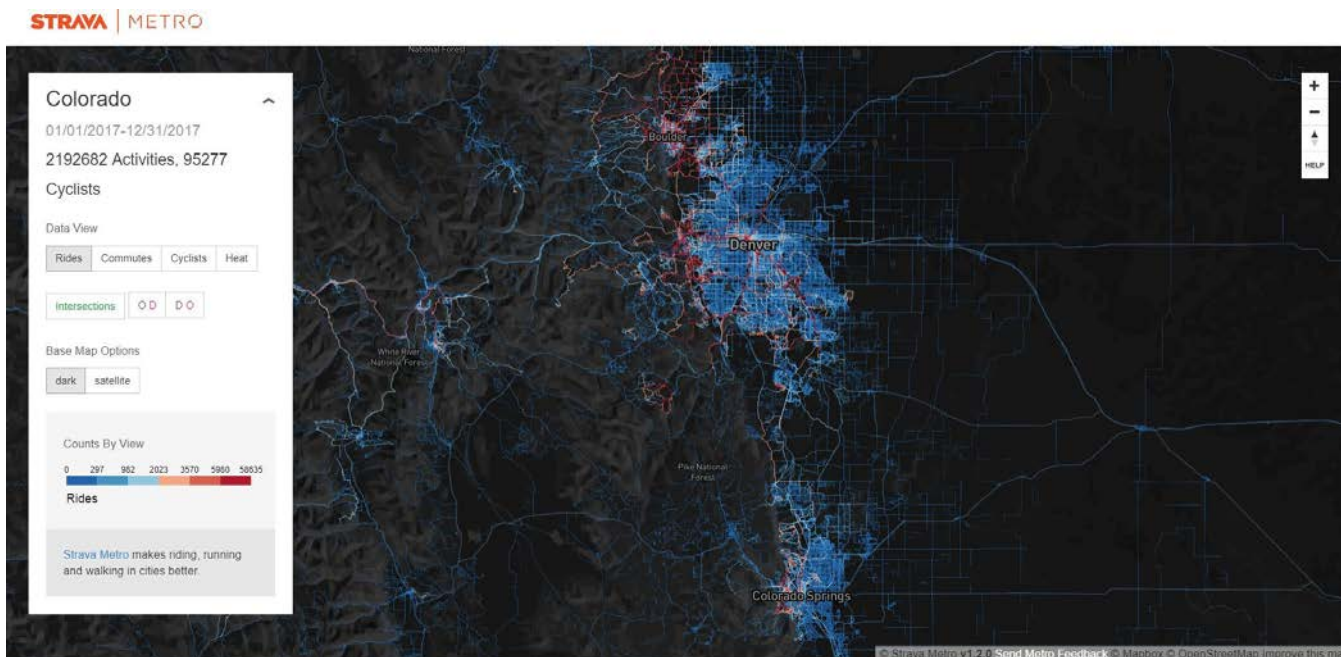


The minute-to-minute data are granular in scale and contain a record for every minute on every road segment upon which a bicyclist has crossed over the defined time (Strava, 2017). Included in the June 2016 to May 2017 Colorado statewide data, the minute-to-minute dataset is a 12 gigabyte (GB) file with more than 228 million records. To reduce data processing requirements, Strava provides "roll-up" files customized to the needs of the data user. Roll-up files provide a set of summarized Strava counts at requested temporal scales. The roll-up data acquired by CDOT consist of weekday/

weekend, monthly, peak riding season (April 1 to October 31), and yearly files. For each of the roll-up files, there is a sum for the pre-determined hourly range (see Table 2). A total of 2.2 million bicycling activities contributed by 95,277 unique users were recorded within Colorado in 2017. The roll-up data on each road segment over the one-year period were used primarily for the correlation analysis described in Section 4.

Table 2. Hourly Ranges "Roll-ups" in the Strava Metro Product Delivered to CDOT

| Label | Hourly Ranges |
|---|---|
| 0 | 12:00 a.m. to 3:59 a.m. |
| 1 | 4:00 a.m. to 5:59 a.m. |
| 2 | 6:00 a.m. to 8:59 a.m. |
| 3 | 9:00 a.m. to 2:59 p.m. |
| 4 | 11:00 a.m. to 12:59 p.m. |
| 5 | 3:00 p.m. to 5:59 p.m. |
| 6 | 6:00 p.m. to 7:59 p.m. |
| 7 | 8:00 p.m. to 11:59 p.m. |
| PEAK SEASON | April 1 to October 31 |

Additionally, the Metro dataset contains a demographic summary file, including average distance (28 kilometers), median distance (21 kilometers), average time (1.7 hours), and median time (1.4 hours). Approximately 75 percent of the users were male and 63 percent of the users were between the ages of 25 and 54 (see Table 3). This study incorporated distance and average time data for quality control processes, as discussed in Section 3, but did not utilize demographic information.

Table 3. Strava Unique User Age and Gender Demographics

| AGE | MALE | PERCENT MALE | FEMALE | PERCENT FEMALE |
|---|---|---|---|---|
| UNDER 25 | 5,063 | 7.1% | 1,446 | 7.3% |
| 25 to 34 | 14,555 | 20.4% | 5,064 | 25.6% |
| 35 to 44 | 16,322 | 22.9% | 4,235 | 21.4% |
| 45 to 54 | 13,882 | 19.5% | 3,024 | 15.3% |
| 55 to 64 | 6,283 | 8.8% | 1,397 | 7.1% |
| 65 to 74 | 1,548 | 2.2% | 265 | 1.3% |
| 75 to 84 | 170 | 0.2% | 14 | 0.1% |
| 85 to 94 | 16 | 0.0% | 3 | 0.0% |
| 95 ABOVE | 50 | 0.1% | 9 | 0.0% |
| BIRTHDAY NOT SPECIFIED | 13,312 | 18.7% | 4,311 | 21.8% |
| TOTAL | 71,201 | 100% | 19,768 | 100% |

## Roadway Data

Two roadway datasets were used for this study. The road segment map included in the Strava Metro data product was derived from Open Street Map (OSM). OSM is a crowd-sourced mapping product with open-source editing features. Strava uses OSM features to create a Geographic Information System (GIS) "Metro edge" layer that contains spatial information for roads and trails, and attributes

such as street names, and a unique identification number for joining to tabular Strava-generated activity data. This allows users to easily combine Strava activity data with the OSM spatial dataset. The roadway segments in OSM range in length from 0.0001 mile to 18 miles, including more than one million features for the state of Colorado. Figure 3 shows a view of the OSM data at Interstate 25 (I-25) and Alameda in Denver, Colorado.

The other roadway dataset was provided by CDOT, which includes a spatial layer of all the state highway facilities (see Figure 4). Both datasets were exported to GIS shapefiles for overlay comparisons and data preparation for the High-Use Bicycle Corridor Analysis.
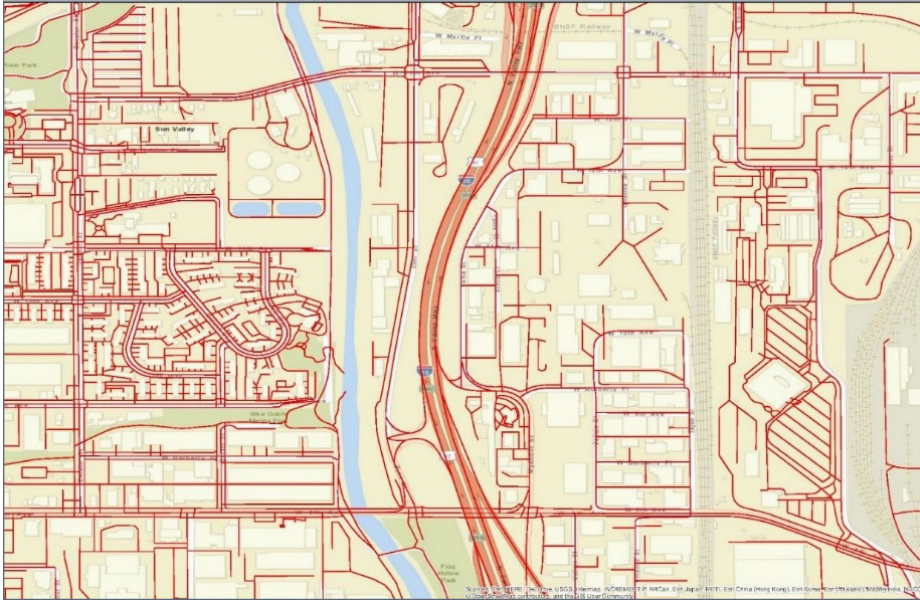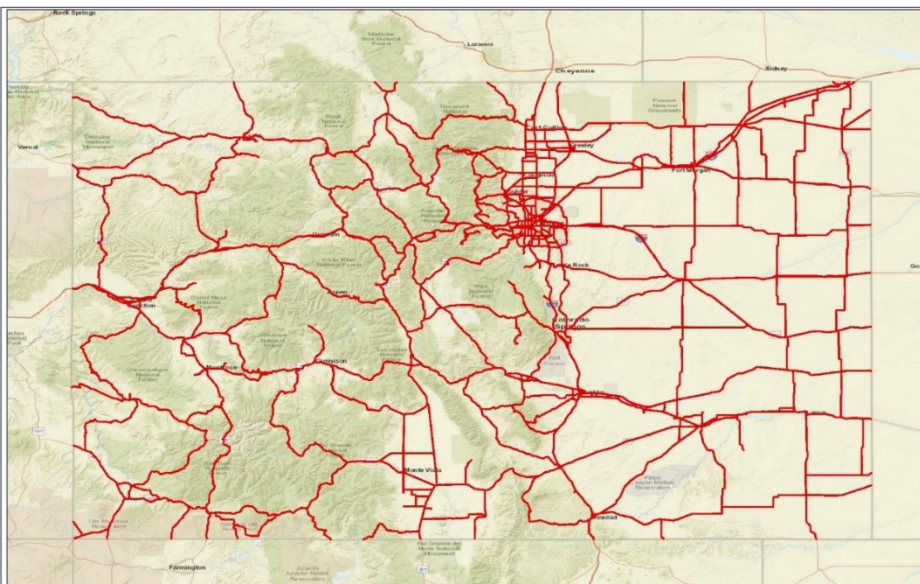
*Figure 3. Zoom-In View of I-25 at Alameda Avenue in OSM*



*Figure 4. Colorado State Highways*

# 3. DATA MANAGEMENT AND PROCESSING

Currently, there are no established and universally accepted methods for processing crowd-sourced bicycling data. An objective of this study is to develop replicable methods for quality control and data management as it relates to crowd-sourced data, using the Strava Metro dataset as a test case. This study has identified and tested methods for preparing the data for analysis within an acceptable level of quality.

## Trip Data Quality Control

Strava data was assessed for quality to increase confidence in analysis results. While Strava data is processed to control locational errors and other invalid data references, some error is anticipated, primarily from users of the application. Strava cannot, for example, guarantee that users of the application have not mistakenly recorded a vehicle trip as a part of a bicycling trip.

While it may be possible to provide additional quality assurances in the user data with trip length, time, or location, speed is considered the most efficient and effective way to identify non-bicycle trip errors and is Strava's recommended practice. Using roll-up data, this method is used to identify the segments that exhibit an unreasonably high median speed over the roll-up time periods. Given the recording rate with GPS devices (5 to 10 seconds), the margin of error for spatial match tends to be higher for a short segment than a long segment. Strava recommends identifying short geometry noise that may cause false values prior to any analysis (Strava, 2017).

Erroneous data for Colorado was identified based on minimum and maximum speed thresholds. For this assessment, any segment longer than 100 meters (approximately 0.06 mile) that exhibits a median speed greater than 45 miles per hour was defined as erroneous. Figure 5 shows the corridors that exhibit an unreasonably high median speed, including I-70, Interstate 76 (I-76), Colorado Highway 470 (C-470), U.S. Highway 36 (US 36), I-25, U.S. Highway 285 (US 285), and Interstate 225 (I-225). Approximately 6 percent of the dataset exhibited erroneous high-speed trip data. These data likely are representative of users who left the Strava application on while driving.

*Figure 5. Median Speed Greater Than 45 mph for All Strava Bicycle Trips in 2017*
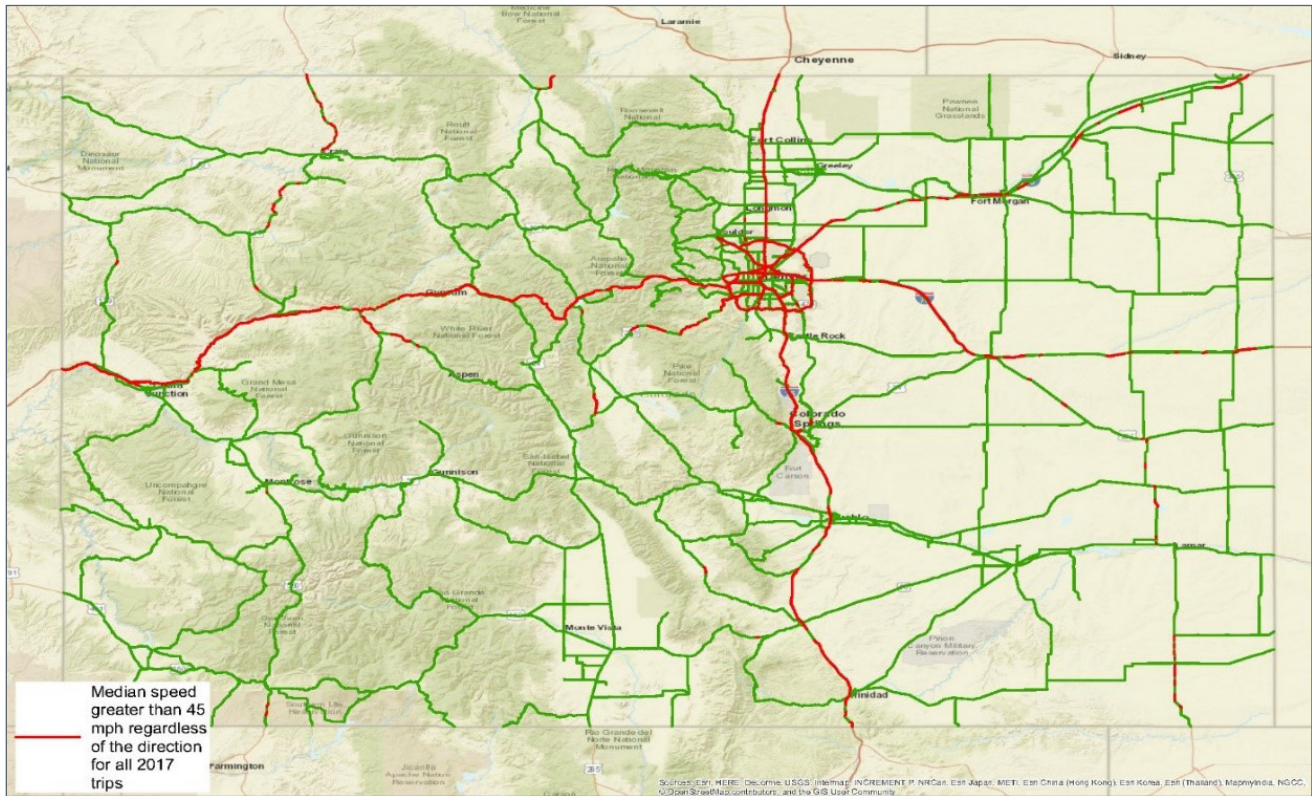


## Spatial Matching

A common error associated with this mobile device data collection method is in the initial data collection process. When mobile device GPS points do not match precisely with the roadway network, it can lead to incorrect facility assignments. Spot checking was performed at a series of locations, including I-70, US 36, and C-470, to ensure that bicycling activities were captured on the correct facilities, particularly regarding the types of facilities. For example, a bicycle trip was **verified to be on the shared-use path (e.g., C-470 Trail) rather than on C-470 where bicycling activity** is restricted.

## Continuous Counter Data Quality

Continuous counter data has been provided by CDOT for this study. Prior to providing this information, counter data are checked by CDOT for accuracy as part of the regular public reporting process. Through this quality control process, some erroneous data are removed; this can result in some of the datasets not containing the entire year of bicycle trip counts. In addition, some sites are **flagged as potentially containing erroneous information because they exhibit total count numbers** that are unlikely based on comparatively similar locations and facilities. These potentially erroneous data are noted in the comparison results.

Finally, due to the low number of available count locations, it is reasonable to suspect that there **is unquantified error in the results of this study.** It has been determined that this level of error is acceptable for the purposes of this study. It is likely that as additional continuous counter information becomes available, the statistical comparison of actual count data to Strava data will increase the accuracy of extrapolations.

This page intentionally left blank.

# 4. COUNTER DATA COMPARISON

A primary goal of this study was to determine how representative Strava data are of total statewide bicycling activity. Stationary counter data provided by CDOT were correlated with Strava data at 16 locations throughout Colorado using simple linear regression. The following sections describe the processes for the correlation analysis.

## Counter Location Matching

The locations of all 16 counters were matched with OSM segments. The nearest OSM segment of each counter was selected as a spatial match. Because of the unreliable spatial matching discussed previously, not all points matched to the exact location. To account for this issue, a spatial selection was made to identify all the links within a buffer zone of 15 feet around each point. Due to the high-level spatial granularity in OSM, it was relatively straightforward to identify all links that match with the locations of CDOT counters for the correlation analysis. Appendix A (*Counter Correlation Analysis*) provides screenshots that show the selected roadway links for counter comparisons.

## Strava and Counter Data Comparisons

CDOT-provided counts were aggregated and summed into the pre-determined hourly intervals as chosen in the roll-up data time periods (see Table 2). Comparisons between the two datasets were made at a yearly level, for weekdays and weekends, and for the peak riding season (April 1 to October 31) during 2017. The overall relationship was evaluated using ordinary least squares regression. $R^2$ values using simple linear regression for each time period provided an indication of the strength of the relationship between CDOT count data and crowd-sourced bicyclist volumes from Strava.

## Correlation Analysis Results

Table 4 is a summary of the counter locations, analysis periods, and total volumes in each dataset. A regression analysis was performed for total annual trips, total annual weekday trips, total annual weekend trips, and "On Season" trips, referring to trips in the months of April through October. The regression analysis results are listed in Table 4. "R Square" values that are close to 1.000 indicate a strong linear correlation between the two compared datasets. The linear

correlations between the counter data and Strava counts were high for all assessed time periods. The percentages of total bicyclists captured by Strava ranged from 0.7 percent to 30.3 percent in these locations. However, it should be noted that the counter resulting in the lowest representation, located at East 8th Street in Pueblo, is likely to contain erroneous data due to unusually high recorded bicycle trip counts. The results show near perfect representation at some locations, such as the Tenmile Canyon Trail, which would likely have higher use by recreational bicyclists utilizing the Strava application. The representativeness of weekday trip data is only slightly less than weekend trip data with some exceptions, such as on the US 36 Bikeway and the C-470 Trail.

*Table 4.  Correlation Analysis Results*

| | Entire Year | | | | Weekday | | Weekend | | On Season[2] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Counter Trips | Strava Trips | % of Total Counts | R Square | % of Total Counts | R Square | % of Total Counts | R Square | % of Total Counts | R Square |
| Mason Street | 81,660 | 2,371 | 2.9% | 0.942 | 2.7% | 0.918 | 3.5% | 0.927 | — | — |
| Poudre River Trail | 93,117 | 5,255 | 5.5% | 0.988 | 5.6% | 0.978 | 5.4% | 0.988 | — | — |
| Remington Street | 162,325 | 4,804 | 2.8% | 0.924 | 2.8% | 0.881 | 3.2% | 0.969 | 2.9% | 0.914 |
| US 36 | 99,817 | 29,992 | 30.3% | 0.994 | 28.8% | 0.990 | 32.0% | 0.997 | 29.4% | 0.989 |
| Highline Canal Trail | 92,596 | 7,674 | 7.4% | 0.969 | 7.6% | 0.918 | 7.4% | 0.993 | 6.9% | 0.953 |
| East 8th Street[1] | 46,547 | 344 | 0.7% | 0.815 | 0.6% | 0.659 | 0.9% | 0.869 | 0.8% | 0.795 |
| South Platte River Trail | 34,088 | 6,651 | 18.8% | 0.966 | 20.7% | 0.900 | 18.1% | 0.988 | — | — |
| Tenmile Canyon Trail | 52,881 | 4,196 | 7.4% | 0.996 | 5.3% | 0.990 | 10.4% | 0.999 | — | — |
| Broadway Avenue | 14,709 | 2,100 | 14.2% | 0.971 | 13.5% | 0.939 | 15.9% | 0.873 | — | — |
| Mason Trail | 355,161 | 14,979 | 4.2% | 0.985 | 4.1% | 0.965 | 4.6% | 0.979 | 4.1% | 0.976 |
| Tejon Street NB/SB | 36,821 | 4,823 | 17.0% | 0.874 | 14.1% | 0.868 | 23.5% | 0.897 | 16.6% | 0.879 |
| US 36 Bikeway (1) | 33,826 | 7,035 | 19.1% | 0.929 | 22.3% | 0.893 | 17.4% | 0.991 | 19.0% | 0.917 |
| US 36 Bikeway (2) | 72,682 | 12,446 | 15.9% | 0.968 | 18.1% | 0.943 | 14.6% | 0.994 | 15.6% | 0.954 |
| US 36 Bikeway (3) | 66,685 | 13,267 | 19.1% | 0.991 | 20.1% | 0.971 | 19.0% | 0.996 | 18.5% | 0.985 |
| Cherry Creek Trail | 405,259 | 24,526 | 5.9% | 0.947 | 5.7% | 0.925 | 6.8% | 0.951 | — | — |
| C-470 Trail | 62,054 | 11,024 | 18.7% | 0.997 | 16.4% | 0.997 | 20.7% | 0.998 | 18.0% | 0.998 |

[1]  While the East 8th Street counter comparison numbers have been retained to maintain some geographic representation in that area, the data are likely to contain erroneous counts due to unreasonably high levels of activity recorded at that site.

[2]  "On Season" comparison, the comparison of counts during months of high activity (April through October), is incomplete for several counter sites that did not have the entire year of data necessary to conduct the comparison.

This page intentionally left blank.

# 5.  HIGH-USE BICYCLE CORRIDOR ANALYSIS

The purpose of the high-use bicycle corridor analysis is to identify the most-used CDOT corridors for bicycling, using Strava user data as a surrogate for total bicycle travel activity. Based on Strava bicycle activity for 2017, this **assessment shows the relative classification of use across the** statewide system, as well as regional sub-areas.

## State Highway System Selection

A large amount of Strava activity occurs off-system, or on facilities and corridors not managed by CDOT.  To complete the assessment of on-system corridors, a state highway layer was created from the OSM spatial dataset. While CDOT maintains a spatial layer of on-system facilities included with its Online Transportation Information System (OTIS), this layer is not an exact match to the OSM data, and is not easily attributed to Strava user data; therefore, a state highway selection process was conducted to select features from the OSM spatial dataset.

The selection processes included a combination of three spatial and attributed data selection methods:

1.  Selection of highway facilities from the OSM layer based on the OSM "CLAZZ" attributes using the following selection query (more information about CLAZZ **identification is included in the** *Strava Metro User Guide*, found in Appendix B):

    "CLAZZ" = 11 OR "CLAZZ" = 12 OR "CLAZZ" = 13 OR "CLAZZ" = 14 OR "CLAZZ" = 15 OR "CLAZZ" = 16 OR "CLAZZ" = 21 OR "CLAZZ" = 22.
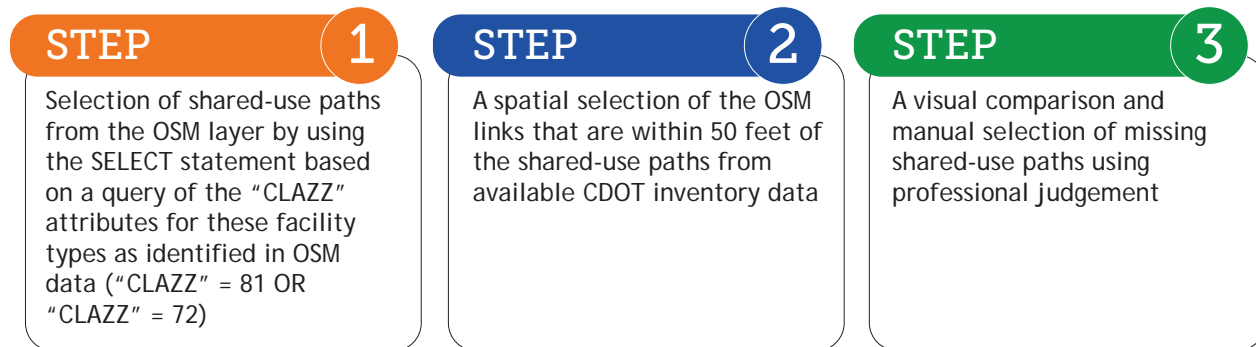
2.  A GIS spatial selection using CDOT's OTIS data and its attributes to select OSM links. The spatial selection query includes OSM links within 1,000 feet of the divided highway segment and 100 feet of an undivided highway.

3.  Visual check and manual selection of state routes based on professional judgment.

Application of this methodology captured all state routes, but it should be noted that some non-state route segments were included in this selection process because those segments intersected with state routes. This issue is more prevalent in denser roadway networks in urban areas. While a more precise state highway layer could be created from the OSM data with additional time and resources, it was determined that, for the purpose of this study, the small number of intersecting off-system segments included do not significantly impact the results of the analysis.

## Shared-Use Path Selection

While the majority of state highway corridors are primarily used for vehicular use, some of these corridors include adjacent non-motorized off-street shared-use paths. CDOT is currently developing a statewide inventory and spatial data collection of shared-use paths. For this study, a subset of the shared-use path inventory data was used to identify OSM segments for trails adjacent to state highway corridors.

The selection process included the following steps:

| STEP 1 | STEP 2 | STEP 3 |
|---|---|---|
| Selection of shared-use paths from the OSM layer by using the SELECT statement based on a query of the "CLAZZ" attributes for these facility types as identified in OSM data ("CLAZZ" = 81 OR "CLAZZ" = 72) | A spatial selection of the OSM links that are within 50 feet of the shared-use paths from available CDOT inventory data | A visual comparison and manual selection of missing shared-use paths using professional judgement |

The layer that was developed out of this selection was merged with the state highways layer into a single layer in ArcGIS for the high-use bicycle corridor analysis.

## High-Use Bicycle Corridor Results

In 2017, the number of annual Strava recorded bicycle trips on state highway corridors ranged from zero to more than 58,000. Compared across the state, approximately 85 percent of the state highway corridor segments have total Strava trips below 1,000, where nearly 2 percent of the segments have total Strava trips numbering more than 10,000. Figure 6 shows a frequency distribution for miles of roadway with recorded Strava bicycle trips.

Given the skewed distribution of segments with higher usage, segments with moderately high use are less observable in the data. It was determined that these outliers should be broken out of the main dataset to allow for a more even distribution and comparison of segments. To categorize statistical outlier segments, the interquartile range (IQR) method was applied. Compared statewide, "very high use" corridors account for 9 percent of the total assessed corridor length and contain 90 percent of the total bicycle trips in Colorado.

The remainder of the data were assessed in GIS using Jenks (Natural breaks) classification to generate three categories of high, medium, and low bicycle use. Jenks classifies data based on "naturally" distributed data groups. An advantage of this method is that it creates classes that have accurate representation of the trends in data distribution. Additionally, Strava recommends using this method of classification in the *Strava Metro User Guide*, found in Appendix B.

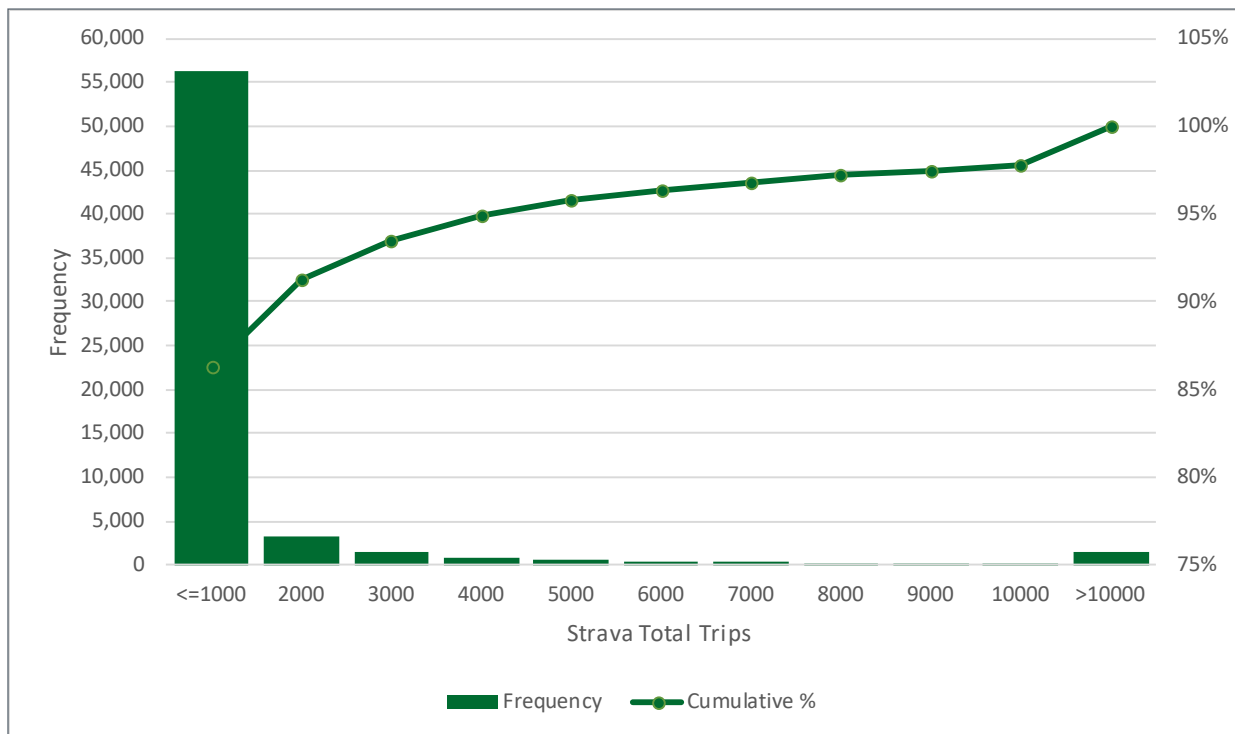*Figure 6. Histogram of Total Strava Bicycle Trips per Highway Segment in 2017*



Figure 7 shows the assigned categories and thresholds on the state highways and adjacent shared-use paths using the Jenks classification method.

Statewide high-use bicycle corridor results show that the predominantly used corridors (very high-use and high-use corridors) include the following, as shown on Figure 7:

- US 50
- US 6
- US 91
- SH 82
- SH 9
- SH 145
- SH 74/Evergreen Parkway
- SH 119/Longmont Diagonal Highway
- Ute Highway
- Cherry Creek Trail
- Clear Creek Trail
- Longmont/Boulder Lobo Trail
- Mineral Belt Trail

- US 550
- US 40
- US 36 Bikeway
- SH 131
- SH 103/Squaw Pass Road
- SH 83
- SH 7/South Saint Vrain Drive
- SH 119/Boulder Canyon Drive
- East Colfax Avenue
- High Line Canal Trail
- Poudre River Trail
- Santa Fe Regional Trail
- Rio Grande Trail

- US 160
- US 24
- US 34/Trail Ridge Road
- SH 72
- SH 5/Mount Evans Road
- SH 74/Bear Creek Road
- North Foothills Highway
- I-25 Frontage Road near Castle Rock
- Broadway Street, Boulder
- Platte River Trail
- C-470 Trail
- Glenwood Canyon Trail
- Broadway Avenue, Grand Junction

*Figure 7. Statewide High-Use Corridors*
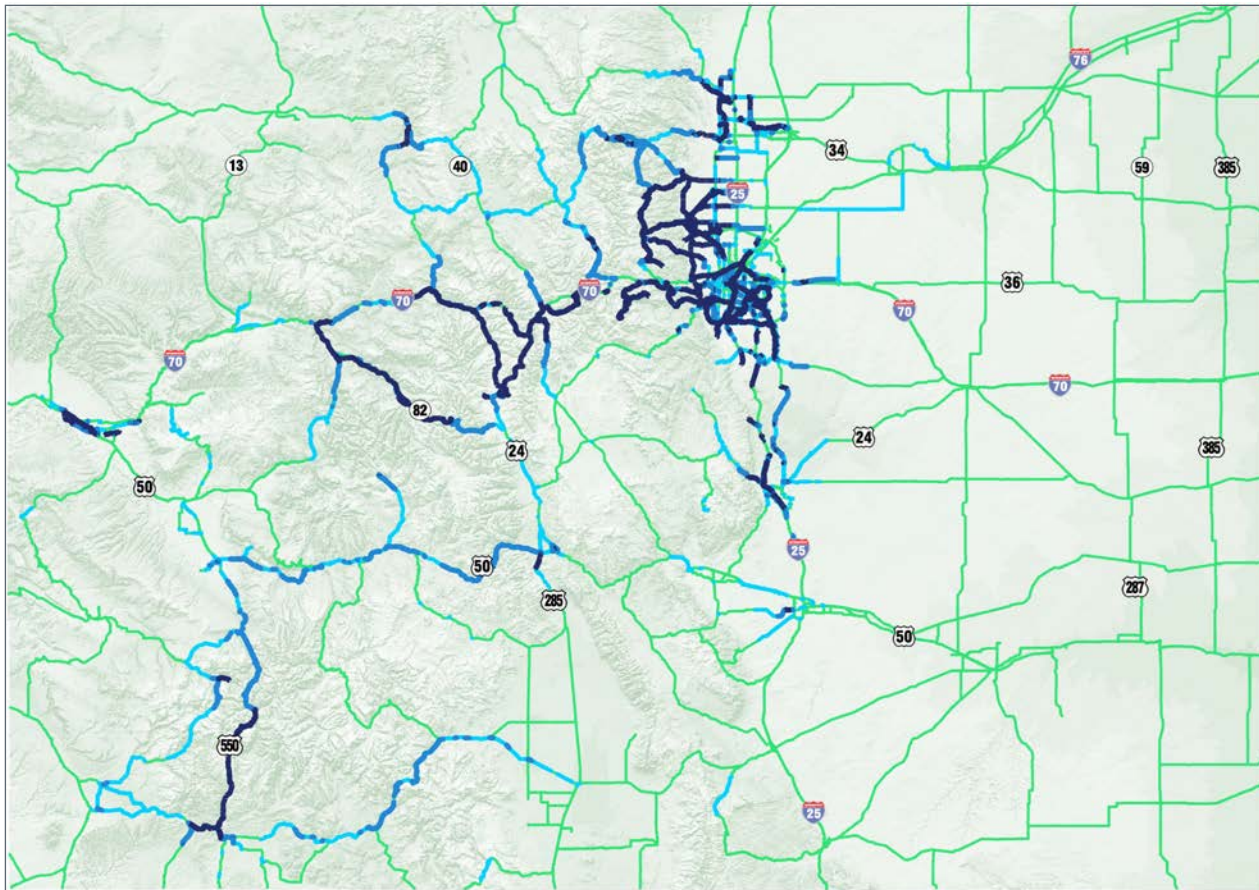


**Colorado Strava Bicycle Activities 1/1/2017 - 12/31/2017**

— Very High (810 - 58,635)  — High (400 - 809)  — Medium (140 - 399)  — Low (0 - 139)

## Regional Bicycle High-Use Results

Bicycle use varies dramatically by location. While assessing bicycle use at a state level provides a comprehensive comparison of the bicycling activities across Colorado, it neglects some of the high-use bicycle corridors in specific regions due to variability in Strava sample size and total users.

To account for this variation, the state was divided into three regions: Western Slope, Front Range, and Eastern Plains (see Figure 8). The regional boundaries were based on geographical characteristics (e.g., mountains versus plains) and general population density (e.g., metropolitan areas of the Front Range). The regionally separated dataset then was classified using the same methodology developed for the statewide bicycle use: outliers were identified using the IQR, and other categories were identified using Jenks.

*Figure 8. Sub-region Boundaries*
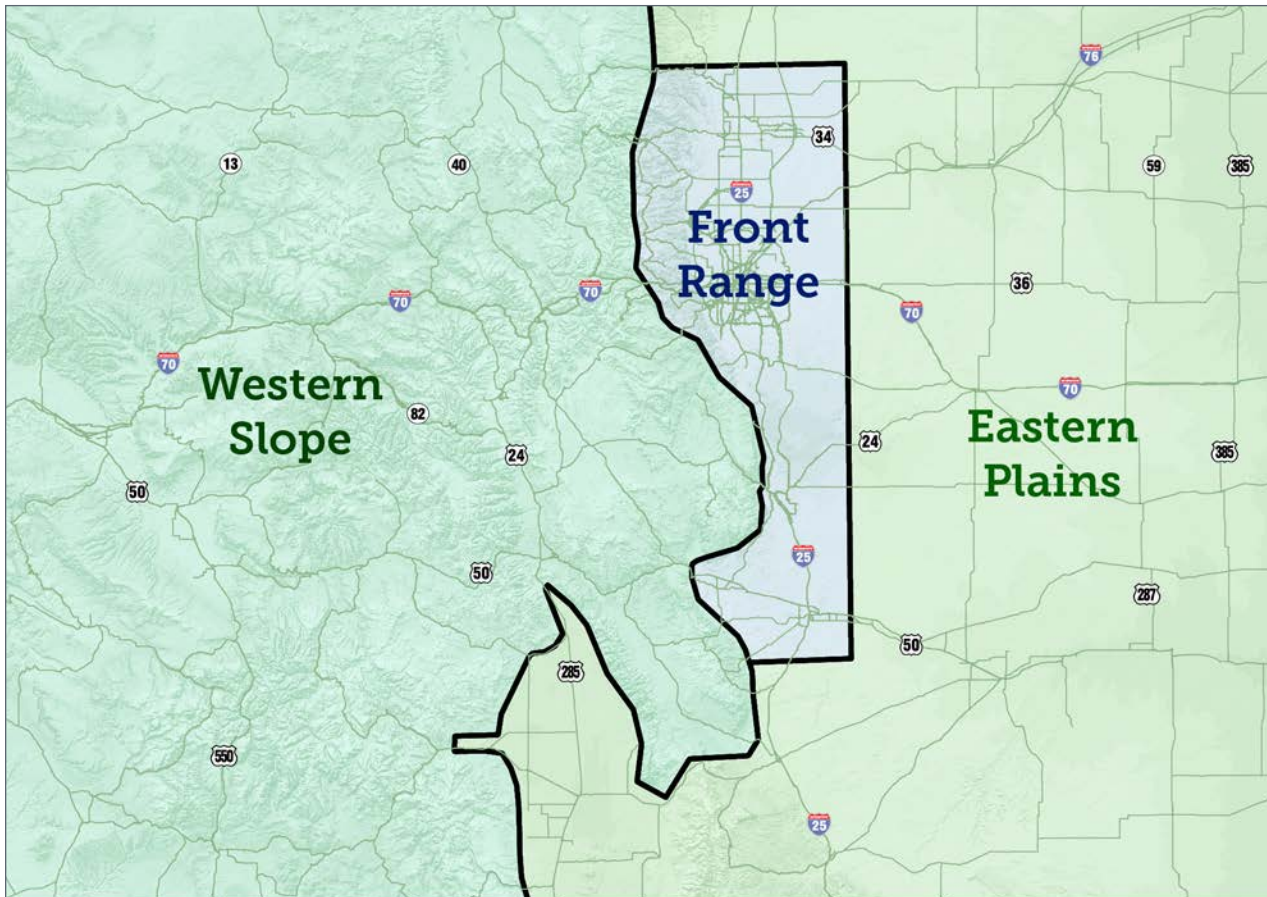


Figures 9, 10, and 11 on the following pages show the usage categories on the state highways and adjacent shared-use paths for each region. The high-use corridors identified in the regional analysis are shown on their respective figures, with the weekday and weekend Strava trip comparisons.

The results of the regional use categories show that the majority of the high-use corridors in the Front Range and Western Slope regions remain unchanged from the statewide analysis. However, due to the regional use criteria, some segments in the identified statewide high-use corridors were not classified as highly used. This is most obvious on corridors including US 50, US 160, US 40, SH 9, and US 550 in the Western Slope region, and SH 7, South SH 83, and East Colfax Avenue in the Front Range region.

In the Eastern Plains region, additional corridors are classified as highly used corridors based on the regional use criteria. These corridors include US 285, SH 12, SH 17, SH 96, SH 52, US 160, SH 144, and US 34.

*Figure 9. Western Slope Region High-Use Corridors*
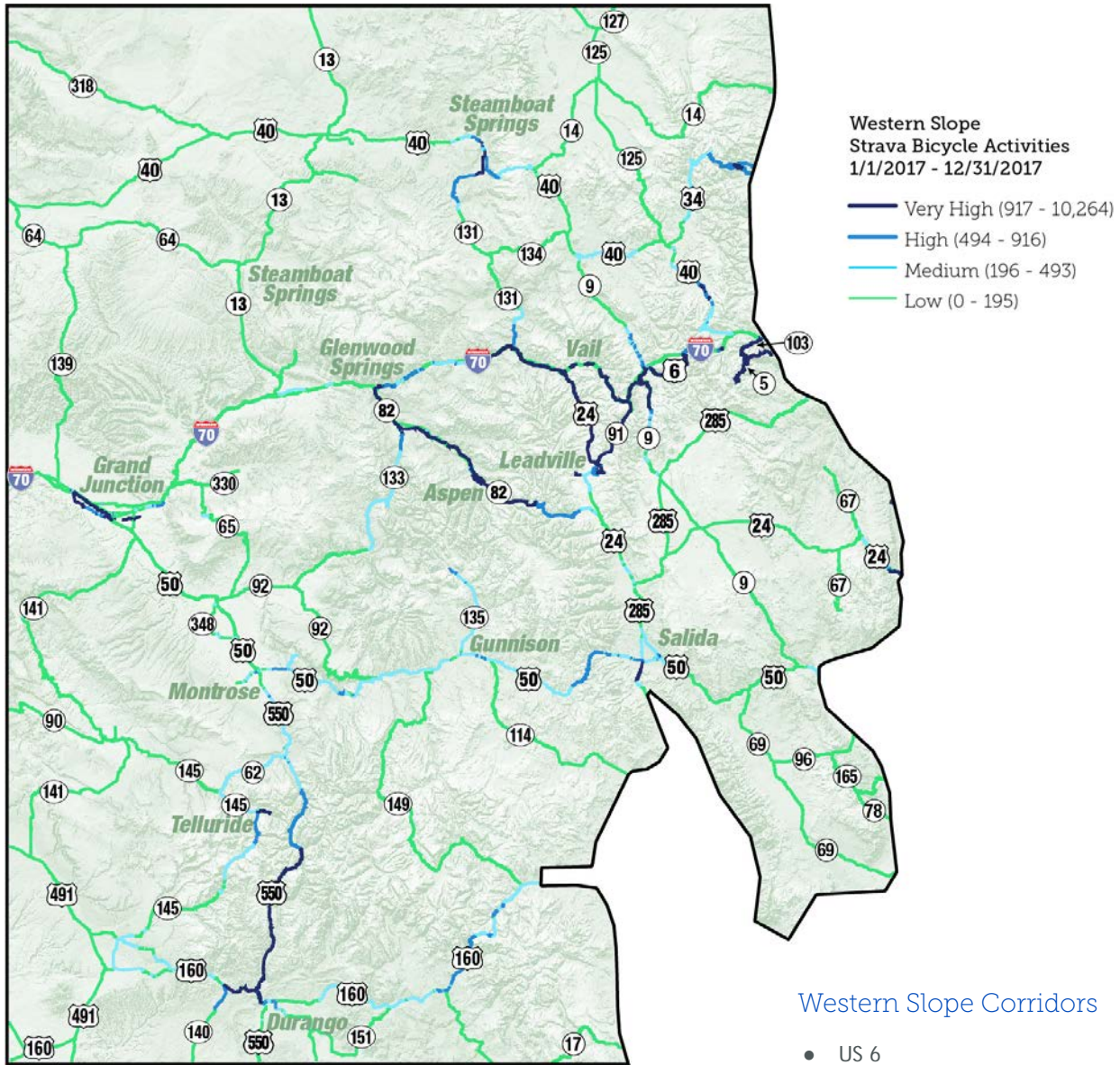


## Western Slope Corridors

- US 6
- SH 82
- US 24
- US 91
- SH 9
- SH 131
- SH 103/Squaw Pass Road
- SH 5/Mount Evans Road
- US 40
- US 34/Trail Ridge Road
- US 550
- US 50
- US 160
- SH 145
- Broadway Avenue, Grand Junction
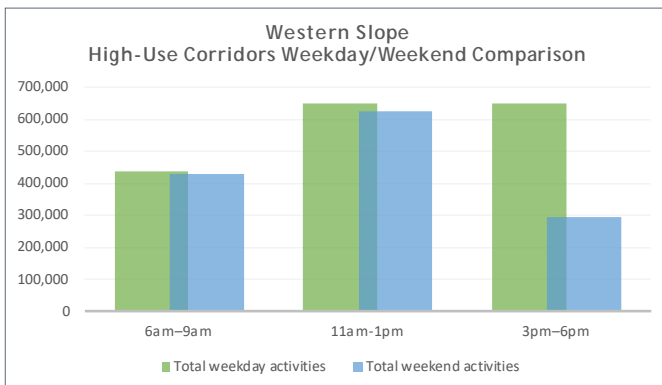- Clear Creek Trail
- Glenwood Canyon Trail

*Figure 10. Front Range Region High-Use Corridors*

## Front Range Corridors

- SH 72
- SH 119/Longmont Diagonal Highway
- SH 119/Boulder Canyon Drive
- SH 7/South Saint Vrain Drive
- North Foothills Highway
- SH 83
- SH 66/Ute Highway
- US 36 Bikeway
- SH 74/Bear Creek Road
- SH 74/Evergreen Parkway
- Broadway Street, Boulder
- I-25 Frontage Road near Castle Rock
- Cherry Creek Trail
- High Line Canal Trail
- Platte River Trail
- Poudre River Trail
- C-470 Trail
- Longmont/Boulder Lobo Trail
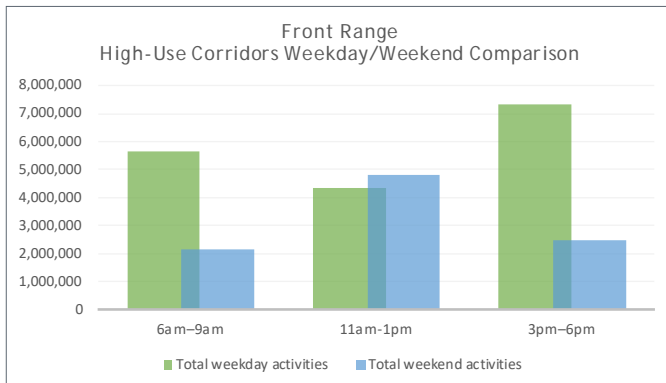- Santa Fe Regional Trail
- Mineral Belt Trail



Front Range
Strava Bicycle Activities
1/1/2017 - 12/31/2017

— Very High (1,265 - 58,635)
— High (682 - 1,264)
— Medium (232 - 681)
— Low (0 - 231)



Front Range
High-Use Corridors Weekday/Weekend Comparison

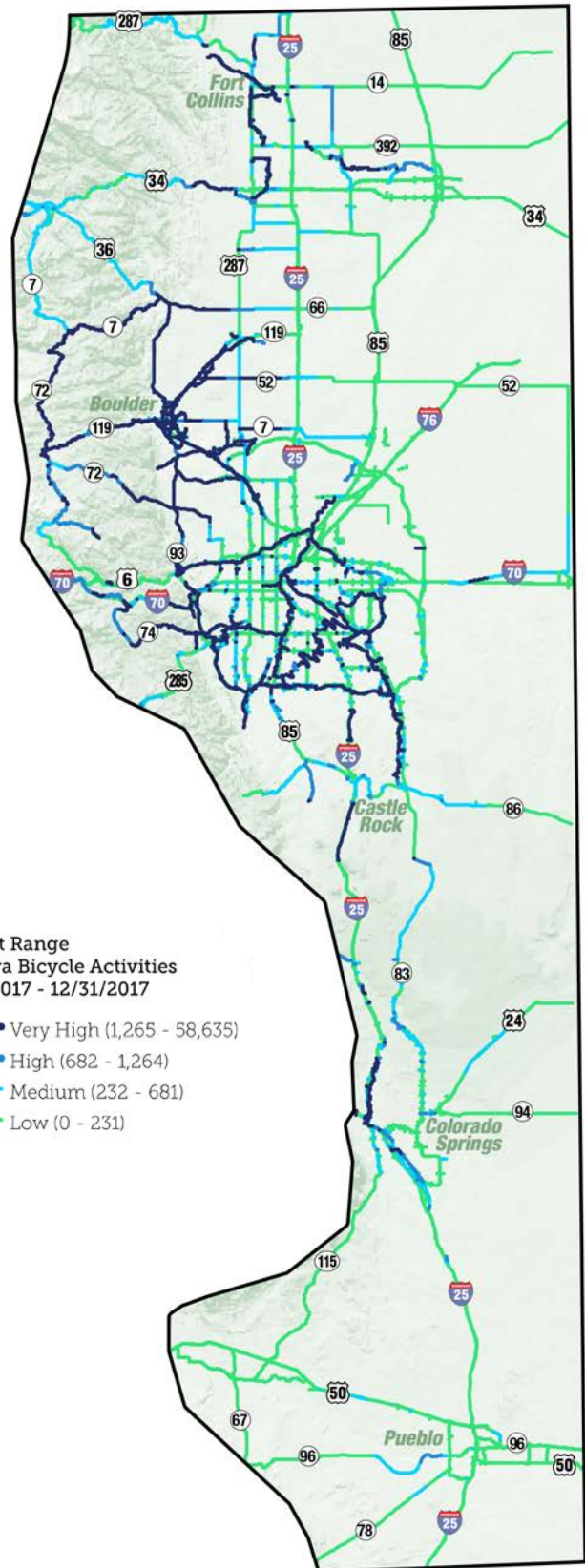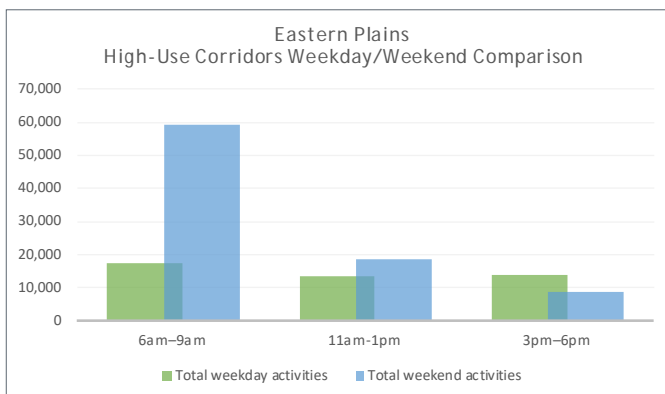■ Total weekday activities ■ Total weekend activities

*Figure 11. Eastern Plains Region High-Use Corridors*

## Eastern Plains Corridors

- US 285
- SH 12
- SH 17
- SH 96
- SH 52
- US 160
- US 34
- SH 144



Eastern Plains
Strava Bicycle Activities
1/1/2017 - 12/31/2017

Very High (71 - 526)
High (40 - 70)
Medium (14 - 39)
Low (0 - 13)



Eastern Plains
High-Use Corridors Weekday/Weekend Comparison

■ Total weekday activities  ■ Total weekend activities
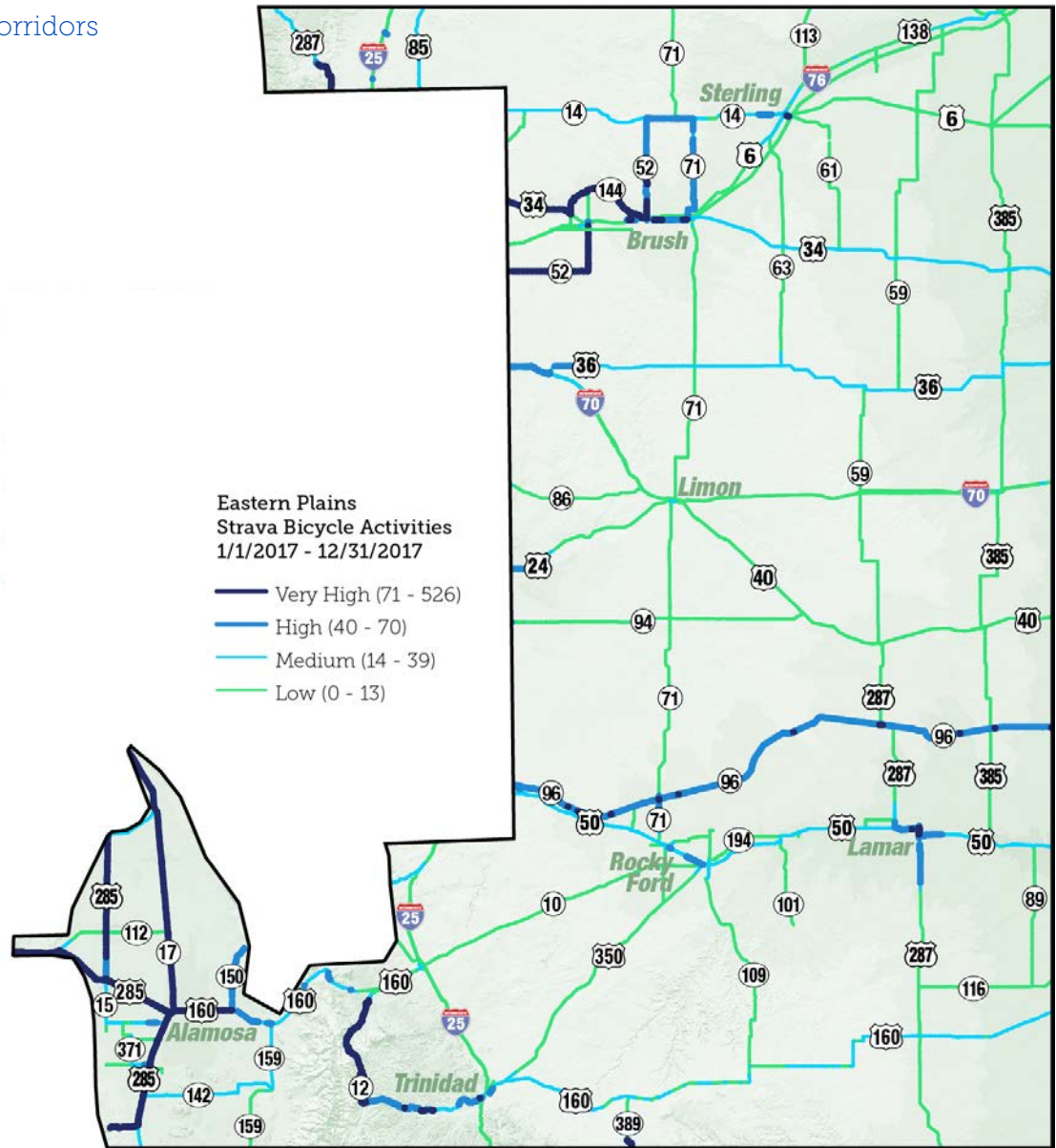
This page intentionally left blank.

# 6. SUMMARY DISCUSSION AND LESSONS LEARNED

The practice of utilizing big data for non-motorized activity evaluation is new and somewhat experimental. The assessment of Colorado Strava data advances bicycle activity analysis and stands as a case example of how a state agency might evaluate similar types of information about bicycling to improve understanding of bicyclist behavior and the non-motorized functions of a transportation system. The identified shortcomings of the assessment (e.g., issues with data quality, sample size, and data integration) are accounted for and considered acceptable for the level of assessment in this study. The following topics are considerations for others interested in conducting similar assessments and furthering this type of work.

## Data Management and Storage

Big data is just that—big. Uncompressed annual minute-by-minute data for the state of Colorado, available in the Strava Metro dataset, includes more than 100 GBs of information and nearly a million individual data points. To access disaggregated data, users must have database capabilities, abundant storage capacity, and sufficient processing power. Strava increases the accessibility of its information with two offerings. Firstly, as discussed in Section 2, they provide a time period roll-up containing aggregated data by select daily time periods. With the exception of some quality control processes, this roll-up information was used almost exclusively in the CDOT analysis. Secondly, users of Strava data are able to request select geographic areas. For the CDOT statewide assessment, it was necessary to utilize the entire Colorado dataset, but with a concentrated evaluation of a specific corridor, city, or region, data are likely to be more manageable. As big data is more readily available and experience using it is more advanced, it is likely that data management and storage will continue to be a key factor in the use of the data.

## Strava Sample Representation Discussions

The results of the counter correlation study show that Strava users represent a range between 3 percent and 30 percent of the total bicycle use on Colorado highway corridors. While this sample of the total bicycling population is not insignificant, in terms of its total bicyclist trip representation, it should be noted that the characteristics of Strava sample population may misrepresent the behavior of the general bicycling population. Typical Strava users tend to be interested in bicycling enough to track regular bicycling activities and they are likely to have a propensity for recreational bicycling. Any representation of bicycling activity as captured by open source data is likely to contain similar bias error regardless of the sample size. Nonetheless, without comparable methods for tracking bicycle activity on a large scale, these data are the best source of information in the current context, and are very useful for planning processes.

# 7. COLORADO STRAVA DATA USER RECOMMENDATIONS

Use of Strava data may be daunting to those without experience in its use, but several key recommendations can be made for practitioners interested in conducting a similar assessment using Strava data.

*Prior to data download:*

1. Preview data—To understand what data are available from Strava for your area of interest, Strava provides a static data viewer online. Through the data viewer, users can preview the information contained in the Strava Metro dataset and toggle through pre-defined data attributes and classifications.

2. User agreement—In the state of Colorado, Strava data has been made available by CDOT to all interested individuals/agencies through a user agreement. Interested individuals/agencies must complete a user agreement between themselves and Strava to access the information. For information about data use and to initiate a user agreement, contact CDOT's Bicycle and Pedestrian section (https://www.codot.gov/programs/bikeped).

3. Gather available bicycle counts—It is possible to develop a high-level understanding of bicycle travel patterns using Strava data without a comparison to actual bike counts, but with counter data, users can make informed assumptions about bicycle trips occurring across the system. Any count data is useful, including continuous or short-duration counts, because Strava data can be compared to any specific time period within a year.

4. Mobilize data storage capacity and technical capabilities—Especially for large areas, it is necessary to have ample data storage capacity. Be prepared to provide tens of GBs in storage space. Large datasets will require the use of a database platform (e.g., SQL, Postgres, BD2, etc.) to access information. Smaller datasets (i.e., Strava roll-up datasets) are manageable in Microsoft Excel. Strava delivers OSM data as shapefiles that are viewed and manipulated in ArcGIS.

5. Define the area of interest—To minimize delay and multiple coordination steps, identify an area of interest that includes the entire area of potential study and then add a little extra. Include an extra buffer around your area of interest so as not to mistakenly exclude any segments of interest.

*Following data download:*

1. Understand the error—This CDOT Strava assessment identified several types of data error observed in the data processing that may induce further erroneous results. Because this type of information is new, misleading conclusions may be easily produced and are potentially damaging to the perceived value of the information. Use best practice judgment to minimize error in the data, and meticulously catalogue all assumptions while producing results.

2. Share results—Misconceptions about open-source data and data captured from mobile devices has the potential to hinder near-term access to information. Data privacy is imperative and use of this data drives the dialogue about how to improve data access within acceptable privacy rules. As the availability and use of this type of information increases, more practitioners will become accustomed to using it as a source of information for making planning and engineering decisions.

# 8. REFERENCES

Boss, D., Nelson, T., Winters, M., Ferster, C., (2018). Using crowd-sourced data to monitor change in spatial patterns of bicycle ridership. Journal of Transport and Health (2018). https://doi.org/10.1016/j.jth.2018.02.008

Chen, C., (2017). Crowd-sourcing Data-driven Development of Bicycle Safety Performance Functions (SPFs): Microscopic and Macroscopic Scales.

Colorado Department of Transportation (CDOT). (2016). Non-Motorized Monitoring Strategic Plan. Available at: https://www.codot.gov/programs/bikeped/documents/2016-10-21-cdot-nonmotorized-monitoring-plan_low-res.pdf.

CDOT. (2015). Bicycle and Pedestrian Master Plan. Available at: https://www.codot.gov/programs/bikeped/building-a-bike-ped-friendly-community/Bike_Ped_Plan/BikePedStatePlan/at_download/file

Griffin, G.P., Jiao, J., (2015). Crowd-sourcing Bicycle Volumes: Exploring the Role of Volunteered Geographic Information and Established Monitoring Methods. URISA Journal. Vol. 27, No.1. http://dx.doi.org/10.1016/j.jth.2014.12.001.

Hochmair, H.H., (2016). Estimating bicycle trip volume for Miami-Dade County from Strava tracking data

Jestico, B., Nelson, T., Winters, M., (2016). Mapping ridership using crowd-sourced cycling data. Journal of Transport Geography 52 (2016) 90–97. http://dx.doi.org/10.1016/j.jtrangeo.2016.03.006.

Moore, M.A., (2015). Modeling Factors Influencing Commuter Cycling Routes: A Study of GPS Cycling Records in Auburn, Alabama

Proulx, F.R., Pozdnukhov, A., (2017). Bicycle Traffic Volume Estimation using Geographically Weighted Data Fusion. Journal of Transport Geography

Strava LLC. (2017). Strava Metro Comprehensive User Guide. Available at: http://metro.Strava.com/wp-content/uploads/2017/04/StravaMetro_5.01_User_Guide.pdf

Vermont Agency of Transportation (Vtrans). (2016). Vtrans On-road Bicycle Plan: Phase 1 Report. Available at: http://vtrans.vermont.gov/sites/aot/files/planning/bikeplan/VTrans%20On%20Road%20Bicycle%20Plan%20Phase%201%20Report%20FULL.pdf

# APPENDIX A:
# COUNTER CORRELATION ANALYSIS

# APPENDIX B:
# STRAVA METRO USER GUIDE