University of Glasgow

Exploring the relationship between Strava cyclists and all cyclists*.

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

*This presentation contains preliminary results which are subject to change. Please do not cite.

Urban Big Data Centre

An ESRC Data Investment

Active travel

- Walking and cycling can generate large benefits
  - Reduced congestion
  - Reduced emissions
  - Improve health
  - Time savings
- Transport Scotland wants 10% of journeys to be made by bicycle by 2020; with cities responsible for achieving this

Do interventions work?

- Evaluating the effectiveness of interventions is difficult due to the lack of data
- Manual counts take place on specific links/points, but these are expensive and hence infrequent
- Automatic counters can be used but these are also expensive and tend to be sparsely located
- Maintenance and calibration is required to keep them working properly

New data

- Activity tracking apps are used by many people and provide valuable new data about activities
- The Strava cycling app uses GPS to track cyclists' journeys
- This offers the possibility of having data at a fine spatial and temporal scale for a large number of people
- The data are already being collected all over the world

University of Glasgow

STRAVA™

- The name is taken from the Swedish word sträva, meaning to strive
- It can be used to track running and cycling activities
- Users can track their activities over time and compare to the activities of their friends or the user community
- Users can also compete in competitions
- The app comes in a free and premium version. The premium version offers extra features and costs £5.99 a month or £49.99 per year

University of Glasgow

STRAVA™

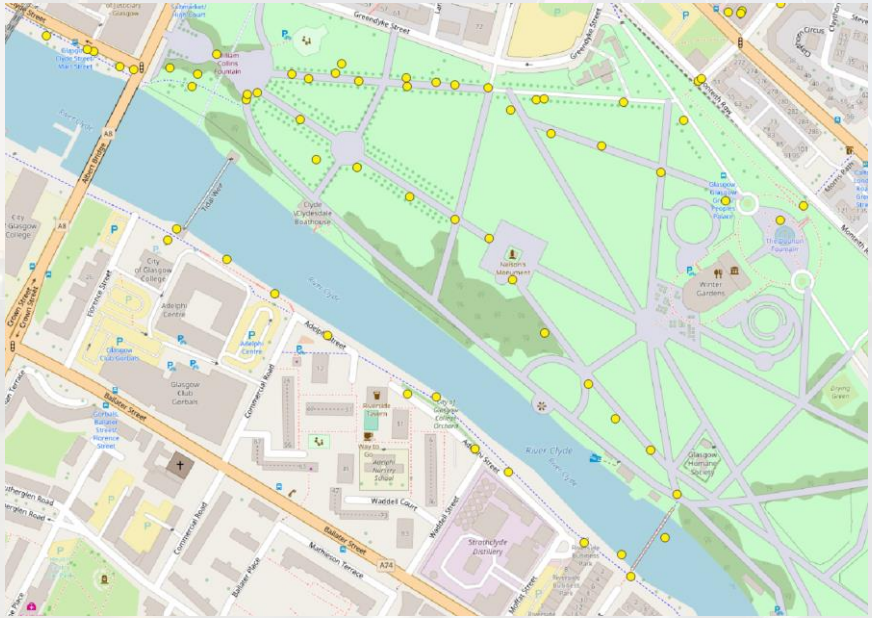
University of Glasgow

STRAVA™

- Users have to start and stop the tracking
- They can tag whether or not their trip is a commute or not
- Strava also gather some demographic information about their users

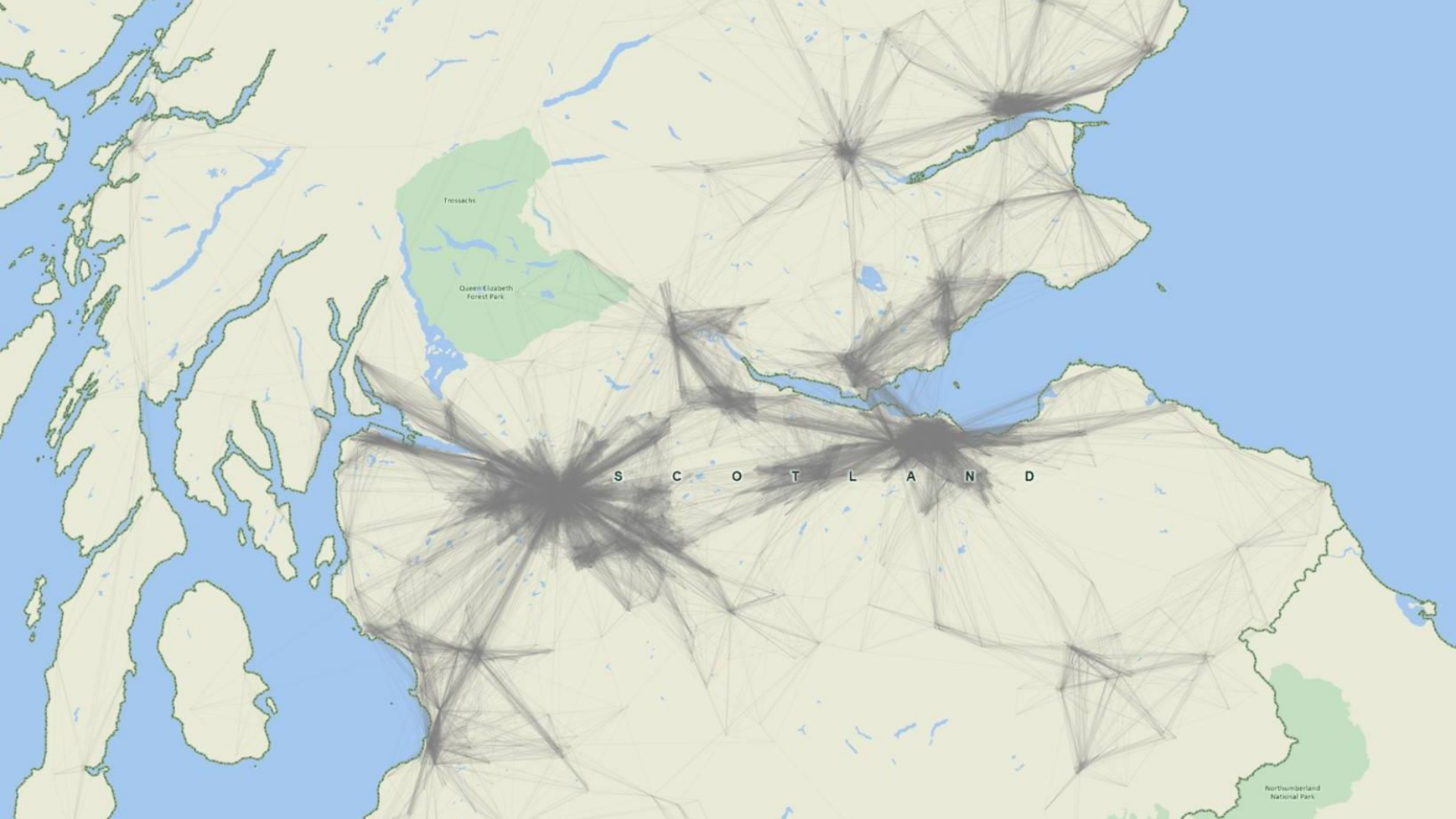
Data

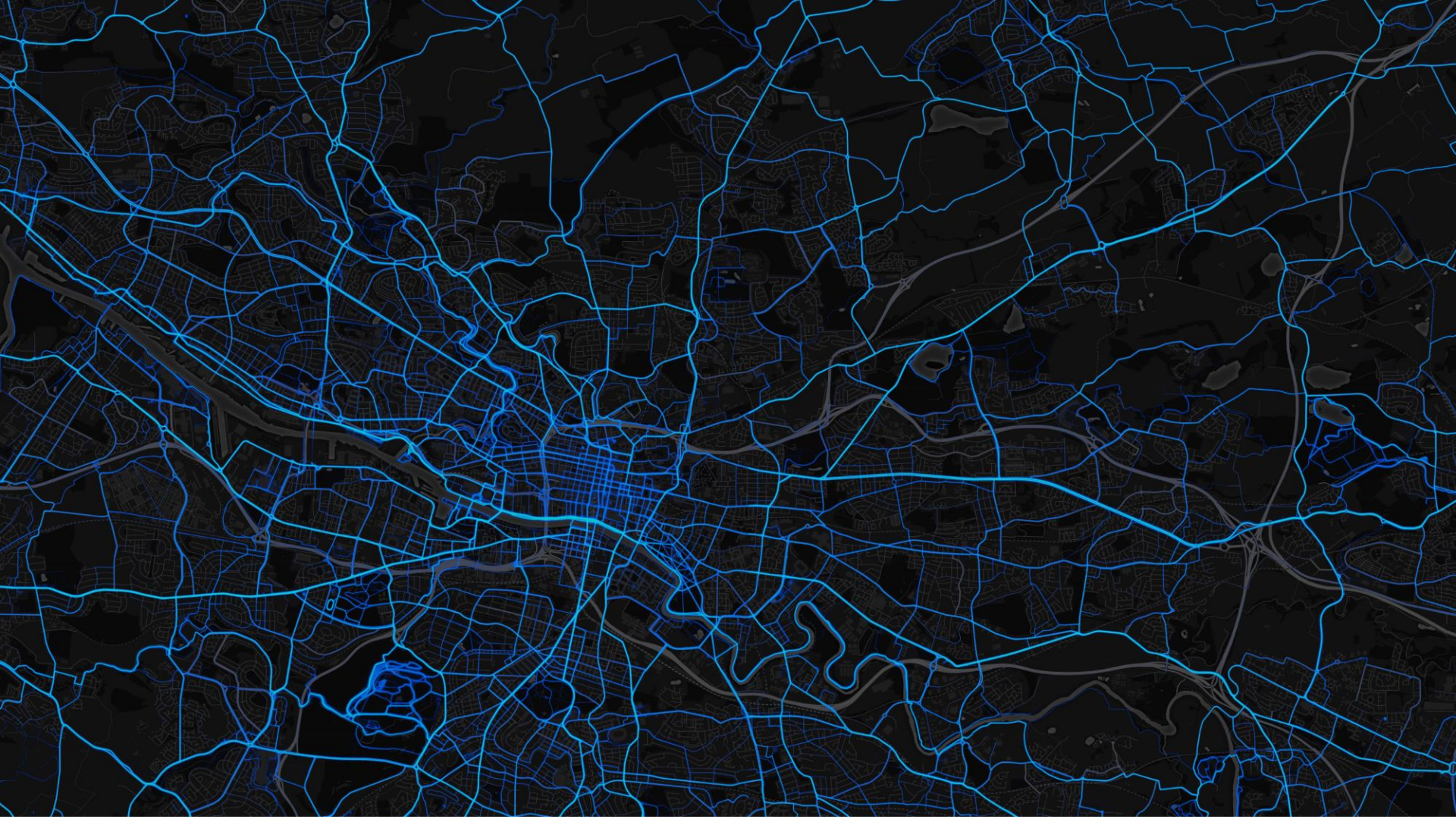
- The movement data collected by the app is raw GPS trajectories represented as a triple (latitude, longitude, timestamp)

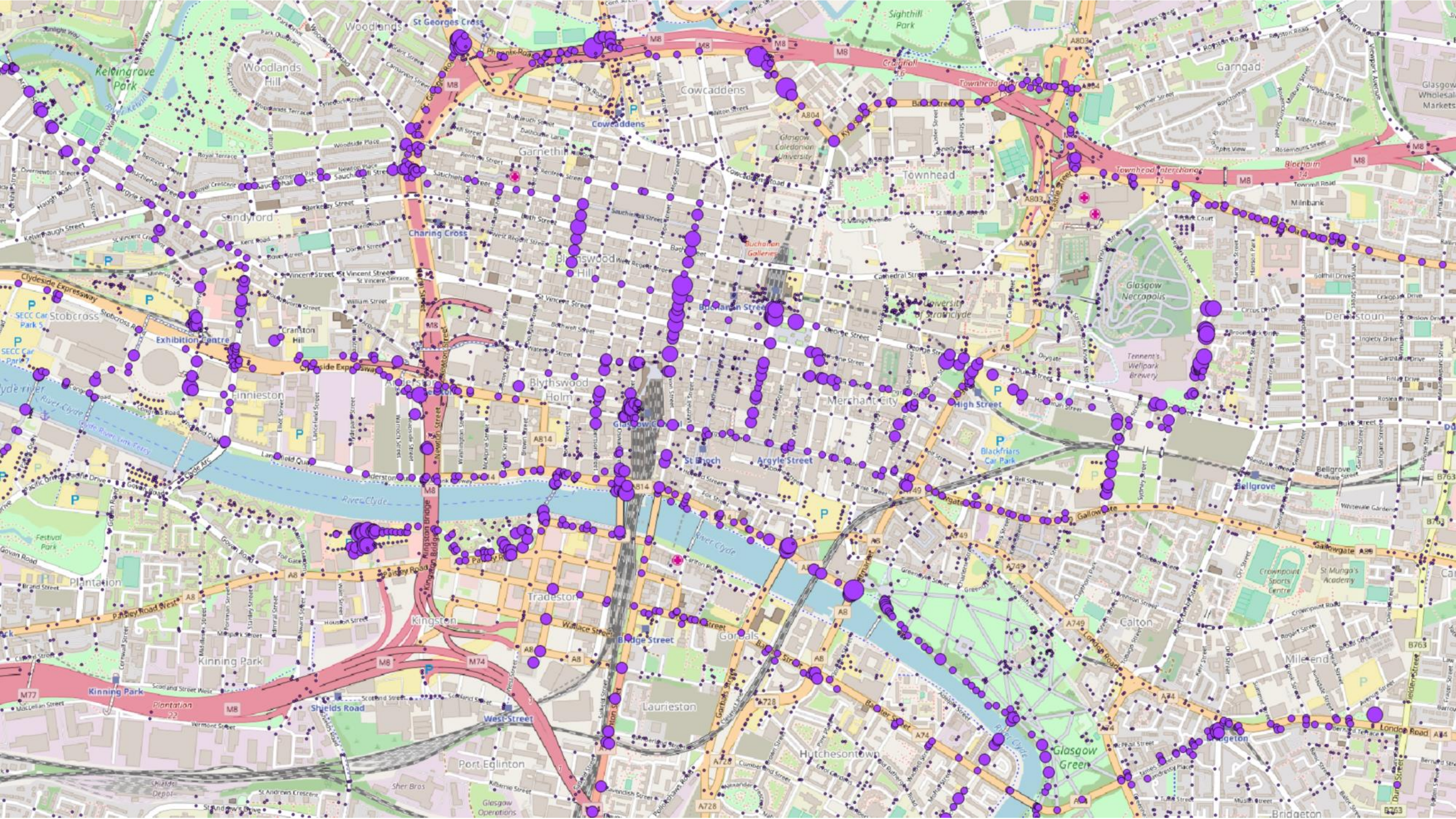


Data

- The GPS trajectories are not made available to researchers
- The data is aggregated and provided to researchers/planners through Strava Metro
- 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


Trossachs
Queen Elizabeth Forest Park
S C O T L A N D
Northumberland National Park





Woodlands
St Georges Cross
Sighthill Park
Kelvingrove Park
Woodlands Hill
Cowcaddens
Garngad
Glasgow Wholesale Markets
Garnethill
Townhead
Townhead Interchange
Blochairn
Sandyford
Charing Cross
Buchanan Galleries
Milnbank
Blythswood Hill
Glasgow Necropolis
Dennistoun
Stobcross
Exhibition Centre
Cranston Hill
Buchanan Street
University of Strathclyde
Tennent's Wellpark Brewery
SECC Car Park 5
Finnieston
Anderston
Blythswood Holm
Glasgow Central
Merchant City
High Street
River Clyde
Blackfriars Car Park
St Enoch
Argyle Street
Bellgrove
Festival Park
Plantation
Tradeston
Kingston
Gorbals
Calton
Crownpoint Sports Centre
St Mungo's Academy
Bridge Street
Mile End
Kinning Park
Laurieston
Shields Road
West Street
Hutchesontown
Port Eglinton
Glasgow Green
Bridgeton
Kingston Bridge
Clydeside Expressway
Gallowgate
London Road
M8
M74
M77
A8
A803
A804
A814
A749
A728
A74
B763

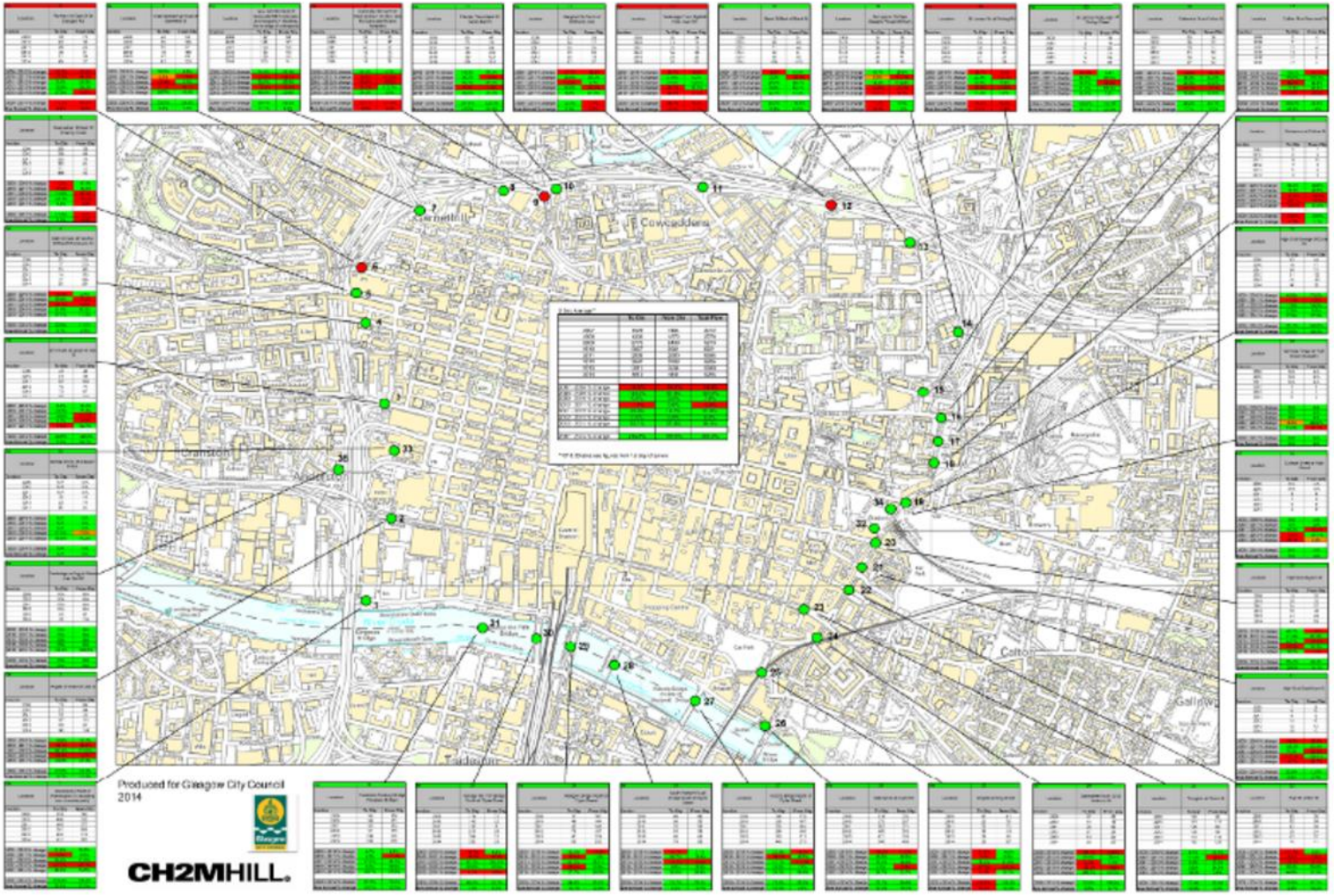
Problems

- We know not all cyclists use the app for every journey
- It is unlikely that a random sample of cyclists use the Strava app
- In Glasgow in 2015 there were 13,684 athletes who recorded 287,833 activities
- The median distance was 14.9 km
- Of this sample 11,216 were male (1,698 female)
- Can the sample tell us anything useful?

Our approach

- Firstly, we can visualise the data and do a basic sanity check; the patterns look like what we would expect given our knowledge of Glasgow
- We can compare it to other sources of data
- We use the annual two-day cordon counts which are conducted in Glasgow city centre
- We match the links where the counts take place to the same link in the Strava data

City Centre Cordon - Cycling Results Summary Plan 2014



Cordon Count (CC)

- Cycle trips are counted in blocks of 30 minutes for 14 hours over two days in September each year
- We use data from 2013, 2014 and 2015
- We aggregate both the CC and the Strava data into four different temporal scales, specifically by:
  - Hour
  - Commuting time (peak hours versus non-peak)
  - Day
  - Two-day (i.e. annual)

University of Glasgow

Correlations

	Sample size	Correlation
Hourly	3192	0.781
Peak Vs Non-peak	684	0.861
One day	228	0.882
Two days	114	0.887

Further work

- We have some additional hypothesis about how these correlations my vary:
  - Does Strava have a higher market share in rich areas e.g. the West End of Glasgow?
  - Is the market share of Strava changing over time?
  - Does the weather affect the percentage of cyclists using Strava?
  - Does the time of day affect the share of cyclists using Strava?

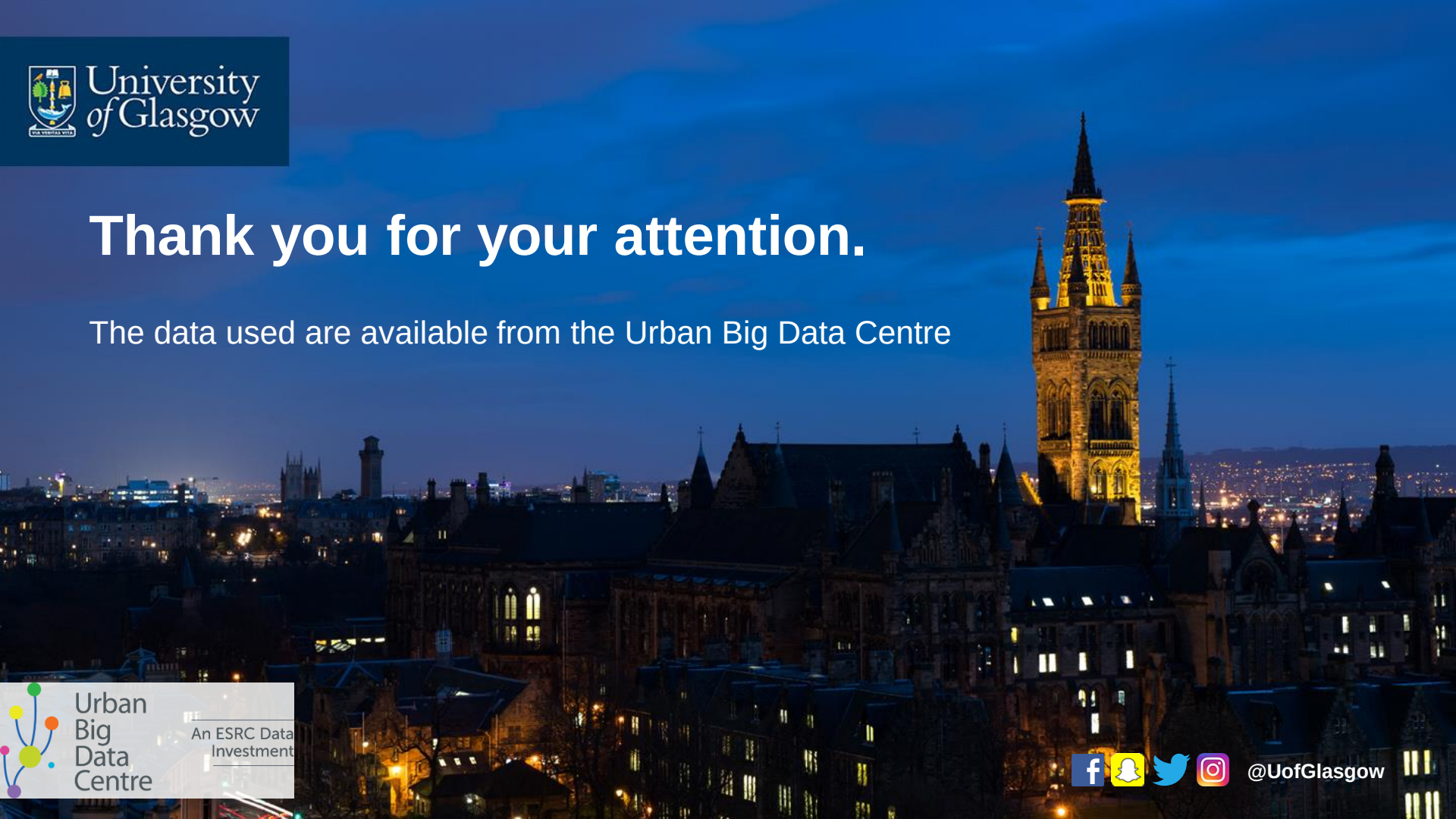
Models

- We have experimented with negative binomial regression models
- The number of total cyclists is modelled as a function of, among other things, the number of Strava cyclists
- This allows us to explore the factors influencing the link between the cycling flows
- It also allows us to adjust the Strava flows to an estimate of total flows across the network

Independent	Model 1		Model 2		Model 3		Model 4	
	Coeffecient	P=	Coeffecient	P=	Coeffecient	P=	Coeffecient	P=
Strava	0.084	0.000***	0.265	0.000***	0.098	0.000***	0.105	0.000***
Commuting (ref non-commuting)								
AM	0.317	0.001***	0.506	0.000***	0.305	0.001***	0.177	0.055
PM	0.553	0.000***	0.823	0.000***	0.542	0.000***	0.449	0.000***
Year (ref:2013)								
Year (2014)	0.162	0.077	0.154	0.086	0.147	0.140	0.125	0.162
Year (2015)	0.046*	0.619	0.001	0.989	0.141	0.152	0.007	0.938
Region (ref:east)								
North	0.074	0.468	0.083	0.409	0.083	0.419	-0.099	0.387
South	0.318	0.002	0.361	0.000***	0.320	0.002**	0.149	0.194
West	0.731	0.000	0.695	0.000***	0.742	0.000***	0.927	0.000***
Interactions								
Strava*am			-0.181	0.000***				
Strava*pm			-0.200	0.000***				
Strava*2014					0.001	0.946		
Strava*2015					-0.030	0.013*		
Strava*North							0.102	0.001**
Strava*South							0.037	0.033
Strava*West							-0.057	0.000***
Intercept (con)	3.057	0.000	2.882	0.000***	3.028	0.000	3.099	0.000***
Dispersion	1.074		1.120		1.081		1.141	

Conclusions

- Strava shows good correlation with observed cycle counts
- The correlation is higher the more we aggregate the observations
- These correlations change depending on different factors
- This seems to correspond with what has been found in the literature


University of Glasgow
Thank you for your attention.
The data used are available from the Urban Big Data Centre
Urban Big Data Centre
An ESRC Data Investment
@UofGlasgow