StravaMetro Product

Documentation: Streets, Nodes and OD

Product Name: StravaMetro

Product Version: v1.03.00

Document Name: StravaMetro Product Documentation

Contents

[Introduction 2](#_Toc406154324)

[Product Description 2](#_Toc406154325)

[Description 2](#_Toc406154326)

[Shipment Contents 3](#_Toc406154327)

[Typical Product Layers and Contents 3](#_Toc406154328)

[Program Versions Required 4](#_Toc406154329)

[Metro Edges 4](#_Toc406154330)

[Metro\_<datatype>: Year, Season, Month 4](#_Toc406154331)

[Metro\_Minute 7](#_Toc406154332)

[Origin Destination Data 8](#_Toc406154333)

[Origin Destination Polygon Data 9](#_Toc406154334)

[Intersections (Nodes) Minute Data 9](#_Toc406154335)

[Intersections (Nodes) Rolled-Up Data 10](#_Toc406154336)

[Demographic Files 12](#_Toc406154337)

[Glossary/Definitions 14](#_Toc406154338)

[Changes for This Release 14](#_Toc406154339)

[Contact Information 14](#_Toc406154340)

[FAQs: 14](#_Toc406154341)

[Appendix: 17](#_Toc406154342)

[Quick Start Guide 17](#_Toc406154343)

[Indexes 17](#_Toc406154344)

[PostGIS Queries 17](#_Toc406154345)

# Introduction

Thank you for using Strava Metro. Strava Metro is the result of years of research and collaboration between Strava GEO and different departments of transportation (DOT) and transportation research groups. It enables cutting edge views into cycling patterns worldwide. Please send feedback and requests to [geo@strava.com](mailto:geo@strava.com).

This document explains the data, table structures and general format of the Strava Metro product. It also contains an FAQ section with answers to common questions.

We are excited to be offering this product with the mission to facilitate a better cycling environment for everyone.

Cheers,

Geo Team

# Product Description

## Description

The Strava Metro product aggregates all of the cycling and/or pedestrian data recorded by Strava members for a given location and time frame aggregated to a street network. This data is provided in a number of different temporal formats from minute to yearly summaries, and produced in classic GIS formats so it can be easily integrated into GIS environments of transportation engineering firms and DOTs.

We’ve gone to great lengths to protect the privacy of our members by anonymizing and aggregating the data to a linear street map such that the value of the data is preserved.

Strava Metro is map agnostic, meaning we can provide the product for use with proprietary maps, Open Street Map etc. Separate data tables are generated which join back to the map street network via primary key. We do not modify the map data directly.

# Shipment Contents

## Typical Product Layers and Contents

This section outlines what is contained in most deliveries of Strava Metro. The Strava Metro product is constantly evolving as we locate and build in key features. The delivery would contain the following data files:

* .csv/.sql raw hourly data file
* .dbf rolled-up views in the cycling data (listed in the table below)
* Origin/Destination data table raw
* Origin/Destination Polygon
* Nodes data table raw
* Nodes point file
* Demographics document
* Product description document
* ESRI .MXD project file
* ESRI .MPK map package file

Below is the list of the core data files that are contained in a typical delivery of StravaMetro.

| Table name | Type |
| --- | --- |
| <identifier>\_edges | Line |
| <identifier>\_nodes | Point |
| <identifier>\_od\_polygons | Polygon |
| <identifier>\_<ride/run>\_rollup\_month\_<year>\_<month>\_weekend | Table |
| <identifier>\_<ride/run>\_rollup\_month\_<year>\_<month>\_weekday | Table |
| <identifier>\_<ride/run>\_rollup\_season\_<month>\_weekend | Table |
| <identifier>\_<ride/run>\_rollup\_season\_<month>\_weekday | Table |
| <identifier>\_<ride/run>\_rollup\_total\_weekend | Table |
| <identifier>\_<ride/run>\_rollup\_total\_weekday | Table |
| <identifier>\_<ride/run>\_rollup\_total | Table |
| <identifier>\_edges\_<ride/run>\_data | .csv/.sql/Table |
| <identifier>\_edges\_nodes\_<ride/run>\_data | .csv/.sql/Table |
| <identifier>\_<ride/run>\_od\_polygons\_data | .csv/.sql/Table |

**Table 1 definition:** There is a naming structure to the files so that a name would look similar to joondalup\_city\_edges\_nodes\_ride\_rollup\_month\_2013\_6\_weekend.

**<identifier>** = Unique name of the region or group that is purchasing the data. For example, joondalup\_city. The identifier is referenced in the email that is sent with the product.

**<year>** = The numeric year of the rollup in conjunction with the following month.

**<month>** = A numeric value of 1 – 12 that represents the months (Jan. – Dec.).

# Program Versions Required

Postgres Version: 9.1+

PostGIS Version: 2.0+

ArcGIS Version: 10.1.+

QGIS

### Metro Edges

**Name:** <identifier>\_edges.shp

**File Type:** Shapefile, PostGIS, OSM

**Datum:** WGS84 [SRID = 4326]

If the map used for the data came from Open Street Map. The OSM street network is converted from the .pbf (xml) format into a PostGIS table during the process. This table is then exported to a shapefile with corresponding related data tables. The id field joins the street network to the rolled up data. Not all OSM attributes are carried through because the Strava Metro process does not require them.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | Id | Double | Unique and permanent OSM Edge ID number for delivery |
| 2 | osm\_id | Double | Preserved OSM street id; this is not stable and is intended to only be used for historical reference |
| 3 | osm\_name | Text | Street name from the OSM data |
| 4 | osm\_meta | Text | Metadata from OSM about the street |
| 5 | Clazz | Double | Street class according to OSM |
| 6 | Flags | Double | Any key flags from the OSM data |
| 7 | source | Double | Starting nodes for the line |
| 8 | target | Double | Ending nodes for the line |
| 9 | Km | Double | Distance of the line in km |
| 10 | length | Double | Distance of the line in meters |
| 11 | the\_geom\*\* |  | Binary internal representation of geometry |

\*\*the\_geom field only in PostGIS formats

### Metro\_<datatype>: Year, Season, Month

**Name:** <identifier>\_<ride/run>\_rollup\_month\_<year>\_<month>\_weekend.dbf

**File Type:** DBF

**Datum:** N/A

To facilitate the use of the Strava Metro product we provide a set of rolled up data files that show pre-determined temporal ranges. These ranges are extracted into monthly, peak riding season and yearly files. These temporal ranges vary with latitude and season. Below are the date and hourly ranges for the AM/PM data used for a typical Strava Metro product.

* On-Season: April 1 – Oct 1
* Very Early AM hours: 12am – 3:59am
  + Labeled as *\_0*
* Early AM hours: 4am – 5:59am
  + Labeled as *\_1*
* AM Peak Hours: 6am – 8:59am
  + Labeled as *\_2*
* Mid-Day Hours: 9am – 2:59pm
  + Labeled as *\_3*
* PM Peak Hours: 3pm – 5:59pm 
  + Labeled as *\_4*
* Early Evening Hours: 6pm – 7:59pm
  + Labeled as *\_5*
* Late Evening Hours: 8pm – 11:59pm 
  + Labeled as *\_6*

For each data file there is a sum for that time frame, and a group for all requested time frame roll-ups. This creates numerous fields per data file and allows for fast pinpointed analysis. If there is a different time frame you are interested in please let us know. These time frames are build-up from what was defined in the contract.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | EDGE\_ID | Double | Unique and permanent Street ID number for delivery |
| 2 | ATHCNT | Double | Count of unique cyclists on the piece of street for the rolled-up date frame. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 3 | RACTCNT | Double | Count of unique cyclists on the piece of street for the rolled-up date frame. This number represents the number of cyclists going against direction the street was digitized (see glossary). |
| 4 | ACTCNT | Double | Count of bike trips (regardless of unique riders) on the piece of street for the rolled-up date frame. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 5 | RACTCNT | Double | Count of bike trips (regardless of unique riders) on the piece of street for the rolled-up date frame. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 6 | TATHCNT | Double | Total number of unique cyclists on the piece of street regardless of direction of travel for the rolled-up date frame. |
| 7 | TACTCNT | Double | Total number of bike trips on the piece of street regardless of direction of travel for the rolled-up date frame. |
| 8 | ACTTIME | Double | Median time in seconds that bike trips on the piece of street for the rolled-up date frame. This number represents the time of cyclists going the direction the street was digitized (see glossary). |
| 9 | RACTTIME | Double | Median time in seconds that bike trips on the piece of street for the rolled-up date frame. This number represents the time of cyclists going against direction the street was digitized (see glossary). |
| 10 | CMTCNT | Double | Total number of commute bike trips on the piece of street regardless of direction of travel for the rolled-up date frame. |
| 11 | ATHCNT\_X | Double | Count of unique cyclists on the piece of street for the rolled-up date frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 12 | RATHCNT\_X | Double | Count of unique cyclists on the piece of street for the rolled-up date time frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the number of cyclists going against direction the street was digitized (see glossary). |
| 13 | ACTCNT\_X | Double | Count of bike trips (regardless of unique riders) on the piece of street for the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 14 | RACTCNT\_X | Double | Count of bike trips (regardless of unique riders) on the piece of street for the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the number of cyclists going the direction the street as digitized (see glossary). |
| 15 | TATHCNT\_X | Double | Total number of unique athletes on the piece of street regardless of direction of travel for rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| 16 | TATHCNT\_X | Double | Total number of bike trips on the piece of street regardless of direction of travel for rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| 17 | ACTTIME\_X | Double | Median time in seconds for bike trips on the piece of street during the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the time of cyclists going the direction the street was digitized (see glossary). |
| 18 | RACTTIME\_X | Double | Median time in seconds for bike trips on the piece of street during the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). This number represents the time of cyclists going against direction the street was digitized (see glossary). |
| 19 | CMTCNT\_X | Double | Total number of commute bike trips on the piece of street regardless of direction of travel for the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| XXX | Pattern Cont. |  | The naming pattern continues for all predefined time frames. For example is the day had 4 time frames then it would be \_0, \_1, \_2, \_3 . |

### Metro\_Minute

**Name:** <identifier>\_edges\_<ride/run>\_data.csv, <identifier>\_edges\_<ride/run>\_data.dbf

**File Type:** Comma-delimited, DBF (only if the file is less than 2 gigs)

**Datum:** N/A

The minute data is the granular format of the Metro product. This file contains a row for every street edge and minute that a cyclist crossed it over the defined time. These tend to be very large data sets in the tens and hundreds of millions of records for an annual review.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | edge\_id | integer | Unique and permanent Open Street Map Edge ID number for delivery |
| 2 | year | integer | Numerical year format (yyyy) |
| 3 | day | integer | Numerical day format (1 - 365) |
| 4 | hour | integer | Numerical hour format (0 - 24) |
| 5 | minute | integer | Numerical minute format (0 - 59) |
| 6 | athlete\_count | ATHLETE\_CO | integer | Count of unique cyclists on the piece of street for the day, hour and minute. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 7 | rev\_ athlete\_count | REV\_ATHLET | integer | Count of unique cyclists on the piece of street for the day, hour and minute. This number represents the number of cyclists going against direction the street was digitized (see glossary). |
| 8 | activity\_count | ACTIVITY\_C | integer | Count of bike trips (regardless of unique riders) on the piece of street for the day, hour and minute. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 9 | rev\_ activity\_count | REV\_ACTIVI | integer | Count of bike trips (regardless of unique riders) on the piece of street for the day, hour and minute. This number represents the number of cyclists going the direction the street was digitized (see glossary). |
| 10 | total\_activity\_count | TOTAL\_ACTI | integer | Total number of bike trips on the piece of street regardless of direction of travel for the day, hour and minute. |
| 11 | activity\_time | ACTIVITY\_T | integer | Median time in seconds that bike trips on the piece of street for the day, hour and minute. This number represents the time of cyclists going the direction the street was digitized (see glossary). |
| 12 | rev\_ activity\_time | REV\_ACT\_1 | integer | Median time in seconds that bike trips on the piece of street for the day, hour and minute. This number represents the time of cyclists going against direction the street was digitized (see glossary). |
| 13 | commute\_count | COMMUTE\_CO | integer | Total number of commute bike trips on the piece of street regardless of direction of travel for the day, hour and minute. |

### Origin Destination Data

**Name:** <identifier>\_<ride/run>\_od\_polygons\_data.csv; <identifier>\_<ride/run>\_od\_polygons\_data.dbf

**File Type:** .csv

**Datum:** N/A

This data has the starting and ending polygon for all the rides that made up the data delivery. It also has an array of all of the polygons that a trip touched along the way. This is a new way of looking at recreation and commute trips while further protecting a user’s privacy. To use the array you cannot load it into a .dbf file as the .dbf has a 255 character limit. It works best in a database that can take advantage of an array. We do not provide roll-ups of the OD because there is not a roll up that we can provide that cannot be built quickly in a GIS program.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | polygon\_id | integer | Unique Starting Polygon ID number |
| 2 | year | integer | Numerical year format (yyyy) |
| 3 | day | integer | Numerical day format (1 - 365) |
| 4 | hour | integer | Numerical hour format (0 - 24) |
| 5 | minute | integer | Numerical minute format (0 - 59) |
| 6 | commute | integer | 0 -1 Flag for locating commutes (1 = commute) |
| 7 | dest\_polygon\_id | integer | Unique Ending Polygon ID number |
| 8 | intersected\_polygons | integer | This is an array of polygons that were intersected during this activity. Only present in the .csv format. |

### Origin Destination Polygon Data

**Name:** <identifier>\_od\_polygons.shp

**File Type:** Polygon Shapefile, postgres sql file

**Datum:** N/A

This data has the starting and ending polygon for all the rides that made up the data delivery. It also has an array of all of the polygons that a trip touched along the way. This is a new way of looking at recreation and commute trips while further protecting a user’s privacy.

### Intersections (Nodes) Minute Data

**Name:** <identifier>\_edges\_nodes\_<ride/run>\_data

**File Type:** Point Shapefile

**Datum:** N/A

The minute data is the granular format of the Metro product. This file contains a row for every node and minute that a cyclist crossed it that node over the defined time. These tend to be very large data sets in the tens and hundreds of millions of records for an annual review.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | node\_id | integer | Unique and permanent Node ID number for delivery |
| 2 | year | integer | Numerical year format (yyyy) |
| 3 | day | integer | Numerical day format (1 - 365) |
| 4 | hour | integer | Numerical hour format (0 - 24) |
| 5 | minute | integer | Numerical minute format (0 - 59) |
| 6 | athletes | integer | Count of unique cyclists at the intersection for the day, hour and minute. This number represents the number of cyclists that meet at the intersection. |
| 7 | activities | integer | Count of activities at the intersection for the day, hour and minute. This number represents the number of activities that meet at the intersection. |
| 8 | median\_wait | integer | Median wait time at the intersection for that minute. |
| 9 | max\_wait | integer | Maximum wait time at the intersection for that minute. This can be very high at times as sometimes cyclists can stop and cause longer than normal wait times. |
| 10 | min\_wait | integer | Minimum wait time at the intersection for that minute. |
| 11 | commute | integer | Count of commute trips at the intersection for the day, hour and minute. This number represents the number of trips that meet at the intersection that are defined as a commute. |

### Intersections (Nodes) Rolled-Up Data

**Name:** <identifier>\_nodes\_<ride/run>\_rollup\_month\_<year>\_<month>\_weekend.dbf

**File Type:** DBF

**Datum:** N/A

To facilitate the use of the Strava Metro Nodes product we provide a set of rolled up data files that show pre-determined temporal ranges. These ranges are extracted into weekend, weekday and yearly files. Below are the date and hourly ranges for the AM/PM data typically used when generating in this product.

* Very Early AM hours: 12am – 3:59am
  + Labeled as *\_0*
* Early AM hours: 4am – 5:59am
  + Labeled as *\_1*
* AM Peak Hours: 6am – 8:59am
  + Labeled as *\_2*
* Mid-Day Hours: 9am – 2:59pm
  + Labeled as *\_3*
* PM Peak Hours: 3pm – 5:59pm 
  + Labeled as *\_4*
* Early Evening Hours: 6pm – 7:59pm
  + Labeled as *\_5*
* Late Evening Hours: 8pm – 11:59pm 
  + Labeled as *\_6*

For each data file there is a sum for that time frame, and a group for all requested time frame roll-ups. This creates many fields per data file and allows for fast pinpointed analysis. If there is a different time frame you are interested in please let us know. These time frames are build-up from what was defined in the contract.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Field** | **Type** | **Description** |
| 1 | NODE\_ID | Integer | Unique and permanent Node ID for delivery |
| 2 | ATHCNT | Integer | Count of unique cyclists at the intersection summarized for the time frame. |
| 3 | ACTCNT | Integer | Count of activities at the intersection summarized for the time frame. |
| 4 | MEDMEDT | Double | This is a bit more complicated as it is the median value of a set of Median values. In this case it the Median of the Median wait time of the activities at the intersection summarized for the time frame. |
| 5 | MEDMAXT | Double | This is a bit more complicated as it is the median value of a set of median values. In this case it the Median of the Maximum wait time of the activities at the intersection summarized for the time frame. |
| 6 | MEDMINT | Double | This is a bit more complicated as it is the median value of a set of Median values. In this case it the Median of the Minimum wait time of the activities at the intersection summarized for the time frame. |
| 7 | CMTCNT | Double | Sum of the commute activities at the intersection summarized for the time frame. |
| 11 | ATHSCNT\_X | Double | Count of unique cyclists at the intersection summarized for the predefined time frames as noted above (where X = the numbered time frame). |
| 12 | ACTSCNT\_X | Double | Count of activities at the intersection summarized for the rolled-up date time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| 13 | MEDMEDT\_X | Double | This is a bit more complicated as it is the median value of a set of Median values. In this case it the Median of the Median wait time of the activities at the intersection summarized between the predefined time frames as noted above (where X = the numbered time frame). |
| 14 | EDMAXT\_X | Double | This is a bit more complicated as it is the median value of a set of median values. In this case it the Median of the Maximum wait time of the activities at the intersection summarized for the rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| 15 | MEDMINT\_X | Double | This is a bit more complicated as it is the median value of a set of Median values. In this case it the Median of the Minimum wait time of the activities at the intersection summarized for rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| 16 | CMTCNT\_X | Double | Sum of the commute activities at the intersection summarized for the time frame for rolled-up time frame between the predefined time frames as noted above (where X = the numbered time frame). |
| XXX | Pattern Cont. |  | The naming pattern continues for all predefined time frames. For example is the day had 4 time frames then it would be \_0, \_1, \_2, \_3 . |

### Demographic Files

**Name:** <identifier>\_Demographics.txt

**File Type:** text files

We understand how important demographic data is to analysis. Therefore, we built a file that is designed to help you have a deeper understanding of the users who make up the data Strava is supplying. The Demographics document shows demographic detail generated by the data sample at the region requested.

Below you can find the fields and their definitions.

|  |  |
| --- | --- |
| **Field** | **Description** |
| County Name | The name of the polygon region to which the data is being summarized at. |
| Athlete ID Count | Number of unique athlete IDs that had a ride start in the region. |
| Activity Count | Number of unique cycling activity IDs that had a ride start in the region. |
| Male Count | Number of male athletes that occupy that region as assigned by Strava. |
| Male Count Under 25 | Number of male athletes that occupy that region as assigned by Strava: Under 25 |
| Male Count 25 - 34 | Number of male athletes that occupy that region as assigned by Strava: 25 – 34 |
| Male Count 35 - 44 | Number of male athletes that occupy that region as assigned by Strava: 35 – 44 |
| Male Count 45 - 54 | Number of male athletes that occupy that region as assigned by Strava: 45 – 54 |
| Male Count 55 - 64 | Number of male athletes that occupy that region as assigned by Strava: 55 – 64 |
| Male Count 65 - 74 | Number of male athletes that occupy that region as assigned by Strava: 65 – 74 |
| Male Count 75 - 84 | Number of male athletes that occupy that region as assigned by Strava: 75 – 84 |
| Male Count 85 - 94 | Number of male athletes that occupy that region as assigned by Strava: 85 – 94 |
| Male Count No Bday | Number of male athletes that occupy that region as assigned by Strava: Unspecified Birth Day |
| Female Count | Number of female athletes that occupy that region as assigned by Strava. |
| Female Count Under 25 | Number of female athletes that occupy that region as assigned by Strava: Under 25 |
| Female Count 25 - 34 | Number of female athletes that occupy that region as assigned by Strava: 25 – 34 |
| Female Count 35 - 44 | Number of female athletes that occupy that region as assigned by Strava: 35 – 44 |
| Female Count 45 - 54 | Number of female athletes that occupy that region as assigned by Strava: 45 – 54 |
| Female Count 55 - 64 | Number of female athletes that occupy that region as assigned by Strava: 55 – 64 |
| Female Count 65 - 74 | Number of female athletes that occupy that region as assigned by Strava: 65 – 74 |
| Female Count 75 - 84 | Number of female athletes that occupy that region as assigned by Strava: 75 – 84 |
| Female Count 85 - 94 | Number of female athletes that occupy that region as assigned by Strava: 85 – 94 |
| Female Count No Bday | Number of female athletes that occupy that region as assigned by Strava: Unspecified Birth Day |
| Blank Gender Count | Number athletes that occupy that region as assigned by Strava with no specified gender. |

# Glossary/Definitions

**Digitization Direction** – The direction the street network was drawn is important to understand when working with StravaMetro. That is how we assign ride counts to the different travel directions. An extra ‘R’ in the header name indicates the opposite of digitization direction. In other words, the direction it is referring to is from the end node to the start node.

**Rolled up** – View showing totals, sums and averages of the minute data for a given temporal range.

**Edge** – This is a piece of road geometry between nodes (intersections)

# Changes for This Release

No changes from the last release.

# Contact Information

Strava

35 S. Main Street

Hanover, NH

**GEO Team -**

Michael Horvath (Strava Co-Founder): [michael@strava.com](mailto:michael@strava.com)

Davis Kitchel (GEO Team Lead): [dkitchel@strava.com](mailto:Davis.Kitchel@strava.com)

Brian Riordan (GIS Lead): [brian@strava.com](mailto:Brian@strava.com)

# FAQs:

#### How do you locate commutes?

Commuter data is derived by 3 methods. The first is the commute flag that is native to the Strava experience. The second is an automated process that locates point to point cycling trips that are within duration and distance constraints. The third is fuzzy name matching from the activity titles.

#### What does time mean? And how can you have 0?

Time is the Median value of all recorded times for that edge. We do not use average time because there are instances where cyclists will stop on a piece of road to go into a store or talk with a friend while their GPS is still running.

A zero value happens when the edge is very small. For example, if the edge is less than 1 meter long and the cyclists is moving at 20mph then the GPS pulse will show no change for that piece of road. This means that from available time data, the cyclists is on both ends of the road at the same time.

#### What does the Strava athlete location mean?

Athletes locations are generated by a process of computing a locus based on the start points of current activities.

#### How well does Strava data represent the cyclists in my community?

The Strava community is made up of all types of cyclists and surveys have shown that most Strava cyclists do not refer to themselves as competitive. In fact, more than one-half of all rides on Strava in denser metro areas are commutes on average so Strava Metro data gives great insight into the needs of those riding for transportation purposes only. Furthermore, in metro areas, nearly everyone is a commuter – either commuting to work, or commuting to the ride they’ll be doing outside the city. Through our own analysis of the data, it is evident that cyclists of all types tend to use the same “best available” roads and paths while cycling in metro areas. Finally, by providing visibility into cycling activity at a high level of granularity, Strava Metro data enables users to analyze patterns by time of day, day of week, season and local geography. It’s simple to filter the data to show only commuter data as well.

#### Does Strava have enough data to provide a meaningful dataset?

Strava represents the most active athlete network in the world. The data set currently includes over 300 billion GPS points. In addition, over 2.5 million GPS-tracked activities, and growing, are uploaded to Strava every week from around the globe. These activities create billions of data points that, when aggregated, enable deep analysis and understanding of real-world cycling and running route preferences.

#### How much does it cost to license Strava Metro data for a city?

License fees are based on the number of Strava members in the requested geographic area and the time span of data required. Pricing is US $0.80 per distinct member in a twelve month period. If you are looking for a heat map image of Strava rides or runs in your area, you can get that free of charge at Strava Labs.

#### Can anyone be part of the Strava Metro dataset?

Anyone using Strava to publicly track their rides, runs and other fitness activities is anonymously contributing to the Strava Metro dataset with every upload. It’s a way for you to vote with your ride or run for better cycling and pedestrian infrastructure in your community.

#### How is Strava Metro data different from bike counters?

Strava Metro takes tracking and measurement of cycling activity into the digital age. It is a supplemental method for the traditional way of collecting data using human counters or electronic gates to survey key intersections within a community. The advantage is that Strava collects data down to the minute at every geographic point on a city grid 24/7.

#### What are some quick ways Strava Metro generates value to DOTs?

Strava Metro enables insight and analysis of bicycling routes and commute patterns around the world. Strava Metro data can be used retrospectively to empower advocacy organizations and government agencies to understand current cycling behavior in local communities and make better-informed decisions when planning, maintaining, and upgrading cycling infrastructure. Strava Metro data can also be used to evaluate efficacy of changes made to this infrastructure within a few weeks after implementation as opposed to waiting months or years to gather data.

#### Does Strava Metro require GIS software?

Strava Metro is designed for use by geographic information system (GIS) professionals familiar with GIS software and engaged in city planning. However, a high-resolution visualization of Strava activity data is available for use for free by cycling advocacy groups and the general public at Strava Labs.

#### How does Strava Metro respect the privacy of Strava members?

The data provided through Metro has been anonymized and aggregated to a linear map so that cycling activity cannot be associated with a specific member of Strava’s community. We are providing this information in anonymous aggregate form to help improve infrastructure and safety for cyclists, runners and pedestrians.

#### What other products is Strava Metro designing?

We see many potential products in the future for Strava Metro. In the near-term we are working on enhanced visualizations and tools for use with Strava Metro data by users without access to and experience in GIS software. We are also looking into providing tools to facilitate the analysis of the data so groups can use it as effectively as possible.

#### How can I help make my community better for alternative transportation?

If you are a cyclist or runner, use Strava to track your activities. In addition to the fun and motivation Strava provides, your cycling and running data will be in the Strava Metro system and can help inform alternative transportation systems in your community. And tell your friends who ride and run. Strava grows by word of mouth.

#### I am part of an advocacy group. How can I spread the word about Strava Metro?

Tell your DOT and city planning colleagues about Strava Metro. We have successfully partnered with several advocacy organizations to jointly approach the planning authorities in their area in presenting what Strava Metro.

#### What is the difference between the Strava heat map and Strava Metro?

The Strava heatmap available free of charge on Strava Labs is a visualization of a large collection of GPS points recorded by Strava members. Strava Metro is a product which contains the data behind the visualization. For example, a popular street for cyclists looks like a bright line in the Strava Heat Map. Strava Metro provides data about how many cyclists rode in which direction on that street minute-by-minute. The Strava heat map is a way to visualize the world of Strava. Strava Metro is for analysis and infrastructure planning.

#### I’m interested in local popular cycling routes. Is Strava Metro what I’m looking for?

No. You’re looking for Strava Routes. We’ve counted the “votes” of millions of runners and cyclists globally and made them available via a route creation and discovery tool.

# Appendix:

## Quick Start Guide

### Indexes

Shapefile indexes often become corrupt and need to be rebuilt. By indexing the ID fields, the joins between all the files will be much faster. In addition, make sure there is a spatial index on the street network.

### PostGIS Queries

The power of this data is at the database level. Being able to generate custom views of the data is key to being able to fully utilize the data. We provide a set of rolled-up views but ultimately these are not going to be the views that will work for all planning groups. Below are a few queries that can be used to build new rolled up views and tables in PostGres/PostGIS.

**Build Cycling Data for Times and Days Example:** This SQL builds a new table that can be exported as a new .dbf file. The key values that can modified here is the hour fieldby changing these values you can change to time range you are rolling up. The other value is at the endof the query ***where day > 0 and day < 32*** by moving the day range you can look at specific weeks or different groups of months. Note Postgres does not have a native Median calculation included. You might need to add this function: <https://wiki.postgresql.org/wiki/Aggregate_Median>

*Create table DevOrangeCountyStravaMetroNodes\_Year as*

*select edge\_id, sum(athlete\_count) as TRiderCnt,*

*sum(rev\_athlete\_count)as TRRiderCnt,*

*sum(activity\_count) as TRideCnt,*

*sum(rev\_activity\_count)as TRRideCnt,*

*sum(total\_activity\_count) as BikeCnt,*

*median(CASE when (activity\_count > 0) then activity\_time end)as BikeTime,*

*median(CASE when (rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeTime,*

*sum(commute\_count) as CommuteCnt, sum(CASE when (hour between 0 and 3) then athlete\_count end) as Rider\_0,*

*sum(CASE when (hour between 0 and 3) then rev\_athlete\_count end) as RRider\_0,*

*sum(CASE when (hour between 0 and 3) then activity\_count end) as Ride\_0,*

*sum(CASE when (hour between 0 and 3) then rev\_activity\_count end) as RRide\_0,*

*sum(CASE when (hour between 0 and 3) then total\_activity\_count end) as BikeCnt\_0,*

*median(CASE when (hour between 0 and 3 and activity\_count > 0) then activity\_time end)as BikeT\_0,*

*median(CASE when (hour between 0 and 3 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_0,*

*sum(CASE when (hour between 0 and 3) then commute\_count end) as Commute\_0,*

*sum(CASE when (hour between 4 and 5) then athlete\_count end) as Rider\_1,*

*sum(CASE when (hour between 4 and 5) then rev\_athlete\_count end) as RRider\_1,*

*sum(CASE when (hour between 4 and 5) then activity\_count end) as Ride\_1,*

*sum(CASE when (hour between 4 and 5) then rev\_activity\_count end) as RRide\_1,*

*sum(CASE when (hour between 4 and 5) then total\_activity\_count end) as BikeCnt\_1,*

*median(CASE when (hour between 4 and 5 and activity\_count > 0) then activity\_time end)as BikeT\_1,*

*median(CASE when (hour between 4 and 5 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_1,*

*sum(CASE when (hour between 4 and 5) then commute\_count end) as Commute\_1,*

*sum(CASE when (hour between 6 and 8) then athlete\_count end) as Rider\_2,*

*sum(CASE when (hour between 6 and 8) then rev\_athlete\_count end) as RRider\_2,*

*sum(CASE when (hour between 6 and 8) then activity\_count end) as Ride\_2,*

*sum(CASE when (hour between 6 and 8) then rev\_activity\_count end) as RRide\_2,*

*sum(CASE when (hour between 6 and 8) then total\_activity\_count end) as BikeCnt\_2*

*,median(CASE when (hour between 6 and 8 and activity\_count > 0) then activity\_time end)as BikeT\_2,*

*median(CASE when (hour between 6 and 8 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_2,*

*sum(CASE when (hour between 6 and 8) then commute\_count end) as Commute\_2,*

*sum(CASE when (hour between 9 and 14) then athlete\_count end) asRider\_3,*

*sum(CASE when (hour between 9 and 14) then rev\_athlete\_count end) as RRider\_3,*

*sum(CASE when (hour between 9 and 14) then activity\_count end) as Ride\_3,*

*sum(CASE when (hour between 9 and 14) then rev\_activity\_count end) as RRide\_3,*

*sum(CASE when (hour between 9 and 14) then total\_activity\_count end) as BikeCnt\_3,*

*median(CASE when (hour between 9 and 14 and activity\_count > 0) then activity\_time end)as BikeT\_3,*

*median(CASE when (hour between 9 and 14 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_3,*

*sum(CASE when (hour between 9 and 14) then commute\_count end) as Commute\_3,*

*sum(CASE when (hour between 15 and 17) then athlete\_count end) as Rider\_4,*

*sum(CASE when (hour between 15 and 17) then rev\_athlete\_count end) as RRider\_4,*

*sum(CASE when (hour between 15 and 17) then activity\_count end) as Ride\_4,*

*sum(CASE when (hour between 15 and 17) then rev\_activity\_count end) as RRide\_4,*

*sum(CASE when (hour between 15 and 17) then total\_activity\_count end) as BikeCnt\_4,*

*median(CASE when (hour between 15 and 17 and activity\_count > 0) then activity\_time end)as BikeT\_4,*

*median(CASE when (hour between 15 and 17 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_4,*

*sum(CASE when (hour between 15 and 17) then commute\_count end) as Commute\_4,*

*sum(CASE when (hour between 18 and 19) then athlete\_count end) as Rider\_5,*

*sum(CASE when (hour between 18 and 19) then rev\_athlete\_count end) as RRider\_5,*

*sum(CASE when (hour between 18 and 19) then activity\_count end) as Ride\_5,*

*sum(CASE when (hour between 18 and 19) then rev\_activity\_count end) as RRide\_5,*

*sum(CASE when (hour between 18 and 19) then total\_activity\_count end) as BikeCnt\_5,*

*median(CASE when (hour between 18 and 19 and activity\_count > 0) then activity\_time end)as BikeT\_5,*

*median(CASE when (hour between 18 and 19 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_5,*

*sum(CASE when (hour between 18 and 19) then commute\_count end) as Commute\_5,*

*sum(CASE when (hour between 20 and 23) then athlete\_count end) as Rider\_6,*

*sum(CASE when (hour between 20 and 23) then rev\_athlete\_count end) as RRider\_6,*

*sum(CASE when (hour between 20 and 23) then activity\_count end) as Ride\_6,*

*sum(CASE when (hour between 20 and 23) then rev\_activity\_count end) as RRide\_6,*

*sum(CASE when (hour between 20 and 23) then total\_activity\_count end) as BikeCnt\_6,*

*median(CASE when (hour between 20 and 23 and activity\_count > 0) then activity\_time end)as BikeT\_6,*

*median(CASE when (hour between 20 and 23 and rev\_activity\_count > 0) then rev\_activity\_time end)as RBikeT\_6,*

*sum(CASE when (hour between 20 and 23) then commute\_count end) as Commute\_6*

*from orange\_county\_edges\_ride\_data*

*where day > 334 group by edge\_id*