

# USO DE DATOS MOVILES PARA EL BIEN SOCIAL

**Loreto Bravo**

Director @ Instituto de Data Science, UDD  
Telefónica R&D Fellow

# Trabajo en Conjunto

- Leo Ferres (UDD)
- Eduardo Graells (UDD)
- Diego Caro (UDD)
- Daniela Opitz (UDD)
- Eric Ancelovici (Head of Telefónica R&D)
- Pablo García (BCI, former Telefónica R&D)
- Ciro Catutto (ISI Foundation)
- Daniela Paolotti (ISI Foundation)
- Mariano Beiró (Conicet)
- Carlos Rodríguez-Sickert (CICS-UDD)
- Tamas David-Barret (Oxford, UDD)
- Janos Kertesz (Budapest University of Technology)
- Sebastián Díaz (CICS-UDD)
- Denis Parra (PUC)
- Anna Rotkirch (Väestöliitto)
- Isabel Behncke (UDD)
- Y muchos más!



**Universidad del Desarrollo**  
Facultad de Ingeniería  
Instituto Data Science

# Nuestro Instituto

*Telefónica*

---

Investigación y Desarrollo Chile

# Datos: sub-producto del uso de servicio de redes

Porcentaje de Mercado  
28%-35%

Los datos se almacenan  
desde el 2015

~1.35 teléfonos por  
persona

## CDRs: registros de llamadas

Fono A	Antena A	Fono B	Antena B	Fecha	Hora	Duración	...
+56985466257	MPPM2U1	+56968544127	MOVALLO	2015-01-02	10:22	67	
+56984677521	ACA3BW1	+56985466257	MPPM2U1	2015-07-19	18:43	159	

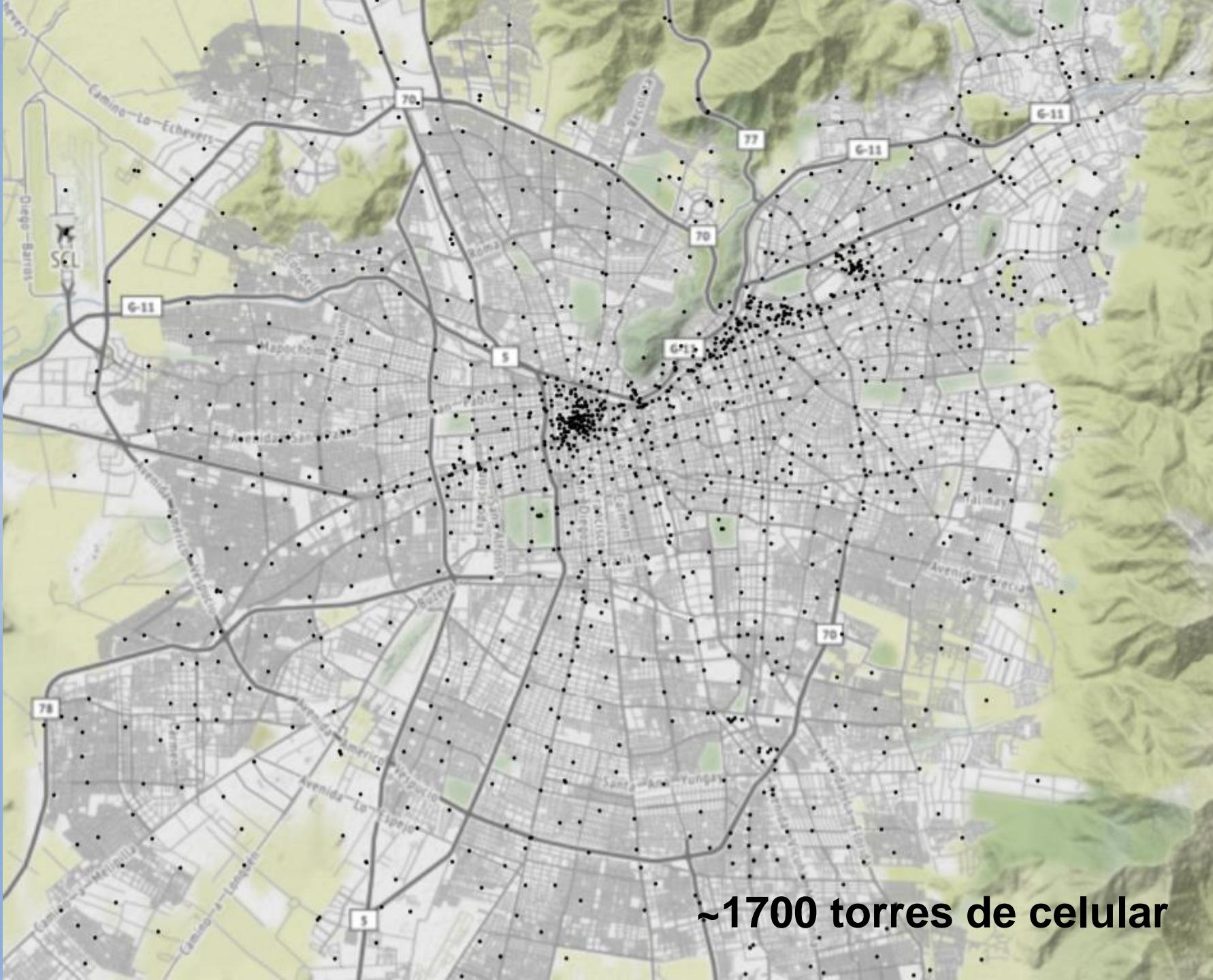
Fono A o Fono B  
pertenece a  
Telefónica



## XDRs: registros de usos de datos

Fono	Antena	Fecha	Hora	Kb	...
+56985466257	MPPM2U1	2016-06-21	7:13	250	
+56984677521	ACA3BW1	2016-06-21	8:05	1324	

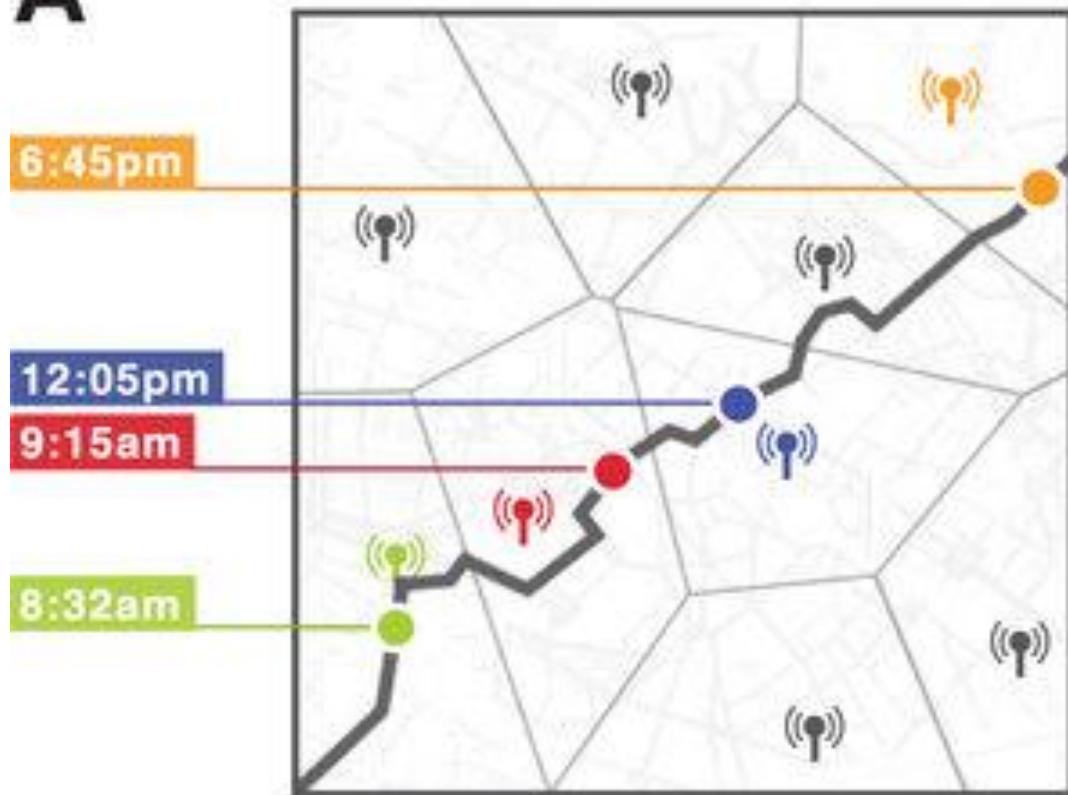
*Para algunos proyectos específicos podemos tener acceso a datos socio-demográficos anonimizados: género, edad, grupo socio-económico*



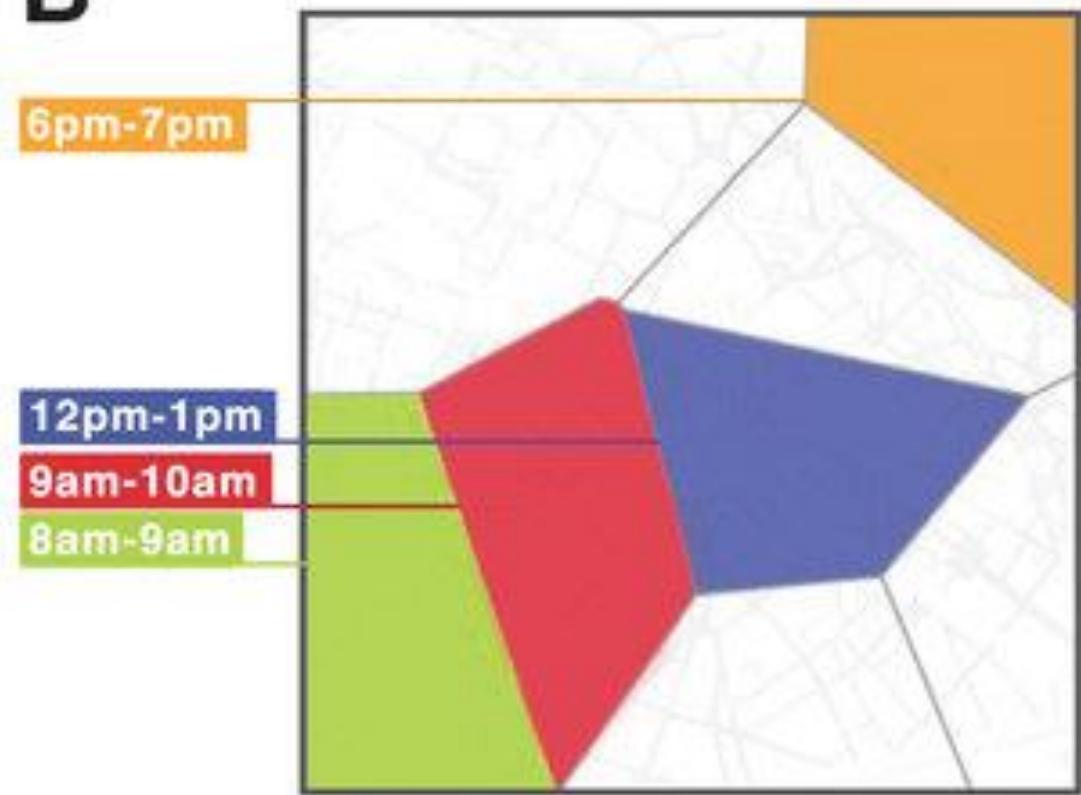
~1700 torres de celular

# Información de Ubicación

A



B

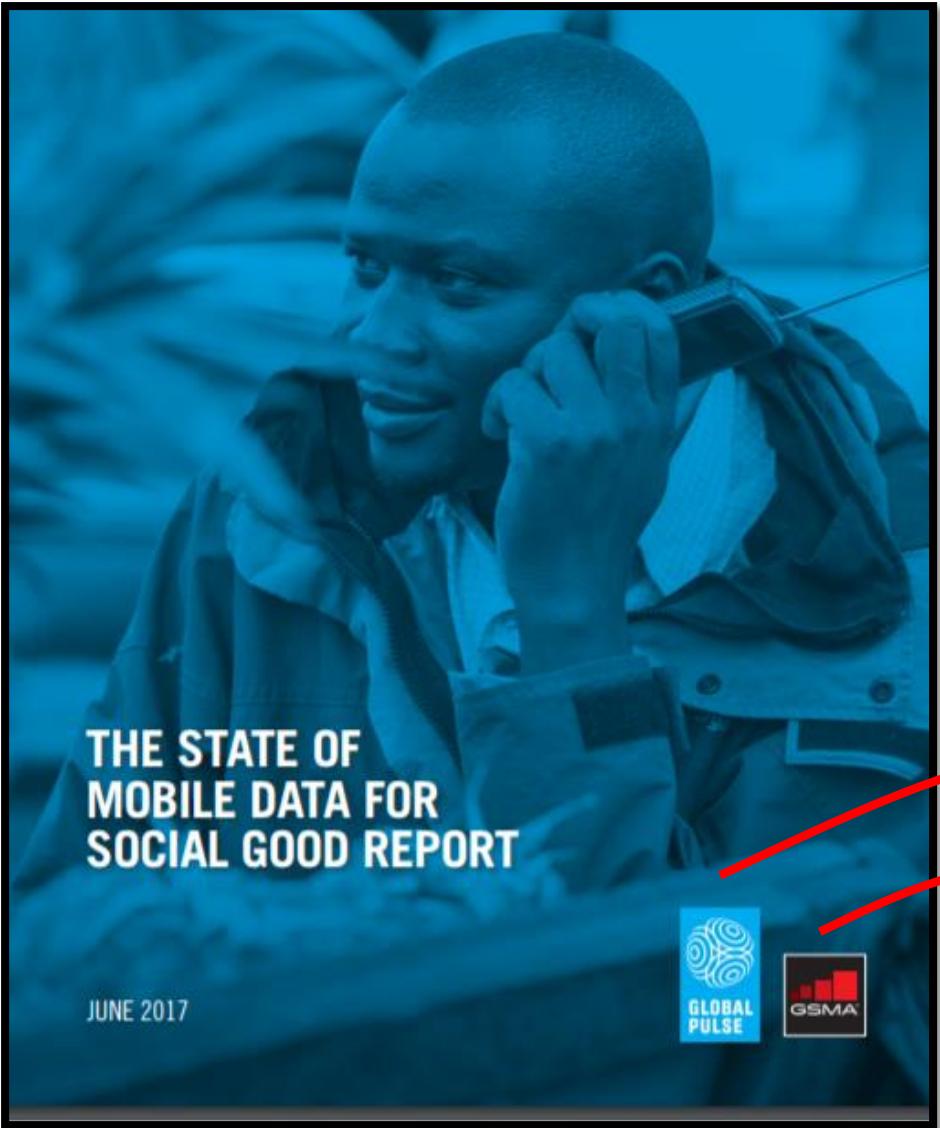


Yves-Alexandre de Montjoye et. al, "Unique in the Crowd: The privacy bounds of human mobility." Scientific Reports

# Objetivos de Desarrollo Sostenible (ODS)



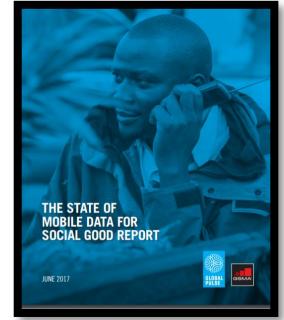
# Datos Móviles para el Bien Social



- Informe:
  - Valor:
    - Movilidad
    - Interacción Social
    - Progreso Económico
  - Desafíos
  - Recomendaciones
- United Nations
- GSMA representa a 750 operadores de telefonía móvil a nivel mundial

# Desafíos

- Falta de fuerte evidencia y comprensión para apoyar la **inversión**:
  - Responsables políticos
  - Profesionales humanitarios
  - Operadores de redes móviles
- Falta de vision compartida y de un **mecanismo de implementación** coherente
- Falta de **capacidades técnicas** a nivel mundial
- Falta de un enfoque común para la **privacidad de datos** y la mitigación de riesgos asociados con el uso de datos
- Riesgos asociados con el uso de datos: **sesgo, reidentificación, daño**

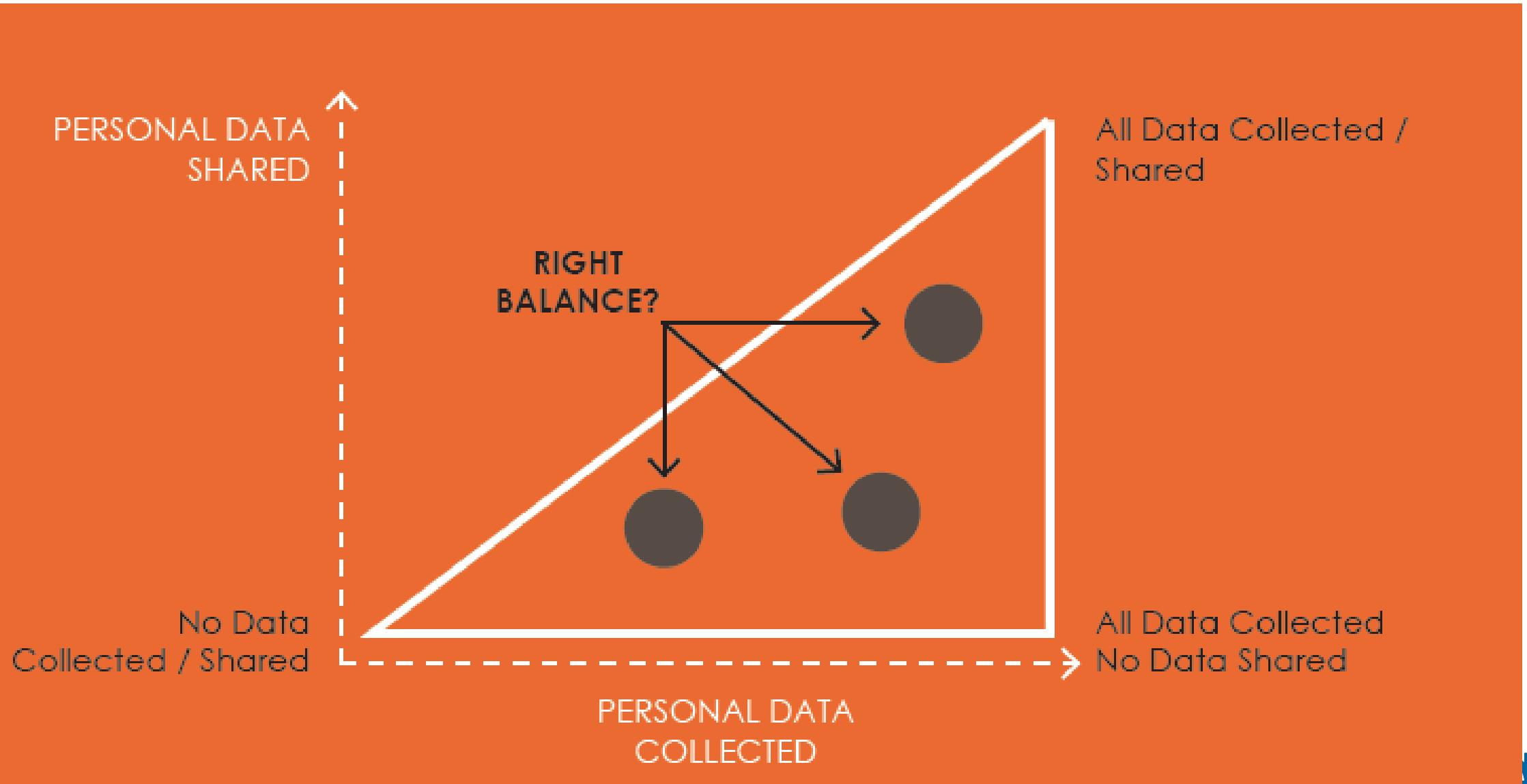


# Riesgos

"... si bien **existen ciertos riesgos y costos** asociados con el uso de datos móviles para producir *insights* para el bien social, **también puede haber riesgos y daños** asociados al no usar esta y otras nuevas fuentes de datos para informar políticas, respuesta humanitaria y otras intervenciones de desarrollo. "



# GDPR: Bien Privado vs Bien Público





“Cuando pienso en el 'big data' de las telecomunicaciones, pienso en los 350 millones de clientes que atendemos en todo el mundo y en los 23.000 millones de eventos móviles que creamos cada día en 21 países. Cuando pienso en el bienestar social, pienso en los compromisos asumidos con la UE cuando se trata de los 17 **objetivos de desarrollo sostenible** (SDGs) para 2030, los cuales representan 169 objetivos con 241 indicadores propuestos. Para nosotros es fundamental forjar una **relación entre nuestro 'big data' y el trabajo en pro del bienestar social**, sobre todo porque el 80% de los 6.000 millones de teléfonos móviles en el mundo corresponden a países en desarrollo. Y es ahí donde podemos tener el mayor impacto.”

—Jose Maria Alvarez Pallete, **Presidente de Telefonica S.A.**

# Datos Móviles para el Bien Social

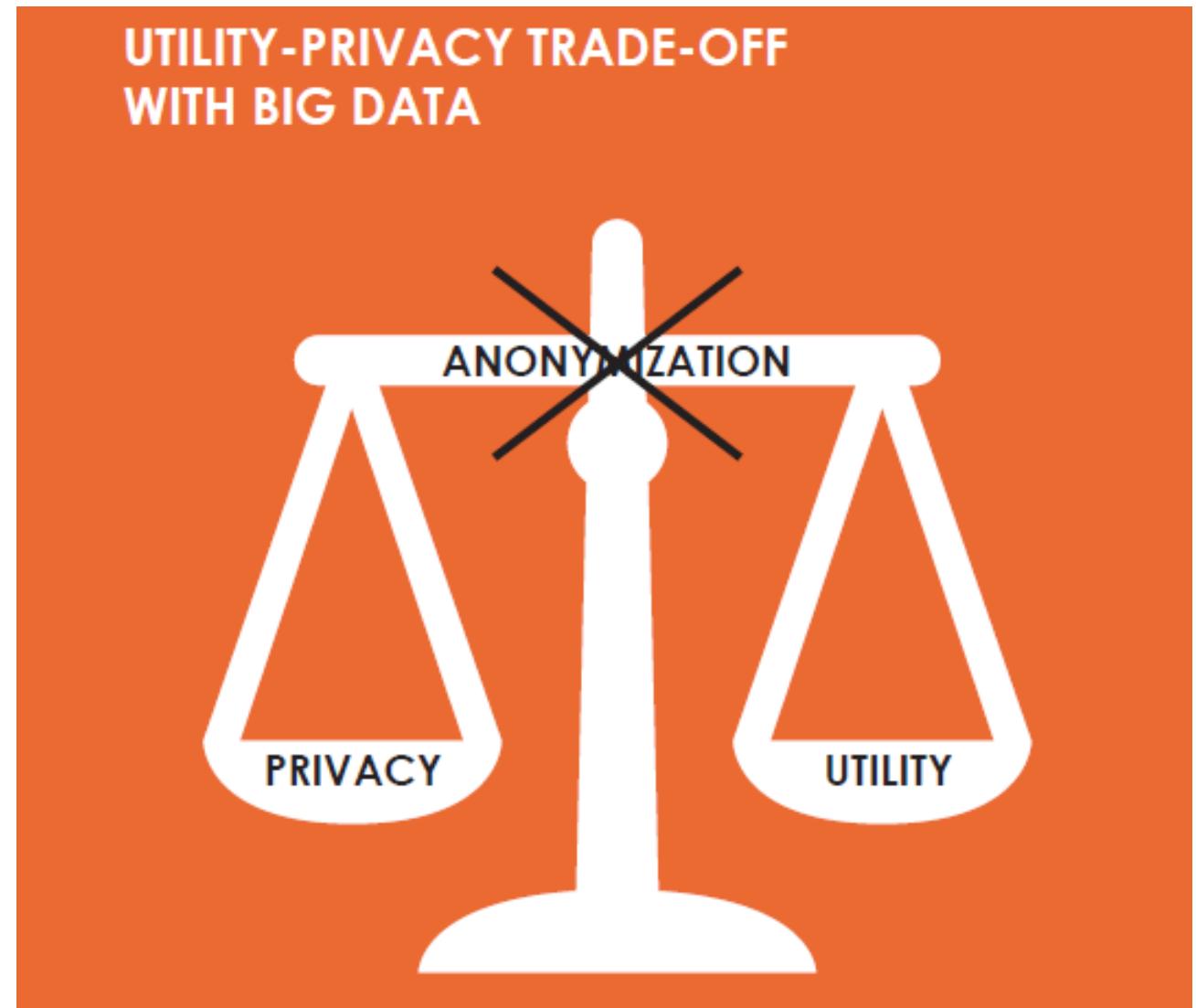
## Recomendaciones:

- Identificar y construir **modelos de negocios sostenibles**
- Abordar las brechas en **privacidad y protección de datos** de tal forma de mitigar los riesgos
- Desarrollar capacidades
- Crear herramientas globales para el bien público.

# ANONIMIZACION EN NUESTROS PROYECTOS

# Anonimización vs Utilidad

- Existen distintas técnicas de (pseudo)anonimización
- En general, mientras más anonimizado un dataset, más inútil es.

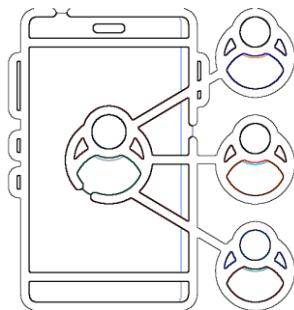
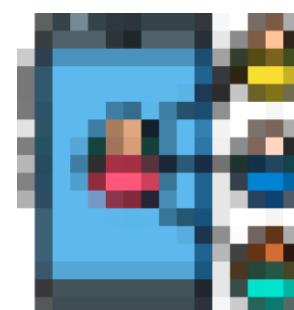
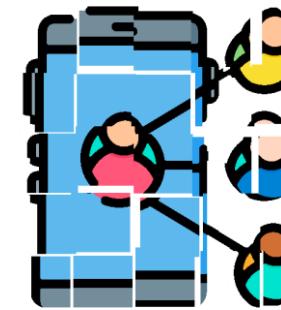
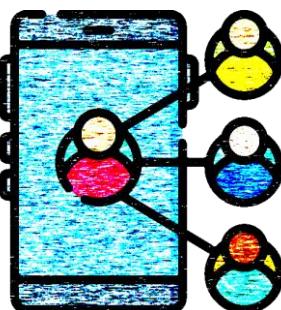
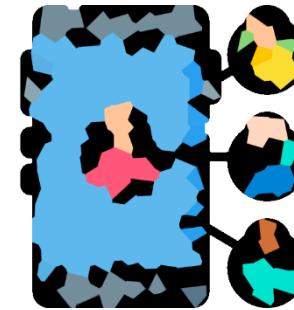
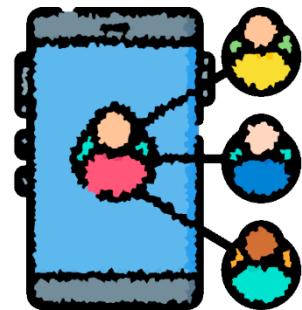
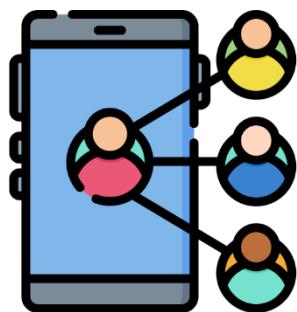


# Anonimización por Proyecto

Para cada proyecto se define un proceso de anonimización y de tratamiento de datos específico



Mitigar riesgos logrando los objetivos del proyecto



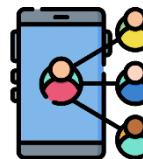
# Anonimización por Proyecto

Para cada proyecto se define un proceso de anonimización y de tratamiento de datos específico



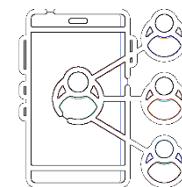
Minimizar Riegos logrando los objetivos del proyecto

CDRs: registros de llamadas



Fono A	Antena A	Fono B	Antena B	Fecha	Hora	Duración	...
+56985466257	MPPM2U1	+56968544127	MOVALLO	2015-01-02	10:22	67	
+56984677521	ACA3BW1	+56985466257	MPPM2U1	2015-07-19	18:43	159	

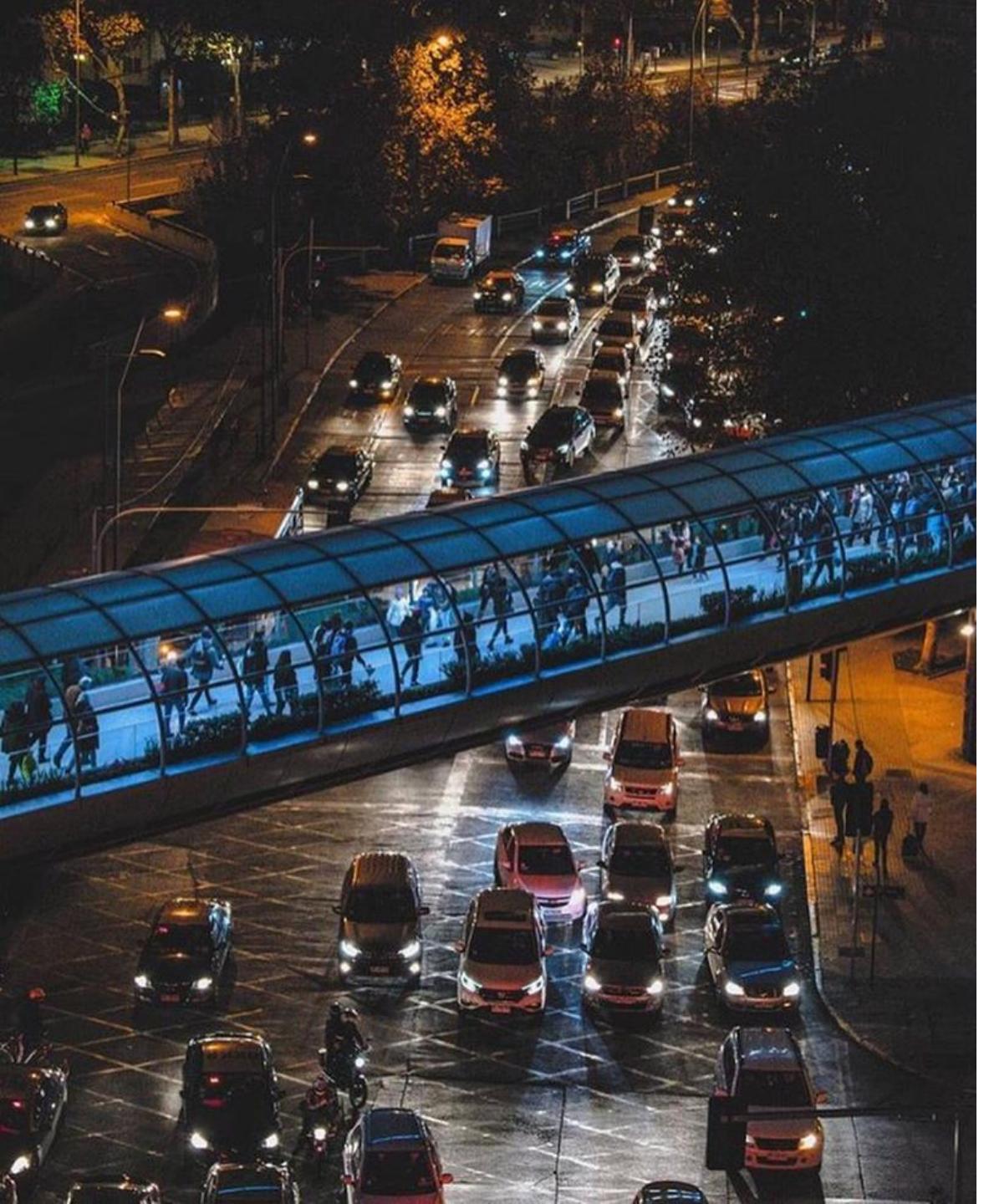
Anonimizado



Fono A	Fono B	Total Llamadas	Duración Total
EGHF564TDFRHIL	LKPHYTRSCOP347	326	22.543
LKPHYTRSCOP347	EGHF564TDFRHIL	105	8.190

1 año  
Todo Chile

ALGUNOS DE  
NUESTROS  
PROYECTOS



# Shopping Mall Attraction and Social Mixing at a City Scale

Beiró, M., Bravo., L, Caro, D., Cattuto, C., Ferres, L., Graells-Garrido, E. EPJ Data Science 7 (28), 2018.

# ¿Pueden los Malls mejorar la integración social?

Louail et al. *Applied Network Science* (2017) 2:11  
DOI 10.1007/s41109-017-0026-3

[Applied Network Science, 2017]  
Applied Network Science

RESEARCH

Open Access



CrossMark

## Crowdsourcing the Robin Hood effect in cities

Thomas Louail<sup>1\*†</sup> , Maxime Lenormand<sup>2\*†</sup>, Juan Murillo Arias<sup>3</sup> and José J. Ramasco<sup>4</sup>

city to another, and directly contribute to shape the spatial distribution of wealth across neighborhoods. By connecting areas, shopping trips also foster metropolitan integration and “social cohesion” (Miciukiewicz and Vigar 2012), whilst the resulting money flows are a key component of the development of territories (Ruault 2014). Large metropolitan areas are characterized by mixed land use in many of their neighborhoods. Every money flows as a first step towards improved spatial equity, we study a bottom-up approach that would rely on a slight evolution of shopping mobility practices. Building on a database of anonymized card transactions in Madrid and Barcelona, we quantify the mobility effort required to reach a reference situation where commercial income is

FR-34093 Montpellier, France  
Full list of author information is available at the end of the article

# ¿Pueden los Malls mejorar la integración social?

[WWW Conference, 2016]

## Measuring Urban Social Diversity Using Interconnected Geo-Social Networks

Desislava Hristova  
Computer Laboratory  
University of Cambridge, UK  
[first.last@cl.cam.ac.uk](mailto:first.last@cl.cam.ac.uk)

Matthew J. Williams  
School of Computer Science  
University of Birmingham, UK  
[m.j.williams@cs.bham.ac.uk](mailto:m.j.williams@cs.bham.ac.uk)

Mirco Musolesi  
Dept. of Geography  
University College London, UK  
[m.musolesi@ucl.ac.uk](mailto:m.musolesi@ucl.ac.uk)

Pietro Panzarasa  
School of Business and  
Management, Queen Mary  
University of London, UK  
[p.panzarasa@qmul.ac.uk](mailto:p.panzarasa@qmul.ac.uk)

Cecilia Mascolo  
Computer Laboratory  
University of Cambridge, UK  
[first.last@cl.cam.ac.uk](mailto:first.last@cl.cam.ac.uk)

### ABSTRACT

Large metropolitan cities bring together diverse interesting opportunities for cultural and intellectual exchange ultimately lead to social and economic enrichment. We present a novel network perspective on the interconnected nature of people and places, allowing us to capture the social diversity of urban locations through the social network and mobility patterns of their visitors. We use a dataset of approximately 37K users and 42K venues in London to build a network of Foursquare places and the parallel Twitter social network of visitors through check-ins.

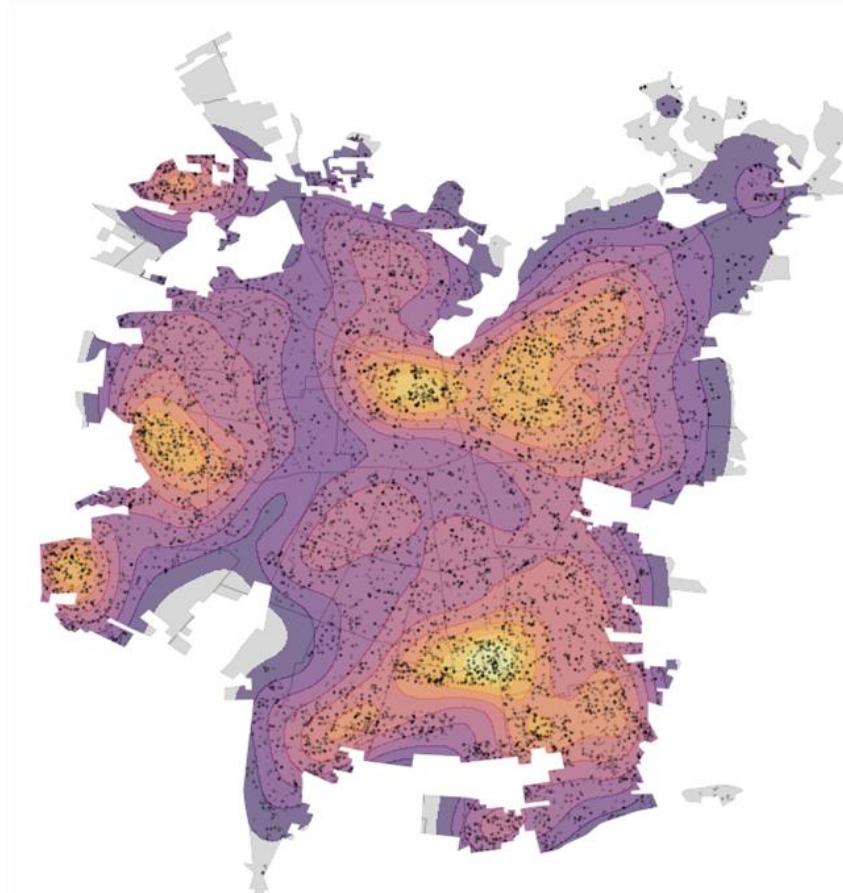
probability than in the other two sub-categories. *Shops* have a more unexpected mixture of top bridging places except for the *Mall* sub-category where it is likely that many strangers cross paths. Bond-inhabitants and measurable through their human mobility network and social network properties [18].

Urban activist Jane Jacobs wrote “[Cities] differ from towns and suburbs in basic ways, and one of these is that cities are, by definition, full of strangers” [21]. In this work we study places that bring together strangers among other types of urban social diversity and measure the relationship with the reciprocity of an user in an in-

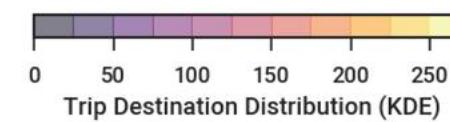
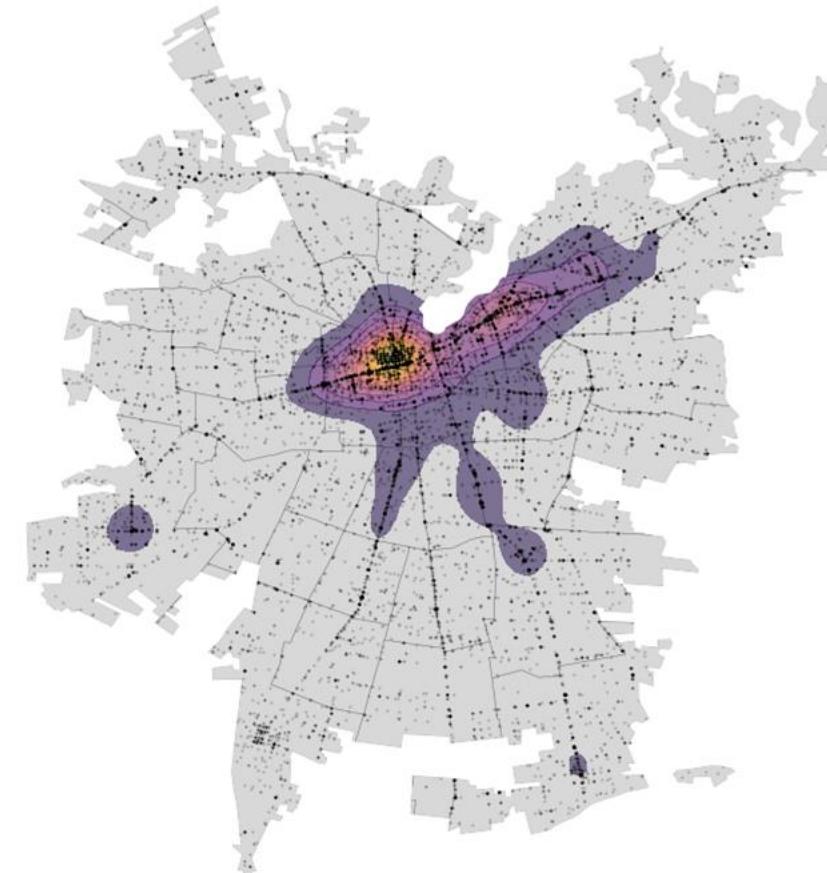
# ¿Pueden los Malls mejorar la integración social?

- Conjunto de datos:
  - XDRs de Agosto del 2016
    - **16 malls** en Santiago
    - **481 torres** interiores
    - **1M** de teléfonos móviles
    - **1.4M visitas** a malls

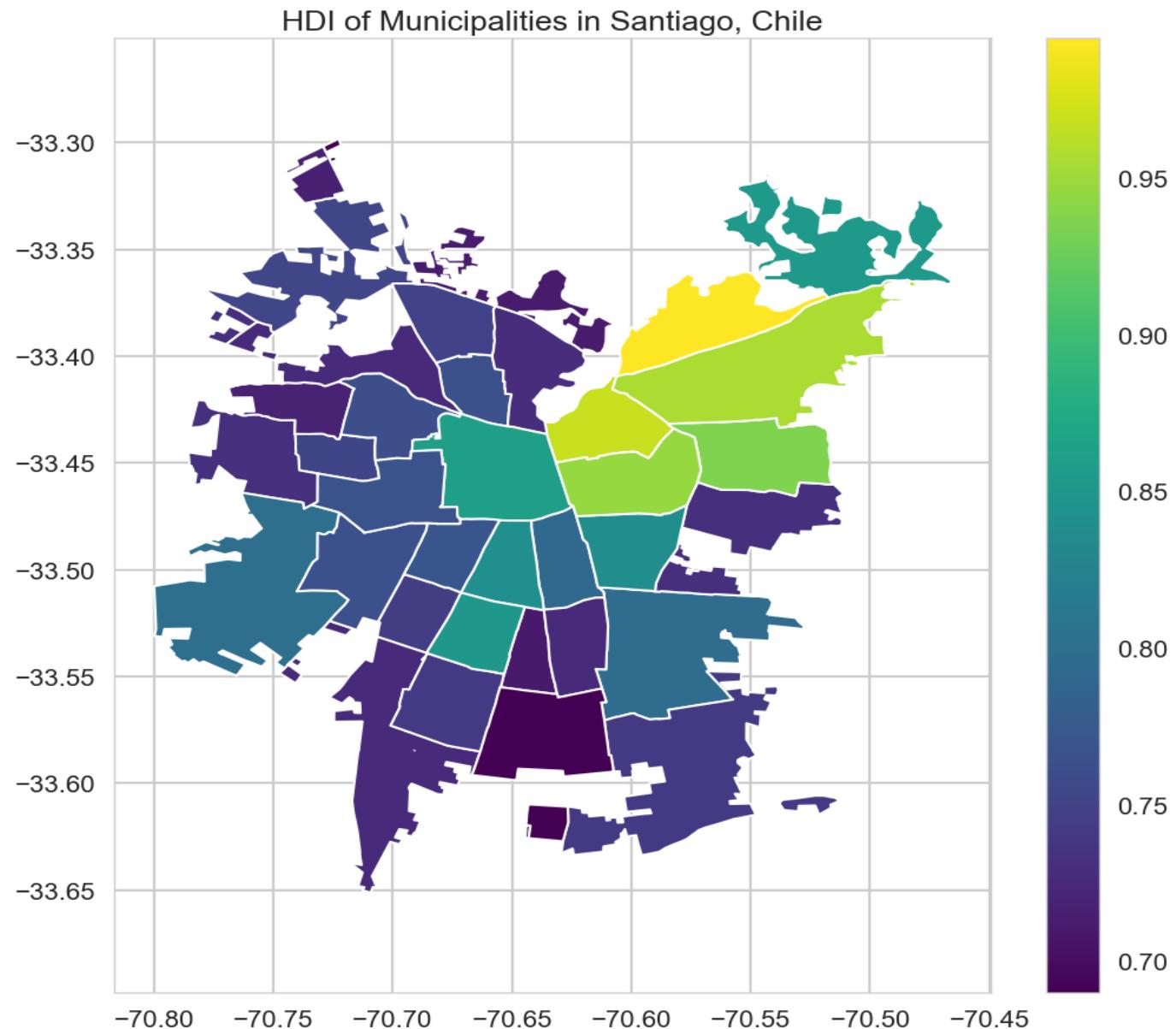
# Distribución de Hogar



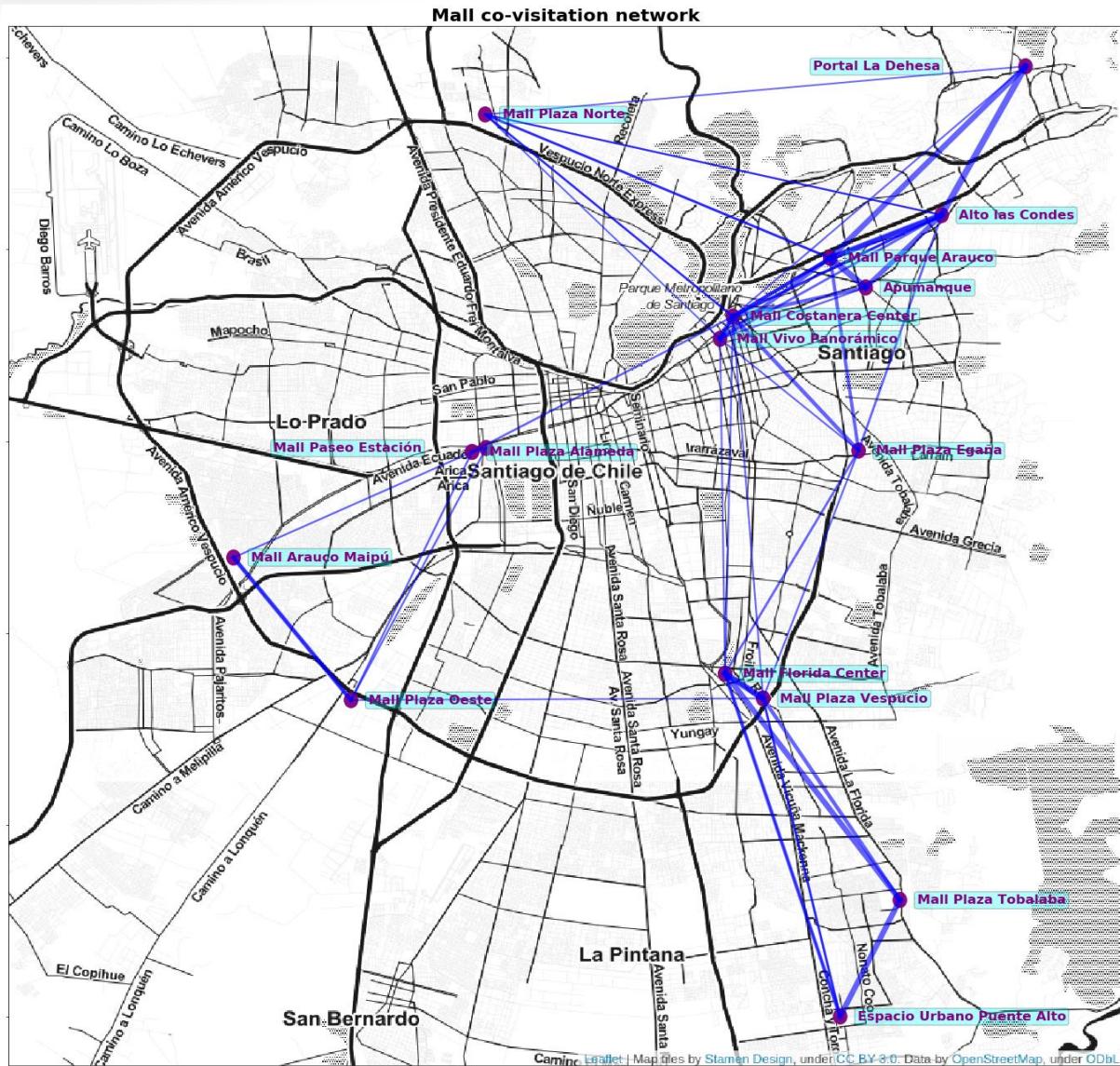
# Distribución de Lugar de Trabajo



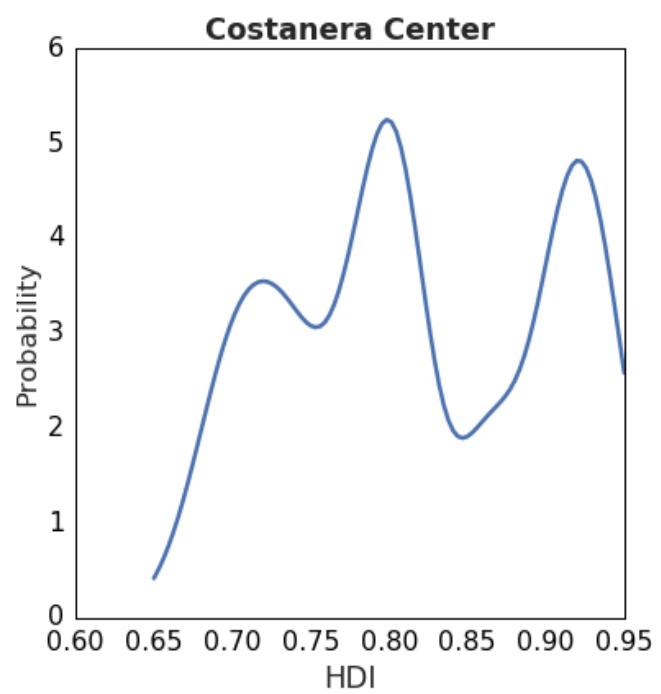
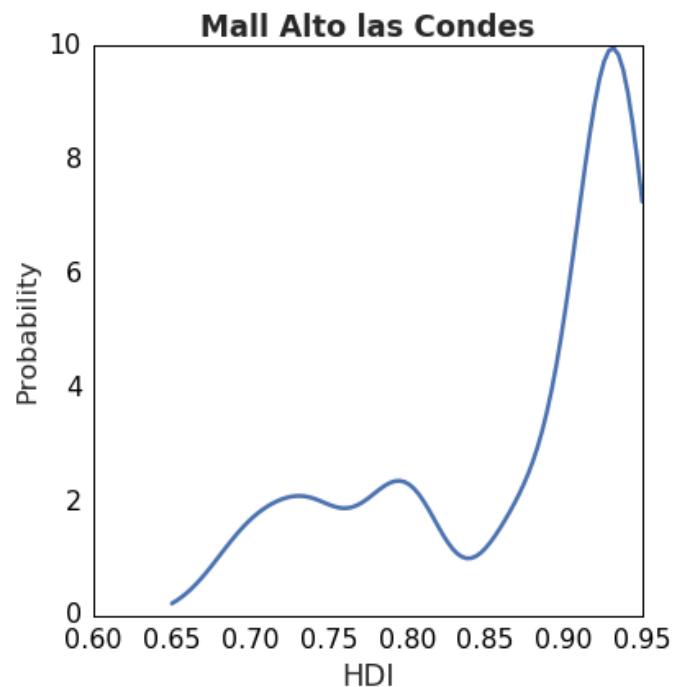
# Índice de Desarrollo Humano (HDI)



# Modelo Probabilístico de Co-visitación



- Resultan 3 clusters considerando similitud según el perfil de HDI:
  - Alta gama:
    - Alto las condes
    - Apumanque
    - Portal la Dehesa
    - Parque Arauco
  - Alta Mezcla Social:
    - Costanera Center
    - Mall Plaza Egaña
    - Panorámico
  - El resto
- Co-visitación:
  - Disminuye con la distancia
  - Aumenta entre perfiles de HDI similares



# Modelo Gravitacional

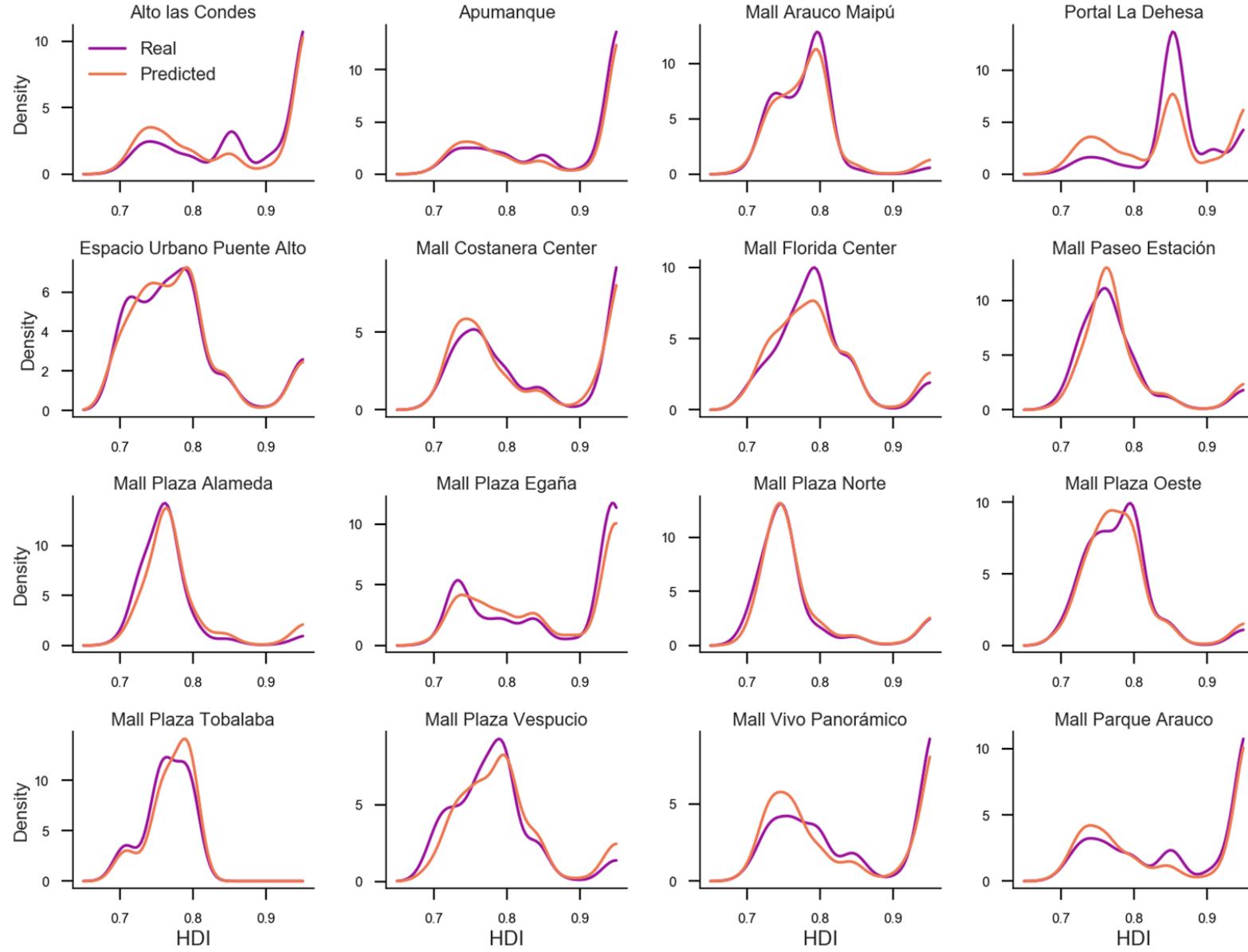
Podemos evaluar qué factores influyen en la visita a los malls usando un **modelo gravitacional**:

$$F_{ij} = G \frac{M_i^\alpha M_j^\beta}{D_{ij}^\gamma}$$

Donde:

- $F$  es la frecuencia con la que personas de la ubicación “ $i$ ” visitan el mall “ $j$ ”
- $M_i$  es la población en la ubicación “ $i$ ”
- $M_j$  es el tamaño del mall “ $j$ ”
- $D$  es la distancia entre la ubicación “ $i$ ” y el mall “ $j$ ”

Los exponentes y constantes consideran otros factores como nivel de desarrollo, ingreso, etc.

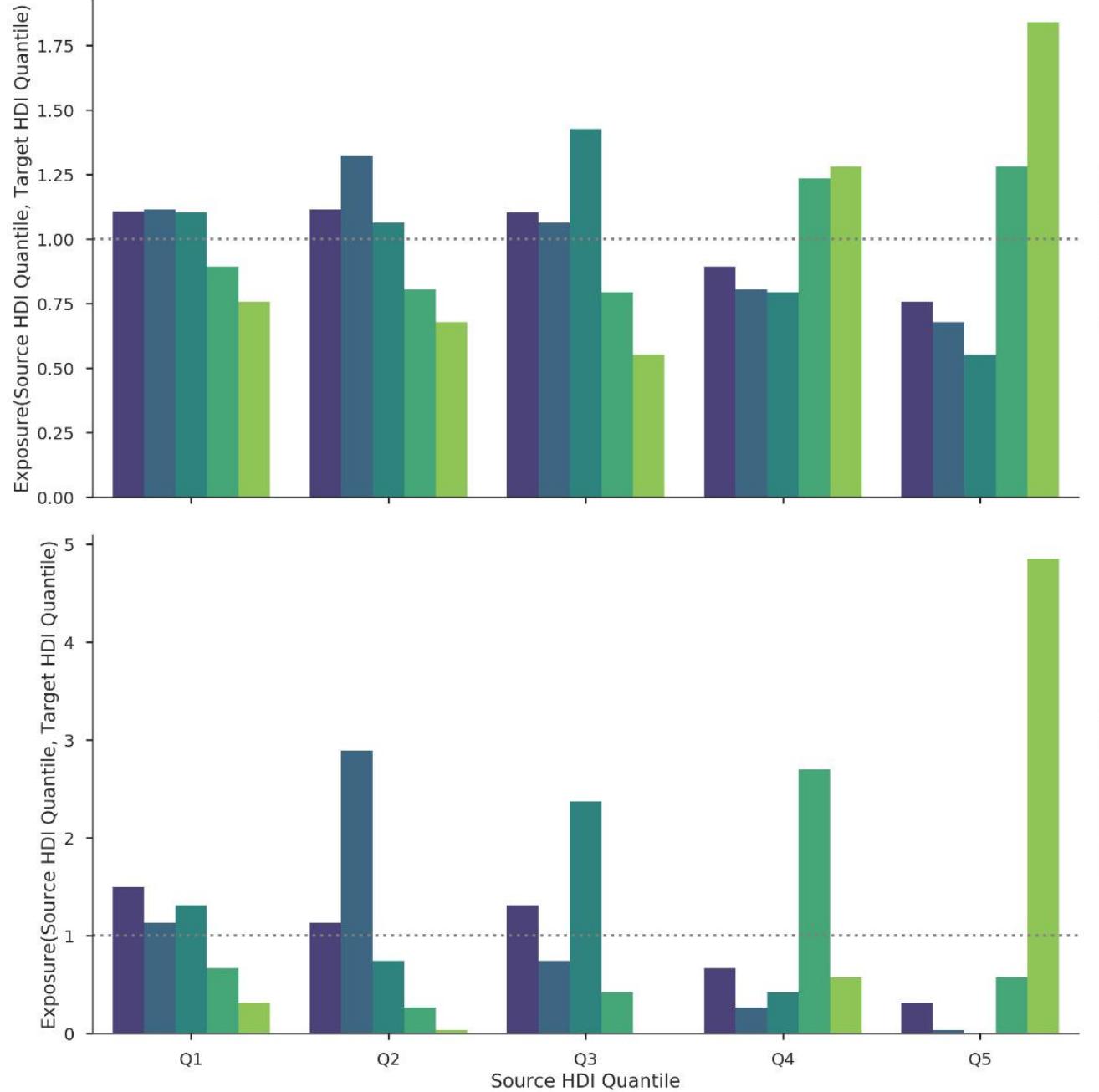


Luego de ajustar el modelo, es posible predecir las características de los vistantes al mall.

Encontramos que **la distancia es el factor más importante a la hora de elegir el mall.**

Si la “mezcla social” se incorpora al modelo, su desempeño mejora:

- “Malls con mayor mezcla social atraen a más personas”



## Modelo de segregación basado en co-locación espacial.

Louf, R., & Barthelemy, M. (2016). Patterns of residential segregation. PloS one, 11(6), e0157476.

model = observed

“Mezcla Social” usando datos de celular

model = null model

“Mezcla social” asumiendo que van al mall más cercano



# El efecto de Pokémon Go en el pulso de la ciudad

Graells-Garrido, E., Ferres, L., Caro, D., & Bravo, L. (2017). *The effect of Pokémon Go on the pulse of the city: a natural experiment*. EPJ Data Science, 6(1), 23.

# ¿Estaba la ciudad más llena después del lanzamiento de Pokémon Go?

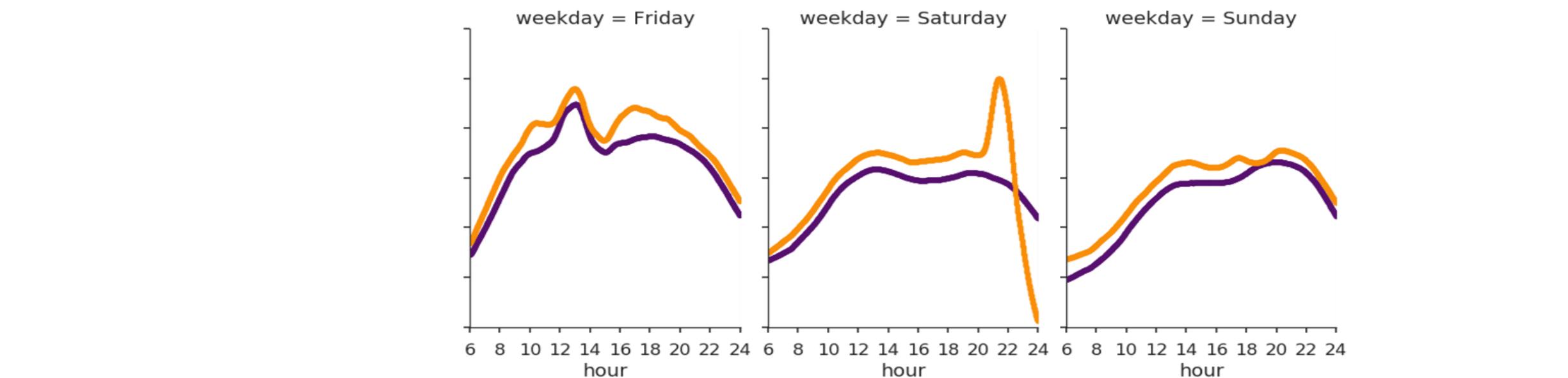
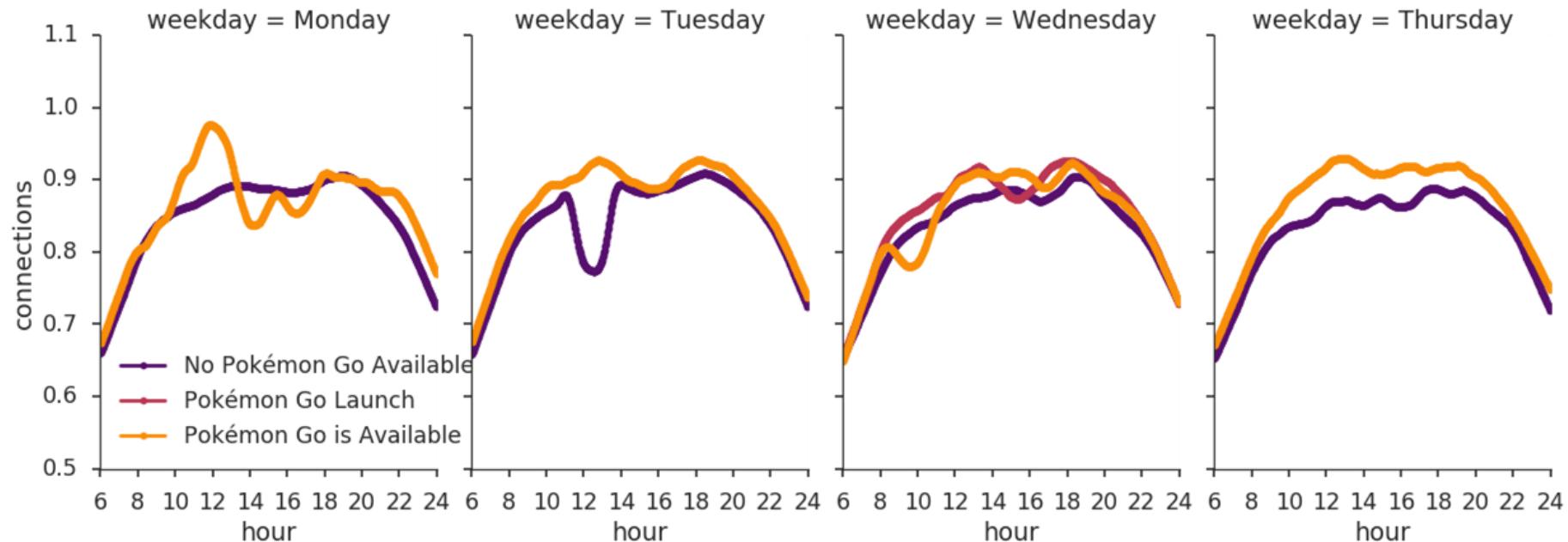


**Propuesta:** Evaluar el pulso de la ciudad (caracterización de la población flotante) **una semana antes y después del lanzamiento de Pokémon Go** en Santiago (3 de agosto del 2017).

**N = 142.988**

Dispositivos estuvieron activos  
todos los días

## XDRs



## Modelo de regresión

Se evaluó el modelo para cada minuto entre las 6AM y medianoche para 700~ áreas de la ciudad definidas en la encuesta de viaje de Santiago del 2012.



$$\log E[X(t)] = \log a + \beta_0 + \beta_1 \text{PoGo} + \beta_2 \text{DayOfWeek} + \beta_3 \text{LandUse} + \beta_4 \text{PokéPoints}$$

Curly braces below the equation map terms to their descriptions:

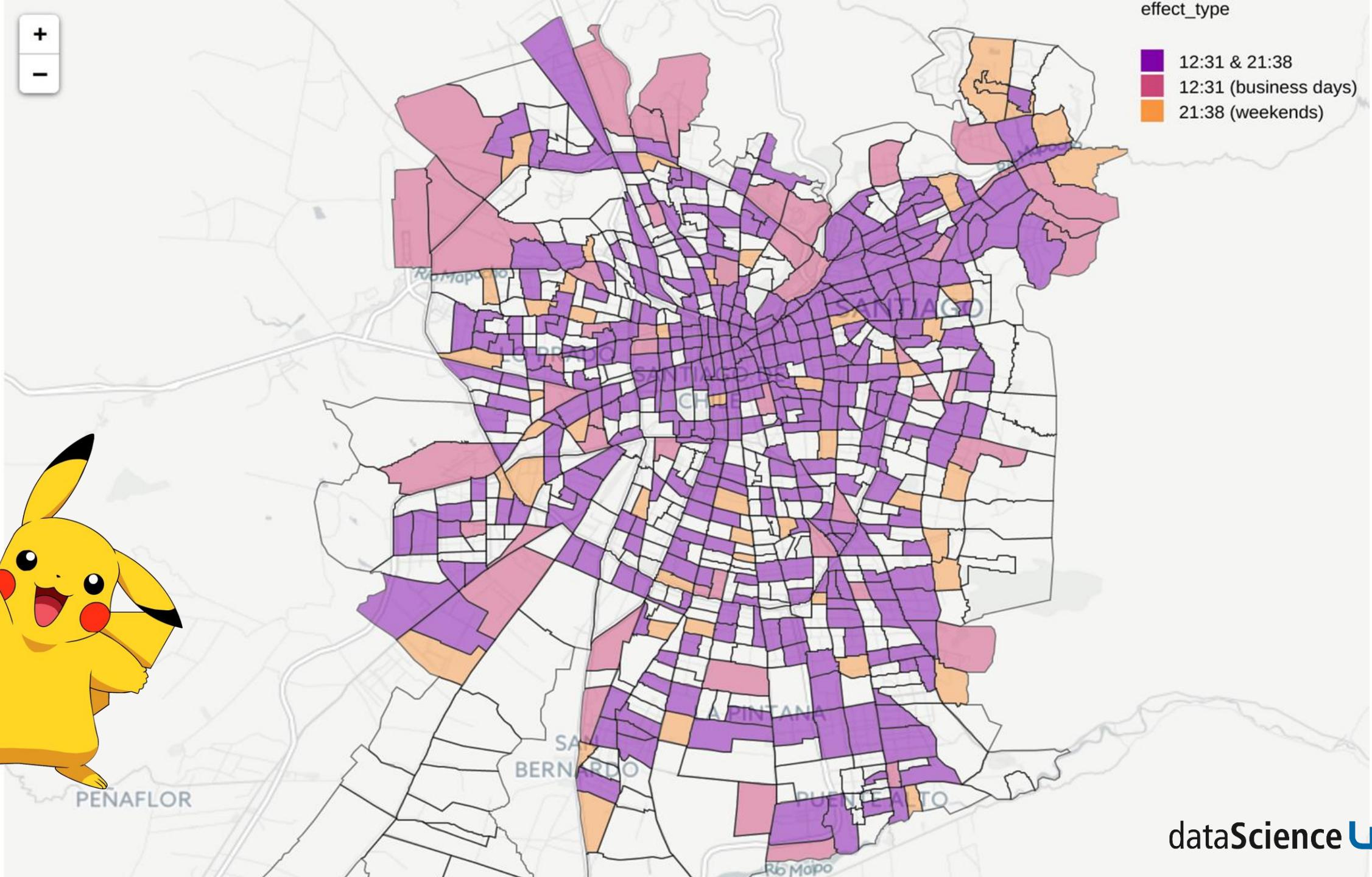
- $\log E[X(t)]$ : Número esperado de personas en el área a tiempo  $t$
- $\beta_0$ : Baseline
- $\beta_1 \text{PoGo}$ : Superficie
- $\beta_2 \text{DayOfWeek}$ : Día de semana/Sábado/Domingo
- $\beta_3 \text{LandUse}$ : Uso de Suelo: residencial, negocio o mixto
- $\beta_4 \text{PokéPoints}$ : # puntos de interés

+

-

effect\_type

- 12:31 & 21:38
- 12:31 (business days)
- 21:38 (weekends)

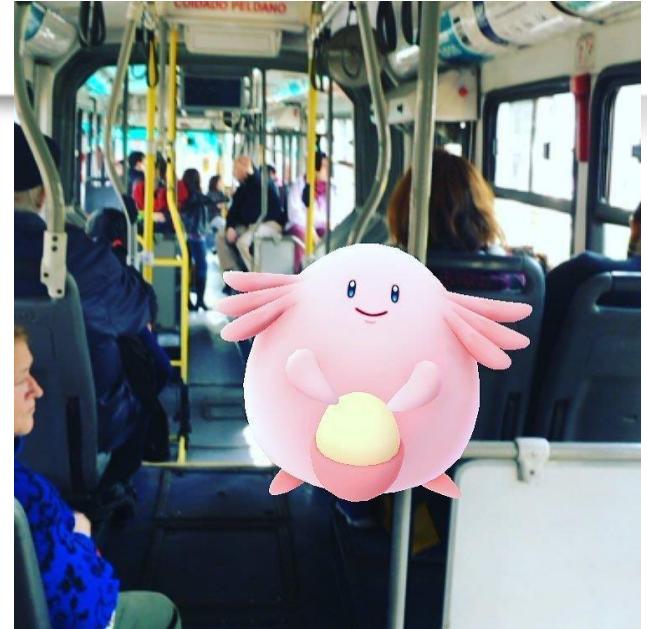


# Conclusiones

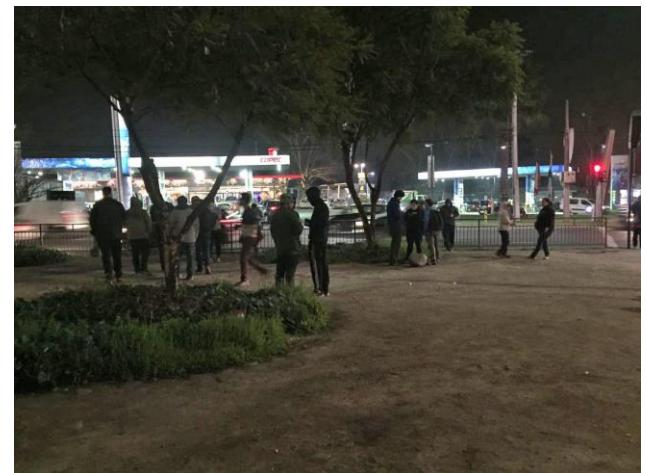
La gente jugo principalmente en durante sus **traslados (13%)**, los que es posible que sean horas estresantes del día.

Se observó que incluso en áreas residenciales, donde es poco probable ver gente de noche, **hubo un aumento significativo en el número de personas (10%)**.

**¿Como se pueden utilizar estos resultados para mejorar el bienestar de la ciudad?**



@aarosenberg



@christian\_21



# Genero Y Movilidad

ANNOUNCING  
**Big Data for Gender  
Challenge Awards**

10 innovations to close the gender data gap.

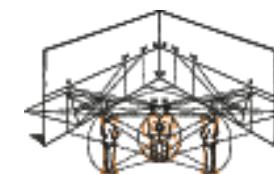
[bit.ly/big-data-challenge-awards](http://bit.ly/big-data-challenge-awards)

September 19, 2017

Data2X is pleased to announce the winners of the [Big Data for Gender Challenge](#), ten projects that apply big data innovations to fill gender data gaps and improve understanding on key aspects of girls' and women's lives. These projects represent 29 researchers, from 20 different institutions, across 8 countries, and were selected from a pool of over 125 proposals.

The winning projects for the Big Data for Gender Challenge are:

1. Gender and Urban Mobility: Addressing Unequal Access to Urban Transportation for Women and Girls
2. Gender and Mobile Money Networks
3. Analyzing Big Data to Understand Uptake and Usage of Financial Services to Advance Women's Financial Inclusion
4. Identifying Women in Mobile Phone Data to Further Financial Inclusion
5. Mining the Web for Insights on Violence Against Women and Conflict-Related Gender-Based Violence in the MENA Region and Arab States
6. Women in the Gig Economy: A Data Gap with Implications for Informal Work, Time Use and



**ISI Foundation**



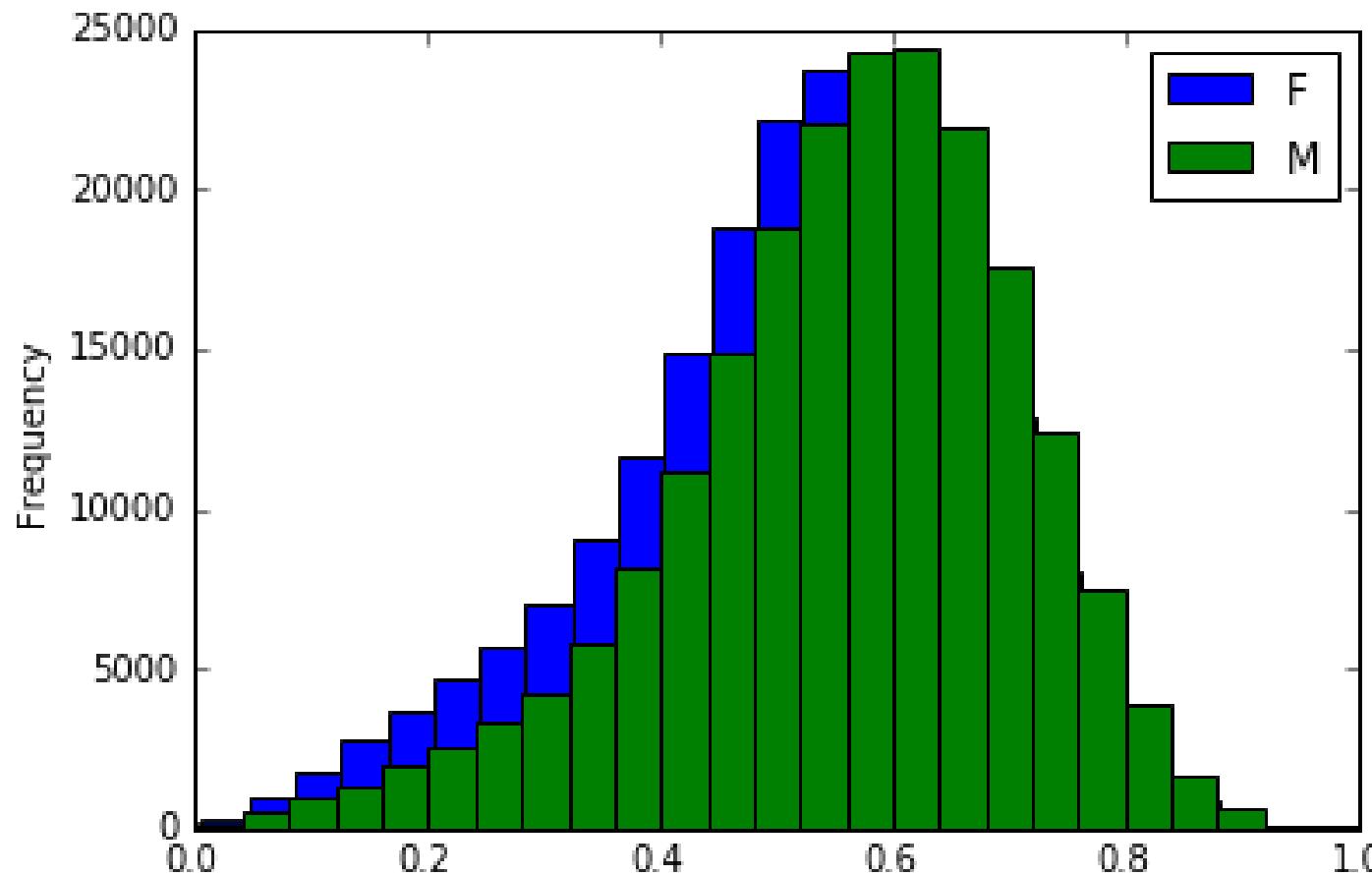
# Data Utilizados

- CDRs anotado con:
  - Genero
  - Grupo Socioeconómico (ABC1, C2, C3, D, E)
  - Número de líneas que posee
  - Prepago o contrato
- Periodo:
  - Junio-Agosto, 2016 (3 meses)
  - 2,148,132,995 filas (CDRs llamadas),
  - 1.06TB

An individual  $i$  “moves more” than an individual  $j$  iff  
 $S^i > S^j$ , where  $S$  is Shannon’s entropy

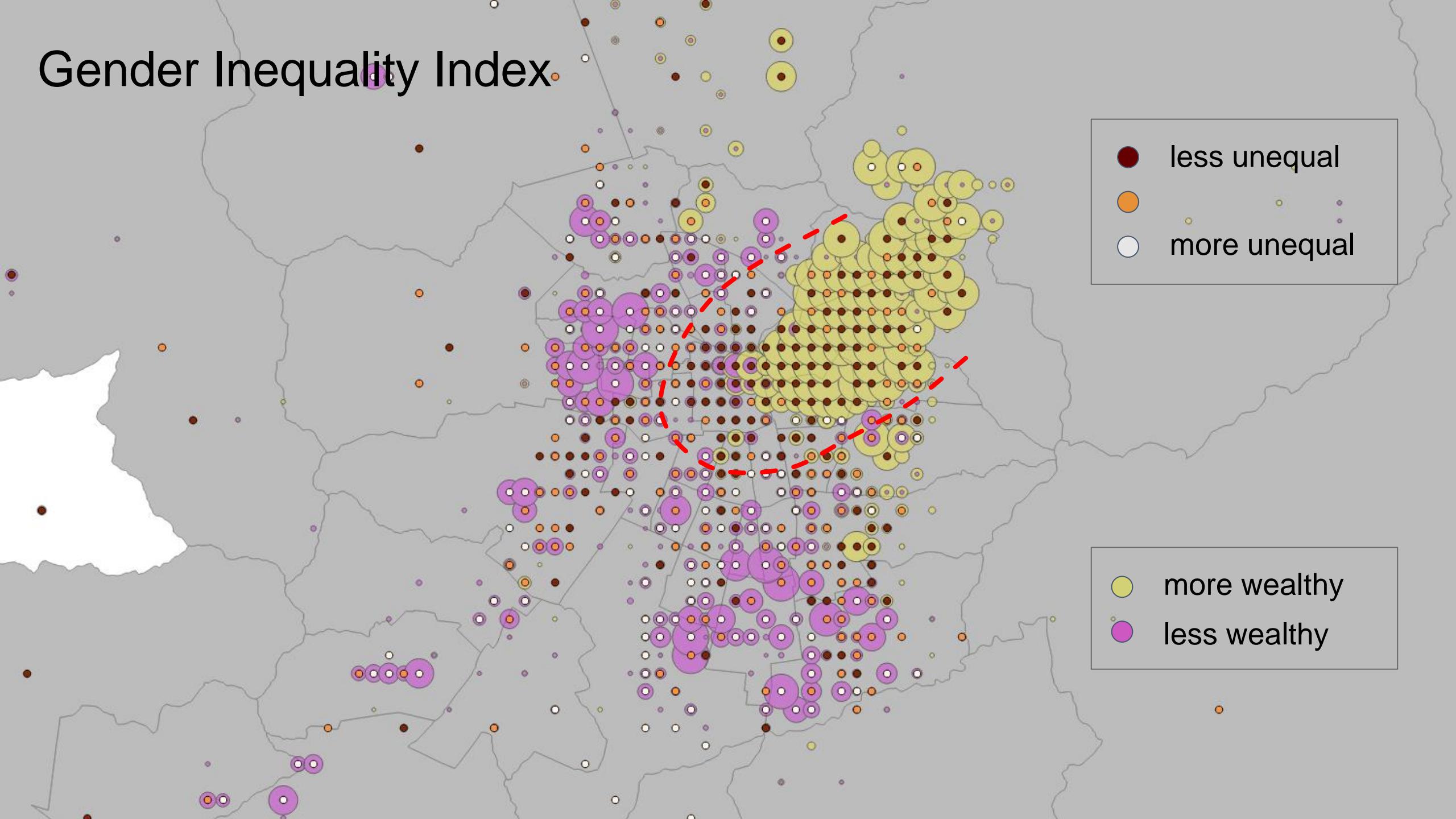
$$S = - \sum_{I \in L} p_I \ln p_I$$

over each individual’s  $i, j$  set of all visited places  $L$ .

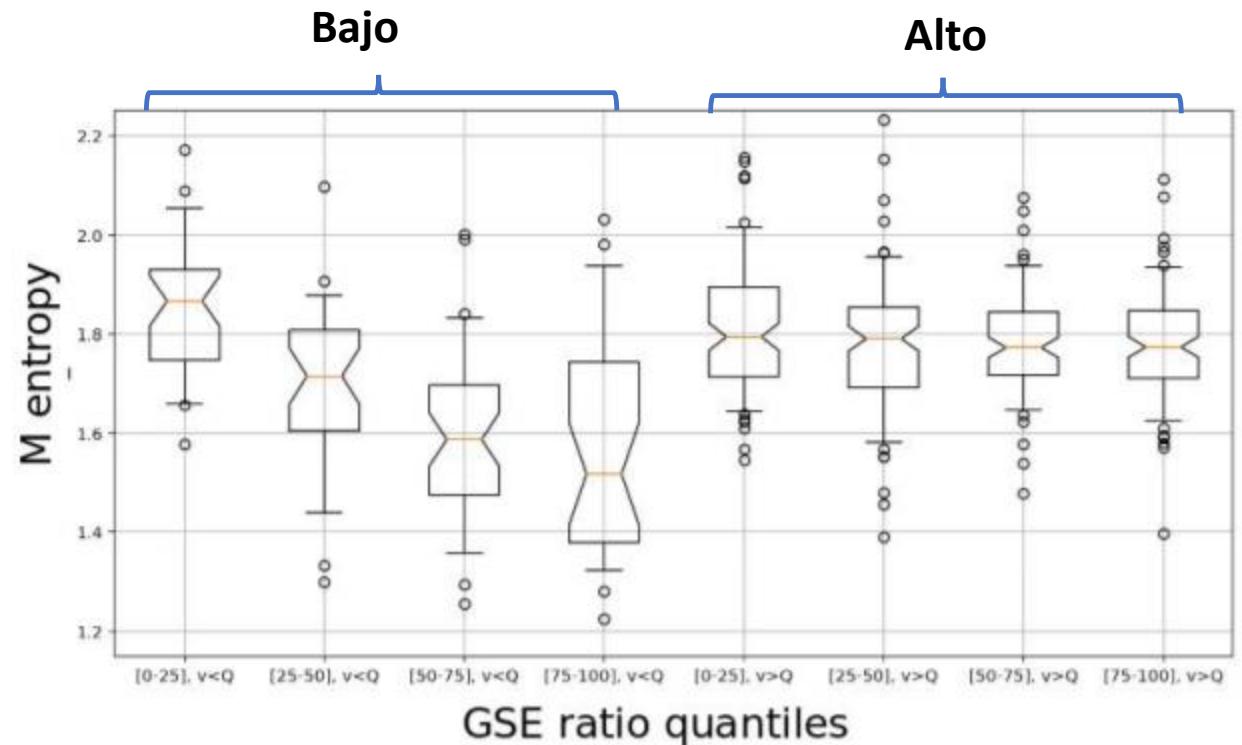


*En general las mujeres tienden a pasar su tiempo en un solo lugar o muy pocos.*

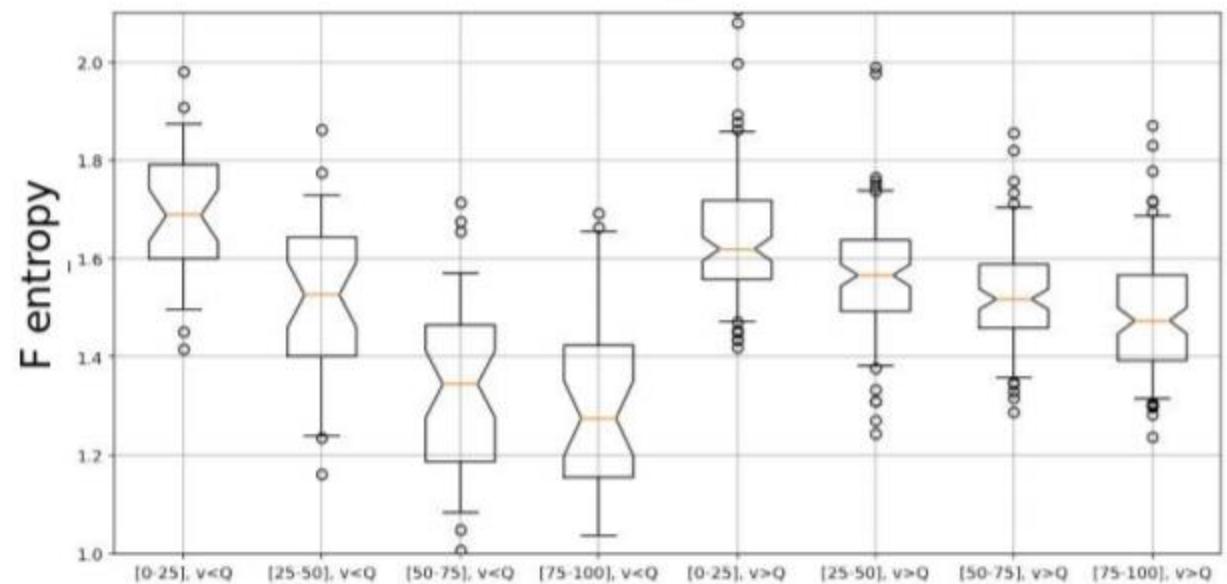
# Gender Inequality Index



## Acceso a Transporte Público

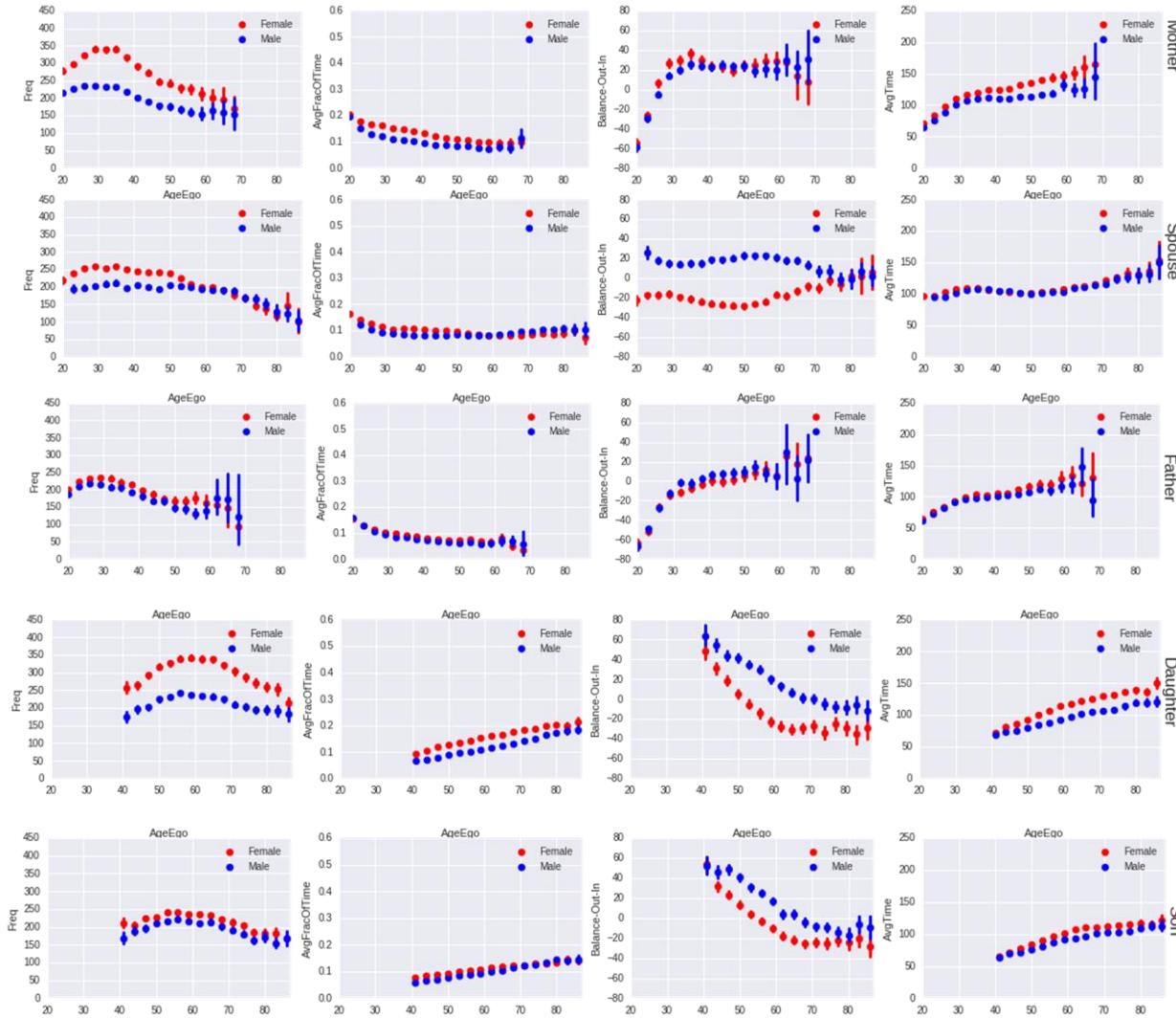


Para las mujeres, el acceso al transporte público solo **mitiga** el efecto del ingreso





# Redes Familiares



# Patrones de comunicación con Madre, Padre, Esposo, Hija, hijo, Hermano

Isabel Behncke Izquierdo, Sebastian Diaz, Loreto  
Bravo, Carlos Rodriguez-Sickert, Anna  
RotkirchJanos Kertesz and Tamas David-Barrett



Università Commerciale  
**Luigi Bocconi**

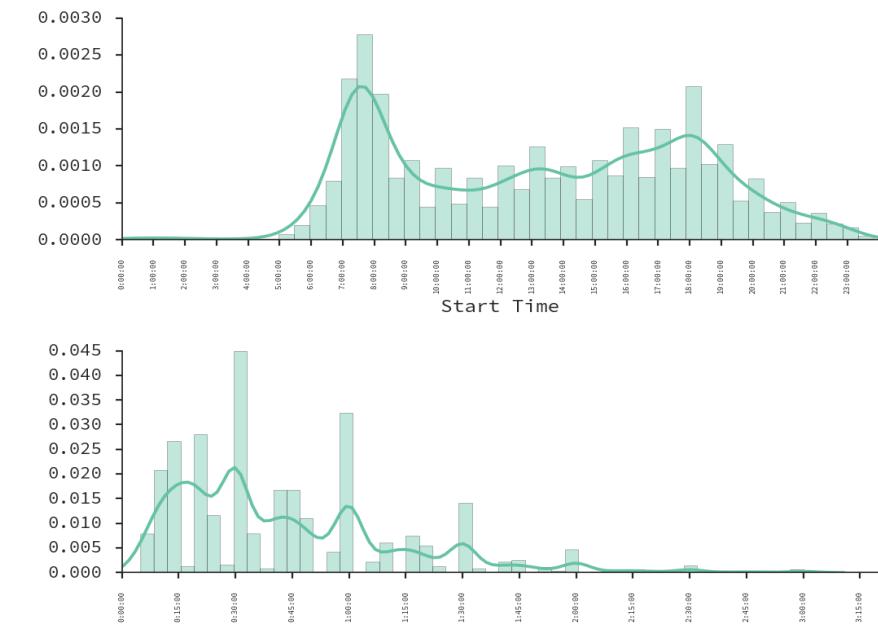
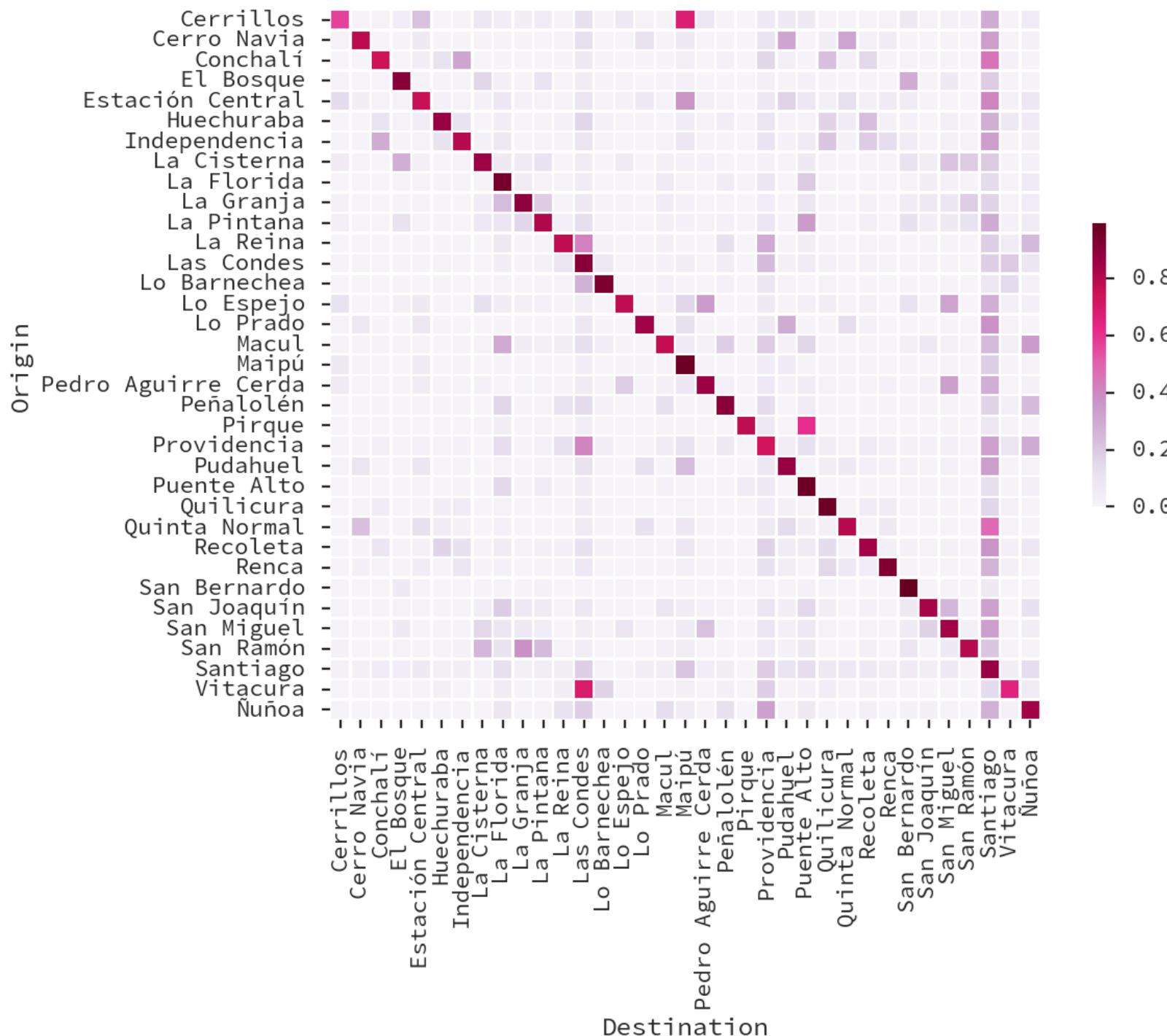


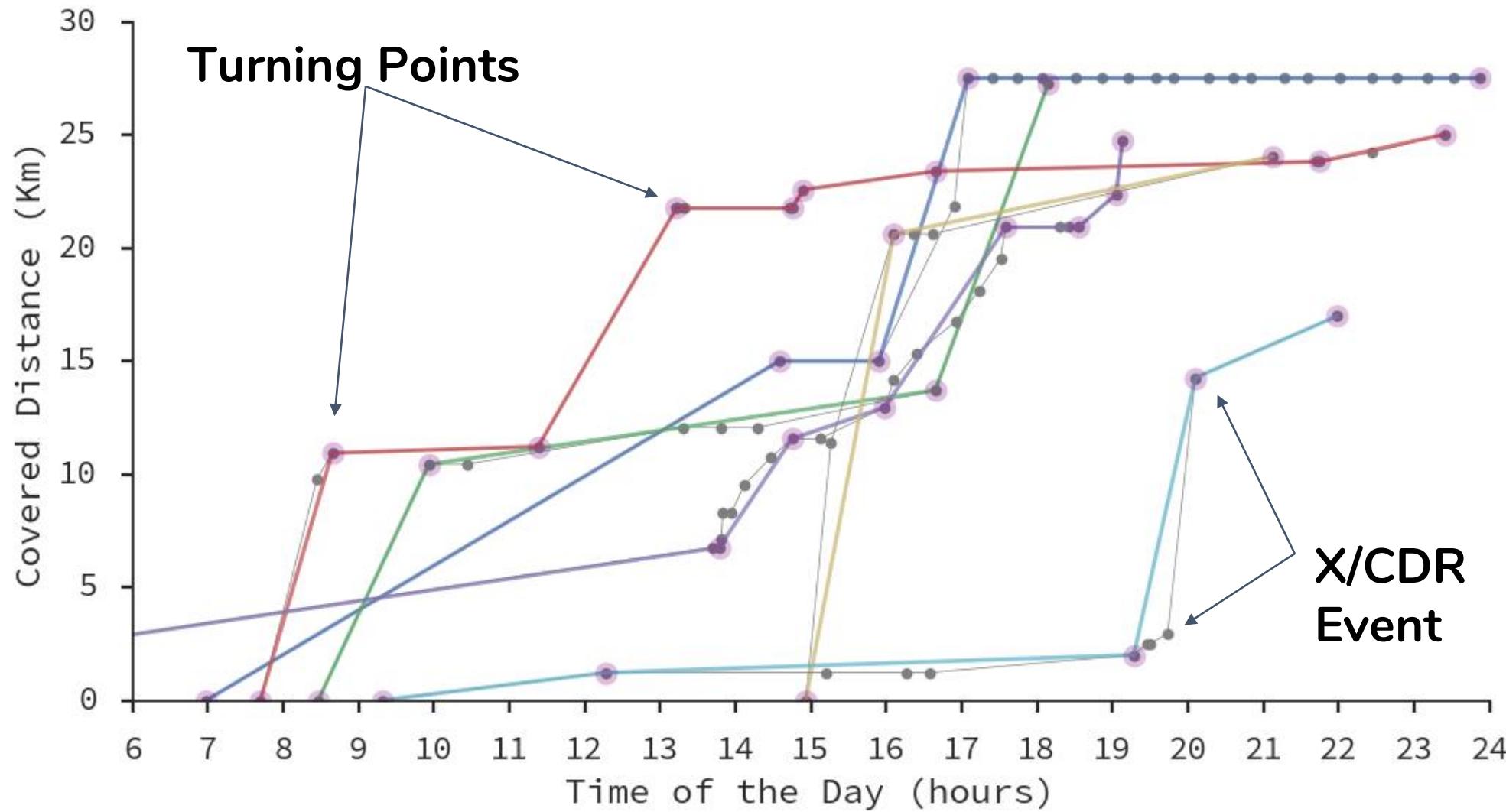
# Detección de Viajes y Modo de Transporte

Graells-Garrido, E., & Saez-Trumper, D. (2016, May). **A day of your days: estimating individual daily journeys using mobile data to understand urban flow.** In Proceedings of the Second International Conference on IoT in Urban Space (pp. 1-7). ACM.

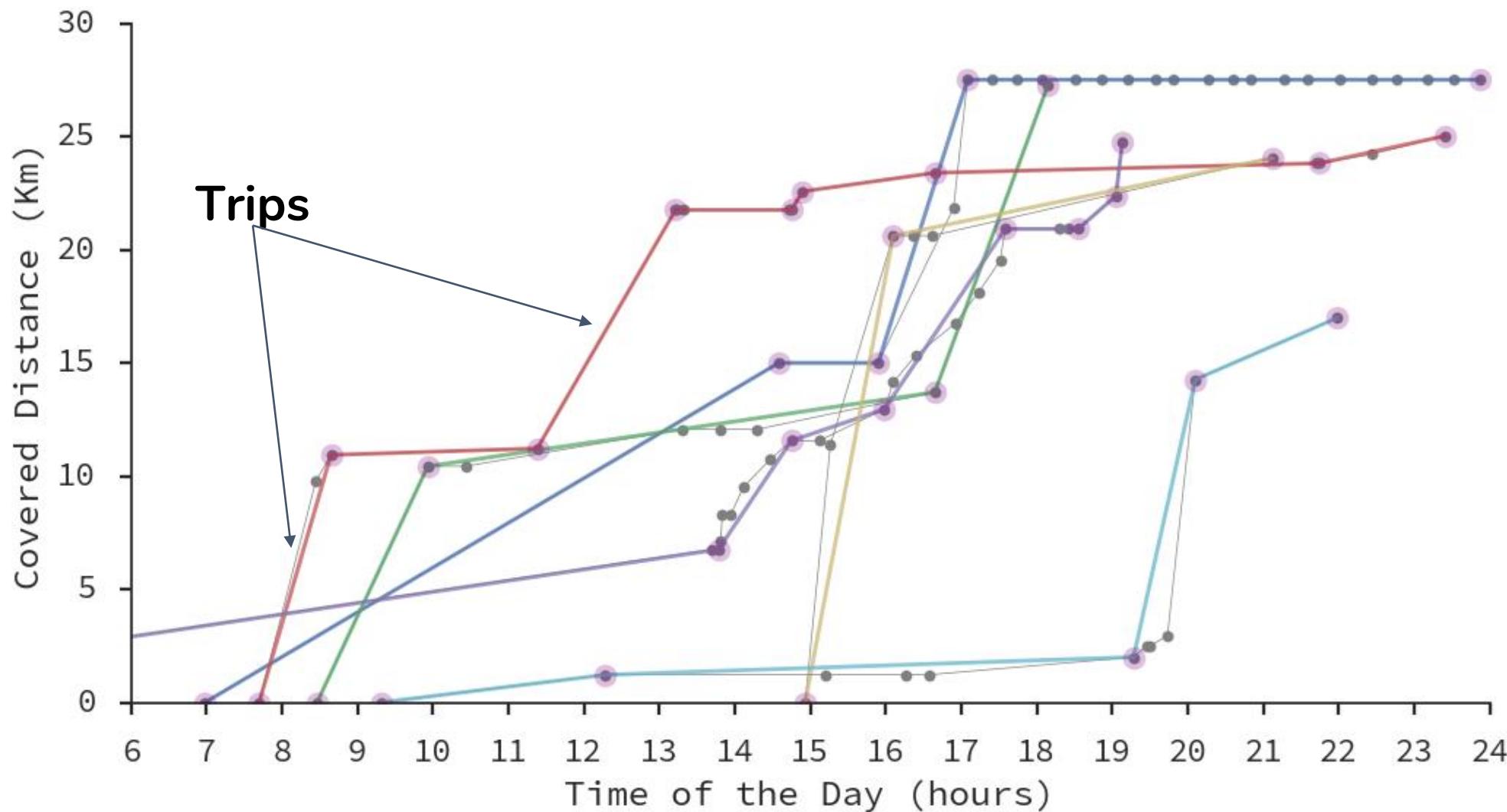
Graells-Garrido, E., Caro, D., & Parra, D. (2017, October). **Toward Finding the Mobilicities within a City with Non-Negative Matrix Factorization.** User Interfaces for Spatio-Temporal Data Analysis (workshop co-located with IUI'2018). Tokyo, Japan.

Inferring modes of transportation using mobile phone dataE Graells-Garrido, D Caro, D Parra. EPJ Data Science 7 (1), 49

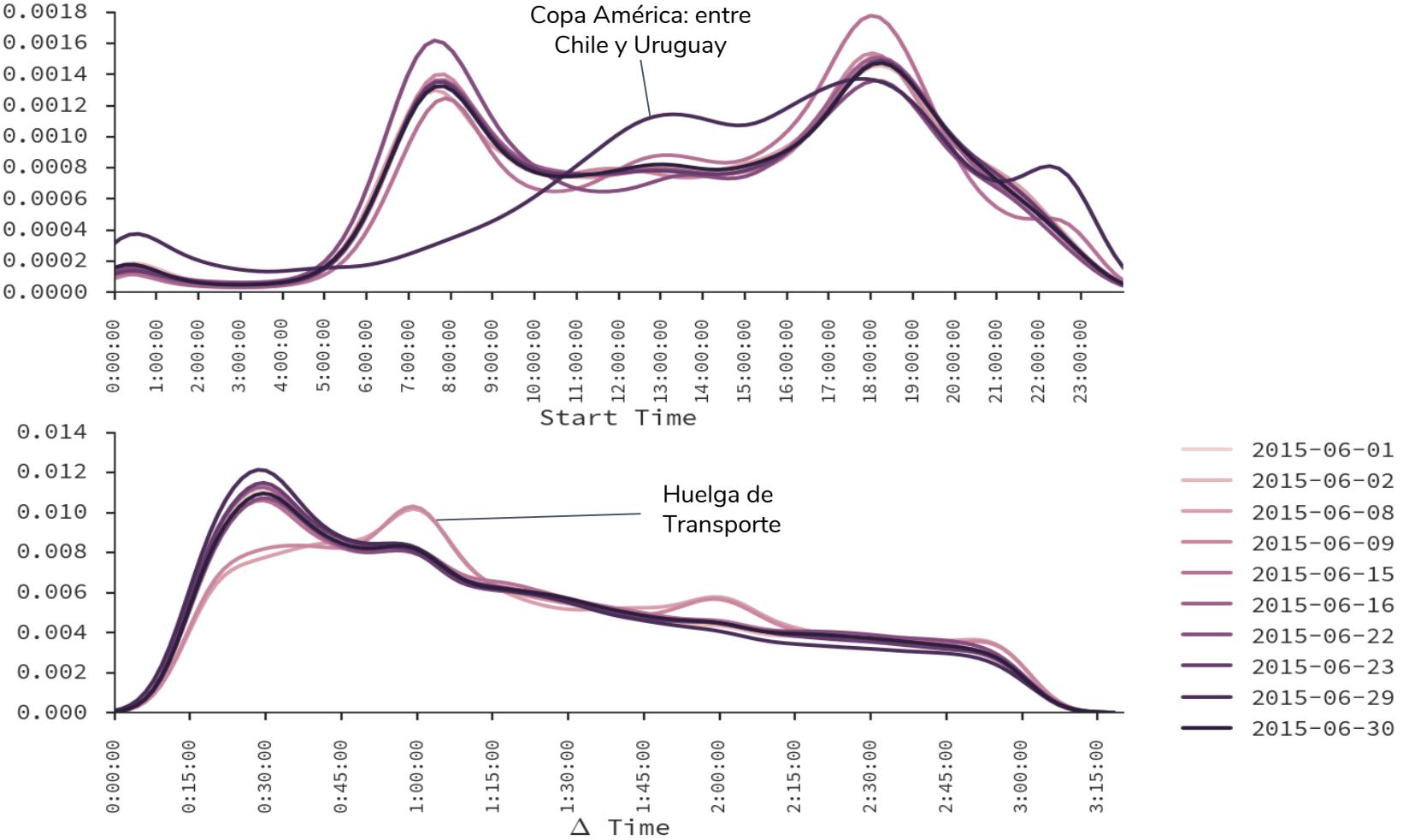




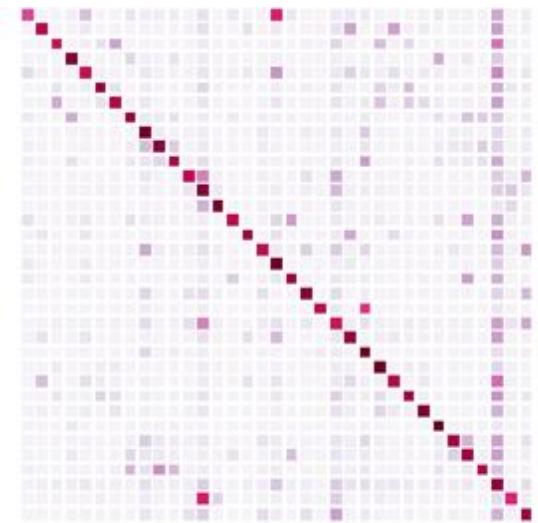
X/CDR Events of a **single device** are positioned in the time-space diagram, with distance relative to their origin positions. We detect **turning points** of the raw trajectories using **computational geometry**.



In an iterative process we classify each segment of the trajectory into **trip** or **stay**. Classification is based on **transportation rules** related to speed, distance and time.



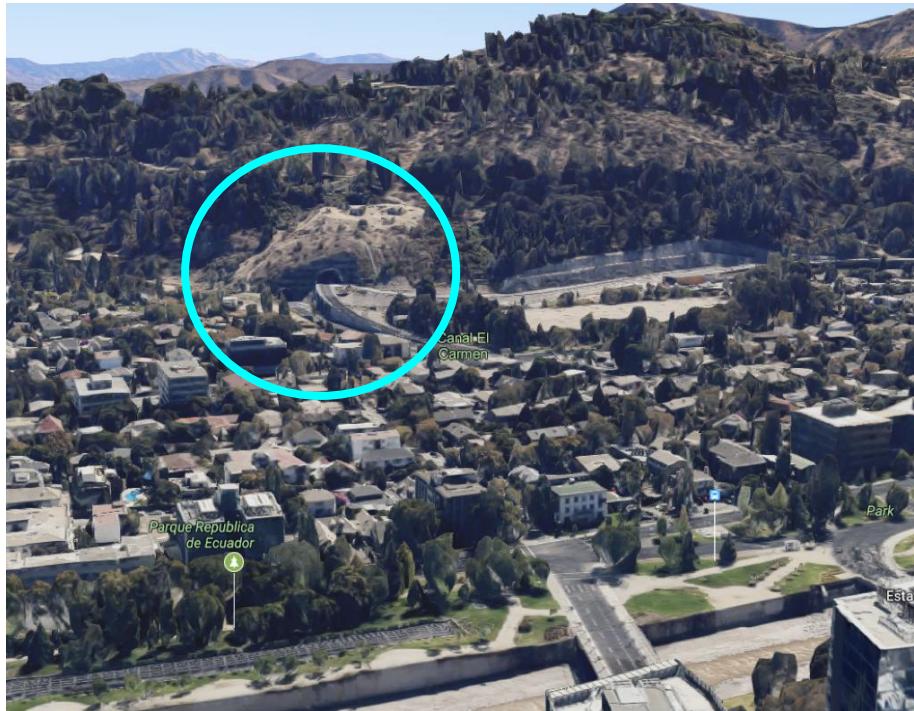
A través de la agregación de viajes podemos construir **matrices de Origen-Destino** a niveles y granularidad no disponibles con métodos tradicionales



# ¿Cómo inferimos la partición Modal de la Ciudad?

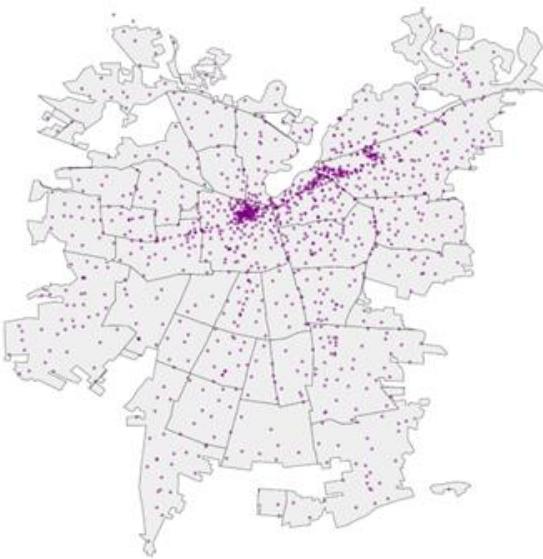
# ¿Cómo inferimos la partición modal de la ciudad?

Algunas torres tienen propiedades interesantes



**El problema:** ¿Cómo identificarlas?

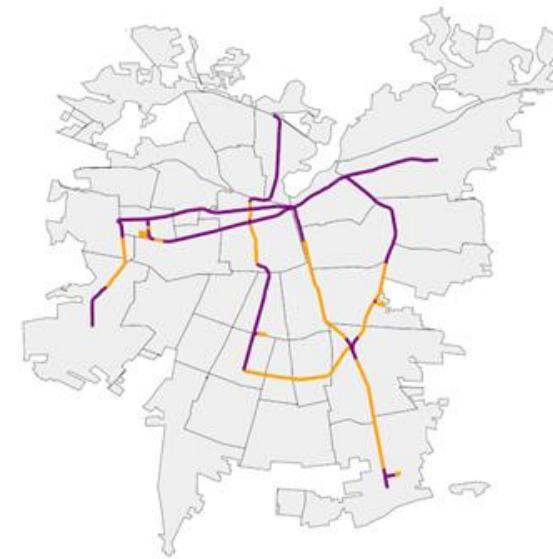
Towers



Highways and Primary Streets (OSM)



Metro Network (OSM)

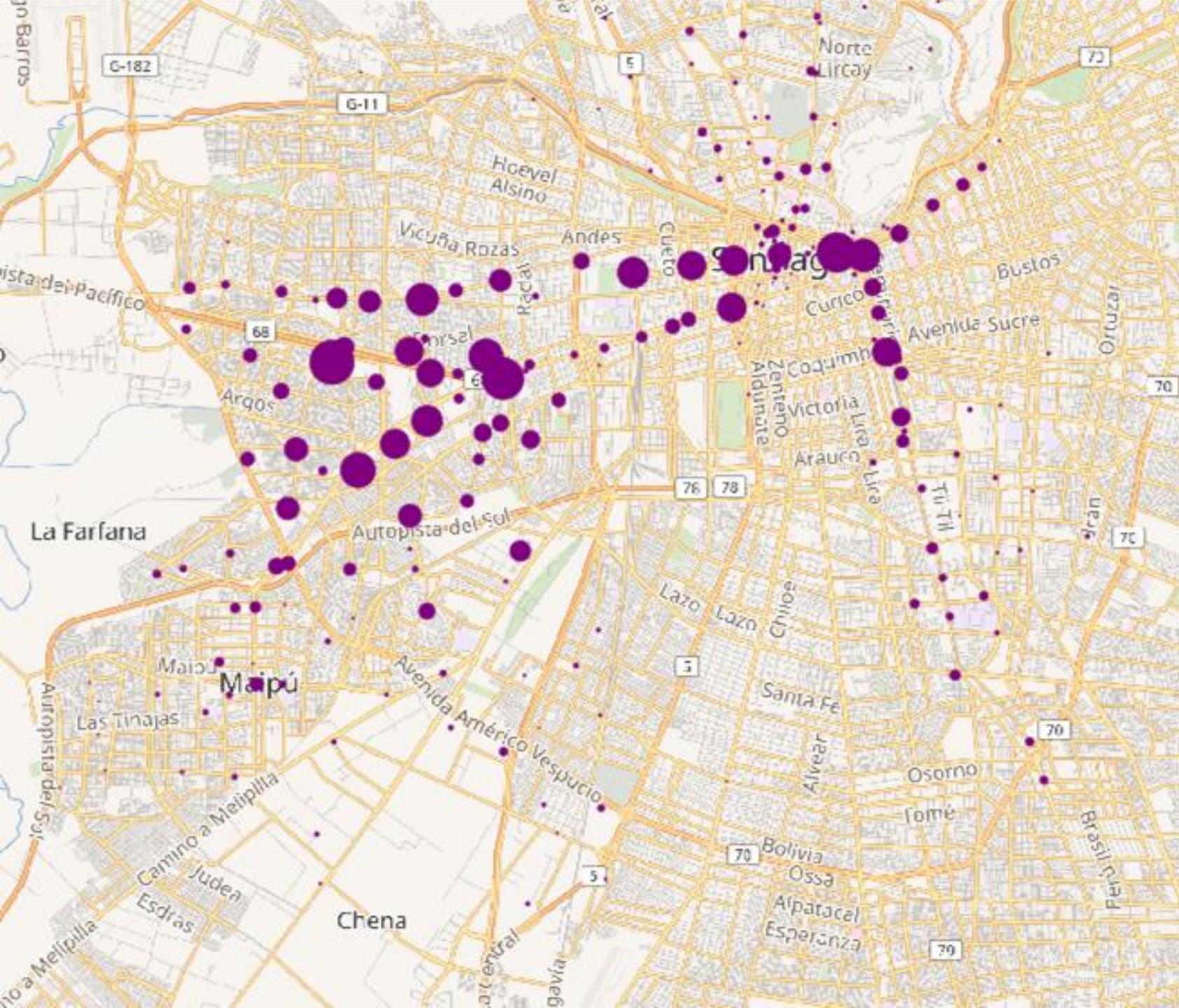


Labeled Towers



Algunas torres están etiquetadas como parte de la **red de metro**.

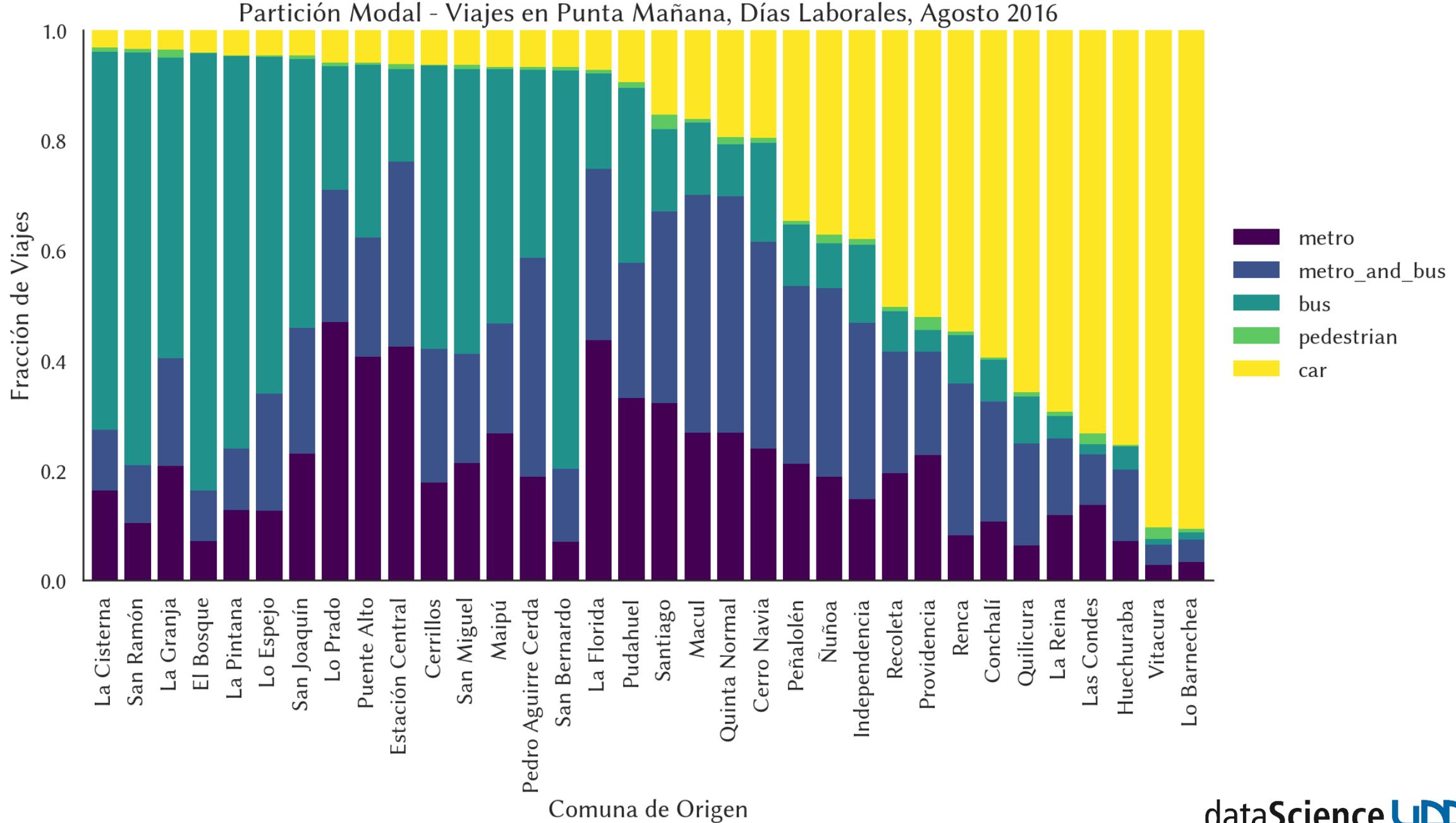
Usando **OpenStreetMap** pudimos etiquetar las cercanas a carreteras y vías principales.



Miramos los datos **agregados** para cada usuario

Nos focalizamos en las **conecciones intermedias** y buscamos **co-ocurrencias**.

En el caso de este dispositivo, **¿Cuál es el modo de transporte más probable para esta persona?**



“... Piensen en cómo comenzaron las cadenas de suministro verdes y sostenibles: algunas empresas fueron líderes en mostrar al mundo que era importante respetar el medio ambiente y la situación de los trabajadores que se encuentran lejos. Ahora, las cadenas de suministro sostenibles son una necesidad comercial. Lo mismo sucederá con los grandes datos para el bien social. Entonces, las compañías que hoy invierten en esta área serán las líderes en el futuro.”

—Richard Benjamins, **Embajador de DATA & AI en Telefónica, LUCA**

# USO DE DATOS MOVILES PARA EL BIEN SOCIAL

**Loreto Bravo**

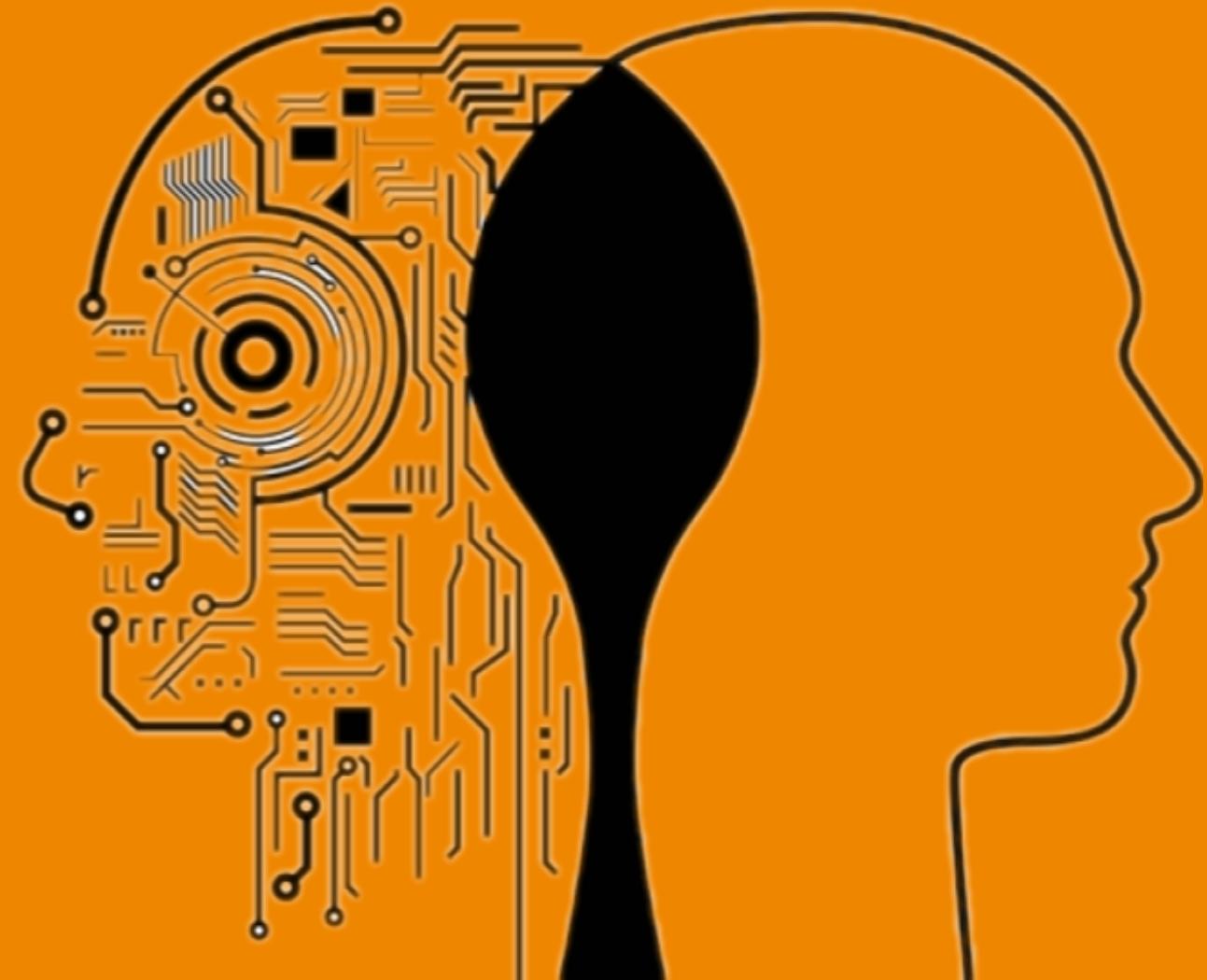
Director @ Instituto de Data Science, UDD  
Telefónica R&D Fellow

## Ai Readiness

Diagnóstico de la adopción de la inteligencia artificial (IA) de empresas en Chile

Informe completo en:

[https://datascience.udd.cl/AI\\_ Readiness.pdf](https://datascience.udd.cl/AI_ Readiness.pdf)



# AI Readiness

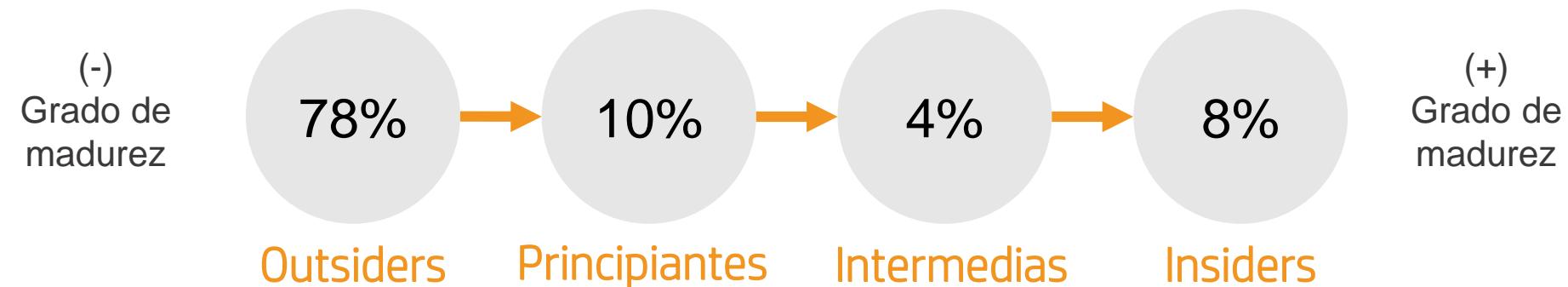
## USANDO IA PARA EVALUAR LA ADOPCIÓN DE LA IA



El objetivo de este estudio es analizar el estado actual de las empresas en Chile respecto de la **incorporación de la IA** en sus procesos, productos y servicios.

Utilizamos algoritmos de **clusterización** para agrupar a las empresas.

Se obtuvieron **4 grupos** ordenados en función de su grado de madurez en la incorporación de Tecnologías de IA





# AUTOPERCEPCIÓN DE LAS EMPRESAS

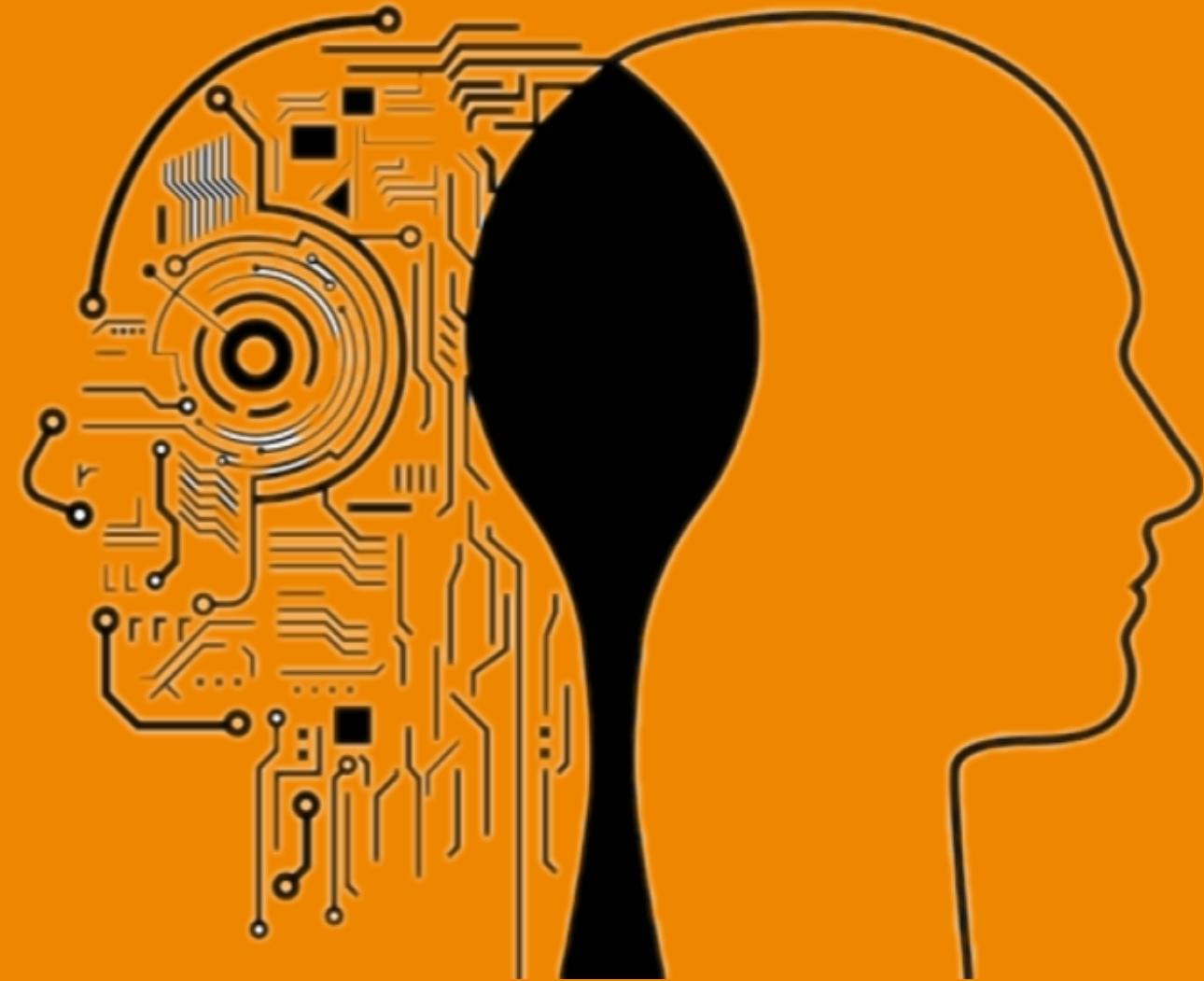
1  
de cada 5 empresas  
consideran que no  
deben incorporar IA a  
su modelo de negocios

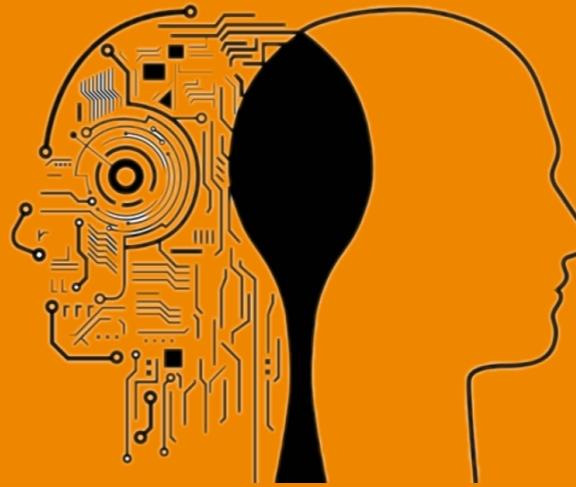
## Ai Readiness

Diagnóstico de la adopción de la inteligencia artificial (IA) de empresas en Chile

Informe completo en:

[https://datascience.udd.cl/IA\\_ Readiness.pdf](https://datascience.udd.cl/IA_ Readiness.pdf)





Dra. Loreto Bravo & Dra. Daniela Opitz  
Instituto de Data Science, Universidad del Desarrollo  
Santiago de Chile, 30 de julio de 2019.

# RECOMENDACIONES



Outsiders



CONCIENCIA

Principiantes



METODOLOGIAS y TALENTOS  
MULTIDISCIPLINARIOS

Intermedias



VIGILANCIA TECNOLÓGICA

Insiders



GESTION DEL CONOCIMIENTO

# CONCLUSIONES

- Incorporar IA es un fenómeno del cual las empresas no podrán desentenderse por mucho tiempo si desean mantener o mejorar su posición competitiva en la economía digital.
- Esta investigación busca arrojar luces sobre algunos puntos críticos sobre los que hay trabajar, para que cada vez sea menor el número de empresas “outsiders” en nuestro ecosistema.

