



University
of Glasgow

The use of crowdsourced cycling data for cycling analyses (Strava)

Dr. Jinhyun Hong, Dr. David McArthur, Dr. Mark Livingston



Urban
Big
Data
Centre

An ESRC Data
Investment

- Large benefits (e.g., reduced congestion and emissions, improved health, etc)
- Quite a lot of car trips are short trips (<2km)
- Transport Scotland wants 10% of journeys to be made by bicycle by 2020; with cities responsible for achieving this



- Several interventions have been applied (e.g., four cycling infrastructure investments before, during and after the Commonwealth Games)
- Understanding cycling behaviour and evaluating the effectiveness of interventions are difficult due to the lack of data
- Manual/Automatic counts take place on specific links/points, but these are expensive and hence infrequent

- 1) Can crowdsourced cycling data be utilised for cycling behaviour studies?
- 2) Where commuting cyclists travel and what are influential factors for their route choice?
- 3) Do the new cycling infrastructure investments in Glasgow produce effective impacts?

- 2013-2016 Strava data



Data are provided as:

- Origins and destinations with route information (at output area level)
- Minute-by-minute link counts of cycling flows
- Information about waiting times at junctions
- Aggregate demographic information

- Manual counts of cyclists from a cordon count carried out in Glasgow in 2013-2015 (38 locations, 2 days per year)
- Glasgow cycling infrastructure data

Research question 1

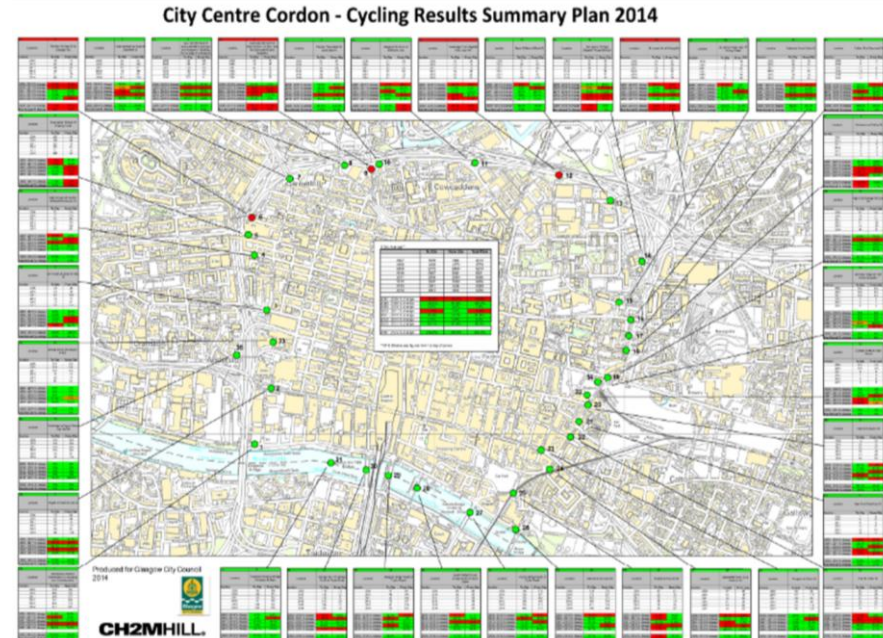
- Data: 2013-15 Strava data and 2013-15 cordon count data
- Analyses: Correlation analysis and simple linear regression

$$y_i \sim N\left(\alpha + \beta_{Strava_{cyc}} x_{stava_{cyc}i}, \sigma\right),$$

for $i = 1, \dots, 684$ (38 locations
* 3 time periods * 2 days * 3 years)

y_i : # cyclists from cordon counts;

$x_{stava_{cyc}i}$: # of Strava cycling trips



Research question 2

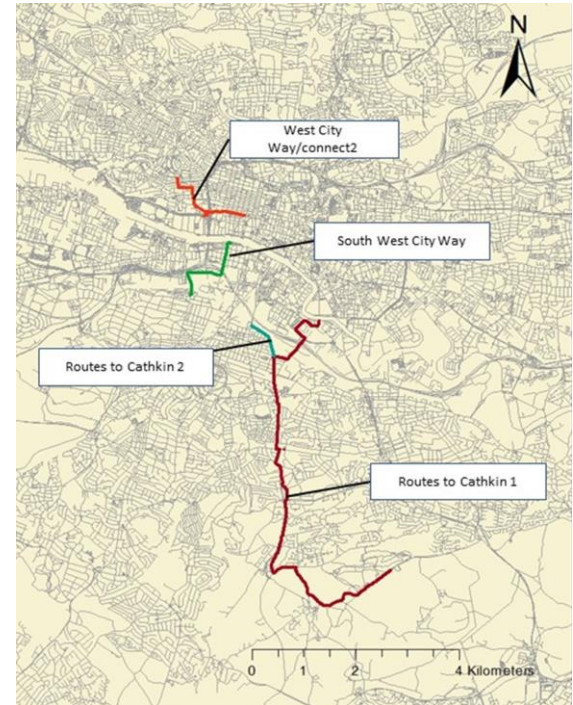
- Data: 2016 Strava data
- Analyses:
 - Compare the routes taken by commuting cyclists with the route they would take if they minimised their travel distance in the city of Glasgow (traffic assignment model)
 - Use Google Maps and local knowledge

Research question 3

- Data: 2013-15 Strava data and cycling infrastructure data
- Analysis: Fixed effect Poisson panel regression

$$\log \lambda_{it} = \beta_{infra1} x_{new\ infra1} + \beta_{infra2} x_{new\ infra2} + \beta_{infra3} x_{new\ infra3} + \beta_{infra4} x_{new\ infra4} + \mu_i + \tau_t, \quad \text{for } i = 1, \dots, 13309 \text{ (output area) \& for } t = 1, \dots, 36 \text{ (month)}$$

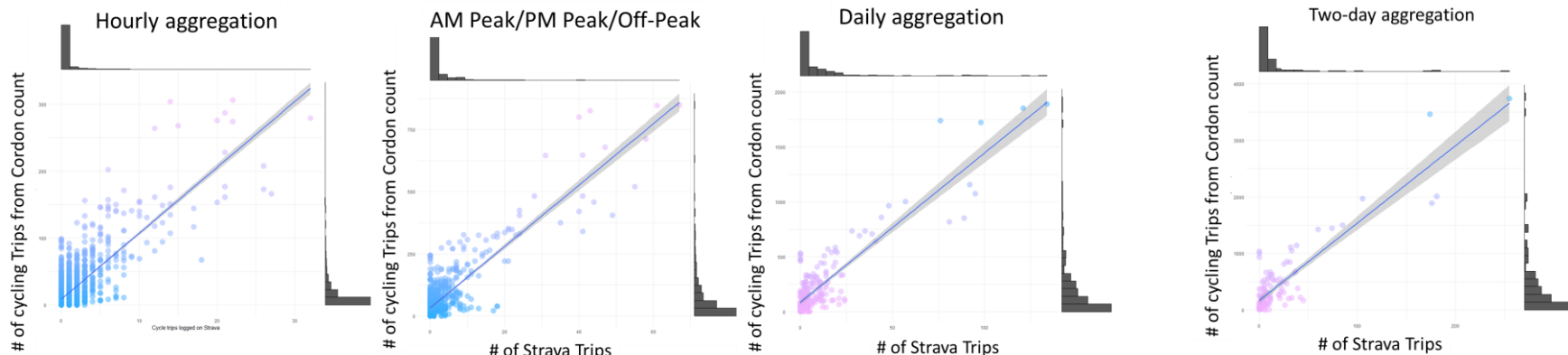
λ_{it} : # Strava cycling trips in area i in month t
 μ_i & τ_t : an out-area-specific effect and series of month fixed effects, respectively



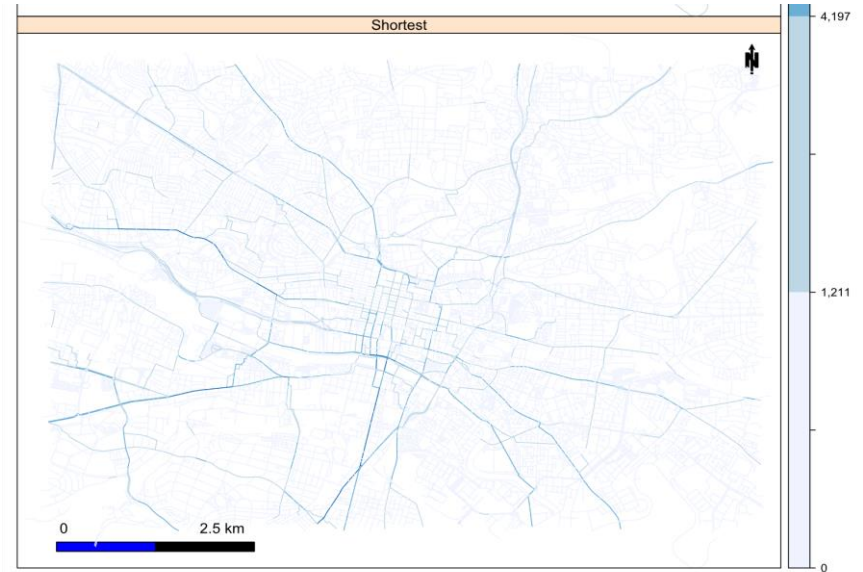
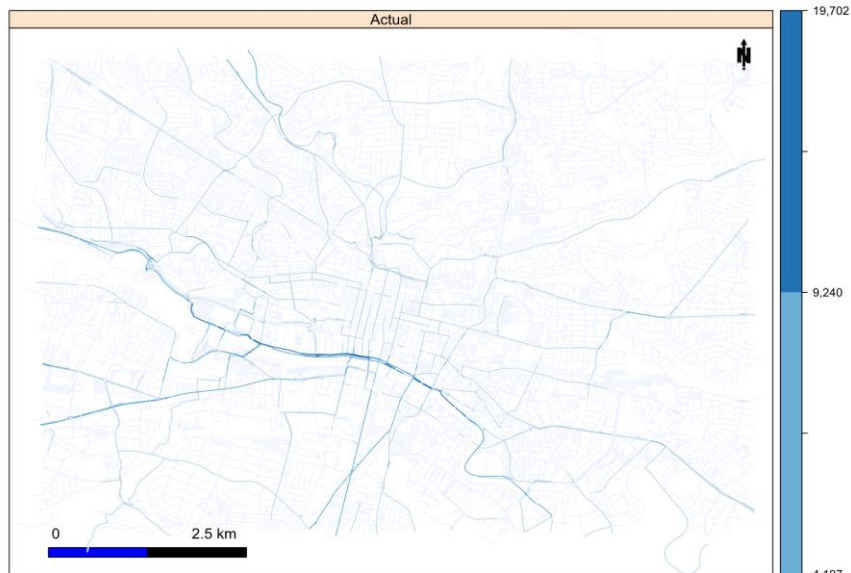
Correlation (# of cordon count vs. # of Strava count) and linear regression analysis (Research Q1)

Level of aggregation	Sample size	Correlation
Hourly	3192	0.781
AM peak, PM peak, off-peak	684	0.861
One day	228	0.882
Two days	114	0.887

Regression	Estimate	SE	P-value
Intercept	32.43	2.34	<2e-16
# Strava trips	12.35	0.28	<2e-16
Adjusted R^2	0.74		

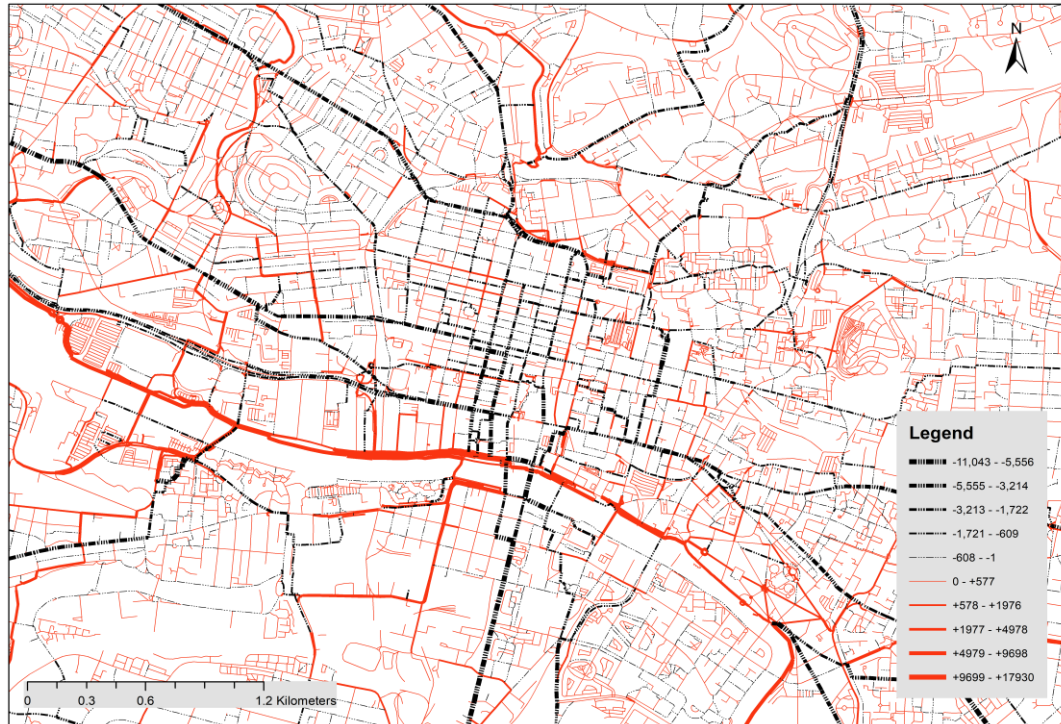


Comparison between shortest routes and actual routes (Research Q2)

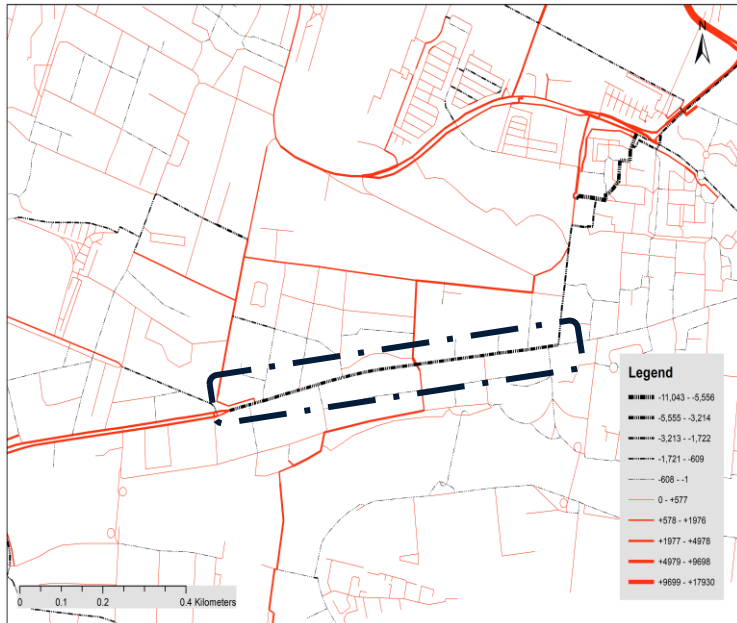


Comparison between shortest routes and actual routes (Research Q2)

Difference between observed flows and predicted flows



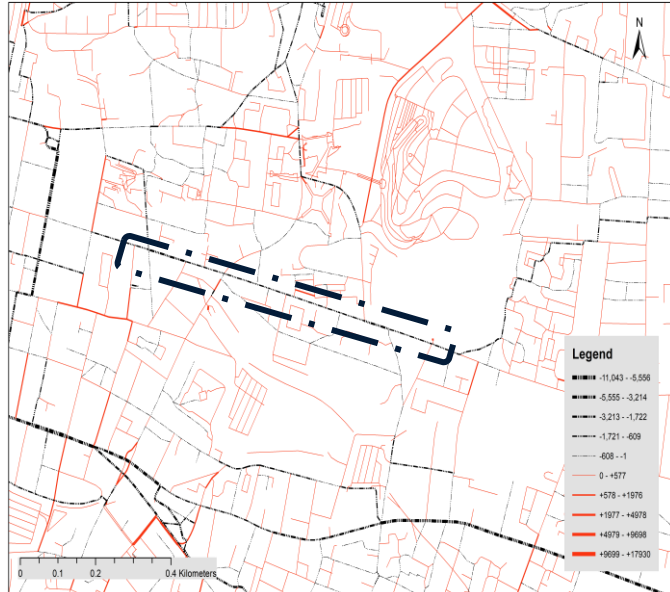
Why less popular? (Research Q2)



- Bus stops, traffic lights with a pedestrian crossing, street parking, no cycling infra

Why less popular? (Research Q2)

This road has cycling infrastructure



- Shared cycle/bus lane, street car parking, built environments (derelict properties – this area is one of the most deprived areas with a high crime rate)

Fixed effects Poisson panel regression (Research Q3)

	Overall effect			Separate effects		
	Estimate	SE	P-value	Estimate	SE	P-value
New Infrastructure (yes= 1)	0.01	0.04	0.85			
Routes to Cathkin 1				-0.10	0.05	0.05
Routes to Cathkin 2				0.08	0.02	0.00
South West City Way				0.12	0.03	0.00
West City Way/Connect 2				0.13	0.04	0.00



Thank you. Any questions ?