Too Old? Updating Survey Transport Mode Estimates Using Mobile Phone Data

Francisco Rowe

Department of Geography and Planning, University of Liverpool, UK

www.franciscorowe.com

@fcorowe

https://doi.org/10.48550/arXiv.2204.09482

Feel Old Yet? Updating Mode of Transportation Distributions from Travel Surveys using Data Fusion with Mobile Phone Data

Eduardo Graells-Garrido[1,2], Daniela Opitz[3], Francisco Rowe[4], Jacqueline Arriagada[5]

[1]Department of Computer Science, University of Chile; [2]National Center for Artificial Intelligence Research; [3]Data Science Institute, Faculty of Engineering, Universidad del Desarrollo; [4]Department of Geography and Planning, University of Liverpool; [5]University of Leeds

For correspondence:
egraells@dcc.uchile.cl

Data availability: The Telefónica Movistar mobile phone records have been obtained directly from the mobile phone operator through an agreement between the Data Science Institute from Universidad del Desarrollo and Telefónica R&D. This mobile phone operator retains ownership of these data and imposes standard provisions to their sharing and access which guarantee privacy. Anonymized datasets are available from Telefónica R&D Chile for researchers who meet the criteria for access to confidential data. The other datasets used in this study are publicly available.

Acknowledgments: We thank GPA Movistar for facilitating the data for this study, and Leo Ferres for insightful discussion. E. Graells-Garrido was partially funded by ANID Fondecyt #1211490. D. Opitz was partially funded by ANID #PAI77190057 and ANID Fondecyt de Iniciación #11220799.

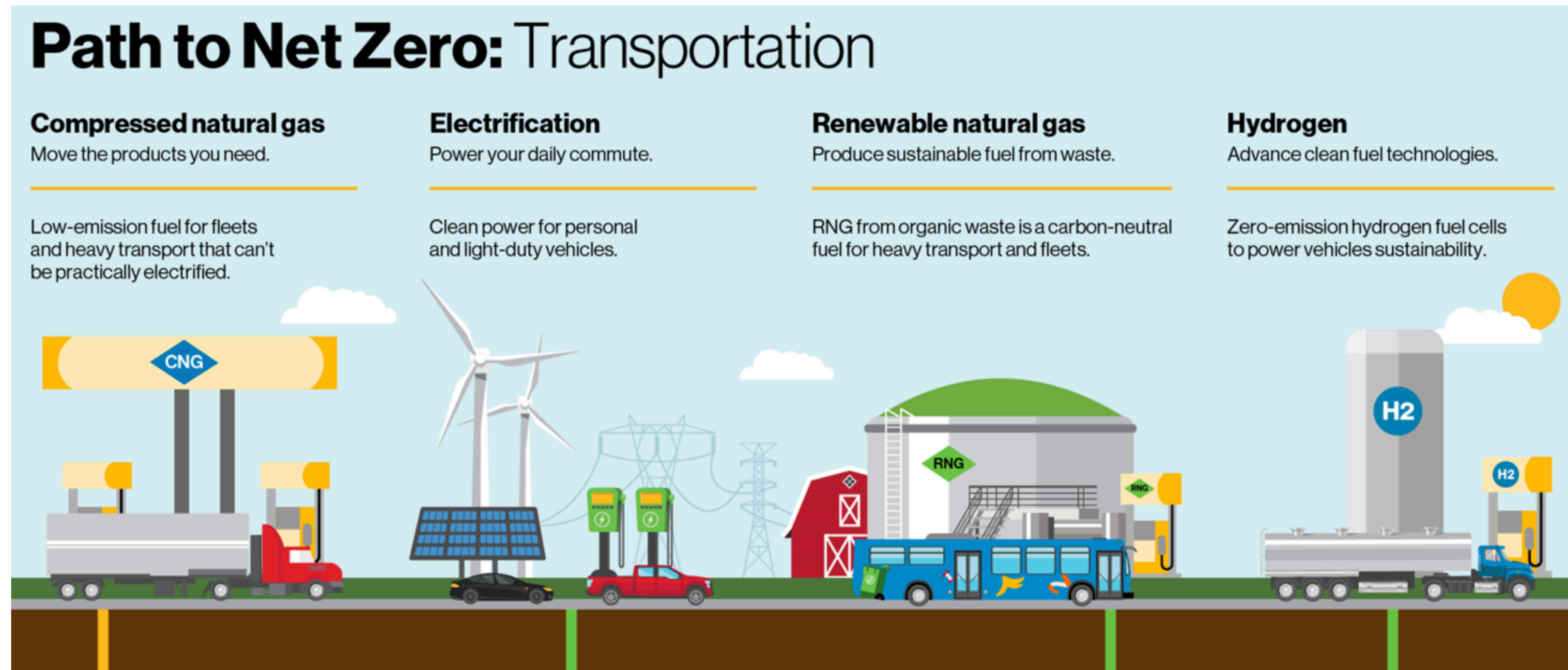
Competing interests: The author declare no competing interests.

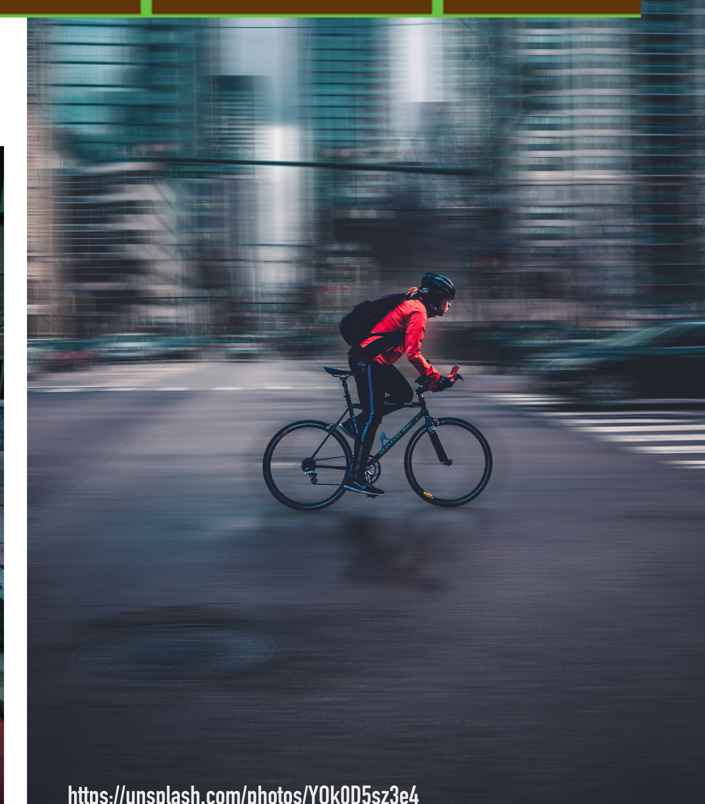
Abstract

Up-to-date information on different modes of travel to monitor transport traffic and evaluate rapid urban transport planning interventions is often lacking. Transport systems typically rely on traditional data sources providing outdated mode-of-travel data due to their data latency, infrequent data collection and high cost. To address this issue, we propose a method that leverages mobile phone data as a cost-effective and rich source of geospatial information to capture current human mobility patterns at unprecedented spatiotemporal resolution. Our approach employs mobile phone application usage traces to infer modes of transportation that are challenging to identify (bikes and ride-hailing/taxi services) based on mobile phone location data. Using data fusion and matrix factorization techniques, we integrate official data sources (household surveys and census data) with mobile phone application usage data. This integration enables us to reconstruct the official data and create an updated dataset that incorporates insights from digital footprint data from application usage. We illustrate our method using a case study focused on Santiago, Chile successfully inferring four modes of transportation: mass-transit (all public transportation), motorised (cars and similar vehicles), active (pedestrian and cycle trips), and taxi (traditional taxi and ride-hailing services). Our analysis revealed significant changes in transportation patterns between 2012 and 2020. We quantify a reduction in mass-transit usage across municipalities in Santiago, except where metro/rail lines have been more recently introduced, highlighting added resilience to the public transport network of these infrastructure enhancements. Additionally, we evidence an overall increase in motorised transport throughout Santiago, revealing persistent challenges in promoting urban sustainable transportation. Findings also point to a rise in the share of taxi usage, and a drop in active mobility, suggesting a modal shift towards less sustainable modes of travel. We validate our findings comparing our updated estimates with official smart card transaction data. The consistency of findings with expert domain knowledge from the literature and historical transport usage trends further support the robustness of our approach.

Context

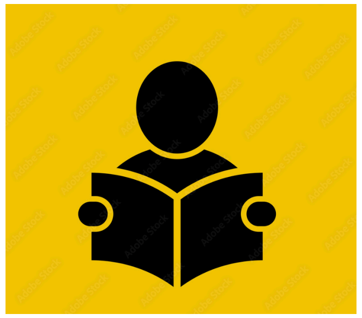
Up-to-date information on different modes of travel is **key but often lacking**





Old Ways

Manual count



Travel surveys

Censuses

BUT...expensive, infrequent and often outdated

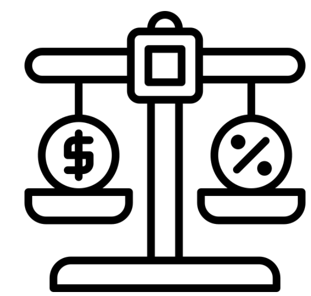
New Ways

Mobile phones

Real-time

High granularity

Cost-effective



BUT...no population attributes

Aim

Develop a data fusion approach to update mode split estimates (mass-transit, taxi, motorised & active transport) by integrating mobile phone data, household survey and census data

Estimate → Update → Validate

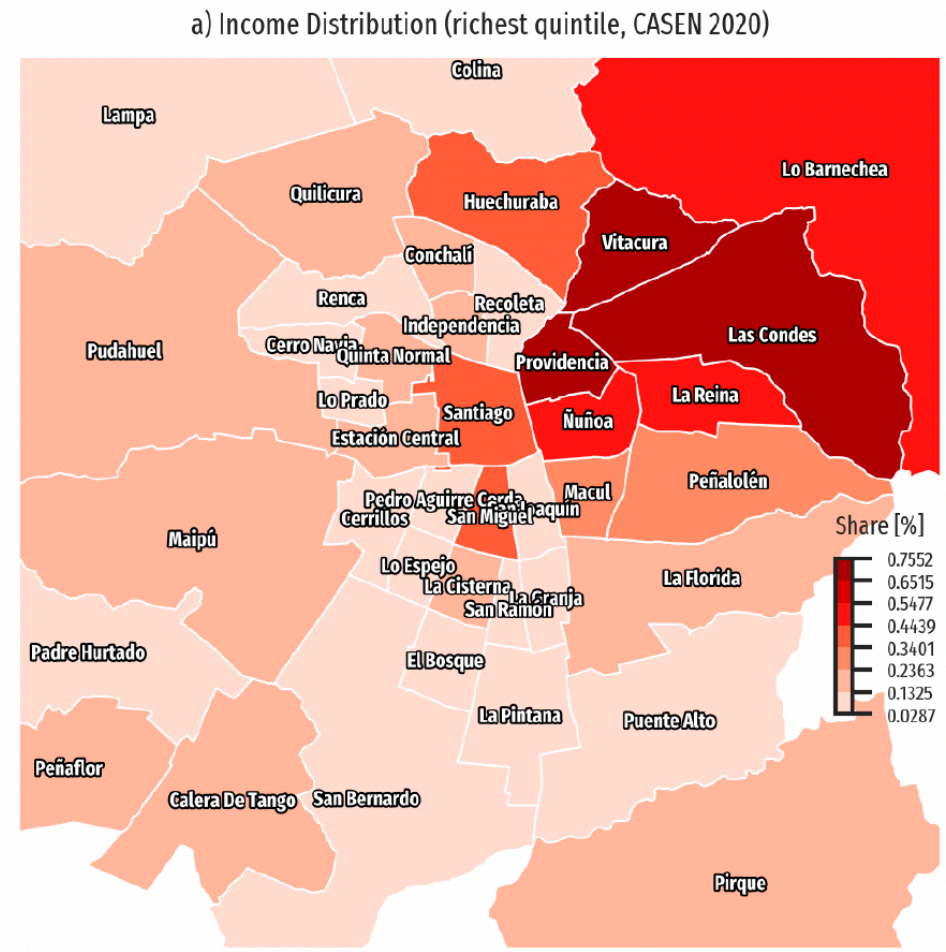
1. Mobility

2. App usage

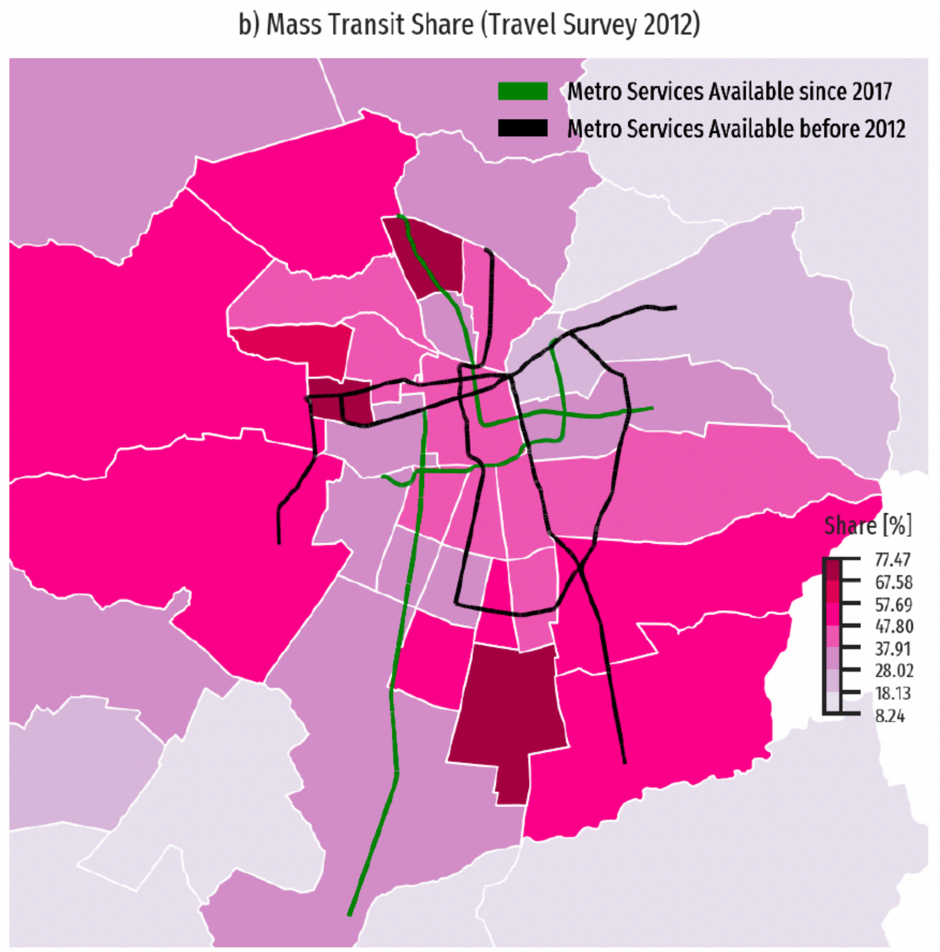
3. Data Fusion

Setting

Santiago



a) Income Distribution (richest quintile, CASEN 2020)
Colina
Lampa
Lo Barnechea
Quilicura
Huechuraba
Vitacura
Conchalí
Renca
Recoleta
Independencia
Las Condes
Pudahuel
Cerro Navia
Quinta Normal
Providencia
Lo Prado
Santiago
La Reina
Ñuñoa
Estación Central
Peñalolén
Macul
Cerrillos
San Miguel
Maipú
Share [%]
0.7552
0.6515
0.5477
0.4439
0.3401
0.2363
0.1325
0.0287
Lo Espejo
La Cisterna
La Florida
San Ramón
Padre Hurtado
El Bosque
La Pintana
Puente Alto
Peñaflor
Calera De Tango
San Bernardo
Pirque

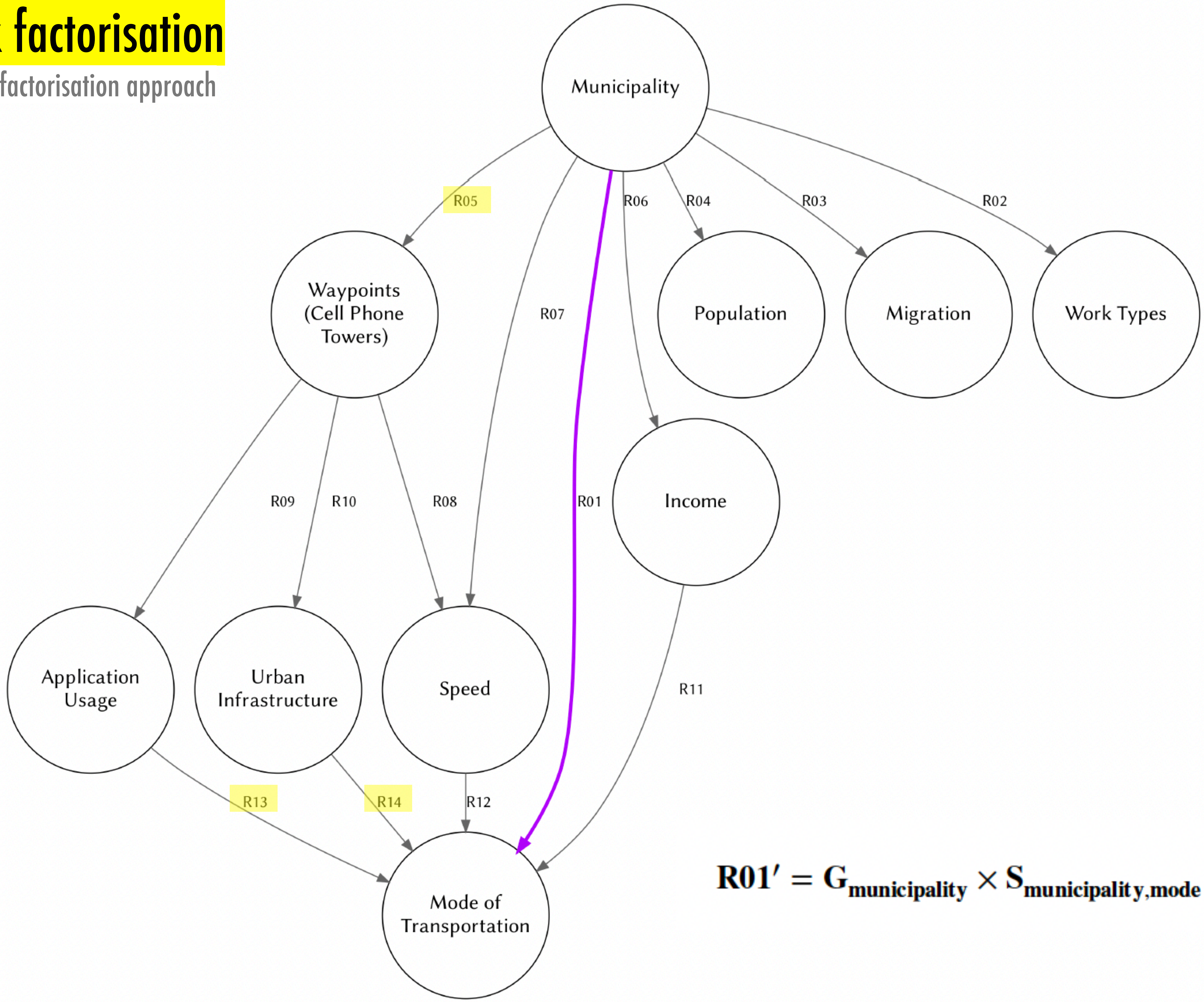


b) Mass Transit Share (Travel Survey 2012)
Metro Services Available since 2017
Metro Services Available before 2012
Share [%]
77.47
67.58
57.69
47.80
37.91
28.02
18.13
8.24

Data Fusion Framework

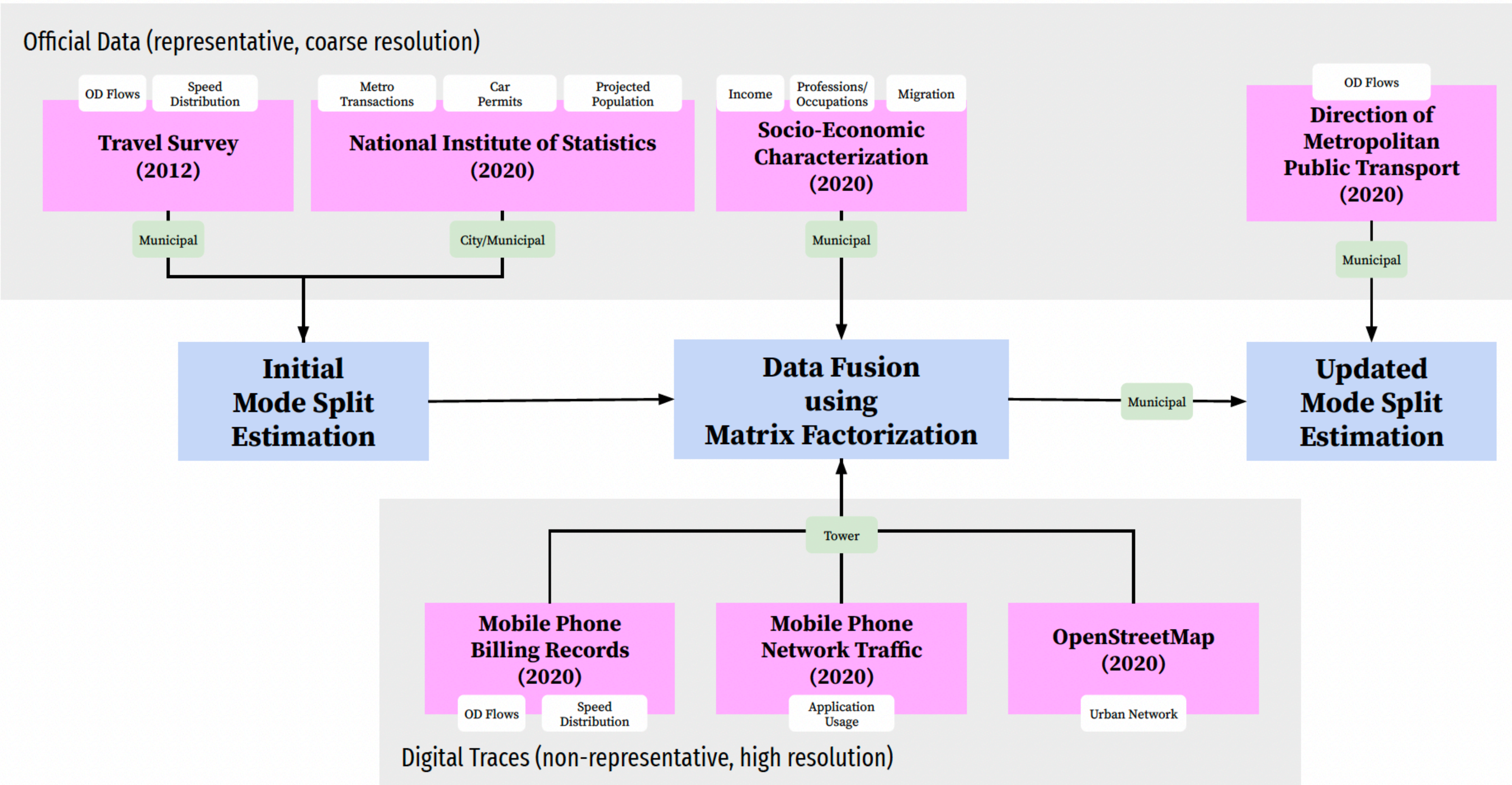
Matrix factorisation

Matrix Tri-factorisation approach



$$\mathbf{R01}' = \mathbf{G}_{\mathbf{municipality}} \times \mathbf{S}_{\mathbf{municipality,mode}} \times \mathbf{G}^{\mathbf{T}}_{\mathbf{mode}}.$$

A Closer Look



Initial Guess

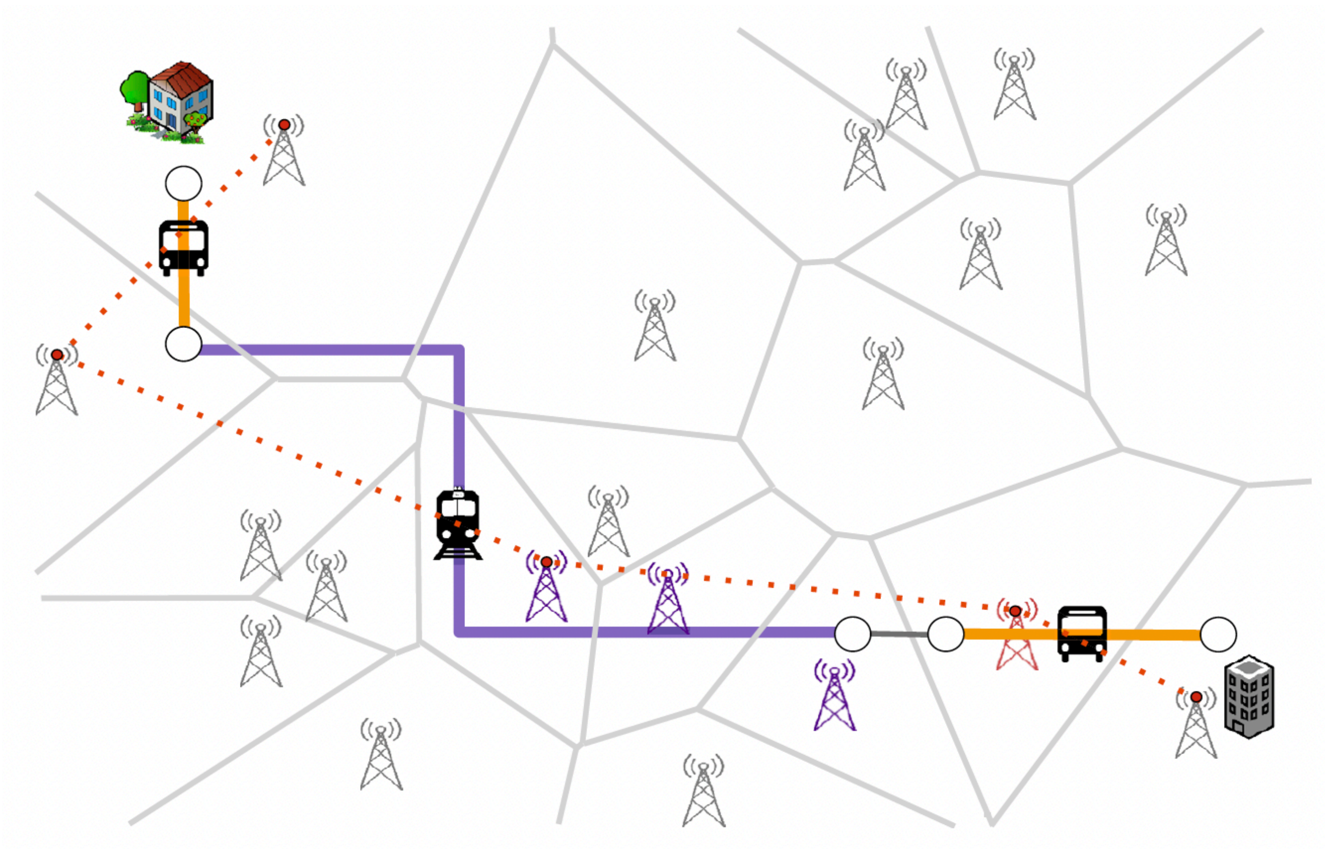
Mass transit	$$MT_{20}(m) = MT_{12}(m) \times \frac{Population_{20}(m)}{Population_{12}(m)} \times \sqrt{\frac{Metro_{20}}{Metro_{12}}}$$
Motorised	$$C_{20}(m) = C_{12}(m) \times \frac{Population_{20}(m)}{Population_{12}(m)} \times \sqrt{\frac{Permits_{20}(m)}{Permits_{12}(m)}}$$
Active	$$A_{20}(m) = A_{12}(m) \times \frac{Population_{20}(m)}{Population_{12}(m)} \times 0.975$$
Taxi	$$T_{20}(m) = (T_{12}(m) + 1) \times \frac{Population_{20}(m)}{Population_{12}(m)} \times 1.09.$$

2012 to 2020

- Metro trips declined in 43%
- Private vehicle circulation permits increased by 15%
- Population numbers have changed
- Social unrest disruption & violent crime
- Ride-hailing services (i.e. Uber)

Defining Waypoints

Telefonica

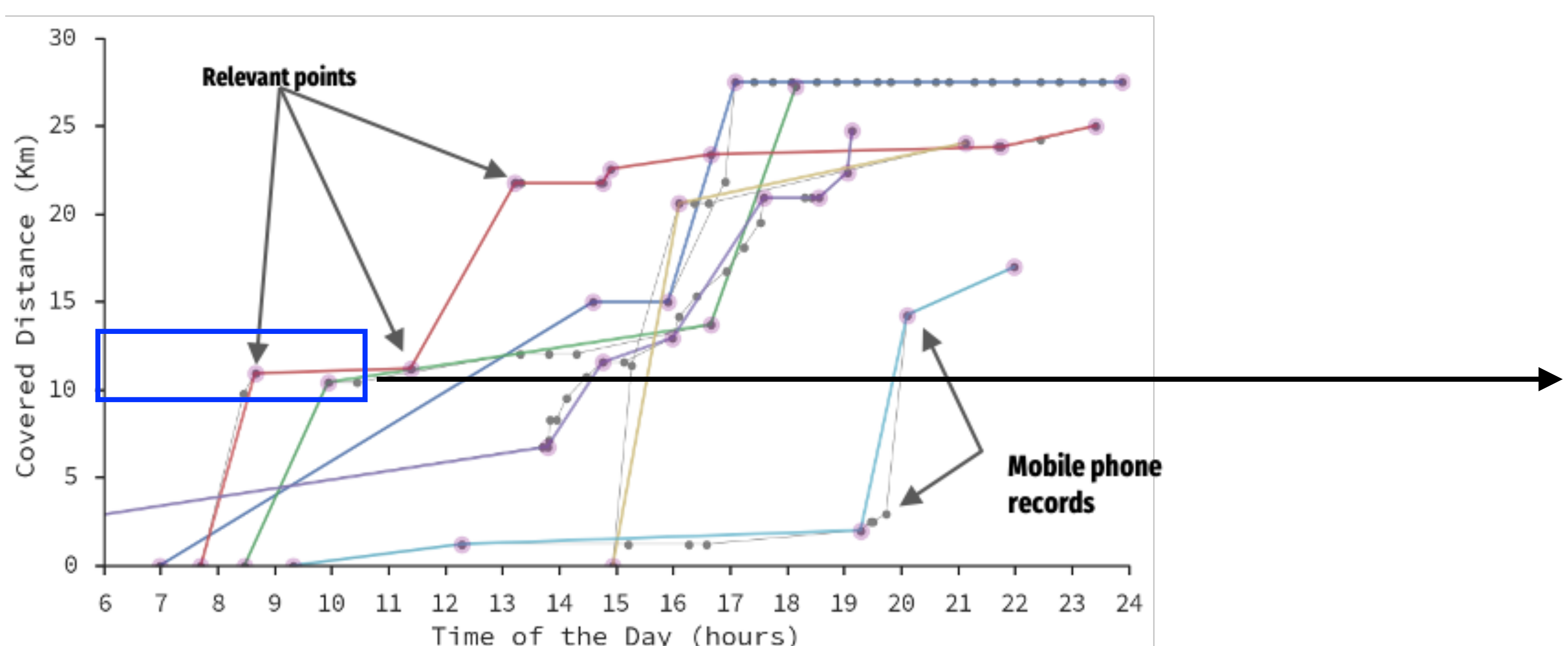


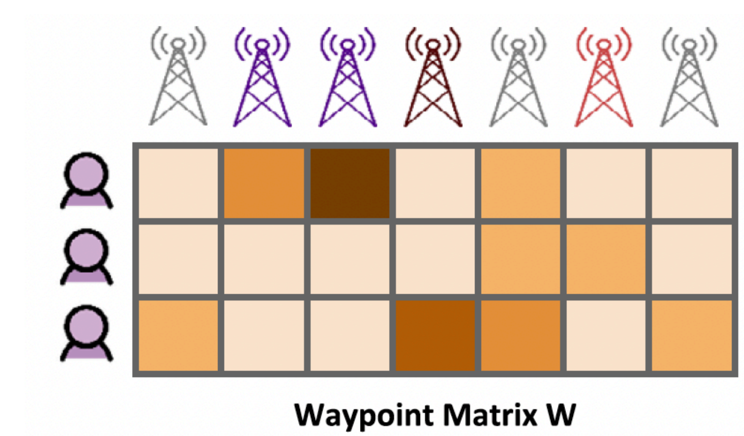
eXtended Mobile Records (XDRs) - call, SMS & downloads

- Device ID (1.8M)
- Timestamp
- Connecting tower (>2k)

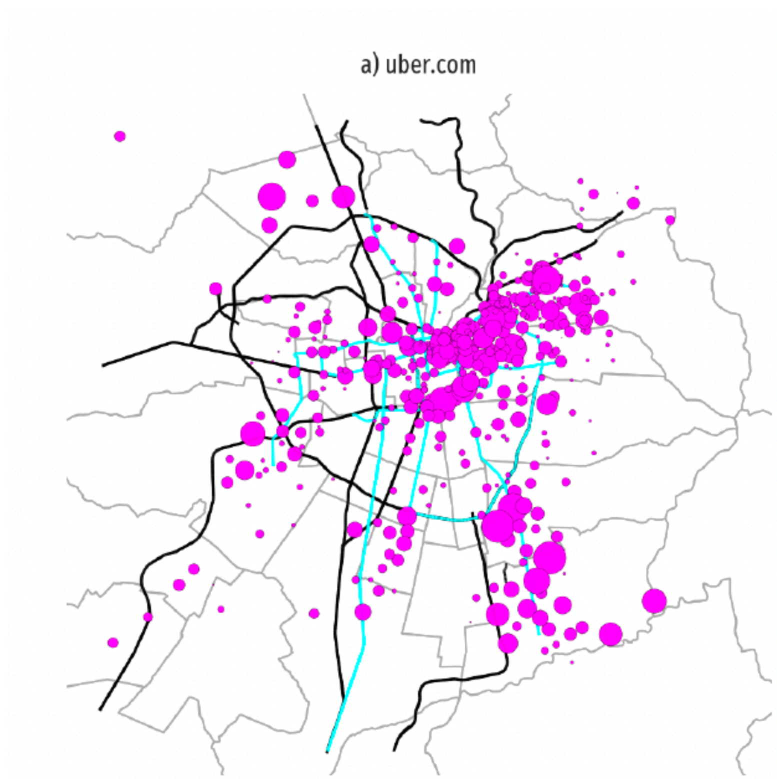
March 9-13th 2020

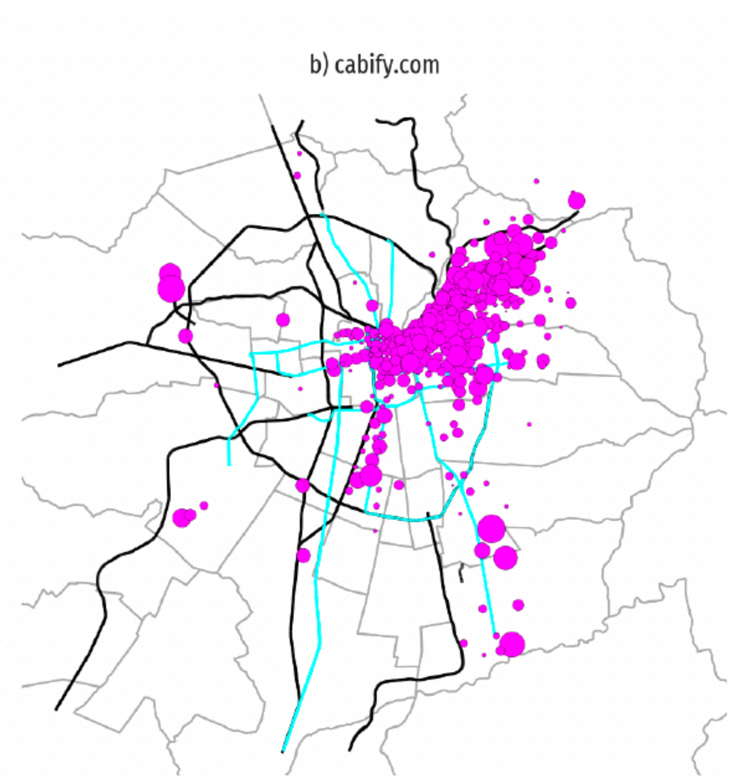
19M trips / 3M 6-9am





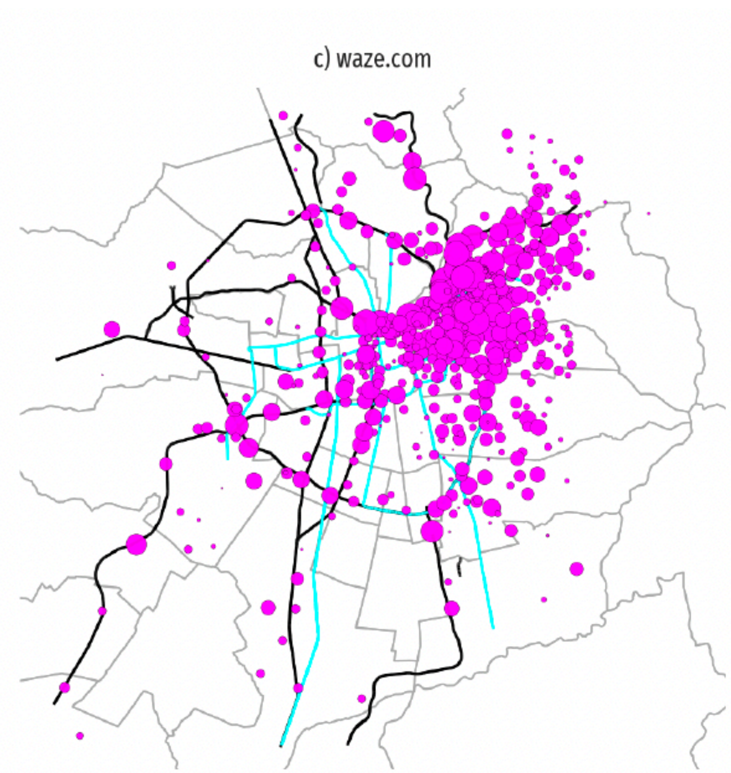
Mobile App Usage to Towers

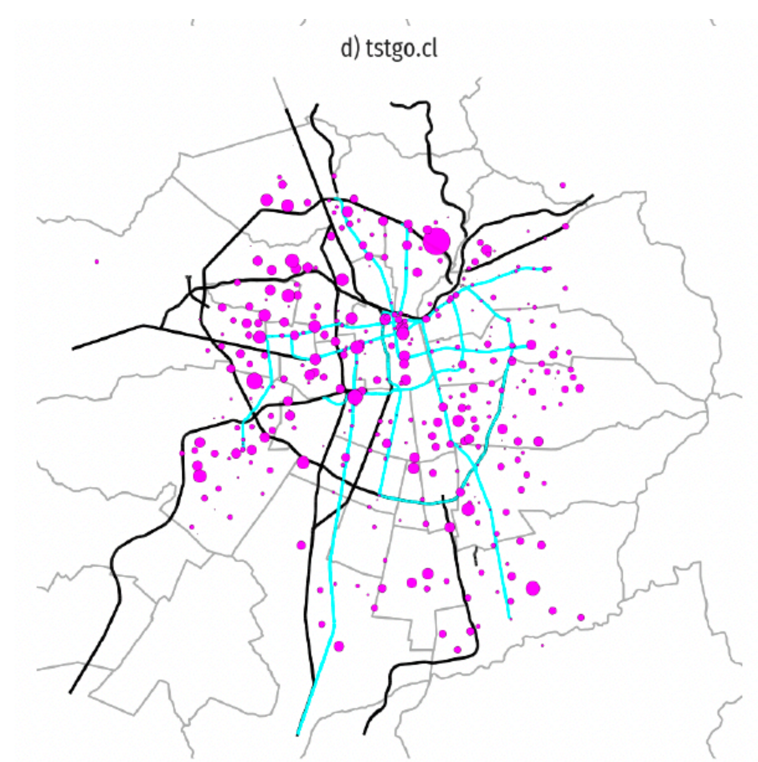




Deep Package Inspection (DPI)

- Web Domain
- Connecting tower





Road Type to Transport Mode

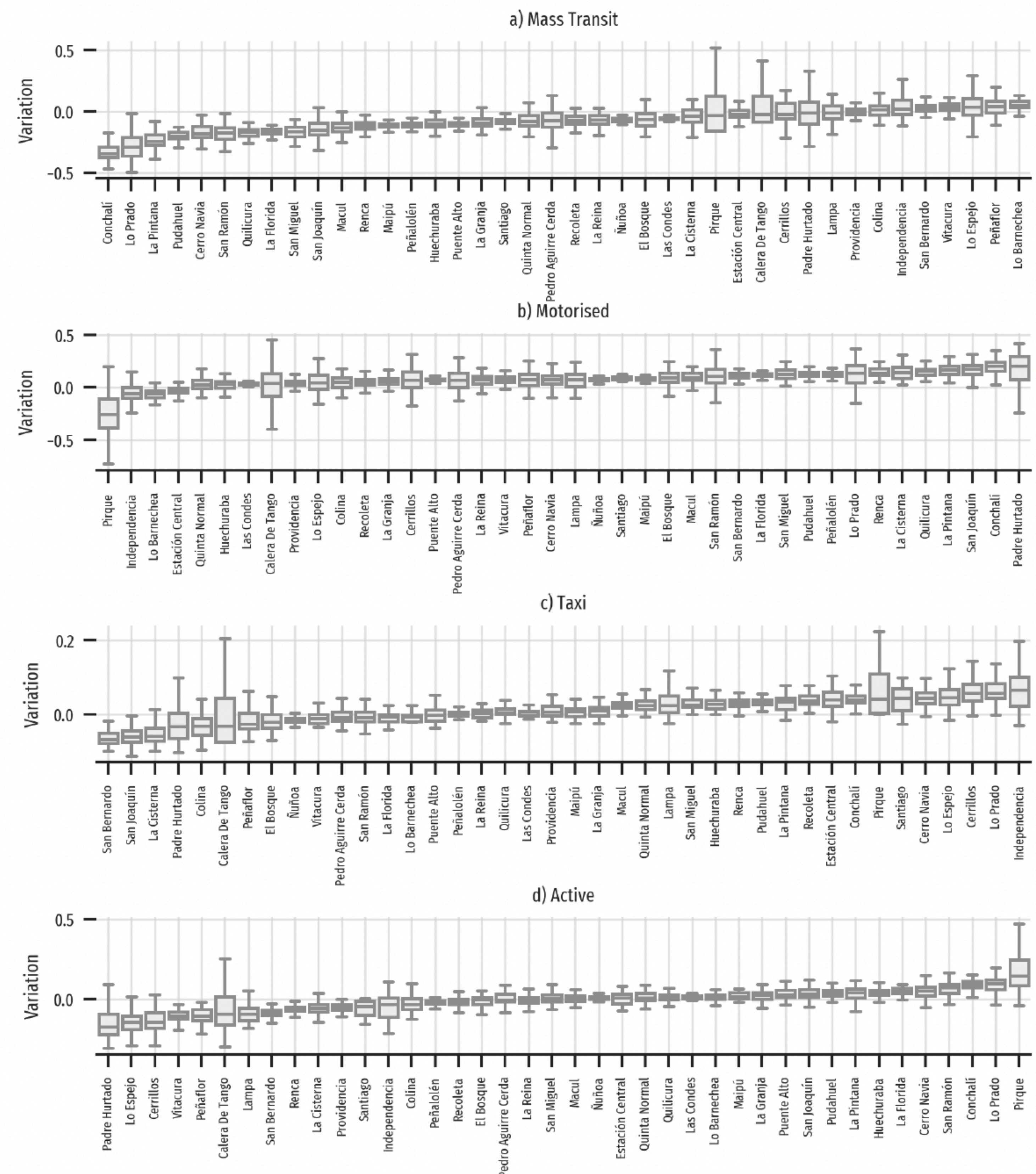
Labelling road types e.g.

- Cycleways & pedestrian streets > active
- Highways > motorised & taxi
- Mass transit > primary streets & rail/metro

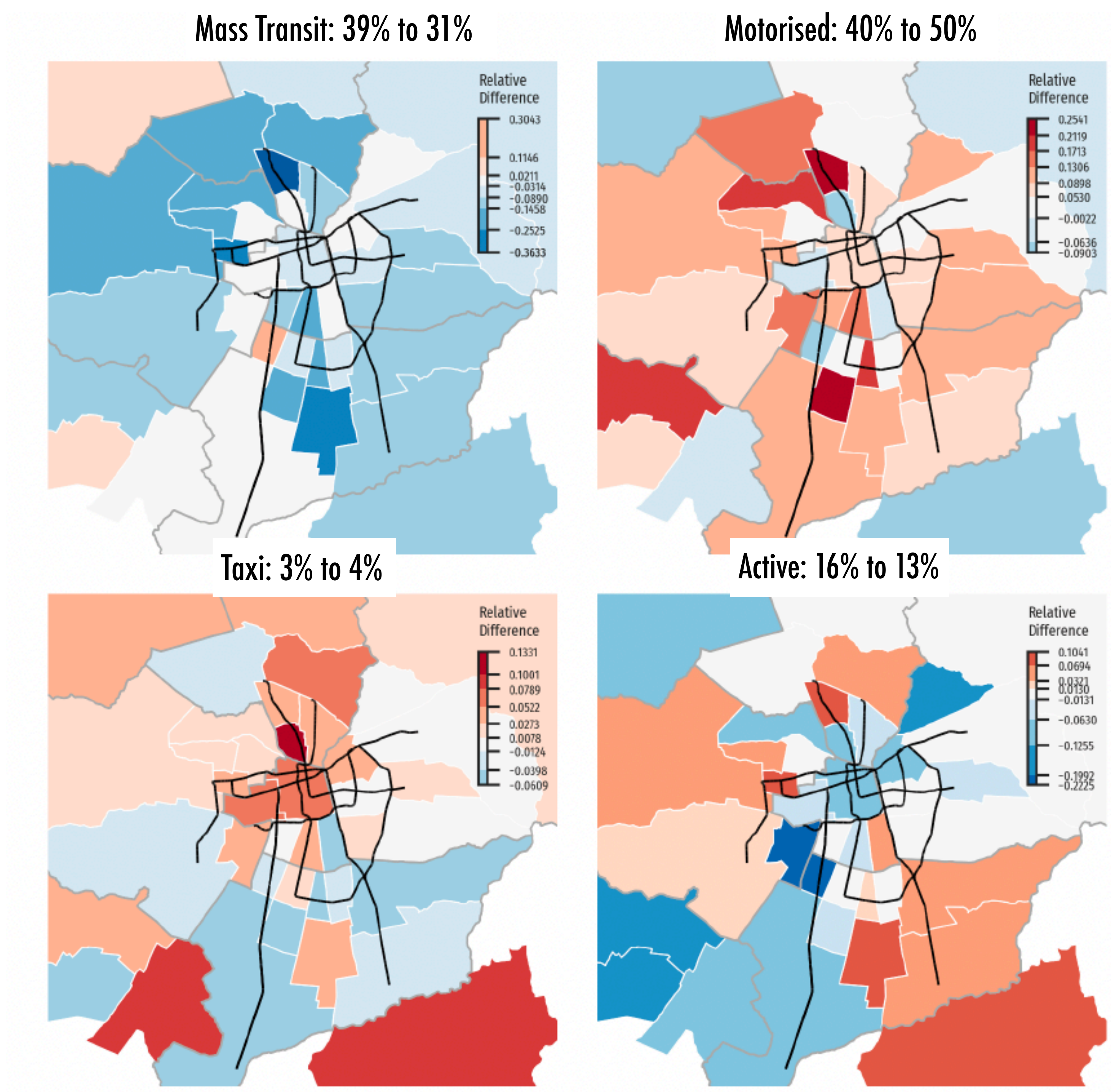
Model Selection

100 iterations

Low municipality-specific variability



Updated Estimates

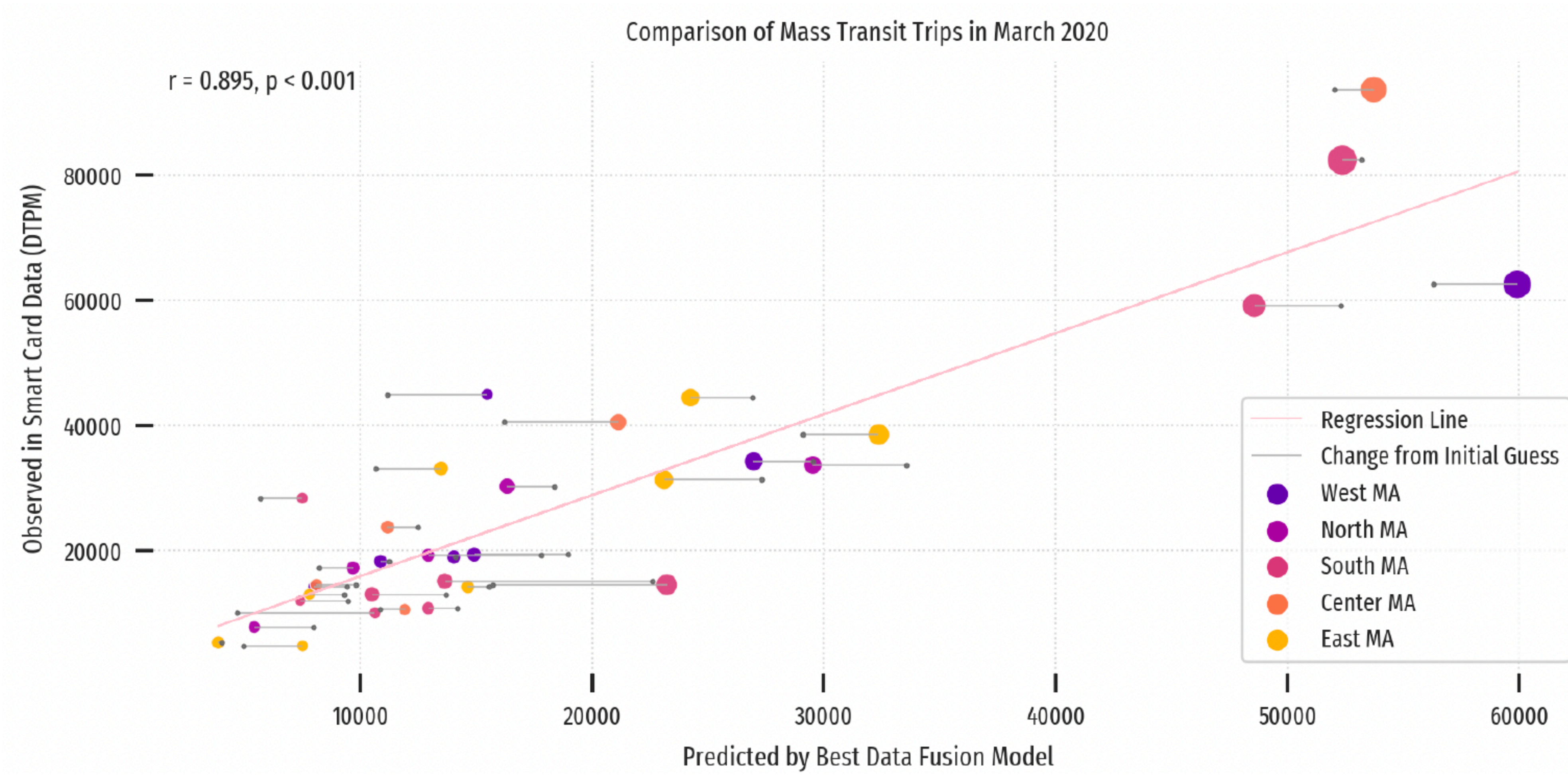


Clear overall trend **but** distinctive local patterns

Model Validation

Validation I

Smart card data

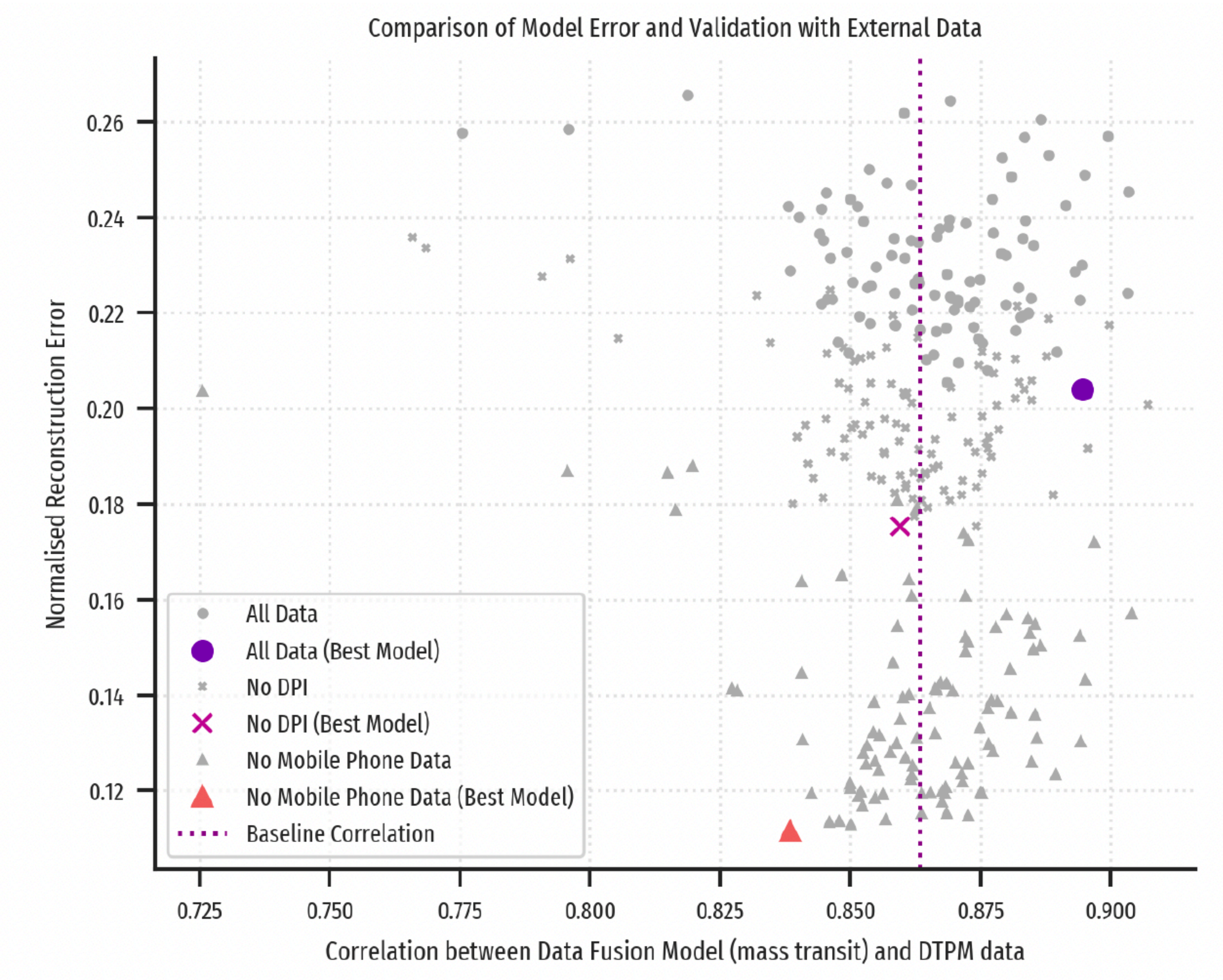


Validation II

What about no mobile phone data?

Model selection?

Error vs Correlation



Conclusion

Proposed a novel data fusion approach to produce updates of mode of transportation estimates

- incorporate multiple transportation modes
- identify bikes and ride-hailing trips

Expose key challenges in the model selection & validation process

Francisco Rowe
F.Rowe-Gonzalez@liverpool.ac.uk

www.franciscorowe.com

@fcorowe

UNIVERSITY OF LIVERPOOL

Geographic Data Science Lab