Better Data for Better Cities

Brian Riordan - CS Lead

metro.strava.com

Agenda

Introductions: 10:15am

Overview of Strava & Metro: 10:15am – 10:45am

Overview of Product: 10:45am – 11:15am

Removing Noise: 11:15am – 11:30am

Comfort break: 11:30am – 12:00pm

Correlation (Crash Points, Counters, Speed): 12:00pm – 12:20pm

Finding Missing Geo:12:30pm – 12:40pm

Delta Analysis: 12:40pm – 1:00pm

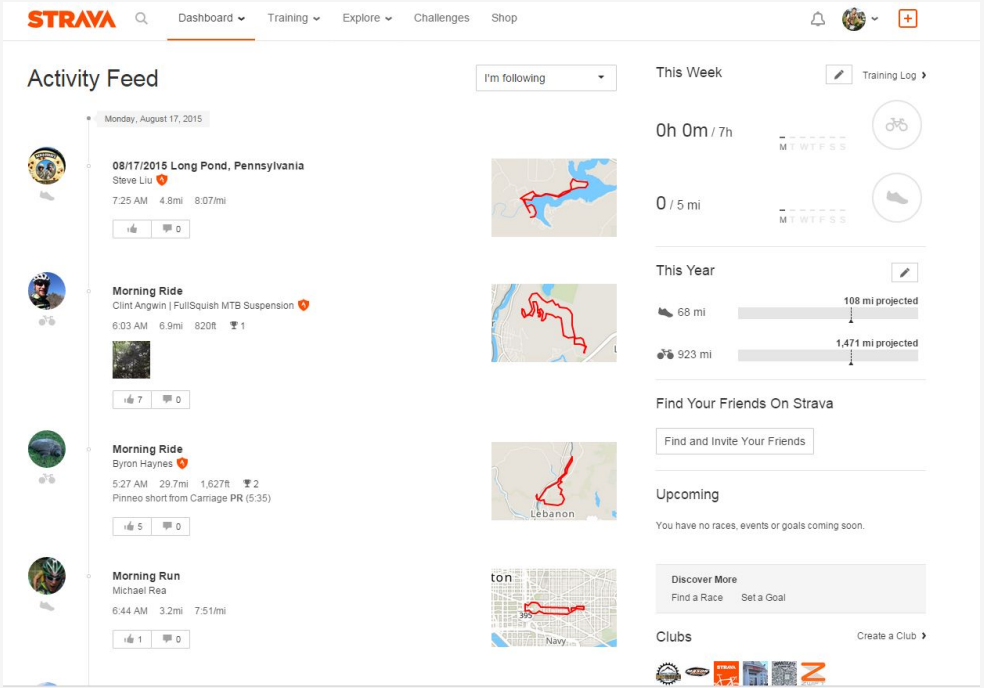
Discussion around potential urban research uses for Strava data: 1:00pm – 1:30pm

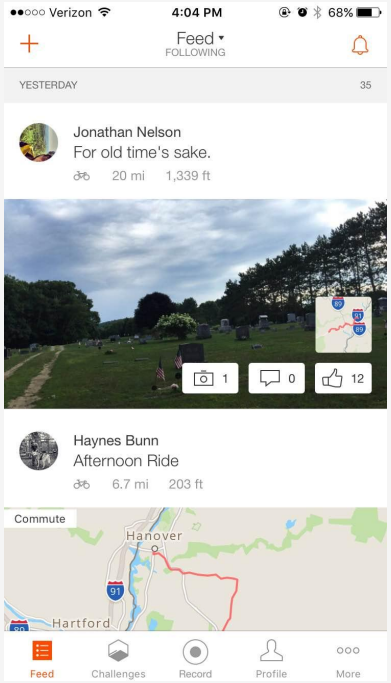
Lunch and networking

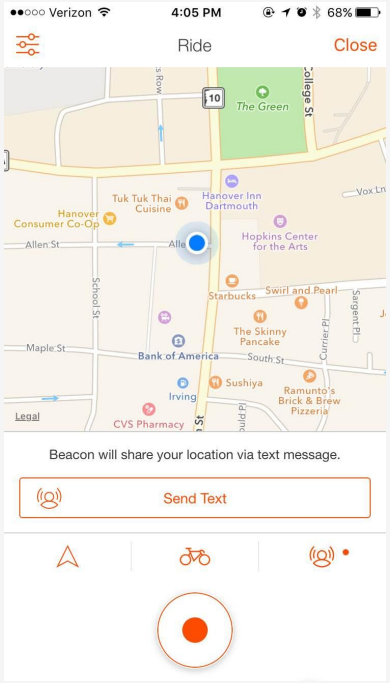

What and Who is Strava?

What is Strava?

The social network for cyclists and runners.







What is Strava Metro?

Data-Driven Bike and Pedestrian Planning

Aggregated, anonymized activity data from Strava's millions of users

Analyze popular or avoided routes, peak commute times, intersection wait times, and origin/destination zones

Processed for compatibility with geographic information system (GIS) environments

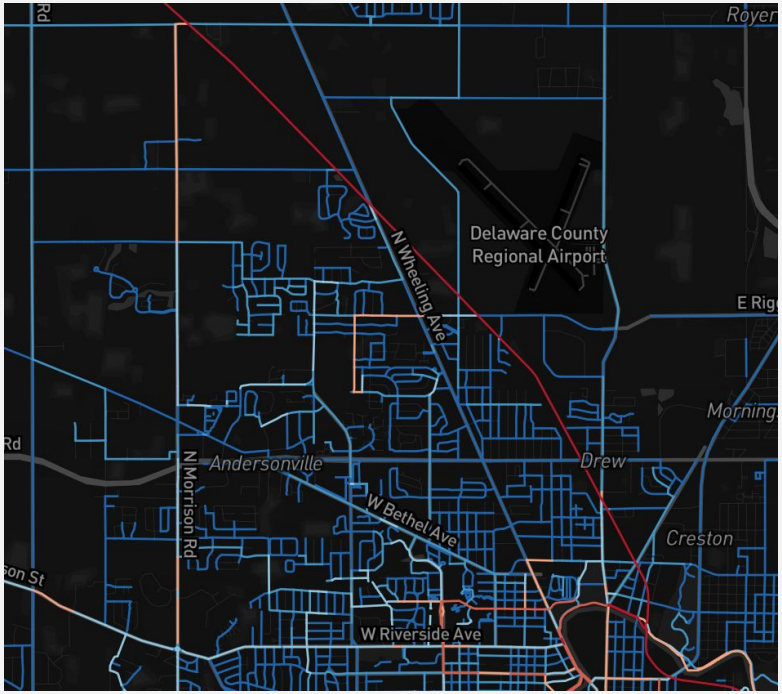
Strava Metro's Mission Statement

To produce state-of-the-art spatial data products and services to make cycling, running, and walking in cities better.

Why Build Strava Metro?

Data-Driven Bicycle and Pedestrian Planning

- Global need for consistent cycling data
- Continues to serve the Strava user
- Further bonds the cycling and pedestrian community
- It’s the right thing to do



How Do We Protect Users' Privacy?

User privacy as the foremost concern

Focused on streets, not individuals

No way to bring a Metro record back to Strava

Opt out switches on Strava (less than 0.01%)

Just use the heatmap?

Not so fast!

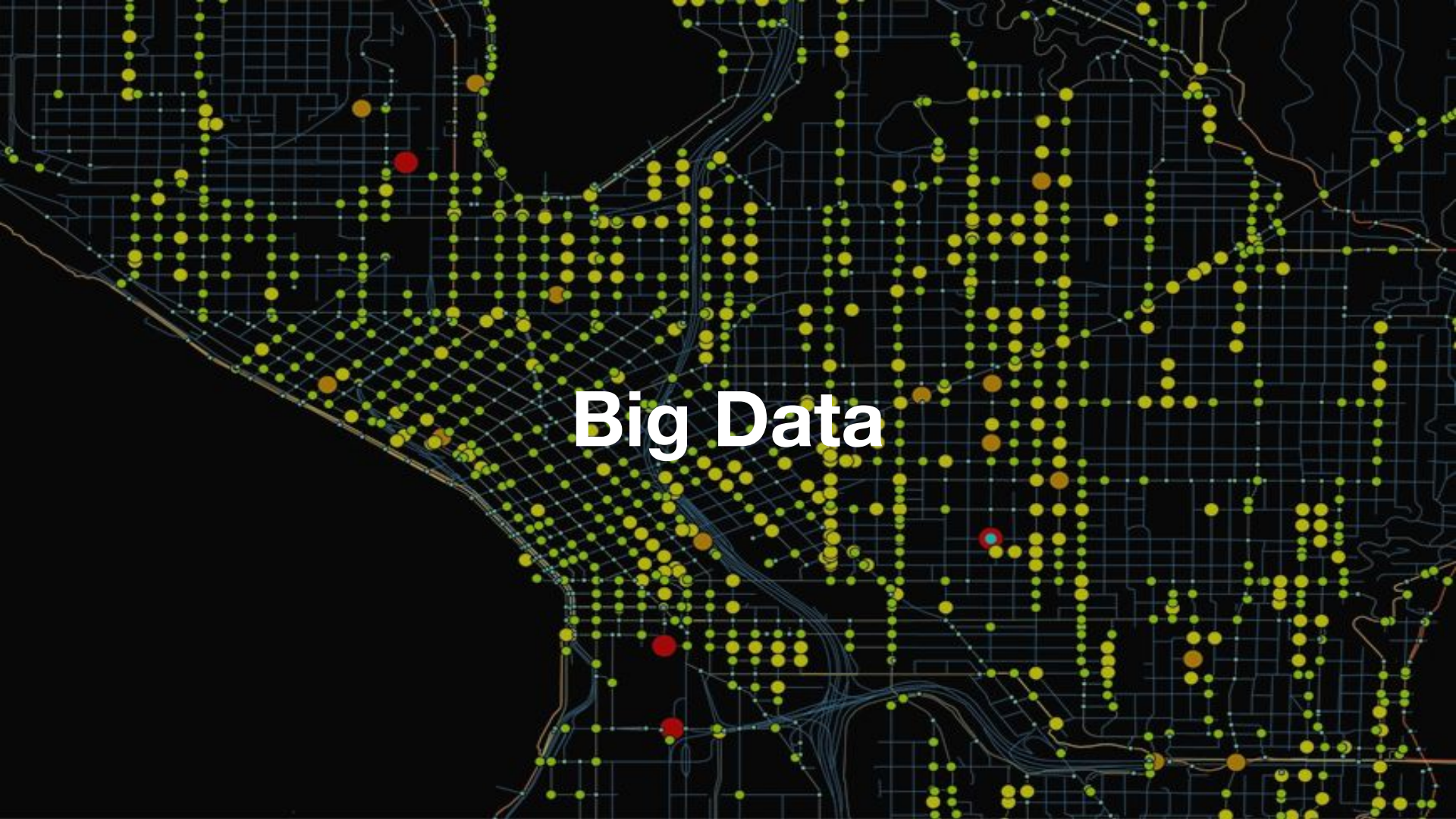
- No temporal scale
- Point Saturation not Use Saturation
- Large cropping of ride start and ends
- Mix of ride types

What is it good for?

- Showing that people ride bikes
- Starting dialogues with city councils
- Keeping track of where you rode this year
- Editing your basemap / finding missing geometry

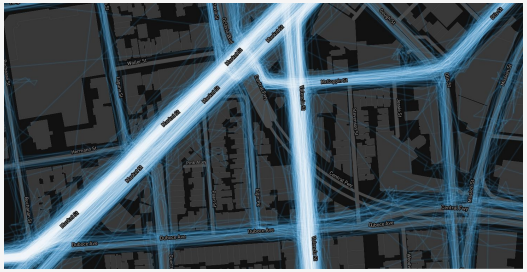
Data Restrictions & Use

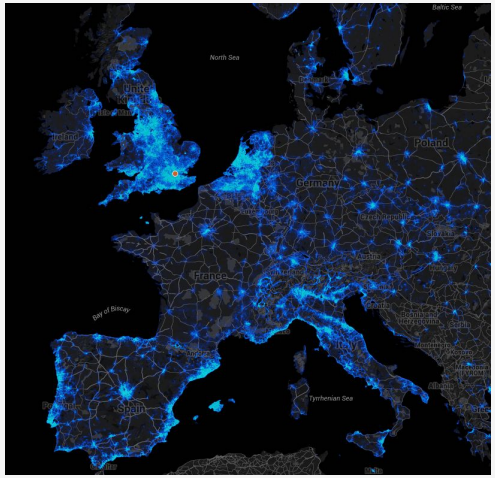
- Contract is built to allow collaboration with the data throughout the county
- Data can be utilized by engineering firms, universities, and other contracted groups
- The data cannot be posted up on a FTP site for public download
- Aggregated results are fine for public display and download

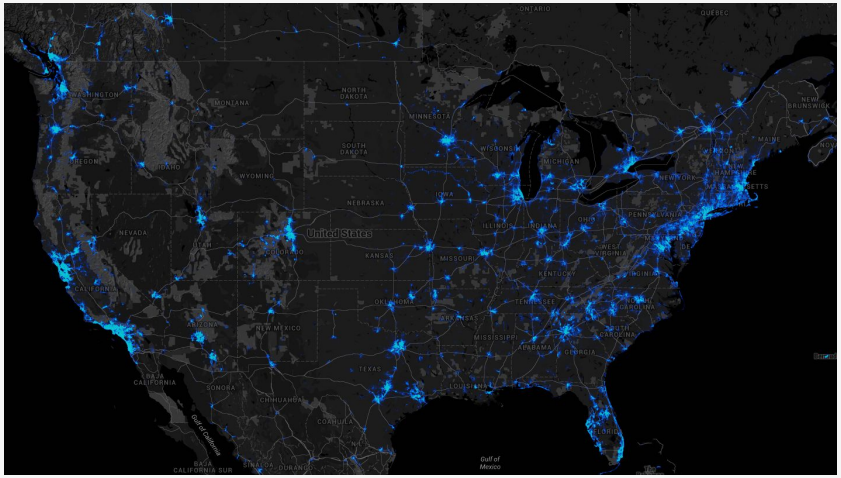

Big Data

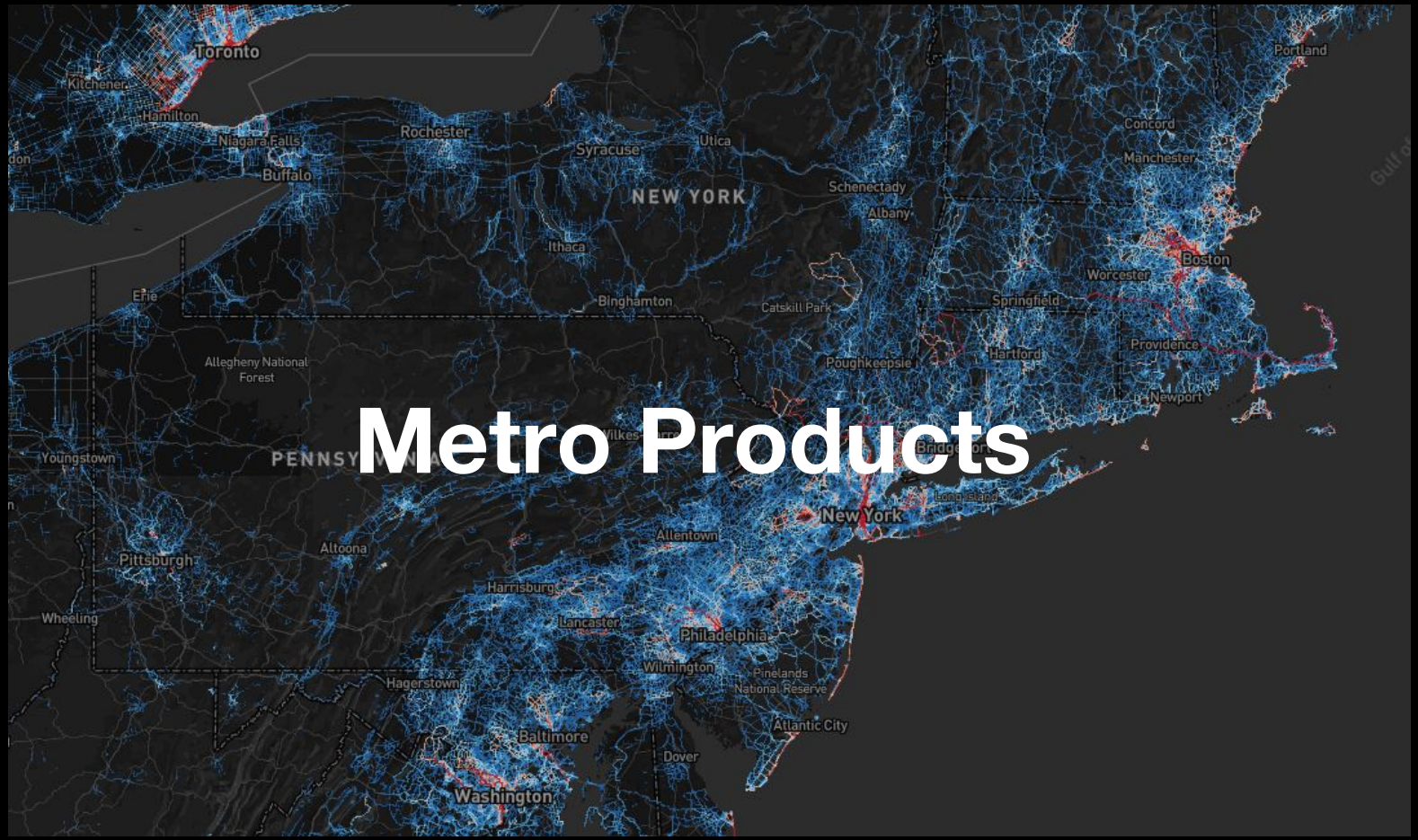
Strava By The Numbers

- Over 8 million activities uploaded per week
- Tens of Millions of Active Users
- 1 Trillion+ second-by-second GPS points globally



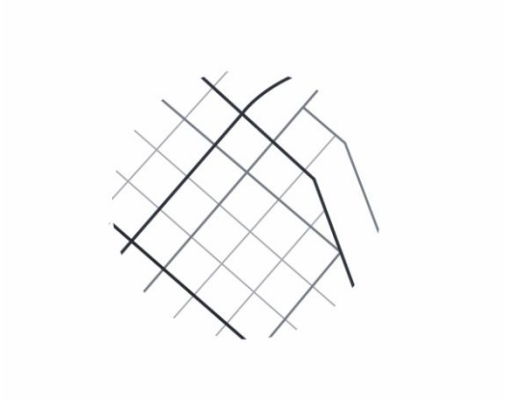






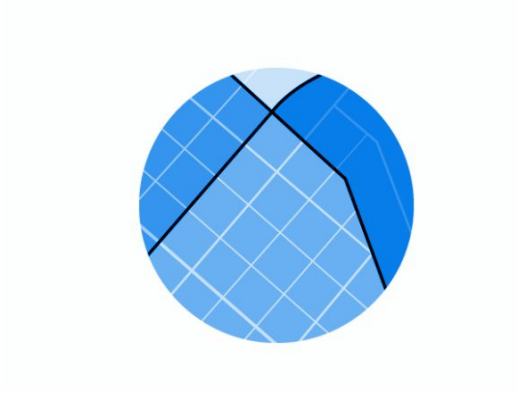
Metro Products

Strava Metro Data



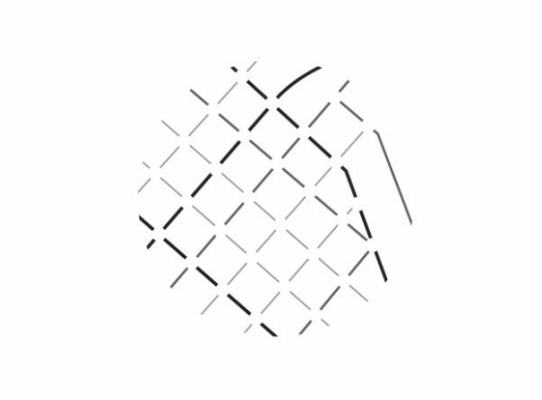
Streets

Minute-by-minute counts across your entire network



Origin / Destination

Understand activity starting and ending points, by region



Intersections

Activity counts and wait times at every intersection

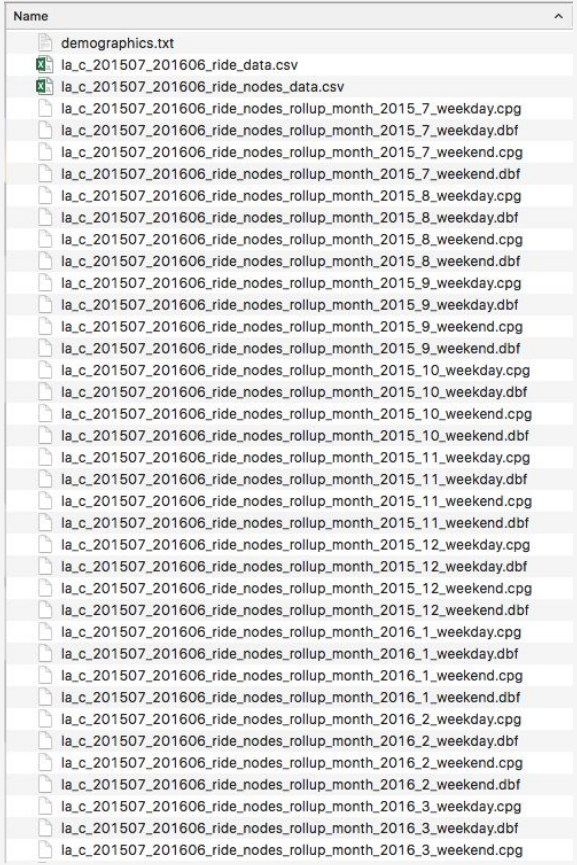
Email Delivery

Contains key information about the data format, location, and coverage

- Time frames and roll-up
    - Below are the date and hourly ranges for the AM/PM data used when generating this product.
        - On-Season: April - October
            - Early AM hours: 12am – 5:59am -- > Labeled as *_0*
            - AM Peak hours: 5am – 9:59am -- > Labeled as *_1*
            - Labeled as *_3*
                - Late Evening Hours: 8pm – 11:59pm
- FTP file location
    - https://stravametro.exavault.com/files#/%2FNA%2FUS%2FCA%2FLA%2FLACounty
- Data Vis
    - http://metro-static.strava.com/dataViewV1/losangeles/losangeles16Q2.html
- User Manual
- Date Range: July 1, 2015 - June 30, 2016

Metro File Structure & Naming Conventions

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



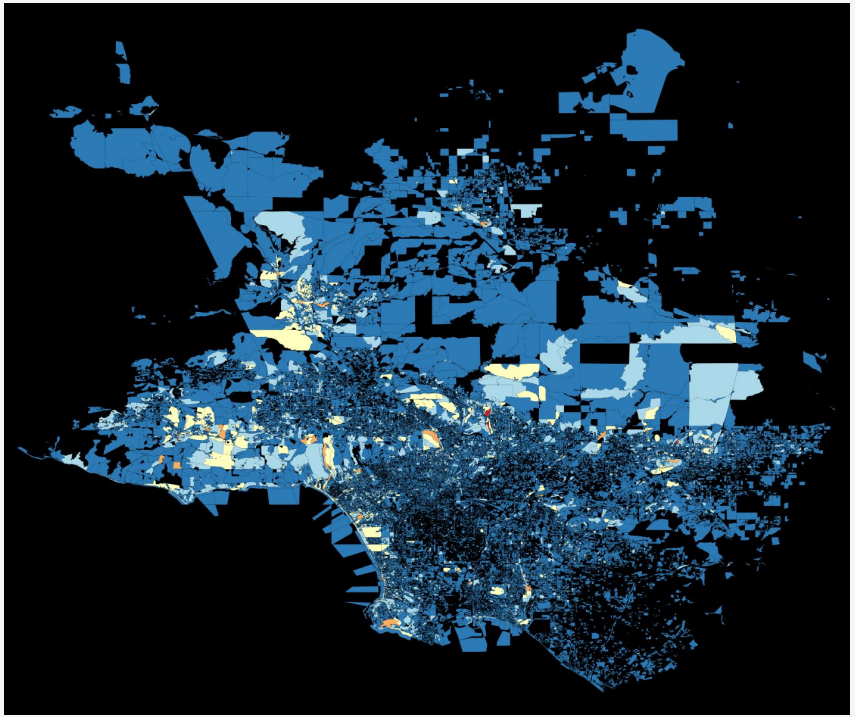
Street Level Data

- Minute by minute tabular data of cycling behavior
- Preferred route direction
- Unique bike trips
- Unique user counts
- Trip purpose: commute and recreation
- Time/Speed (seconds)

	EDGE_ID	ATHCNT_0 ▼	RATHCNT_0	ACTCNT_0	RACTCNT_0	TATHCNT_0	TACTCNT_0	ACTTIME_0	RACTTIME_0	CMTCNT_0	RCMTCNT_0	ATHCNT_1
406645	139288	367	8	421	28	373	449	8.9805374620	17.45127726…	256	5	219
406624	139267	363	10	418	25	368	443	4.6754569378	3.5172259635	247	8	213
406647	139290	348	12	392	32	356	424	18.70039557…	13.06984566…	247	12	185
406646	139289	346	13	397	27	353	424	13.79808979…	11.1145850376	251	8	189
396263	127740	345	0	512	0	345	512	15.99356673…	*NULL*	294	0	681
406648	139291	343	36	394	53	348	411	0.9767840987	0.9307139932	241	22	224
406633	139276	340	14	394	41	353	435	7.0942751252	6.3904346934	249	27	191
403775	136242	334	0	487	0	334	487	12.38249701…	*NULL*	282	0	654

Origin/Destination Polygons

- Polygonal starting and ending points of trips
- Reported by the minute
- Trip purpose flag
- Array of intersected polygon IDs

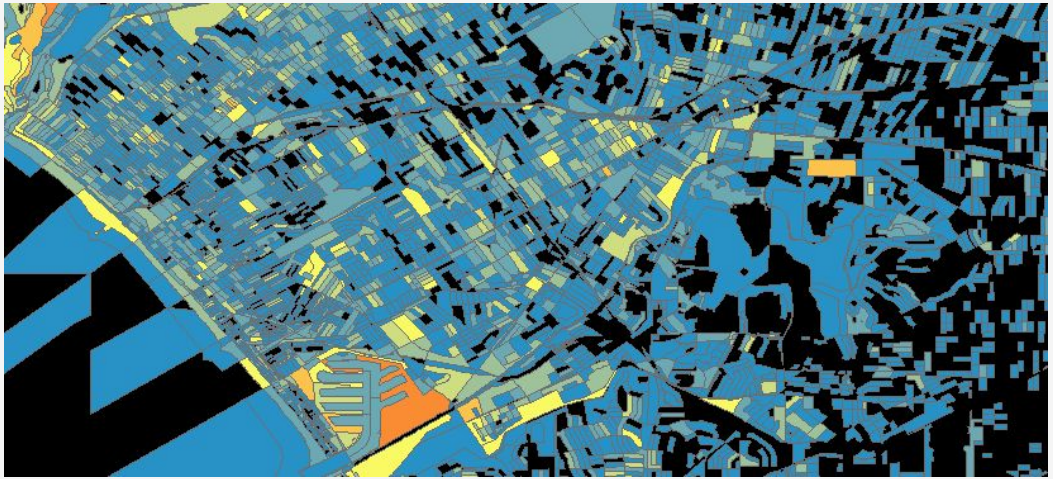


	polygon_id	year	day	hour	minute	commute	dest_polygon_id	intersected_polygons
140549	14788	2015	248	15	6	1	104454	{99989,104454,136853,153895,8204,11743,14788,17975,20130,20782,33169,50832,51824,82805,89430,149026}
841131	147589	2016	72	10	57	1	118584	{99988,140404,147589,118596,120849,1825,2494,35351,93492,100816,118444,140412,150127,41753,118584,130523,78910,2532,2450,138086}
838300	147589	2016	79	12	53	1	118584	{99988,140404,147589,118596,120849,1825,2494,2492,35351,93492,100816,118444,140412,150127,152854,41753,118584,78910,2450,138086}
1023200	128227	2016	159	19	18	0	128227	{99988,120849,128227,140404,147589,2494,2492,35351,118444,150127,138086,78910,93492,2532,2450,118596}
535232	147589	2015	362	10	38	0	78249	{99988,104120,140404,147589,138086}
786902	59374	2016	17	15	29	0	59374	{99988,104120,140404,118596,120849,2494,93492,118444,78910,2532,73559,34956,2450,138086,132970}
1195688	65995	2016	158	11	52	1	73185	{9998,12896,15153,15679,17276,17683,18517,20182,20766,21544,25152,32392,35076,36474,39661,41154,42365,53611,57392,58010,58098,65995,66293,67603,68036,73185,73718,76682,77740,78798,79405,8186...

OD Data

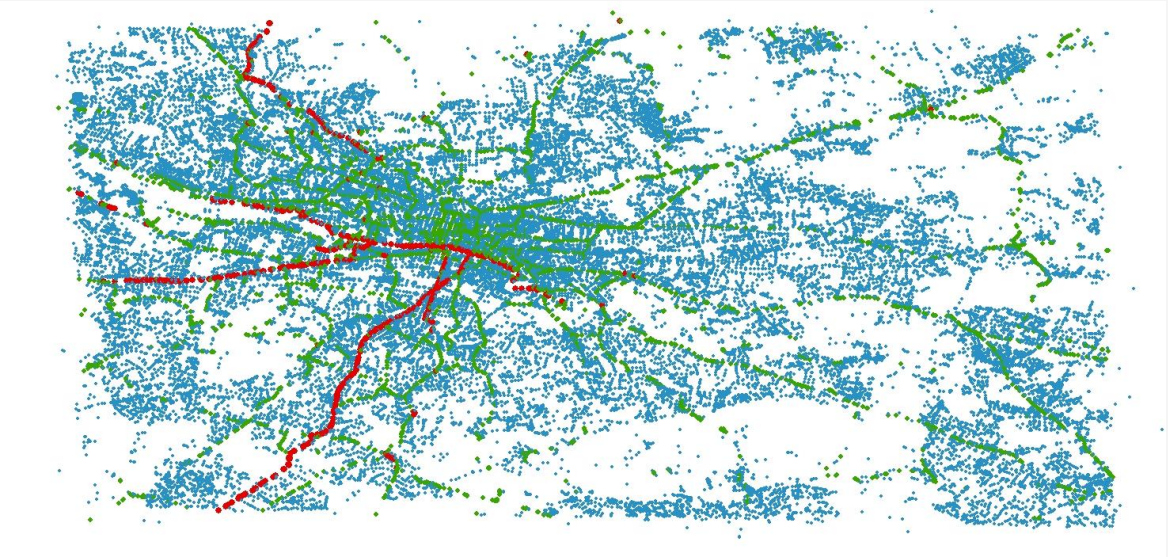
Use the starting and ending points to locate key zones

- Select all polygons that start in the morning
  - "HOUR" < 10 and "HOUR" > 4
- Summarize and export
- Join to polygon table
- Display



Intersections

- Wait times at intersection
- Congregation of users at intersections
- Minute by minute with purpose flag



	NODE_ID	ATHCNT_0	ACTCNT_0	PCT_0_0	PCT_025_0	PCT_05_0	PCT_075_0	PCT_1_0	CMTCNT_0	ATHCNT_1	ACTCNT_1	PCT_0_1	PCT_025_1	PCT_05_1	PCT_075_1	PCT_1_1	CMTCNT_1	ATHCNT_2	ACTCNT_2	PCT_0_2	PCT_025_2	PCT_05_2	PCT_075_2	PCT_1_2
192863	148229	443	728	0	2	4	6	244	500	598	1892	0	3	5	7	67	714	845	2311	0	3	5	8	84
192682	148200	437	720	0	2	4	6	244	495	593	1867	0	3	5	7	66	695	844	2309	0	3	5	8	86
202203	151191	434	679	0	3	5	8	69	458	684	2137	0	2	4	6	137	696	937	2854	0	2	4	6	146
4122	99930	433	623	0	2	4	5	55	395	1009	5462	0	1	3	5	998	2474	978	2393	0	1	3	6	113
209678	153424	431	855	0	3	4	12	2575	541	863	3716	0	2	4	7	167	2049	1236	4084	0	2	4	6	158
192862	148228	430	709	0	2	4	6	241	484	590	1872	0	3	4	7	68	709	829	2264	0	3	4	7	81
316067	79319	430	699	0	1	2	3	156	397	1166	7252	0	1	2	3	101	3445	1262	3282	0	1	2	4	108
4263	99948	429	629	0	2	3	4	55	395	1056	5744	0	2	3	5	804	2540	1031	2530	0	2	3	6	112

Demographics Data

- Rolled-up counts of users in the data files
- Breakdown of age and gender from users

```
Metro Demographics
Date Run: 2016-07-30 09:21:00 +0000

Athlete ID Count: 10503
Activity Count: 175410
Average Distance: 23376.521919394116
Median Distance: 14421
Average Time: 4742.8601873617256637
Median Time: 2773
Male Count: 8612
Male Count Under 25: 539
Male Count 25 - 34: 1543
Male Count 35 - 44: 2275
Male Count 45 - 54: 1674
Male Count 55 - 64: 377
Male Count 65 - 74: 61
Male Count 75 - 84: 3
Male Count 85 - 94: 1
Male Count No Bday: 2136
Female Count: 1255
Female Count Under 25: 114
Female Count 25 - 34: 310
Female Count 35 - 44: 280
Female Count 45 - 54: 167
Female Count 55 - 64: 34
Female Count 65 - 74: 2
Female Count 75 - 84: 0
Female Count 85 - 94: 0
Female Count No Bday: 348
Blank Gender Count: 0
Average Uploads: 210.652
Commute Counts: 116136
```

Strava Metro Data - Commutes

Commutes are the #1 requested data feature in Strava Metro.
75% of all Strava users upload Commute data.

Activities in urban areas are commutes 60% - 40% of the time (High of 90% in London).

Commutes and recreation rides in urban areas have very high correlation in route choice.

Commutes > 6 miles have the same street choices as < 6 miles they are just longer trips.

Strava Metro Data - Commutes Cont’

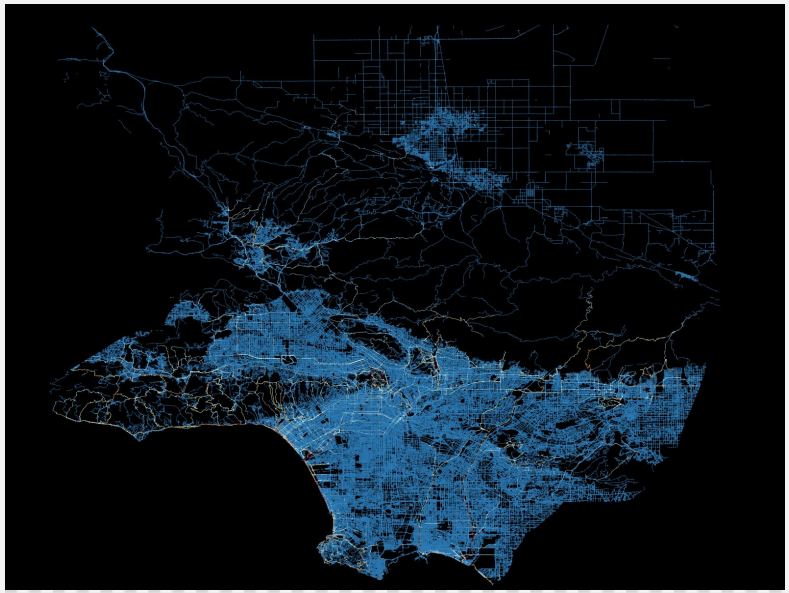
Users tend to not select the commute flag on Strava.com

3 ways to detect commute activities:
Activities tagged as a commute on Strava.com

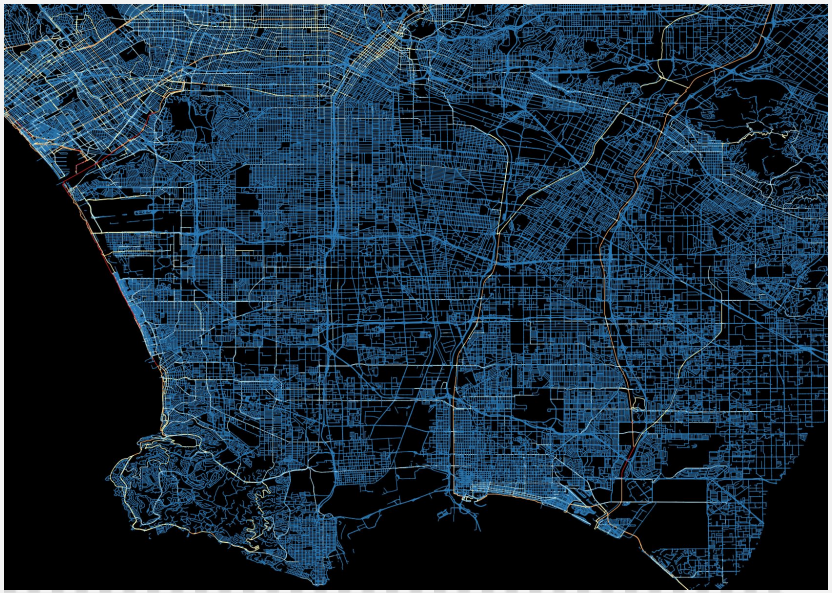
Activities with key words in the titles “To Work”, “Commute To Store”

Starting and Ending Points more than 1km apart within a distance and time threshold *(This can be user defined but we find 30 miles and 90 minutes to be a good top threshold)*

Macro & Micro Levels of Spatial and Temporal Details



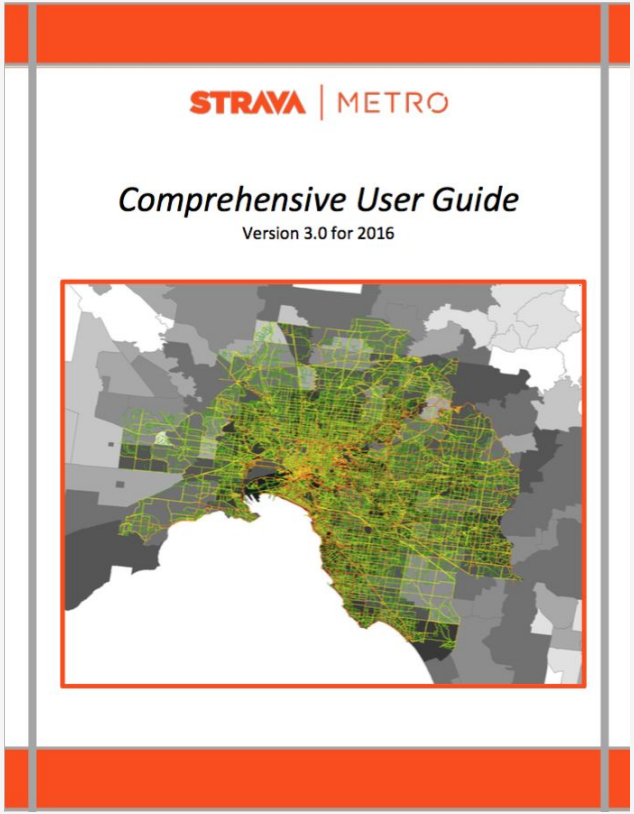
LA Full Core Routes



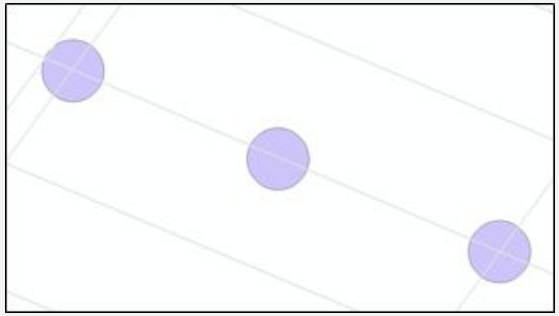
Key Commute Routes and Corridors

Working with the Rollups & Data Structures

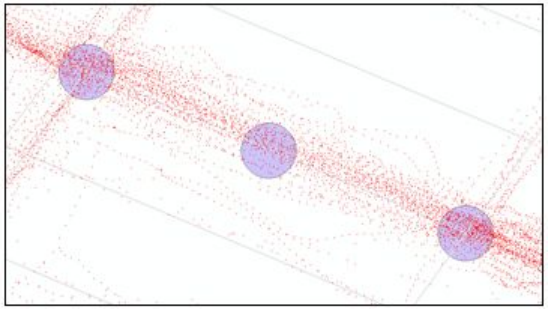
Refer to the User Guide for naming conventions and field definitions



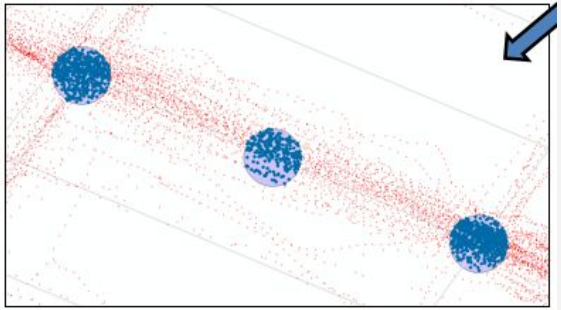
Metro Matching Process



Node structure w/ false node



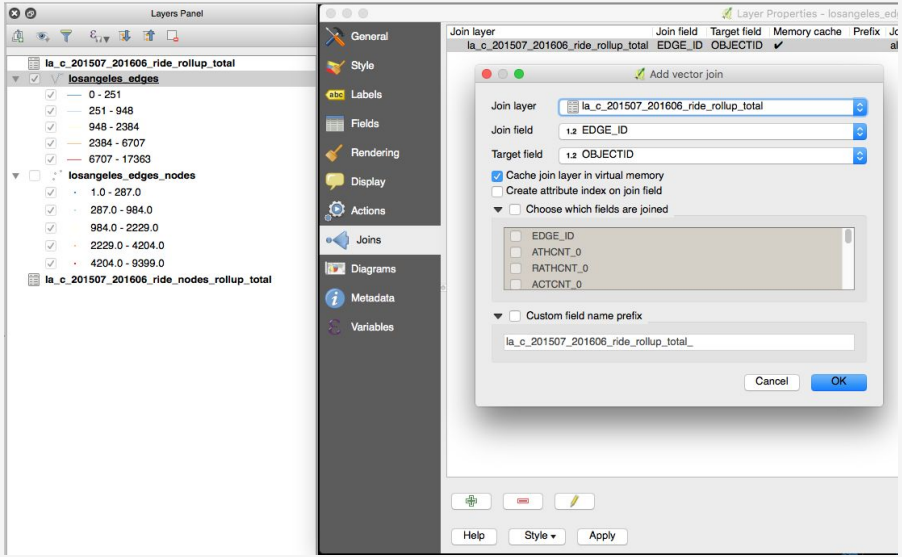
Overlay points

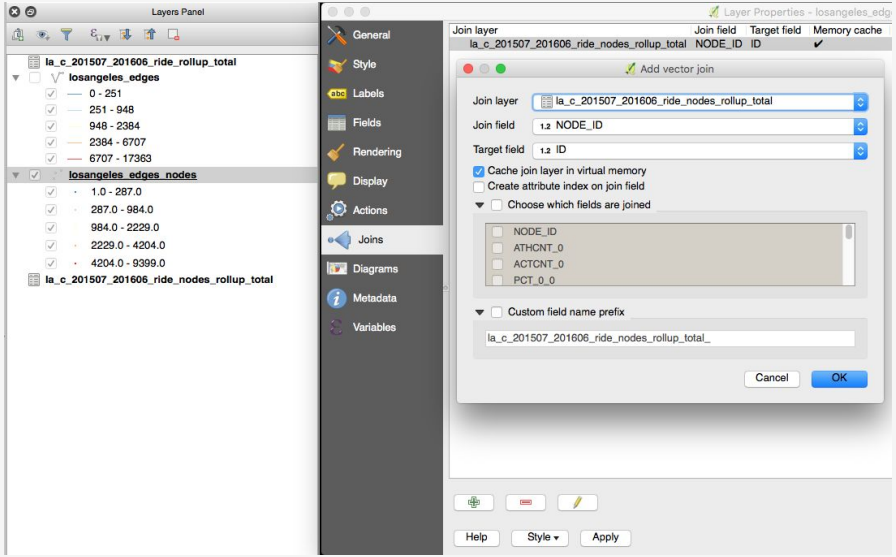


Intersection of points and nodes

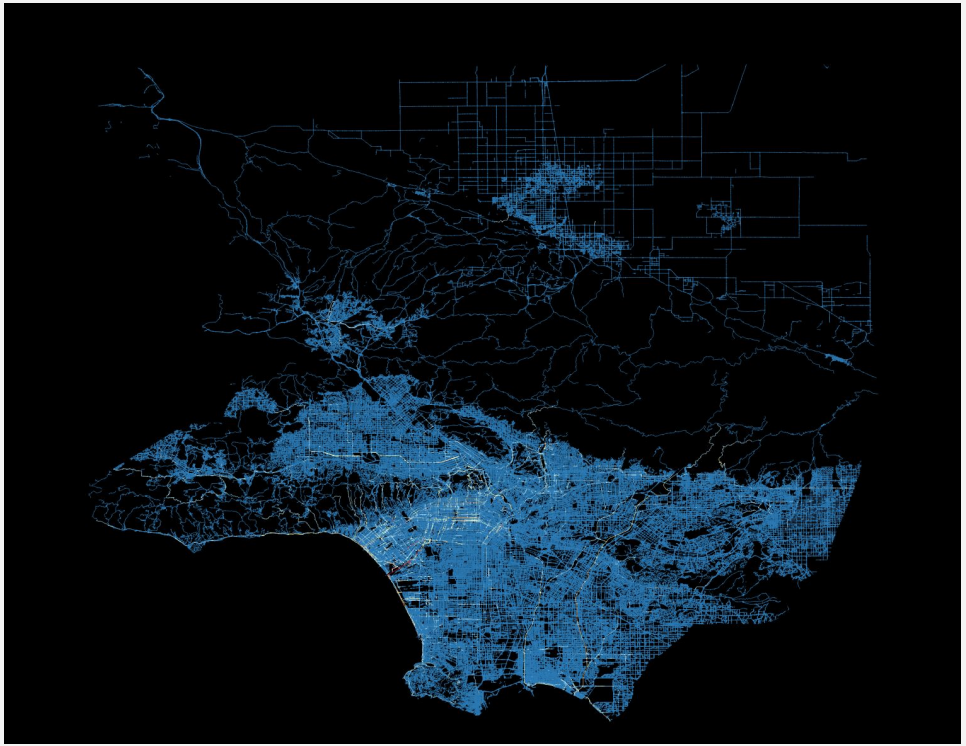
Working with Joins

Entire system is built on joining the data files back to the Node, Edges, and OD Spatial layers



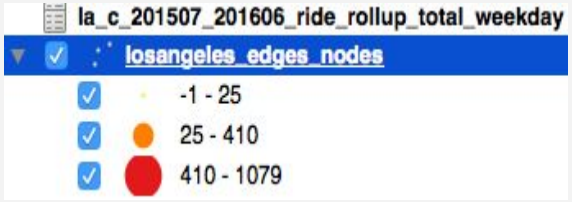


Working with Edge Data

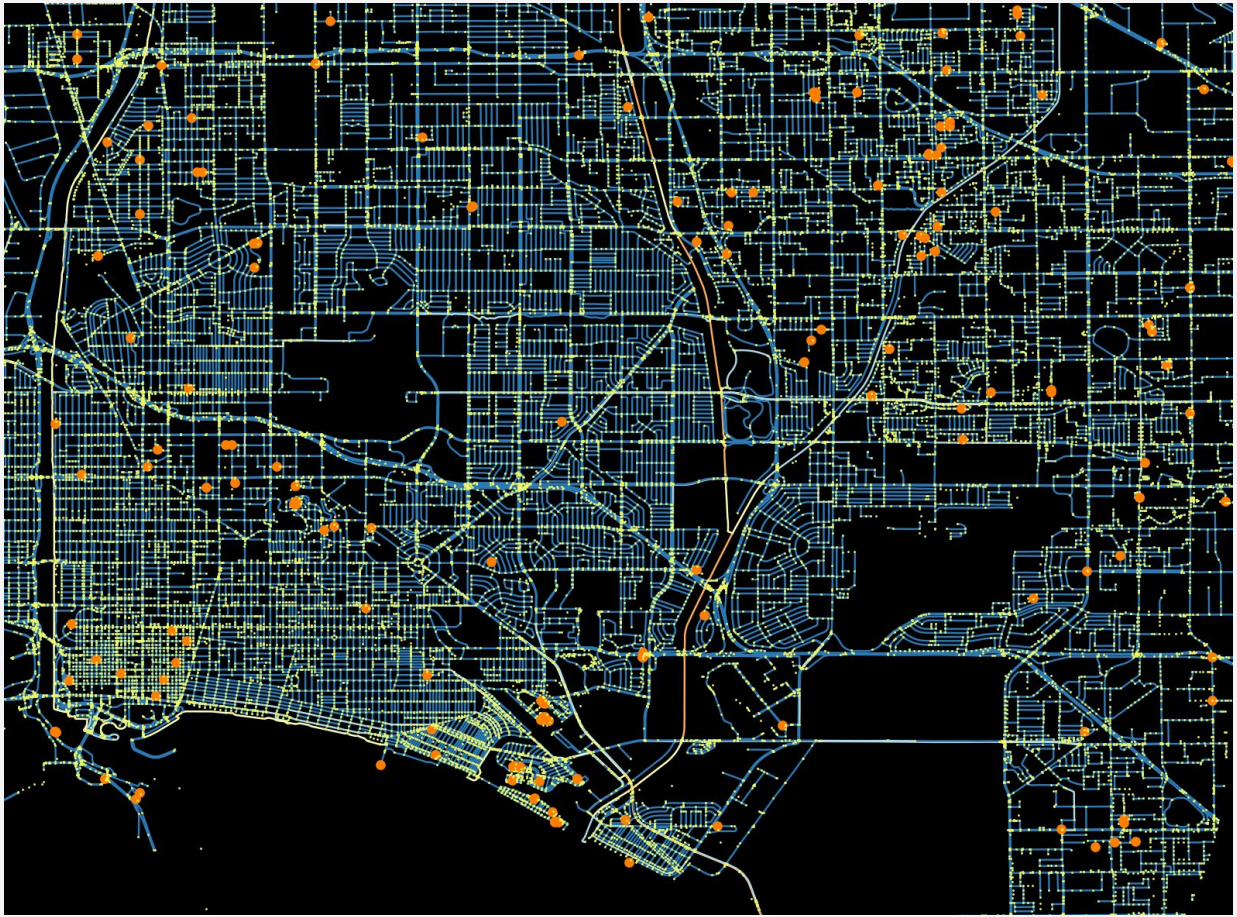


	EDGE_ID	ATHCNT_0 ▼	RATHCNT_0	ACTCNT_0	RACTCNT_0	TATHCNT_0	TACTCNT_0	ACTTIME_0	RACTTIME_0	CMTCNT_0	RCMTCNT_0	ATHCNT_1	RATHCNT_1	ACTCNT_1	RACTCNT_1	TATHCNT_1	TACTCNT_1	ACTTIME_1	RACTTIME_1
16733	1649	98	87	1022	680	120	1702	14.29631829…	12.44831730…	97	110	1097	608	4841	1872	1278	6713	14.78032182…	12.92645410…
628	211	91	78	1001	467	112	1459	4.0861936660	4.0794579934	109	62	1001	535	4156	1516	1171	5642	4.1512001191	4.1510622487
3340	263	90	97	598	1067	115	1484	1.8674234841	1.8781990943	67	117	699	980	2325	4113	1158	5672	1.8965579933	1.9013753879
133190	160744	89	62	276	264	127	540	6.1568697840	6.5354635586	112	180	1825	1432	7326	7397	2471	14723	6.4943662965	6.7710672707
629	212	87	75	757	573	104	1267	2.3056942429	2.3902203700	93	55	936	636	3518	1883	1132	5010	2.3577182890	2.4233445049
2087	214	87	72	799	554	102	1353	26.0311187449	21.85510567…	87	52	1022	474	3961	1585	1172	5546	24.50270776…	22.40810721…
2727	223	87	61	712	584	98	1296	6.5035315734	7.6077083843	108	49	867	652	3876	1763	1145	5639	6.1952490827	7.7455562325
2086	213	85	69	646	632	100	1278	35.30838540…	34.49075711…	77	75	938	504	3540	1521	1113	5061	35.31599401…	34.79528824…

Working with Nodes Data



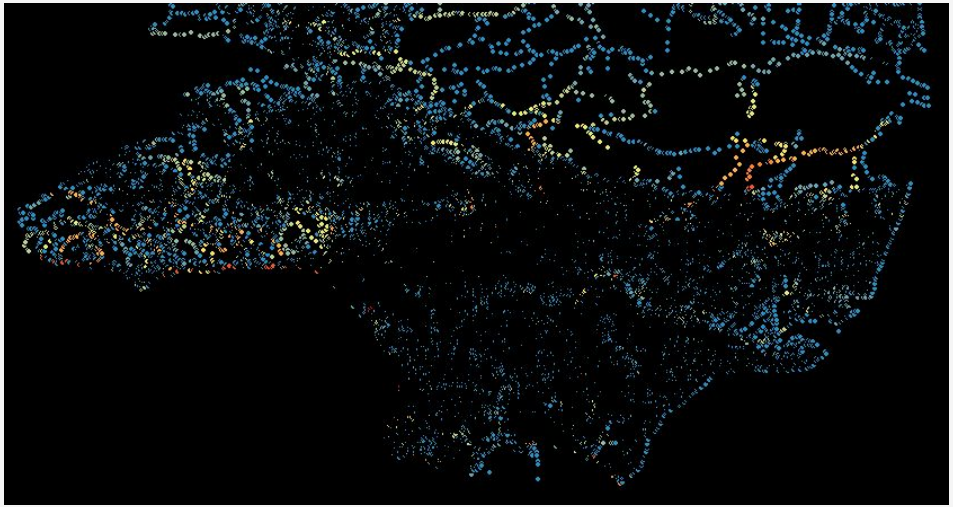
	EDGE_ID	ATHCNT_0 ▼	RATHCNT_0	ACTCNT_0	RACTCNT_0	TATHCNT_0
16733	1649	98	87	1022	680	120
628	211	91	78	1001	467	112
3340	263	90	97	598	1067	115
133190	160744	89	62	276	264	127
629	212	87	75	757	573	104
2087	214	87	72	799	554	102
2727	223	87	61	712	584	98
2086	213	85	69	646	632	100



Node Data

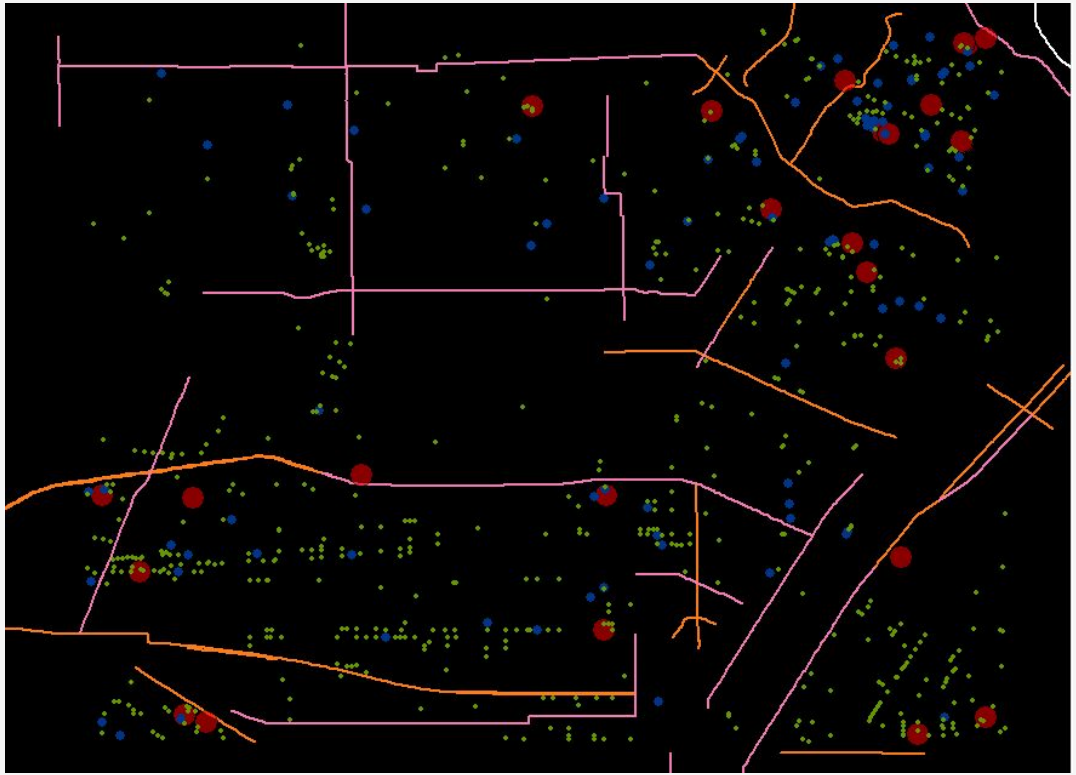
Use the intersections to locate slow down times

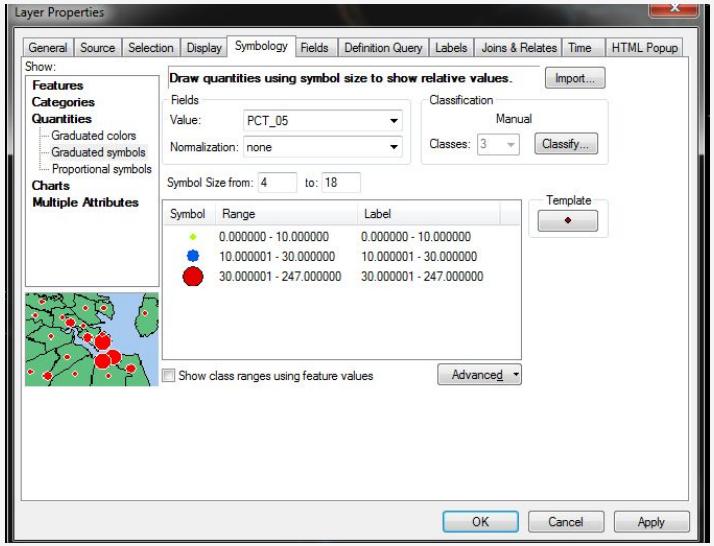
- Join ID --> node_id
- Sort by total % .75
- Any Noise? (5914 seconds....)
    - Remove all values with aths < 5

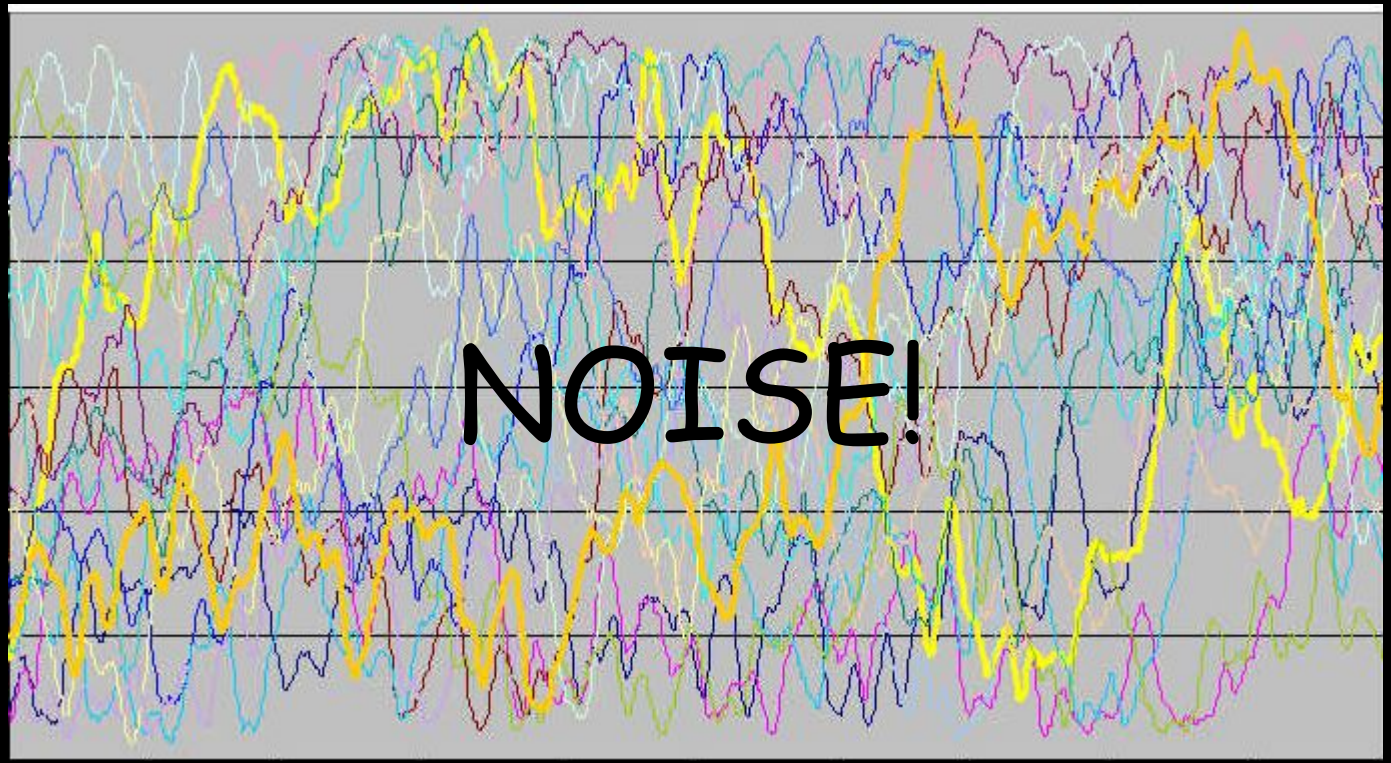


Node Data

Use the intersections to locate slow down times

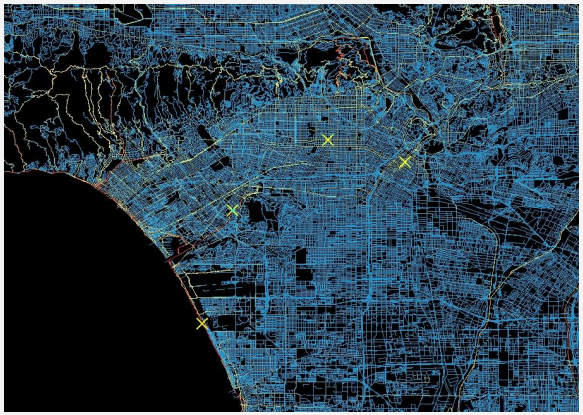




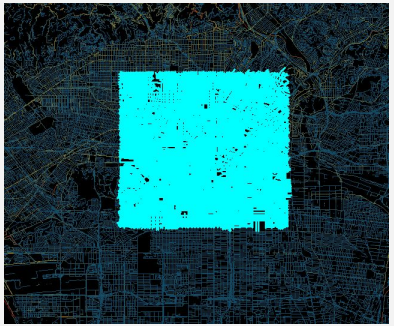

NOISE!

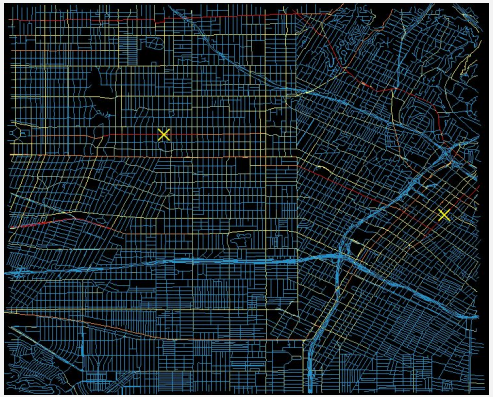
Strava Metro: Data Clean-up

Remove the noise!



1. Join the edges total to edges
    a. Symbolize by TATHCNT
    b. Lots to get lost in...
2. Select a subset in a region you're interested in
    a. Export & re-symbolize





Strava Metro: Data Clean-up

What are these crazy values....

1. Calculate Speed
    a. Recalculate length in miles
    b. Create 2 new fields: spd & rspd as double
    c. Open field Calc: length/((activity time in seconds/60)/60)
    d. Beware the divide by zero rules....
    e. Locate and remove erroneous data

Strava Metro: Data Clean-up

What are these crazy values….

1. Hints
    a. Very small segments (<10 meters) /w high counts
    b. Very high speeds with low counts < 5

Helpful Calculations for Field Calculator

Calculate Speed Per Edge from seconds:

length/((activity time in seconds/60)/60)

Filter Points:

start_date_local < '2015-09-1' AND start_date_local > ' 2015-08-01'

Percent Change:

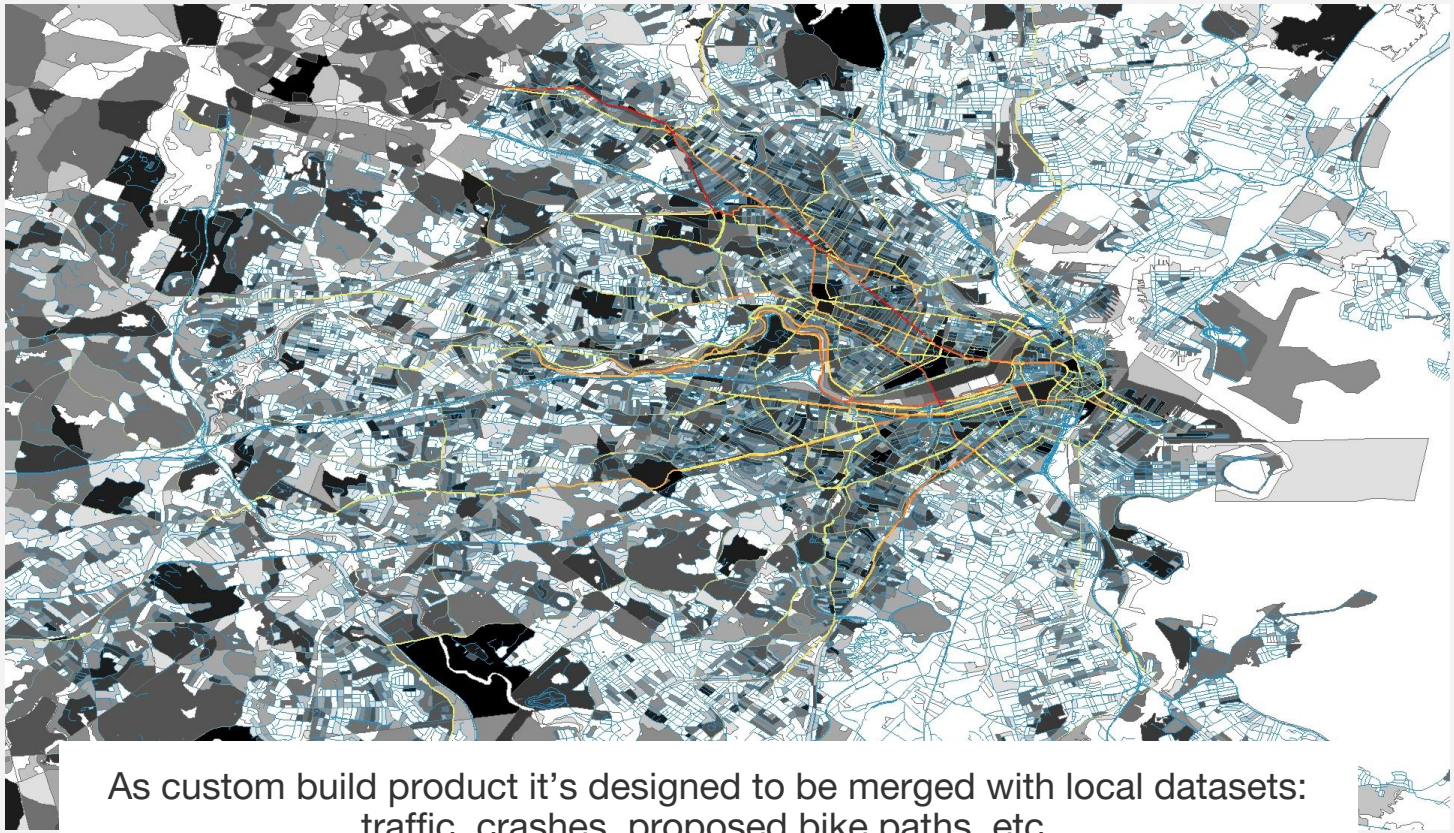
(([TACTCNT_12_13]-[TACTCNT])/ [TACTCNT])*100

Break!

Part 2

Correlation (Crash Points, Counters, Speed): 12:00pm – 12:20pm
Finding Missing Geo:12:30pm – 12:40pm
Delta Analysis: 12:40pm – 1:00pm

Strava Metro Bringing Data Layers Together

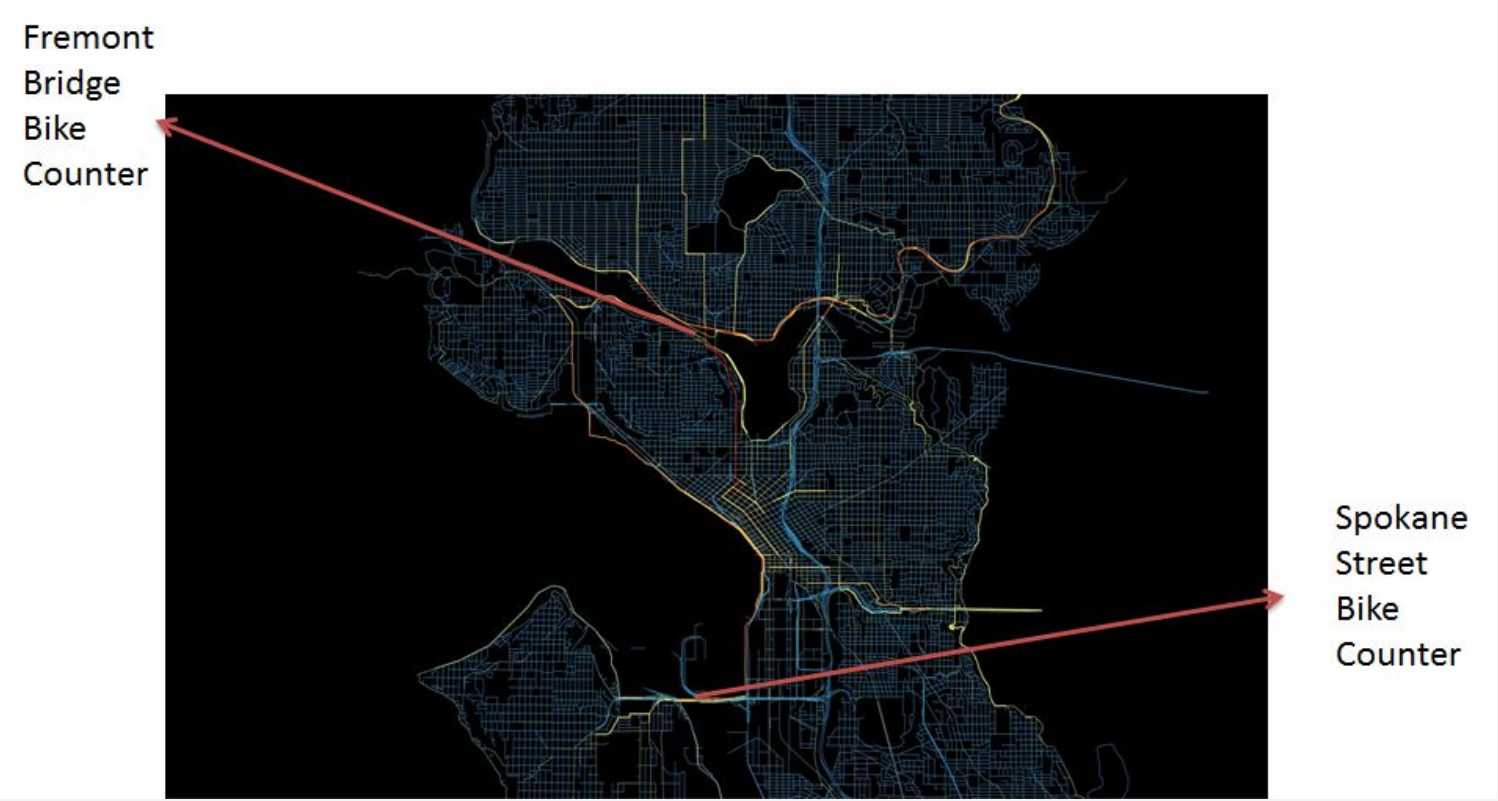


As custom build product it’s designed to be merged with local datasets: traffic, crashes, proposed bike paths, etc.

Strava Metro Correlation to Counting Programs

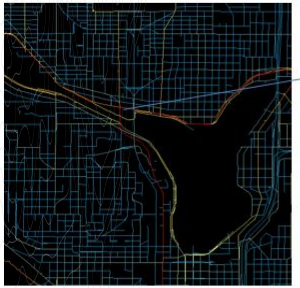
- Strava Metro’s use and impact is multiplied when the data is used in conjunction with an established counting program
- Counting programs only show saturation at a single point and dilute from there
- Metro shows the rest of the network - it’s like a counter on every corner

Strava Metro Correlation to Counting Programs Cont’

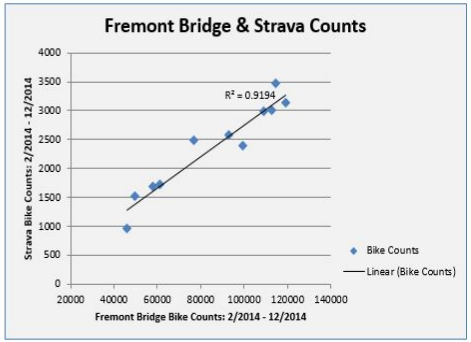


Correlating Strava to Counting Programs Cont’

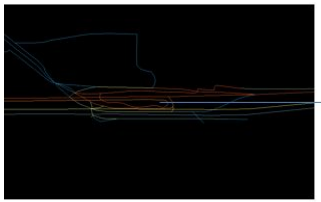
Fremont Bridge Bike Counts



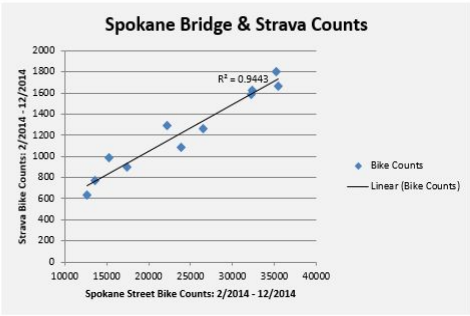
Strava: 25,980
Fremont Counter: 939,386
Percent of Strava to Population: 2.77%
R2: 0.9194



Spokane Bridge Bike Counts



Strava: 13,602
Fremont Counter: 266,850
Percent of Strava to Population: 5.10%
R2: 0.9443



Using counting programs with the Metro data allows the data to become even more useful. Strava correlation with counting programs is statistically amazing, with r-squared values typically around 0.8.

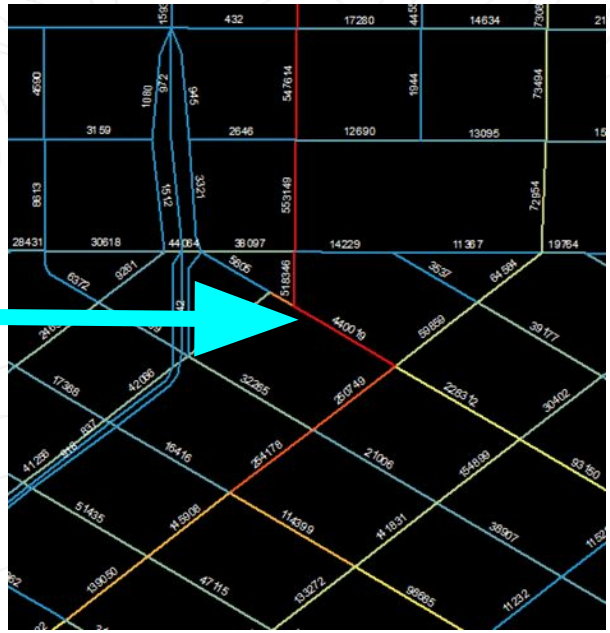
Correlating Strava to Counting Programs Cont’

16,297 Strava Bike Trips

X 27 Multiplier

= **440,019** year bike trips

(199,476 6- 9am)



How far can we push this? ---> Total Miles Traveled in SDOT by Bike in 2014:

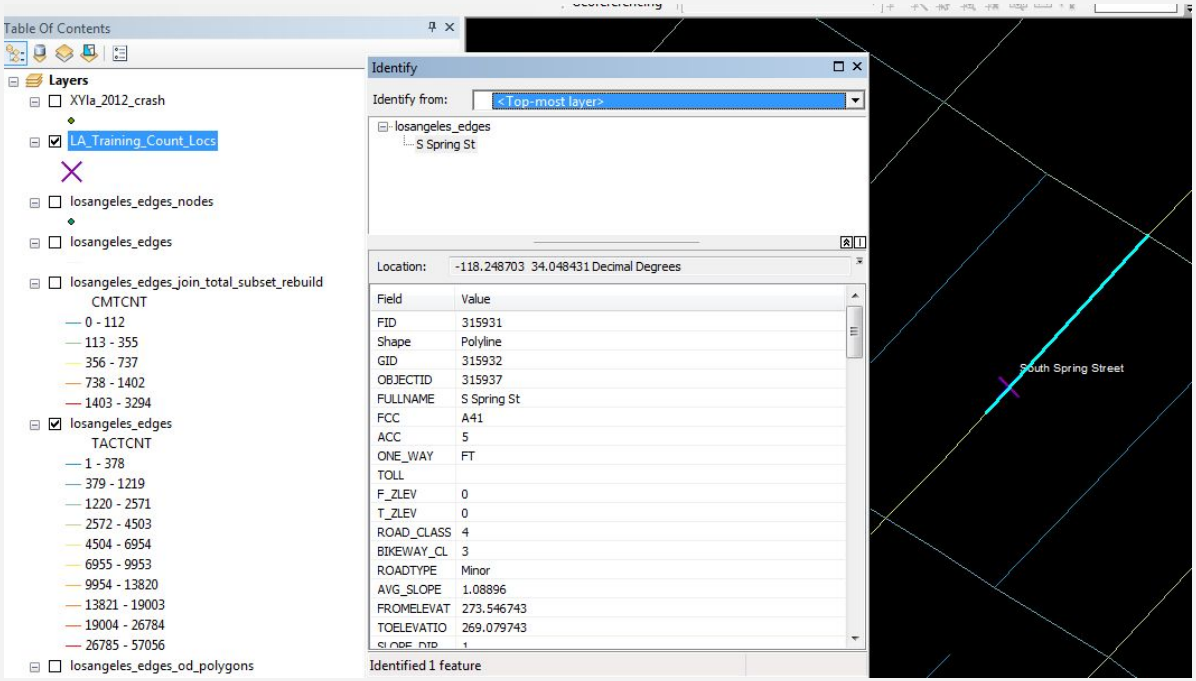
63,253,198

Bike Count Correlation: Merging data feeds

1. You need specific count data
    a. We will use the mock data to show how to do this
    b. 4 Locations /w 12 months of Counts

Bike Count Correlation Cont’

1. Locate Street objectIDs for Metro Counts
    a. Zoom to streets with bike count spots
    b. Grab the objectID (x4)
    c. Beware of dual count locations vs one-sided



Bike Count Correlation Cont’

Grab the month counts per street to Merge back with counter locations

```
select edge_id, EXTRACT(YEAR FROM datetime)|| ',' ||EXTRACT(MONTH FROM datetime) g_date,
sum(total_activity_count) acts,
sum(CASE when (EXTRACT(DOW FROM datetime) not in (0,6)) then total_activity_count end) as acts_weekday,
sum(CASE when (EXTRACT(DOW FROM datetime) in (0,6)) then total_activity_count end) as acts_weekend
from la_edges_ride_data
where edge_id in (315937,520926,146497,520706)
group by edge_id,g_date
order by edge_id, g_date
```

Data Output | Explain | Messages | History

	edge_id integer	g_date text	acts bigint	acts_weekday bigint	acts_weekend bigint
1	146497	2015,10	328	263	65
2	146497	2015,11	257	202	55
3	146497	2015,12	161	135	26
4	146497	2015,7	362	324	38
5	146497	2015,8	387	302	85
6	146497	2015,9	311	267	44
7	146497	2016,1	187	139	48
8	146497	2016,2	243	186	57
9	146497	2016,3	296	244	52

Bike Count Correlation Cont’

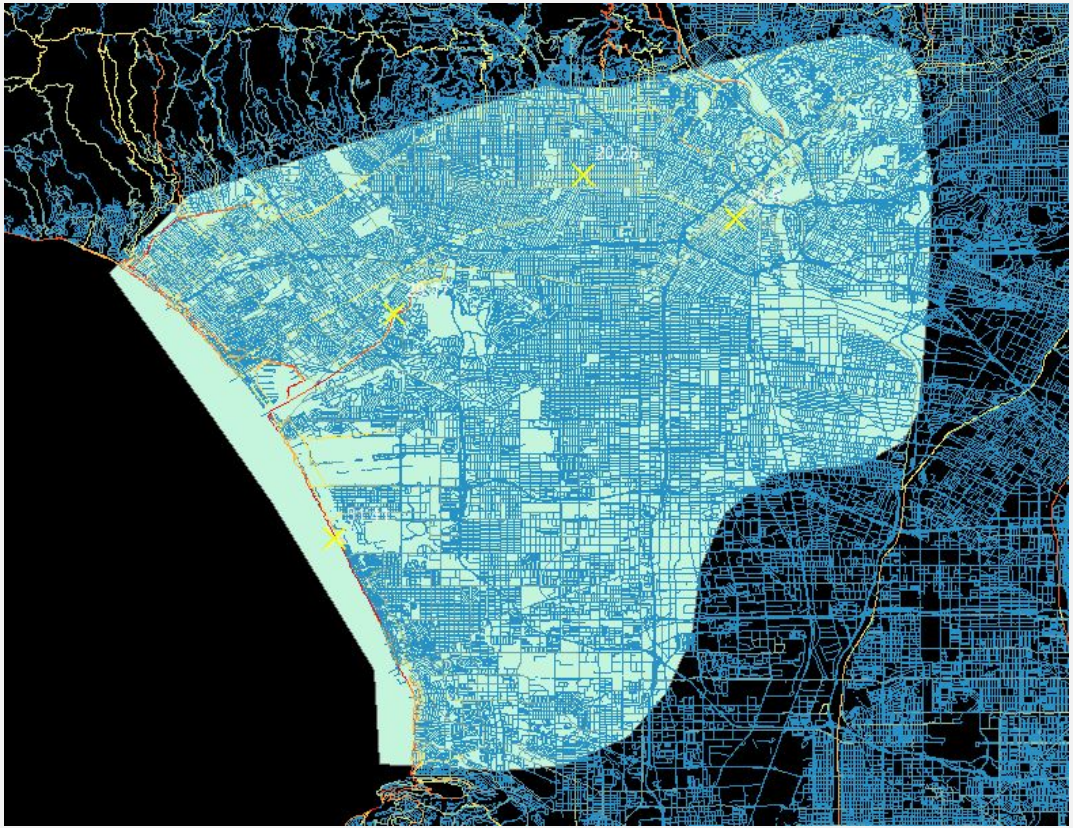
Create a regression table to understand R2 and % Correlation

- This is typically done in a spreadsheet program like excel
- The key here is to end up with you site specific multiplier for extrapolation

	A	B	C	D	E	F	G	H	I
1	edge_id	date	acts	acts weekday	acts weekend	bike counter	%	Ext Factor	
2	146497	2015,7	362	324	38	10498	0.034482759	29	
3	146497	2015,8	387	302	85	11416.5	0.033898305	29.5	
4	146497	2015,9	311	267	44	8863.5	0.035087719	28.5	
5	146497	2015,10	328	263	65	9594	0.034188034	29.25	
6	146497	2015,11	257	202	55	7710	0.033333333	30	
7	146497	2015,12	161	135	26	4347	0.037037037	27	
8	146497	2016,1	187	139	48	5516.5	0.033898305	29.5	
9	146497	2016,2	243	186	57	7047	0.034482759	29	
10	146497	2016,3	296	244	52	8658	0.034188034	29.25	
11	146497	2016,4	319	258	61	9729.5	0.032786885	30.5	
12	146497	2016,5	288	235	53	7992	0.036036036	27.75	
13	146497	2016,6	288	241	47	8856	0.032520325	30.75	

Bike Count Correlation Cont’

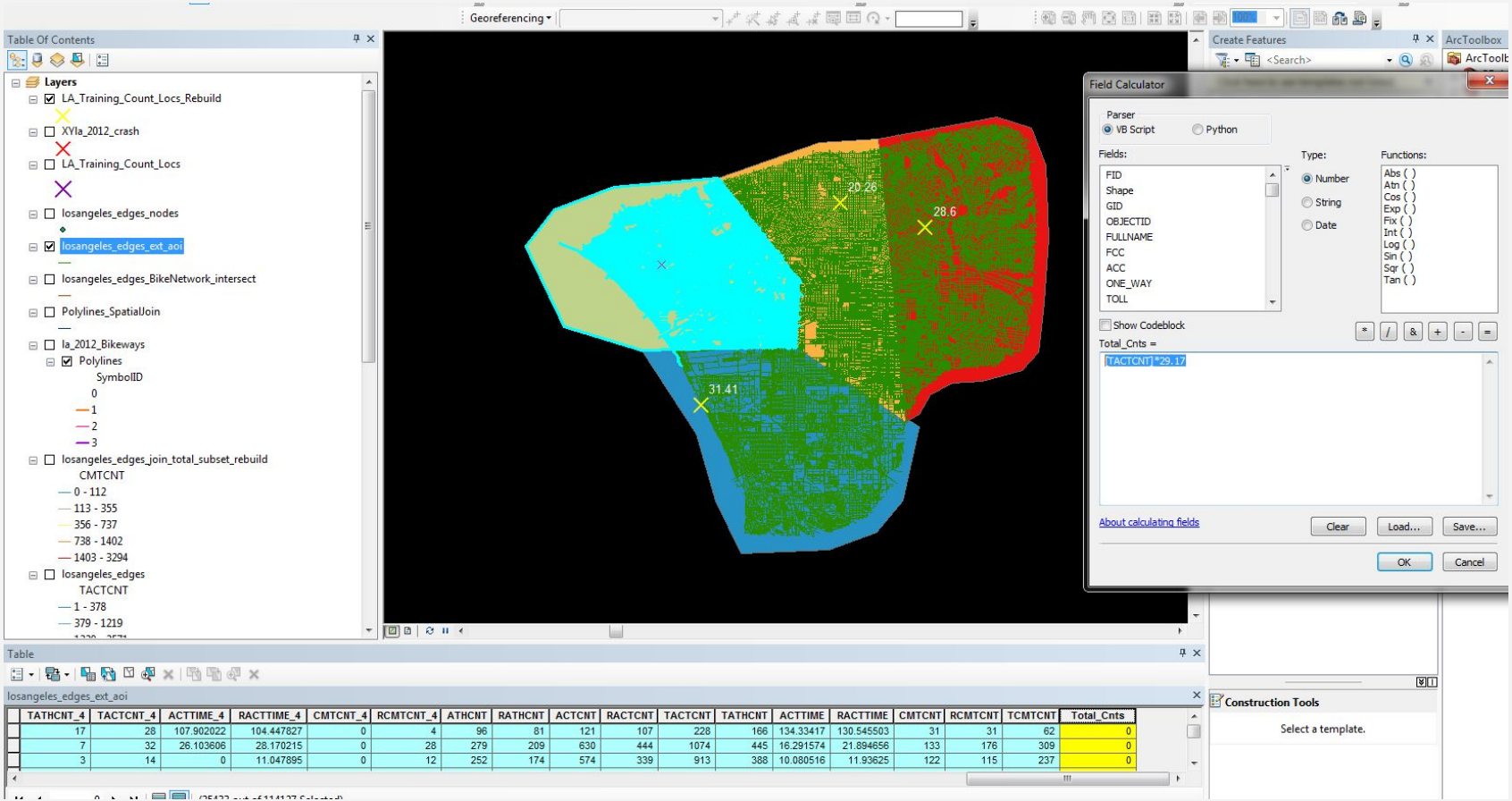
Match this back to the street network via AOI Polygons



Bike Count Correlation Cont’

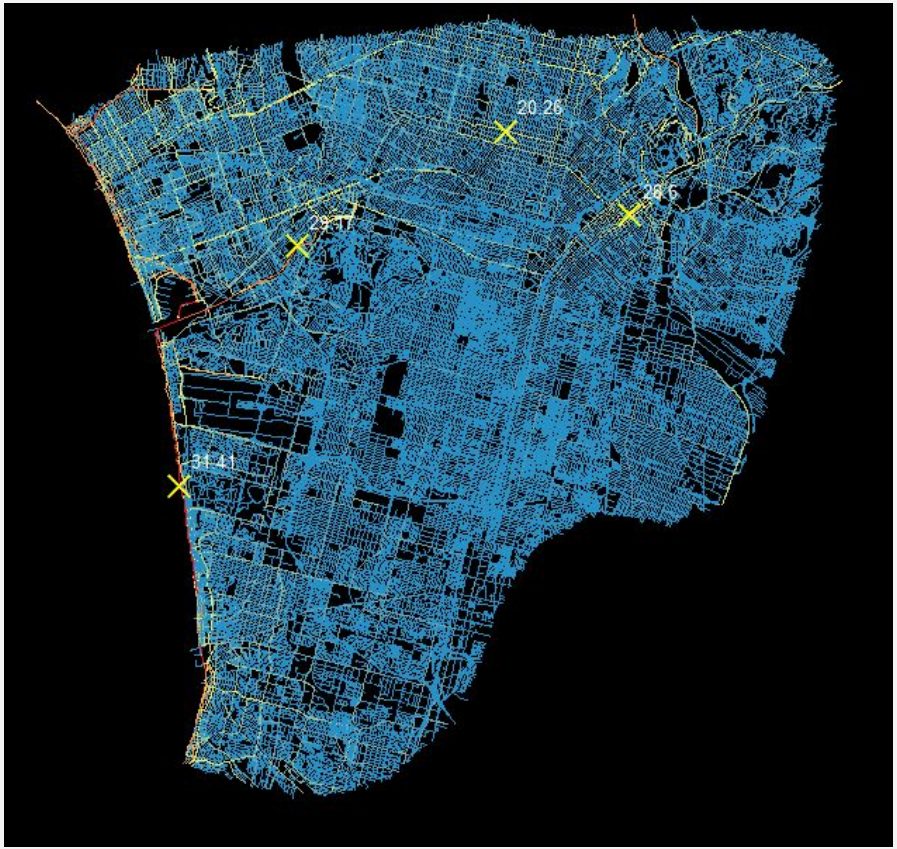
- Create 4 Extrapolation Zones
- Create new Extra Field
- Calculate Total Counts
- Display via jenks
- Create Summed Values
  - Total Miles
  - Total Time
  - Carbon offset?

Bike Count Correlation Cont’



Bike Count Correlation Cont’

Total Miles: 100,872,889

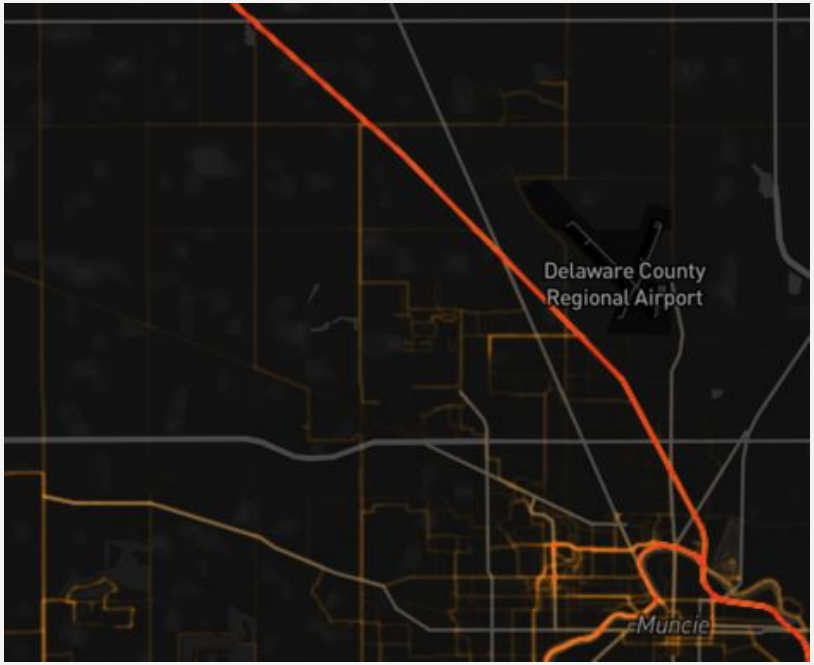


Basemaps

Important to think about multi-modal transportation

- Should include all:
  - Streets
  - Roads
  - Trails
  - Paths

- And should break at all intersections (decision points)



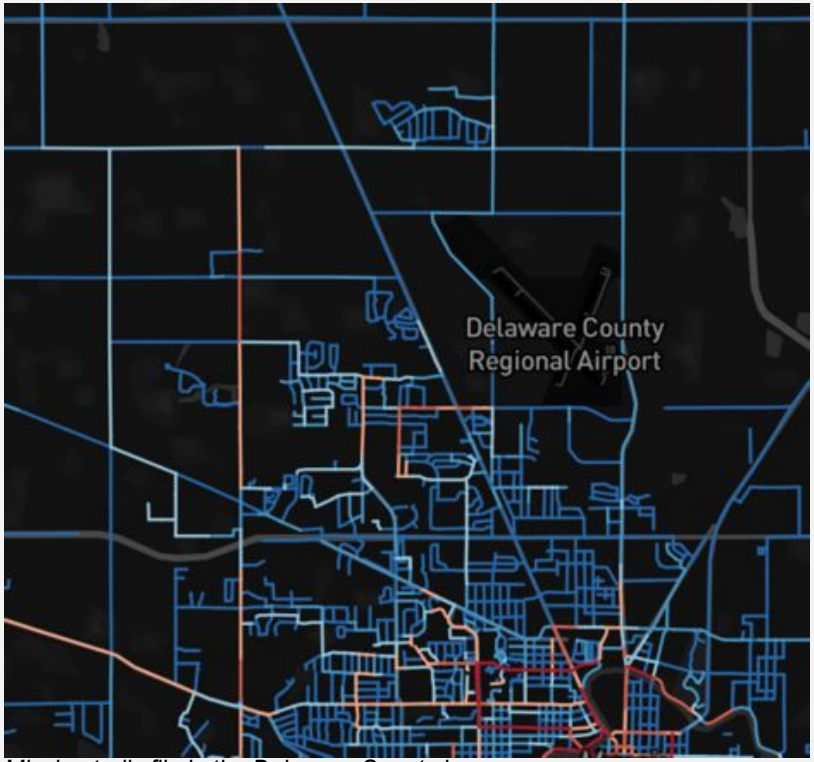
Strava Heatmap, Delaware County, Indiana

Basemaps

Important to think about multi-modal transportation

- Should include all:
  - Streets
  - Roads
  - Trails
  - Paths

- And should break at all intersections (decision points)



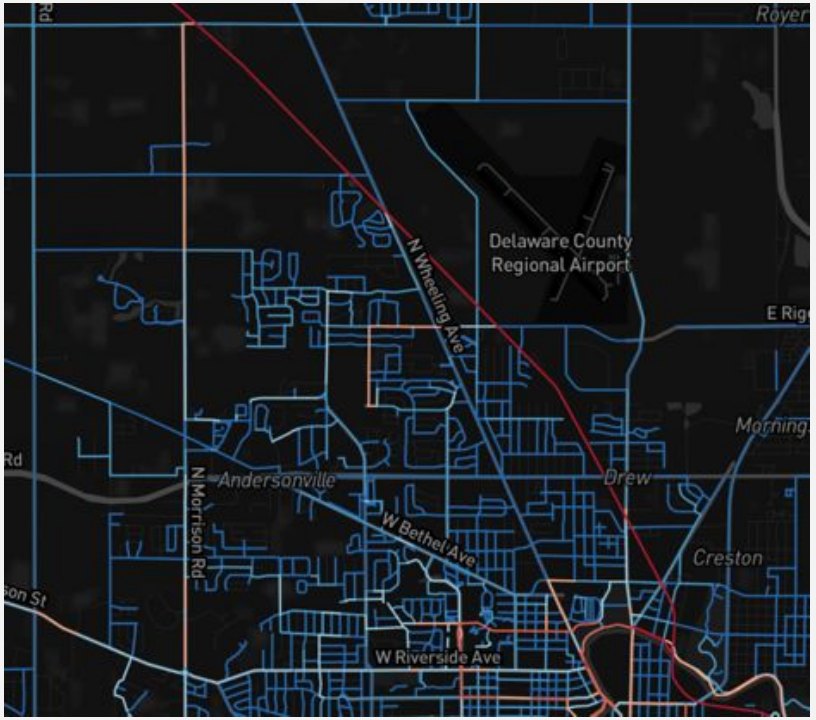
Missing trails file in the Delaware County basemap

Basemaps

Important to think about multi-modal transportation

- Should include all:
  - Streets
  - Roads
  - Trails
  - Paths

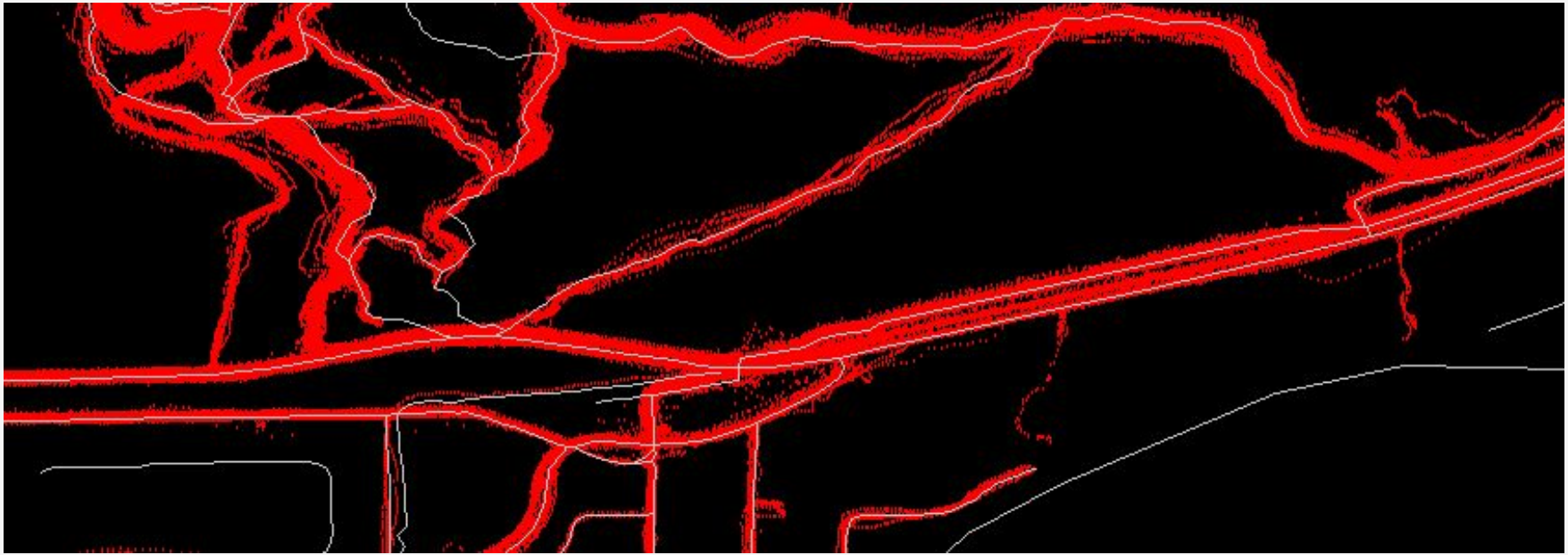
- And should break at all intersections (decision points)



Streets and trails merged into one basemap

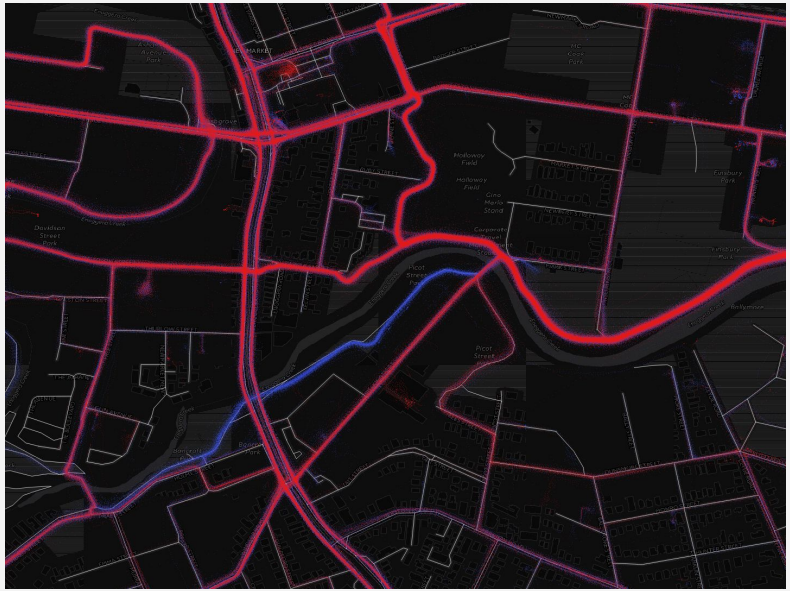
Basemaps

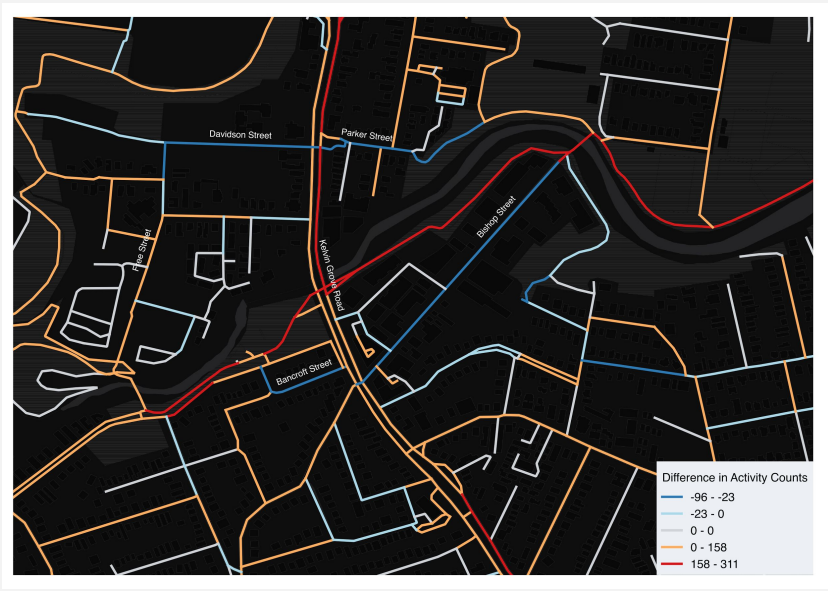
Missing/Inaccurate Geometry



Misaligned basemap geometry will cause count discrepancies in the Strava Metro data

Strava Metro Validating Investment in Cycling Infrastructure



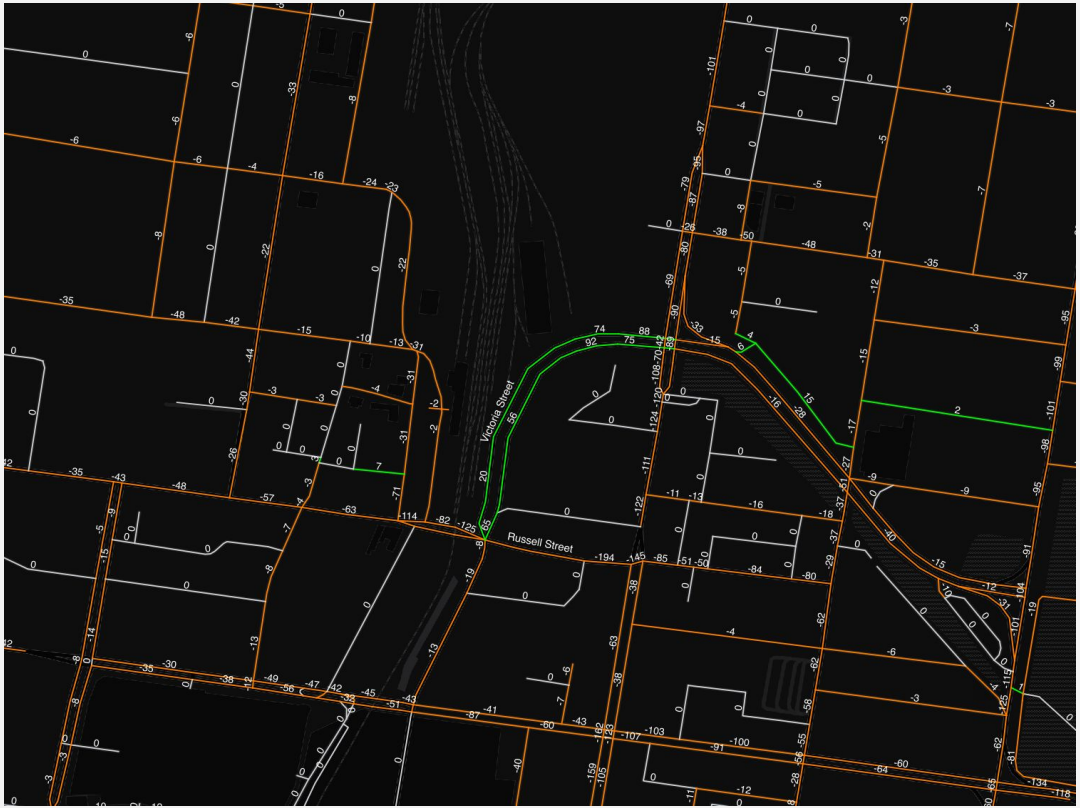


Metro provides key insights into how the cycling population is adapting to new cycleways, protected lanes and surging car populations. The left image shows the GPS points before (red) and after (blue) a new section of cycleway was opened. The Metro data on the right shows the actual change in percent with blue losing trips and red gaining trips.

Strava Metro Validating Investment in Cycling Infrastructure

Further exploration of investment in dedicated cycleways.

Strava Metro Validating Investment in Cycling Infrastructure



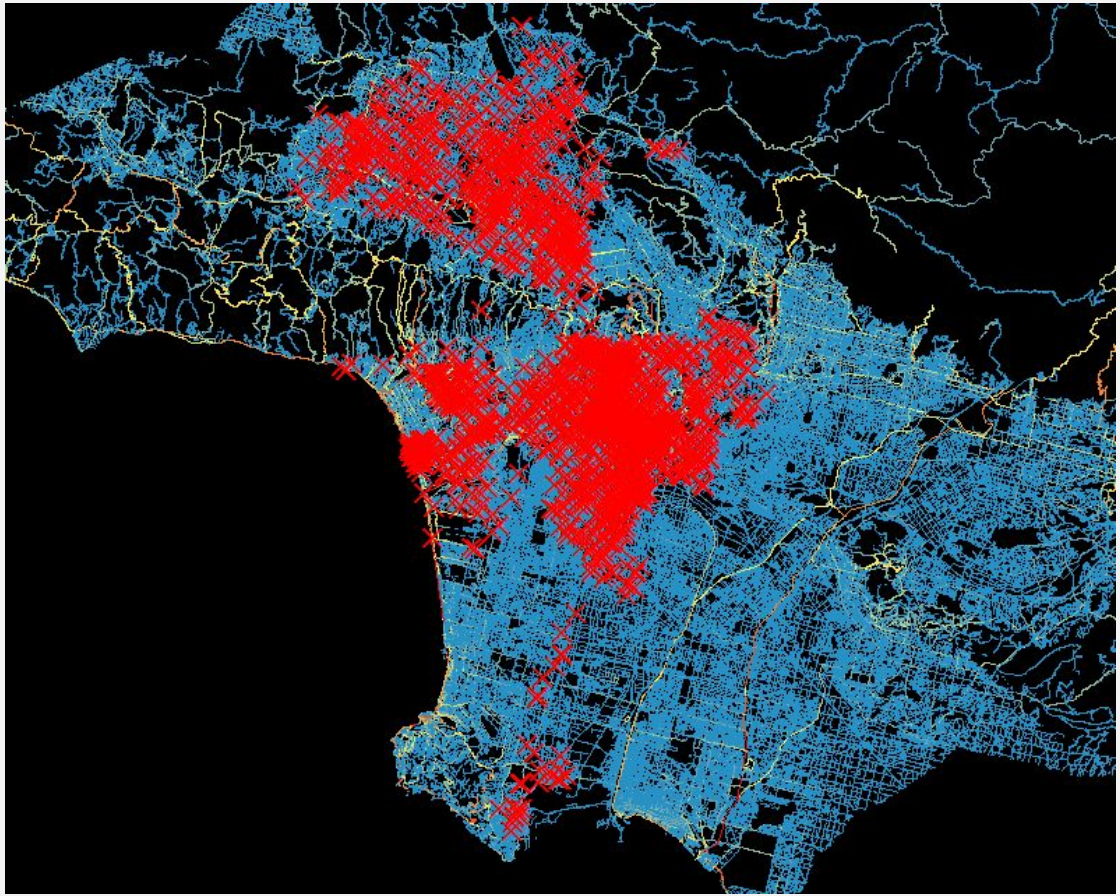
Changes to cycling infrastructure impact routing throughout the network

Strava Metro: Crash Cluster Analysis

2012 Crash Dataset

Locate Streets touched

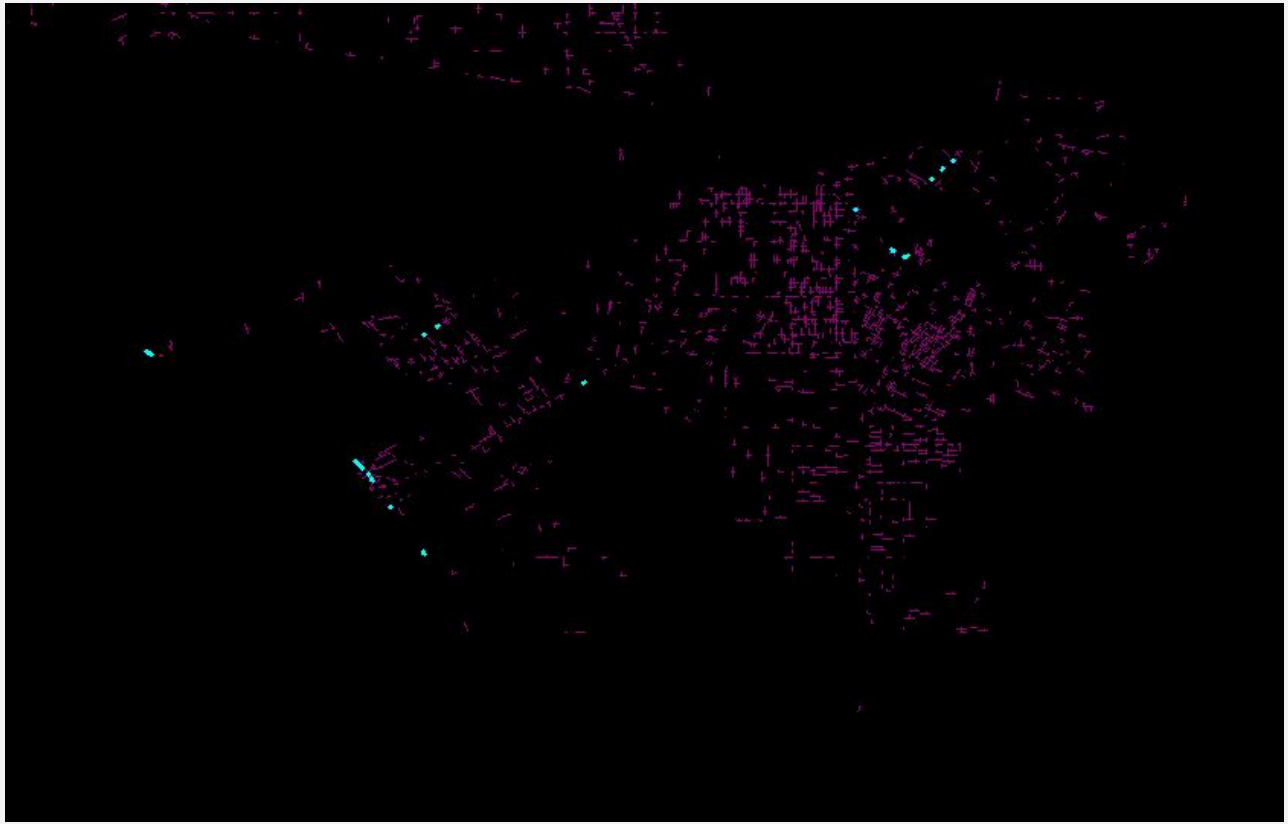
ID Trends and focus streets



Strava Metro: Crash Cluster Analysis

Subset of streets /w high

Commute counts



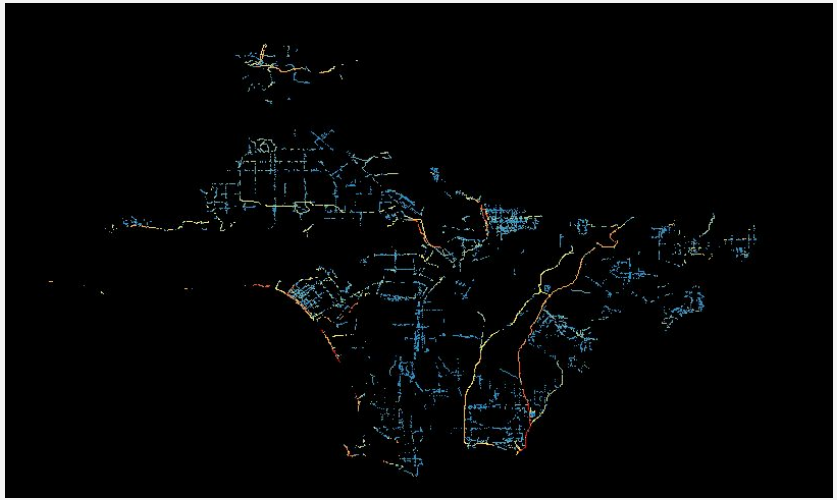
Strava Metro: Bike Facility Hot Spots

Deep Dive & Out of the Box Thinking

What happens when we take just the network and look at it with use?

How do we do this spatial match?

Can you find the top points were connectivity challenges are causing issues?



Part 3

Discussion around potential urban research uses for Strava data: 1:00pm – 1:30pm

Afternoon Session - Hands On

1) Removing Noise
2) Correlation (Crash Points, Counters, Speed)
3) Finding Missing Geo
4) Delta Analysis

Working with the Raw Data

Many-to-One Relationships

OID	EDGE_ID	YEAR	DAY	HOUR	MINUTE	ATHLETE_CO	REV_ATHLET	ACTIVITY_C	REV_ACTIVI	TOTAL_ACTI	ACTIVITY_T	REV_ACT_01	COMMUTE_CO
0	378156	2015	174	12	22	0	1	0	1	1	0	754	0
1	378156	2015	172	15	58	1	0	1	0	1	1395	0	0
2	378156	2015	178	10	47	1	0	1	0	1	550	0	0
3	377065	2014	193	16	40	0	1	0	1	1	0	12	0
4	377065	2014	223	13	4	1	0	1	0	1	19	0	0
5	377066	2014	223	13	4	1	0	1	0	1	115	0	0
6	377053	2015	113	15	25	1	0	1	0	1	4472	0	0
7	377053	2015	144	12	35	1	0	1	0	1	4348	0	0
8	377053	2015	147	17	13	1	0	1	0	1	4145	0	0
9	377053	2015	153	11	42	1	0	1	0	1	3972	0	0
10	377053	2015	159	14	43	1	0	1	0	1	4106	0	0

Relational databases

CREATE TABLE la_edges_ride_data

(edge_id integer, year integer, day integer, hour integer, minute integer, athlete_count integer, rev_athlete_count integer, activity_count integer, rev_activity_count integer, total_activity_count integer, activity_time numeric, rev_activity_time numeric, commute_count integer)

WITH (OIDS=FALSE);

ALTER TABLE la_edges_ride_data

OWNER TO postgres;

Working with the Raw Data

Copy Data In:
psql -U postgres -d strava -c "Copy la_edges_ride_data FROM
'C:\LA\la_c_201507_201606_ride_data.csv' DELIMITERS ',' CSV HEADER;"

Update Table with Datetime Field and Populate:
ALTER TABLE la_edges_ride_data ADD COLUMN datetime timestamp without time zone;

Update Date Field:
UPDATE la_edges_ride_data set datetime = cast(timestamp '2014-12-31' + interval '1 day' * day + interval '1 hour' * hour + interval '1 minute' * minute as timestamp) where year = 2015;

UPDATE la_edges_ride_data set datetime = cast(timestamp '2015-12-31' + interval '1 day' * day + interval '1 hour' * hour + interval '1 minute' * minute as timestamp) where year = 2016;

Indexes:
CREATE INDEX la_c_ride_data_datetime_idx ON la_edges_ride_data USING btree (datetime);

CREATE INDEX la_c_ride_data_edge_id_idx ON la_edges_ride_data USING btree (edge_id);

CREATE INDEX la_c_ride_data_hour_idx ON la_edges_ride_data USING btree (hour);

Simple SQL with Metro

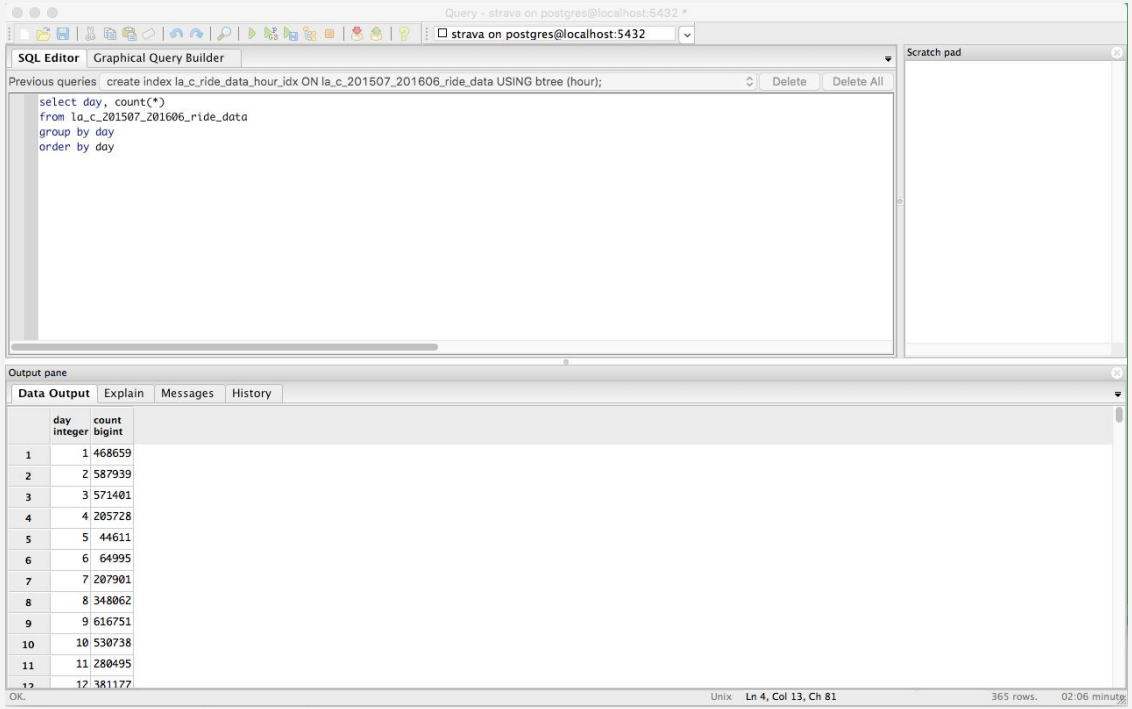
Counts by Day

SELECT day, count(*)

FROM la_c_201507_201606_ride_data

GROUP BY day

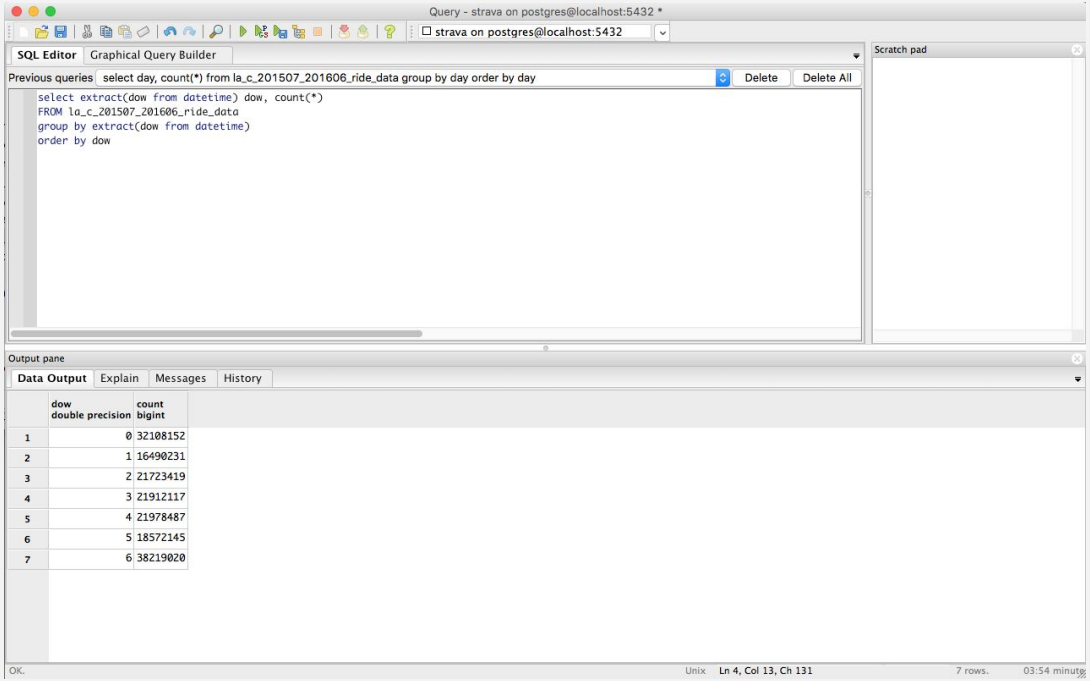
ORDER BY day



Simple SQL with Metro

Counts by Day Of Week

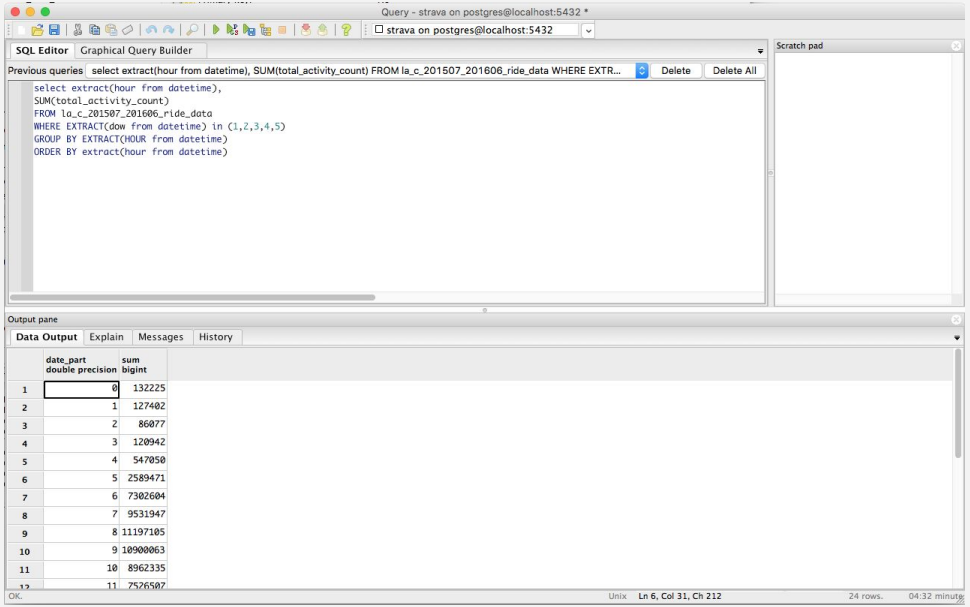
```
SELECT EXTRACT(dow from datetime) dow, count(*)
FROM la_c_201507_201606_ride_data
GROUP BY EXTRACT(dow from datetime)
ORDER BY dow
```



Simple SQL with Metro

Hour Pulses

```
SELECT EXTRACT(HOUR from datetime),
SUM(total_activity_count)
FROM la_c_201507_201606_ride_data
WHERE EXTRACT(dow from datetime) in (1,2,3,4,5)
GROUP BY EXTRACT(HOUR from datetime)
ORDER BY EXTRACT(HOUR from datetime)
```



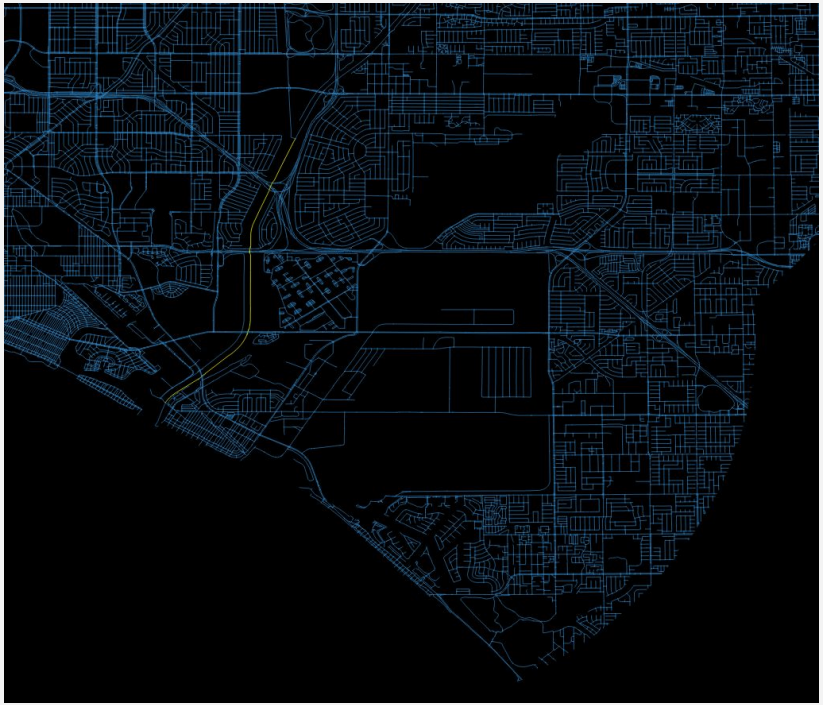
Simple SQL with Metro

Use for a Group of Streets

```
SELECT EXTRACT(YEAR FROM datetime )|| ',' ||EXTRACT(MONTH FROM datetime )|| ',' ||EXTRACT(day FROM
datetime ) || ',' ||EXTRACT(hour FROM datetime) g_date,
SUM(athlete_count) aths, SUM(activity_count) acts, SUM(rev_activity_count) rev_acts, median(activity_time) time1,
median(rev_activity_time) time2
FROM la_c_201507_201606_ride_data
WHERE edge_id IN (521911,522002,521913,521914,521915,521918,521921,521961,521971)
and datetime > '2015-09-07' and datetime < '2015-09-17'
GROUP BY g_date
ORDER BY g_date
```

Simple SQL with Metro

Use for a Group of Streets



Query - strava on postgres@localhost:5432 *

strava on postgres@localhost:5432

SQL Editor | Graphical Query Builder

Previous queries | Delete | Delete All

```
SELECT EXTRACT(YEAR FROM datetime )|| ',' ||EXTRACT(MONTH FROM datetime )|| ',' ||EXTRACT(day FROM datetime ) || ',' ||EXTRACT(hour FROM
SUM(athlete_count) aths, SUM(activity_count) acts, SUM(rev_activity_count) rev_acts, median(activity_time) time1, median(rev_activity_t
FROM la_c_201507_201606_ride_data
WHERE edge_id IN (521911,522002,521913,521914,521915,521918,521921,521961,521971)
and datetime> '2015-09-07' and datetime < '2015-09-17'
GROUP BY g_date
ORDER BY g_date
```

Scratch pad

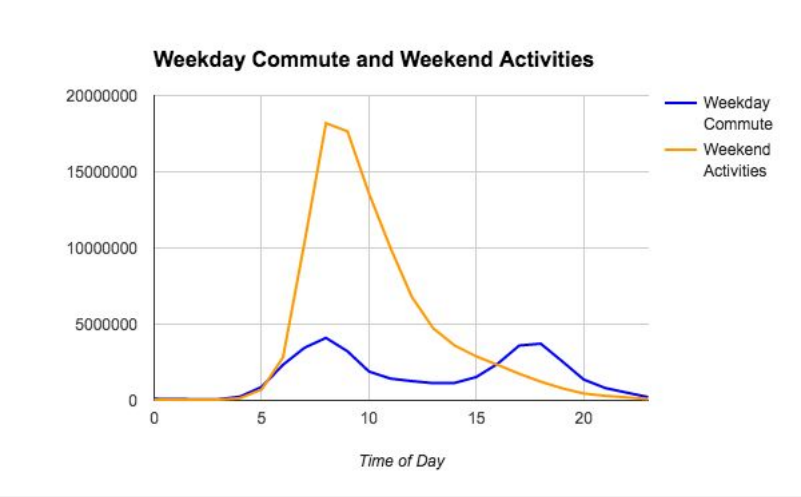
Output pane

Data Output | Explain | Messages | History

	g_date text	aths bigint	acts bigint	rev_acts bigint	time1 numeric	time2 numeric
1	2015,9,10,10	15	15	51		17.4052629136853000
2	2015,9,10,11	2	2	8		130.8924282025825000
3	2015,9,10,12	2	2	5		202.1942227486020000
4	2015,9,10,13	1	1	0	384.1310000000000000	
5	2015,9,10,14	18	18	9	185.5710000000000000	
6	2015,9,10,15	1	1	5		81.9796073179688500
7	2015,9,10,16	0	0	4		120.4382939024936000
8	2015,9,10,17	22	22	28		306.4934721079930000
9	2015,9,10,18	46	46	95		163.3830259376650000
10	2015,9,10,19	51	51	25	196.7675000000000000	
11	2015,9,10,21	7	7	0	9.8270000000000000	
12	2015,9,10,4	1	1	0	465.1220000000000000	

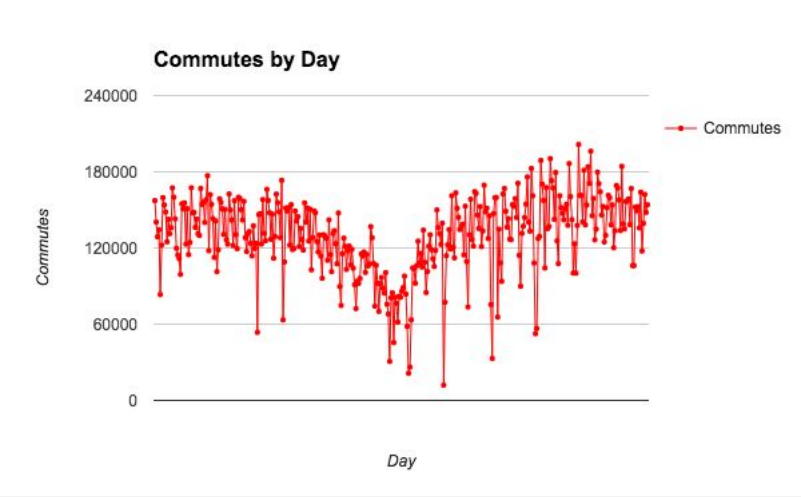
OK. Unix Ln 5, Col 13, Ch 433 156 rows. 1.6 secs

Using Metro to Gain Temporal Movement Trends



Week to Week Trends

Explore the two distinct user groups on Strava: weekday commuters and weekend recreation cyclists

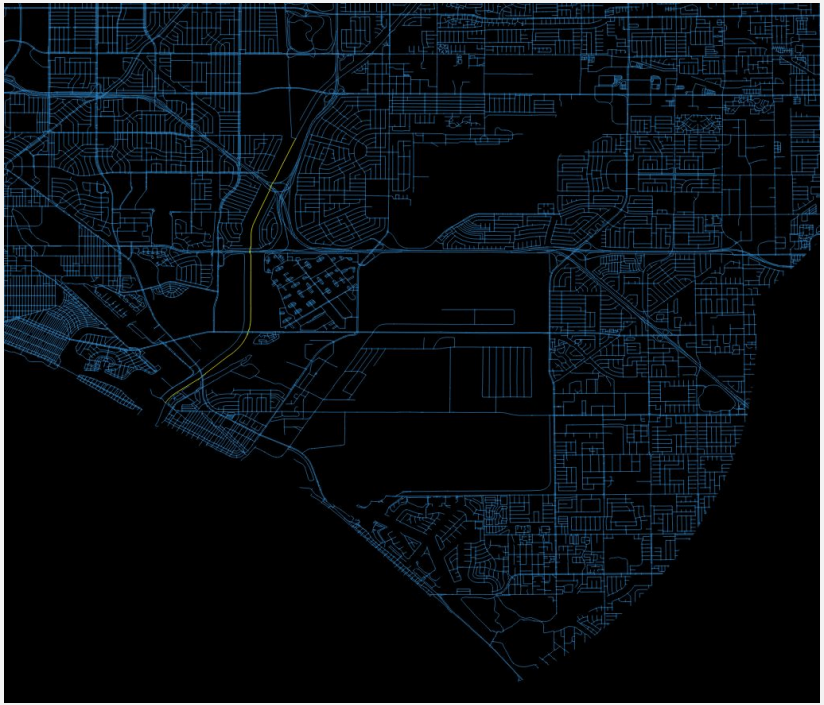


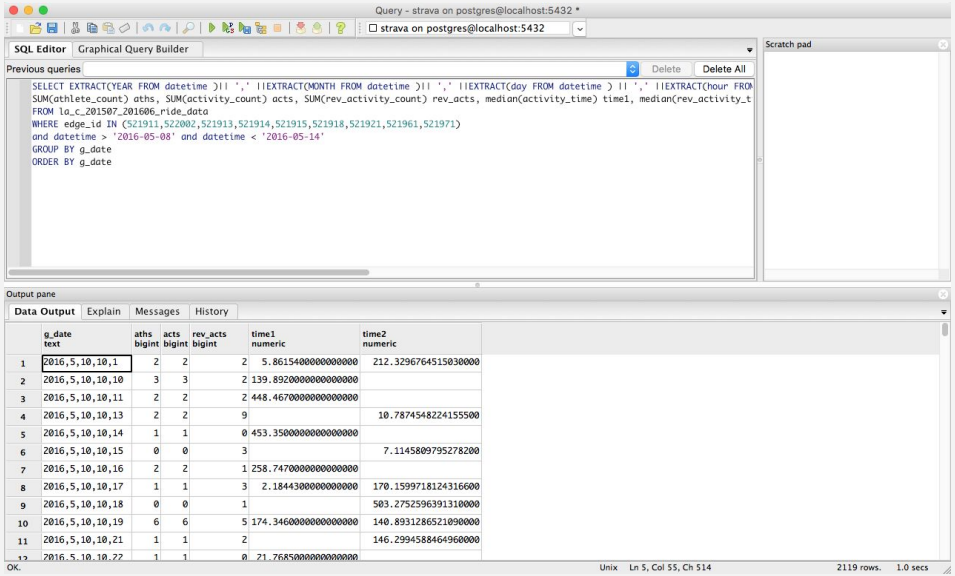
Yearly Commute Trends

Explore seasonality patterns - bike to work month, summer lulls, holidays, etc.

Strava Metro: Corridor Analysis

Lower San Gabriel Trail





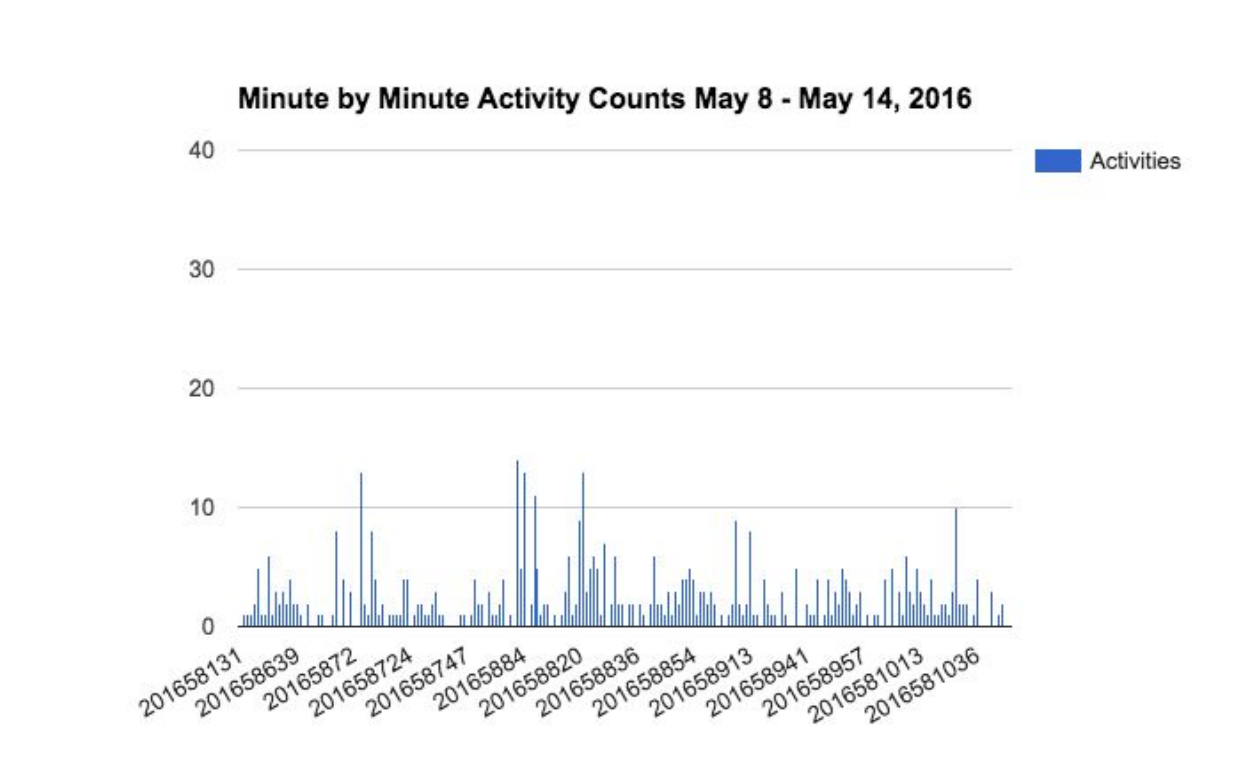
Strava Metro: Corridor Analysis

Lower San Gabriel Trail

```
SELECT EXTRACT(YEAR FROM datetime )|| ',' ||EXTRACT(MONTH FROM
datetime )|| ',' ||EXTRACT(day FROM datetime ) || ',' ||EXTRACT(hour FROM
datetime) || ',' ||EXTRACT(minute FROM datetime) g_date,
SUM(athlete_count) aths, SUM(activity_count) acts, SUM(rev_activity_count)
rev_acts, median(activity_time) time1, median(rev_activity_time) time2
FROM la_c_201507_201606_ride_data
WHERE edge_id IN
(521911,522002,521913,521914,521915,521918,521921,521961,521971)
and datetime > '2016-05-08' and datetime < '2016-05-14'
GROUP BY g_date
ORDER BY g_date
```

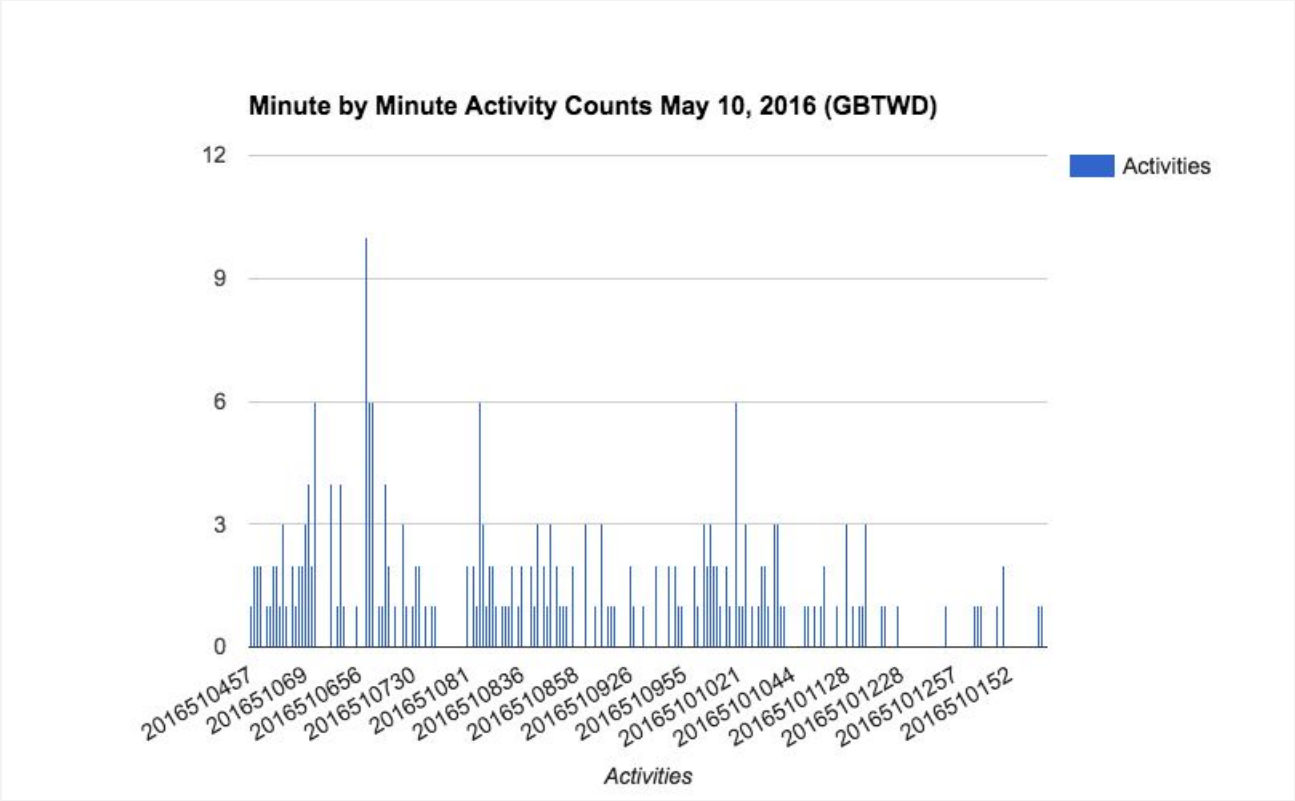
Strava Metro: Corridor Analysis

Lower San Gabriel Trail



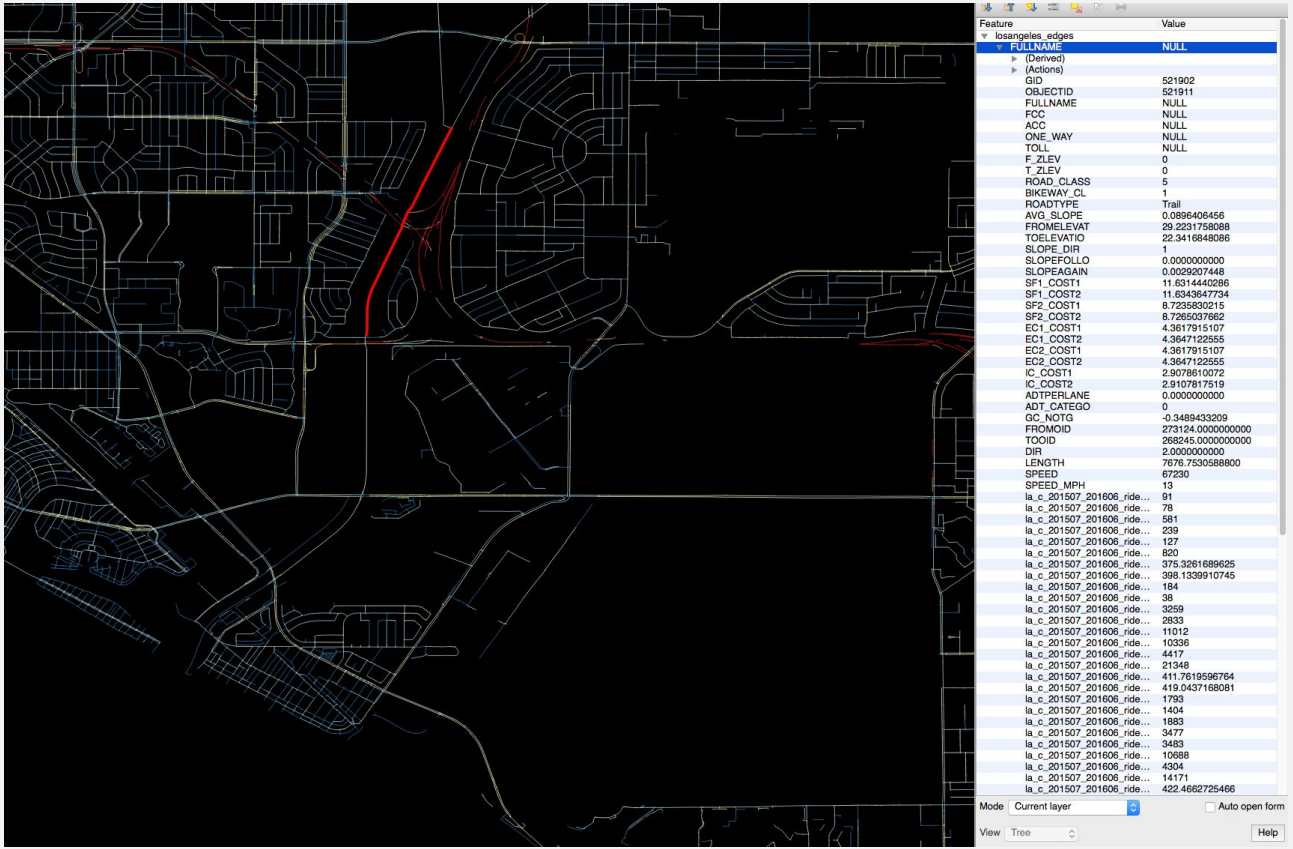
Strava Metro: Corridor Analysis

Lower San Gabriel Trail



Strava Metro: Corridor Analysis

Lower San Gabriel Trail



Feature	Value
losangeles_edges	
FULLNAME	NULL
(Derived)	
(Actions)	
GID	521902
OBJECTID	521911
FULLNAME	NULL
FCC	NULL
ACC	NULL
ONE_WAY	NULL
TOLL	NULL
F_ZLEV	0
T_ZLEV	0
ROAD_CLASS	5
BIKEWAY_CL	1
ROADTYPE	Trail
AVG_SLOPE	0.0896406456
FROMELEVAT	29.2231758088
TOELEVATIO	22.3416848086
SLOPE_DIR	1
SLOPEFOLLO	0.0000000000
SLOPEAGAIN	0.0029207448
SF1_COST1	11.6314440286
SF1_COST2	11.6343647734
SF2_COST1	8.7235830215
SF2_COST2	8.7265037662
EC1_COST1	4.3617915107
EC1_COST2	4.3647122555
EC2_COST1	4.3617915107
EC2_COST2	4.3647122555
IC_COST1	2.9078610072
IC_COST2	2.9107817519
ADTPERLANE	0.0000000000
ADT_CATEGO	0
GC_NOTG	-0.3489433209
FROMOID	273124.0000000000
TOOID	268245.0000000000
DIR	2.0000000000
LENGTH	7676.7530588800
SPEED	67230
SPEED_MPH	13
la_c_201507_201606_ride…	91
la_c_201507_201606_ride…	78
la_c_201507_201606_ride…	581
la_c_201507_201606_ride…	239
la_c_201507_201606_ride…	127
la_c_201507_201606_ride…	820
la_c_201507_201606_ride…	375.3261689625
la_c_201507_201606_ride…	398.1339910745
la_c_201507_201606_ride…	184
la_c_201507_201606_ride…	38
la_c_201507_201606_ride…	3259
la_c_201507_201606_ride…	2833
la_c_201507_201606_ride…	11012
la_c_201507_201606_ride…	10336
la_c_201507_201606_ride…	4417
la_c_201507_201606_ride…	21348
la_c_201507_201606_ride…	411.7619596764
la_c_201507_201606_ride…	419.0437168081
la_c_201507_201606_ride…	1793
la_c_201507_201606_ride…	1404
la_c_201507_201606_ride…	1883
la_c_201507_201606_ride…	3477
la_c_201507_201606_ride…	3483
la_c_201507_201606_ride…	10688
la_c_201507_201606_ride…	4304
la_c_201507_201606_ride…	14171
la_c_201507_201606_ride…	422.4662725466

Mode: Current layer — Auto open form

View: Tree — Help

Average speed ranges

from 12 - 18 mph