

Gender Gaps in Urban Mobility

Leo Ferres

@DataScienceUDD & Telefónica R+D, Chile
lferres@udd.cl | Tw, IG, GH: @leoferres

b: 2018-11-10 11:02:24 -0300 - f: 2018-12-01 13:58:30 -0300
EuroCSS 2018, Cologne, Germany

Motivation

1. Mobility is central to the human species,
2. 68% of the world population will live in cities by 2050 (@UNReport)¹
3. public transportation is central to urban mobility,
4. data inform decisions and policies around the world, including mobility,
5. **most data are not gender disaggregated** (@UN, @Data2x)², and thus gender bias is pervasive,

So, let's **assume** that

- ▶ Mobility is a gendered problem (@Law1999, @Rosenbloom2009)³.

¹<https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>

²<http://data2x.org>

³Our bibliography at <https://github.com/leoferreis/bibliography>

How is mobility a gendered problem?

- ▶ Women “move less” (Adeel2016),
- ▶ Women “move more” (@Psylla2017),
- ▶ They “move about the same” (in terms of predictability, @Song2010)

Few large-scale data sets are available.

Research question

- ▶ Does access to public transport affect women different than men?

Mobility

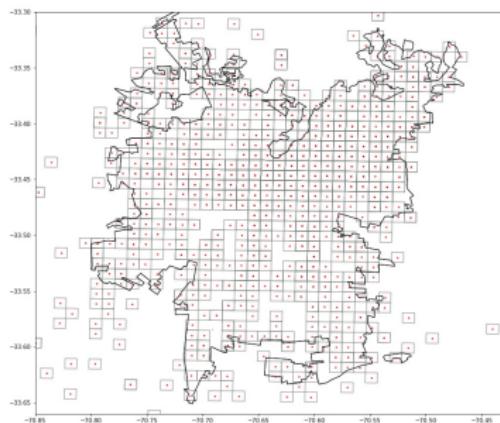
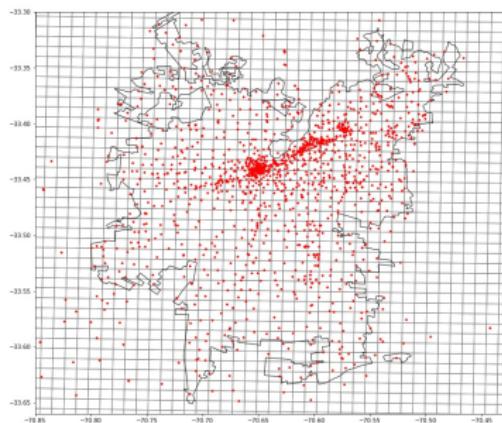
Definition: An individual i “moves more” than an individual j iff $S^i > S^j$, where S is Shannon’s entropy

$$S = - \sum_{l \in L} p_l \ln p_l \quad (1)$$

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

Place

Definition: A place $l \in L$ is a 1Km² cell in a square grid⁴ where there is at least one Base Transceiver Station (BTS, fancy for cell tower).



~1300 cell phone towers, grouped in ~700 spatial clusters, excellent coverage of Santiago metro area

⁴If you'd like to know why I hate Voronois, catch me after the talk and change my mind.

Call Detail Records (CDRs)

Definition: A CDR c is a tuple $\langle n_a, n_b, l_a, l_b \rangle$, where

- ▶ n_a and n_b are the caller and the callee, respectively;
- ▶ l_a and l_b are the “in” and “out” towers, respectively, and
- ▶ t is time of call.

In our case, each c has been enriched with:

- ▶ g_a, g_b , the gender of n_a, n_b (binary, male/female),
- ▶ e_a, e_b , the socio-economic group of n_a, n_b (ABC1, C2, C3, D, E)⁵,
- ▶ u , the number of lines owned by n_a .
- ▶ c_a, c_b whether “prepaid” or “contract”

⁵See <https://bit.ly/2BvONTI>, in Spanish :(

Santiago de Chile

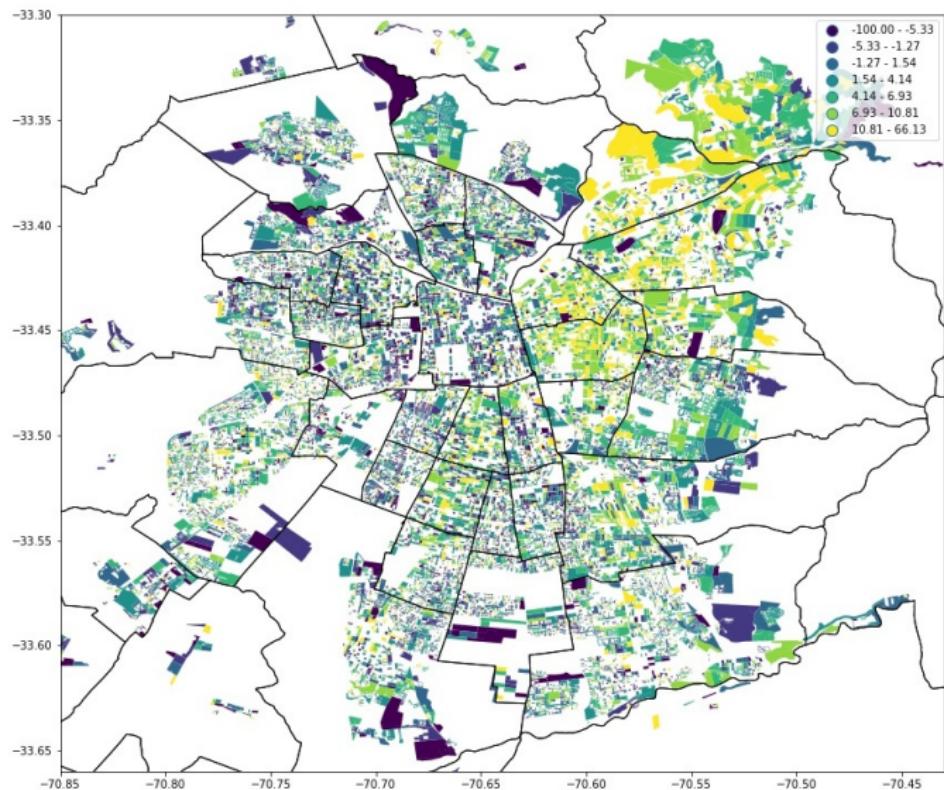


Figure 1: Santiago's comunas and ratio female-males/population per block

Data: GSE (socio-econ)

	ABC1_survey	C2_survey	C3_survey	D_survey	E_survey
ABC1_cdr	0.920507	0.274798	-0.831116	-0.749976	-0.614479
C2_cdr	0.494589	0.437208	-0.434237	-0.506650	-0.529060
C3_cdr	-0.442193	0.057356	0.442734	0.310214	0.105707
D_cdr	-0.779982	-0.454446	0.694044	0.692539	0.699729
E_cdr	-0.750934	-0.488173	0.597858	0.760358	0.694927

Figure 2: Socio-economic group (CDR vs CASEN pearson correlation)

- ▶ (ABC1, C2)/(C3, D, E), roughly cutting at USD \$1000 for C3

Data: CDRs

We analyzed

- ▶ Period:
 - ▶ June-August, 2016 (3 months)
 - ▶ 2,148,132,995 rows (CDRs, calls), 1.06TB
- ▶ Pre-processing, for each (unique) n (in origin and destination):
 - ▶ $u = 1$,
 - ▶ g and e are not null,
 - ▶ $\text{events}(n) > 91$, and finally
 - ▶ $\text{home}(n)$ is not null
 - ▶ c is “contract”

for a total of **372,152** individuals, **50.9%** female.

Data: Public transport (gtfs.org)

- ▶ access to public transport
- ▶ number of reachable stations
- ▶ average velocity to reach other nodes in the network

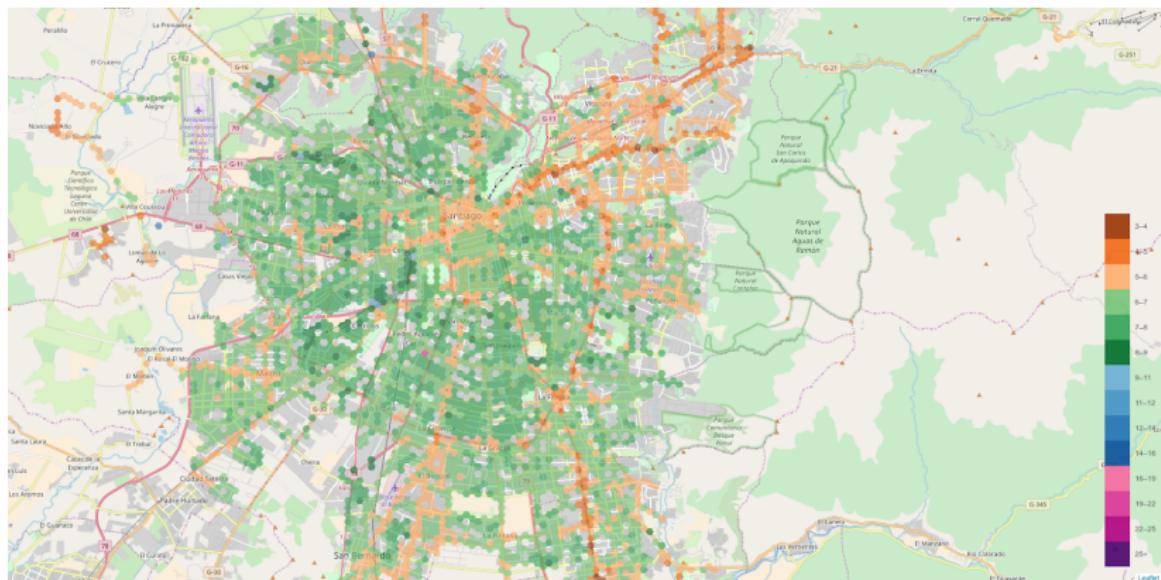


Figure 3: Accessibility map from GTFS public transport data

Results: Entropy

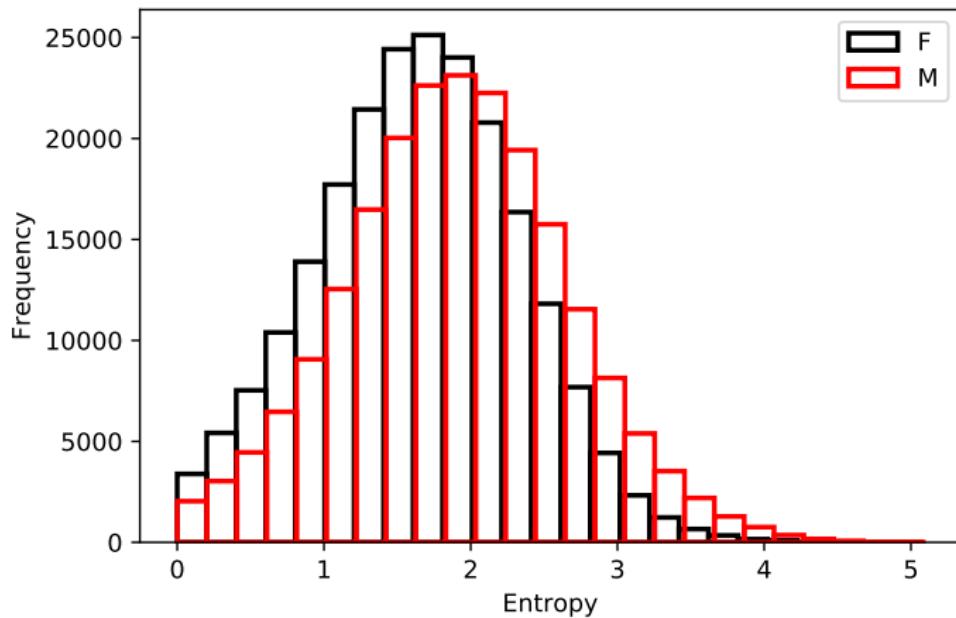


Figure 4: Entropy is different (statistically significant) for males vs females
Ks_2sampResult(statistic=0.144, pvalue=0.0), so in general, women appear to spend more time at fewer locations

Meanwhile, in the third most sold Chilean newspaper...

LATERCERA Temas/ TORNEO NACIONAL ELECCIONES UDI TELETON 2018



Isabel B.
Investigadora SUMO
Observatorio de Movilidad

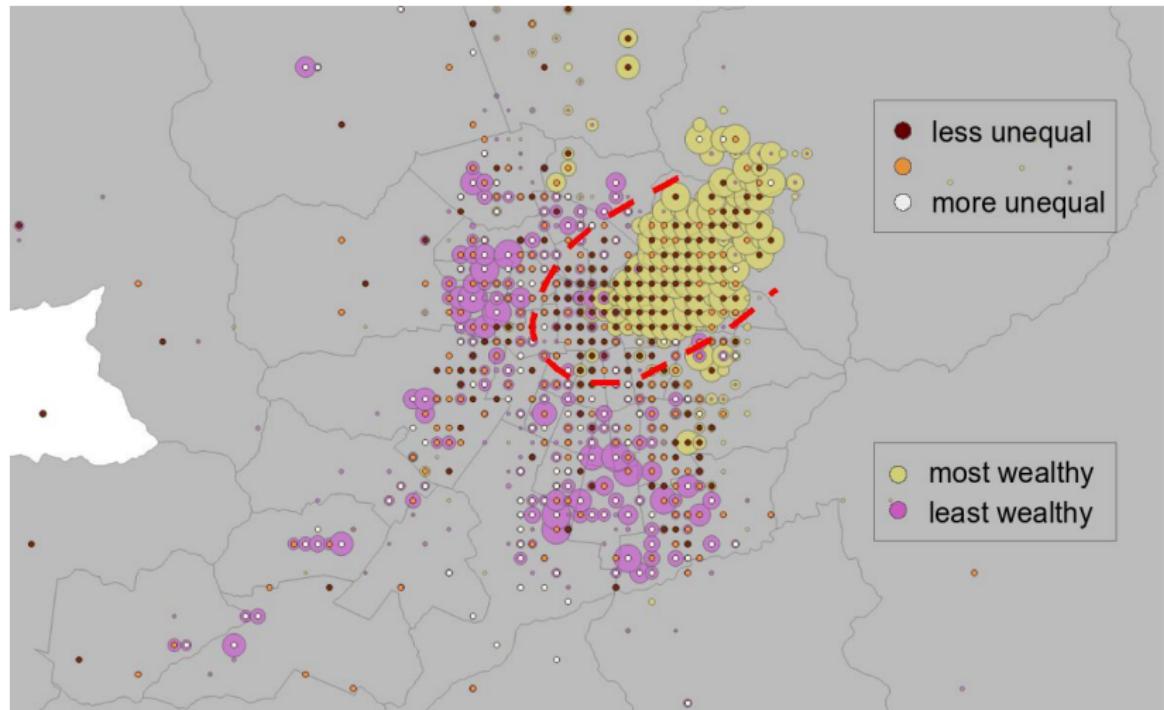
Opinión

Nosotras nos movemos más

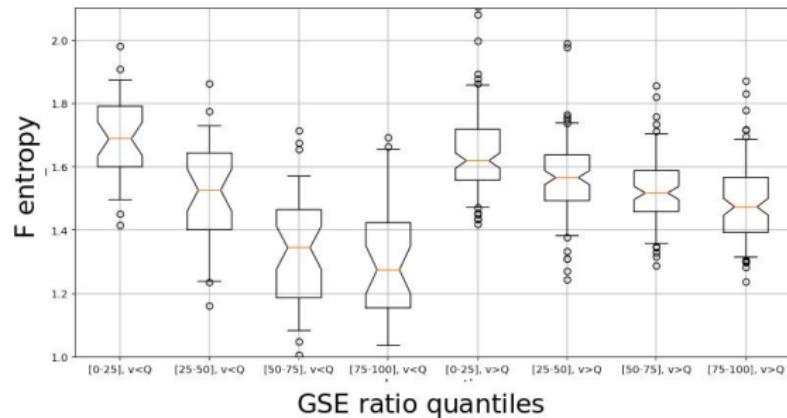
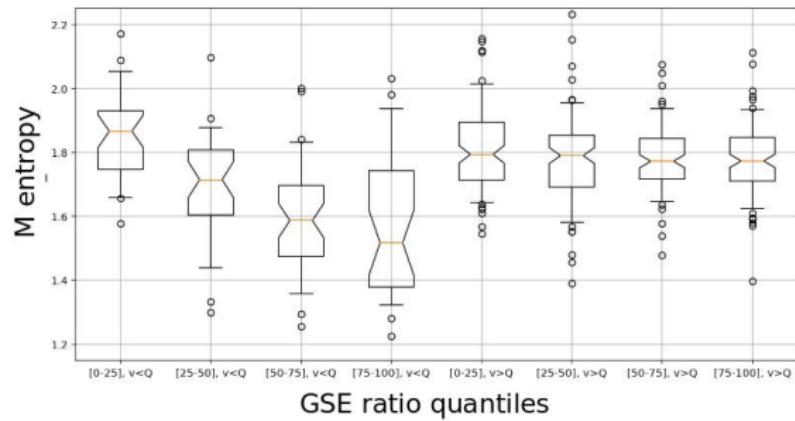
JUE 3 MAY 2018 | 05:07 AM



Results: A gender inequality index



Results: Effect of access to public transport



Conclusions

- ▶ Mobility is (strongly) gendered,
- ▶ public transport accessibility affect women different than men,
- ▶ we need a smarter way to think about mobility, particularly when it comes to gender, and

BUT...

- ▶ Should policies be implemented?
 - ▶ small, incremental,
 - ▶ A/B testing when possible,
 - ▶ involve social scientists and policy makers early on.

In any case,

- ▶ without data equality, there's no gender equality

Cast of characters⁶

- ▶ **Natalia Adler** (UNICEF)
- ▶ **Ciro Cattuto** (ISI Foundation)
- ▶ Pablo García (BCI, formerly Telefónica R& D)
- ▶ Laetitia Gauvin (ISI Foundation)
- ▶ André Panisson (ISI Foundation)
- ▶ Simone Piaggesi (ISI Foundation)
- ▶ Michele Tizzoni (ISI Foundation)
- ▶ Karim Touma (Fallabella, formerly Telefónica Big Data)
- ▶ **Stefaan Verhulst** (The GovLab, NYU)
- ▶ Andrew Young (The GovLab, NYU)

Special gratitude to all our collaborators at Telefónica Chile, without whom all these studies would be (*literally*) impossible, and funding from the United Nations Data2X group.

⁶**bold** is PI on Data2x

Thank you!



Privacy

It's a common question:

- ▶ Numbers are hashed,
- ▶ towers are aggregated at the 1Km² level,
- ▶ no reporting of towers with < 3 different (hashed) numbers,
- ▶ data do not leave TEF servers (except at very high levels of aggregation),
- ▶ We **do not** report on (or care about, really) individuals (only aggregations).

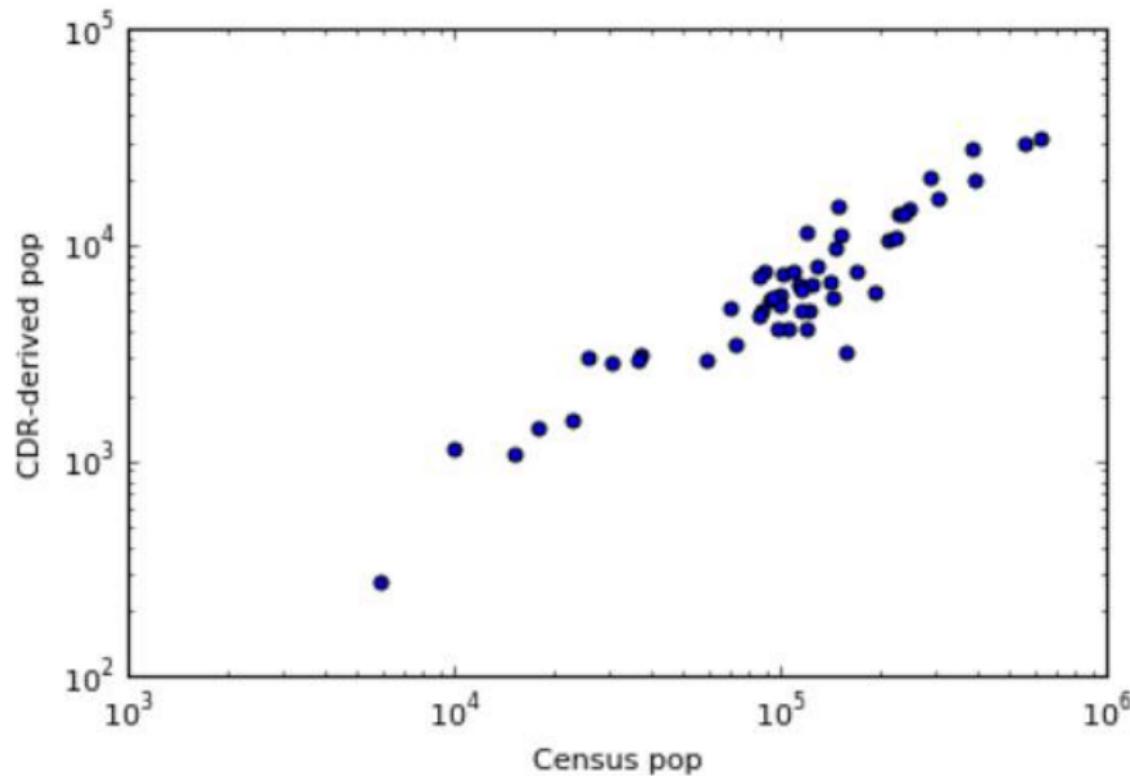
BUT...

Privacy

The **uncommon** question is the following:

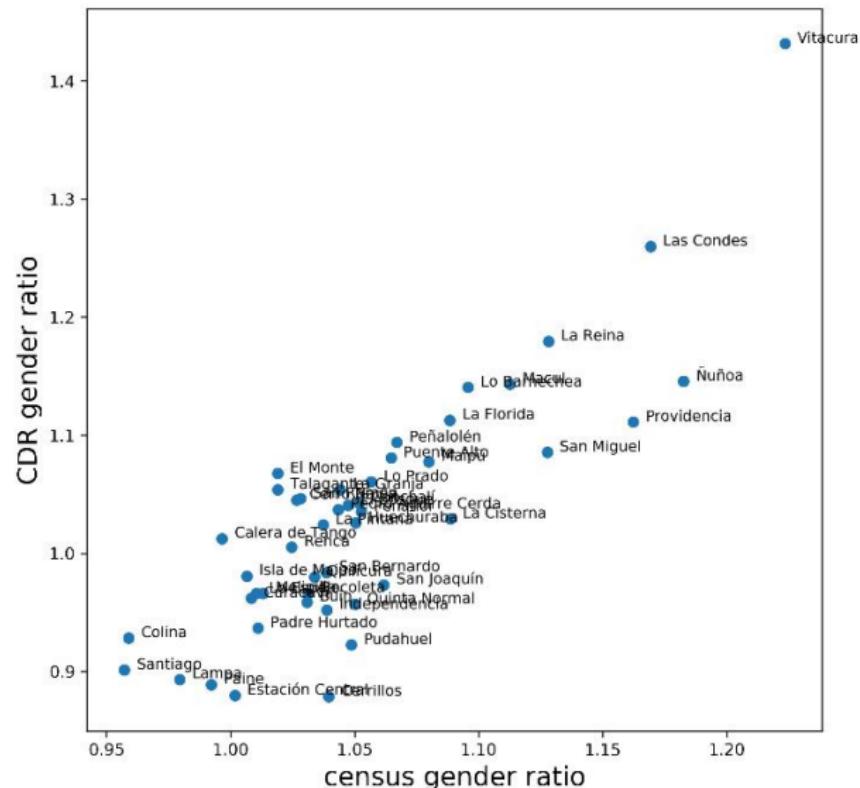
- ▶ what are the social costs of **not** doing these studies?

Representativeness (comuna level)



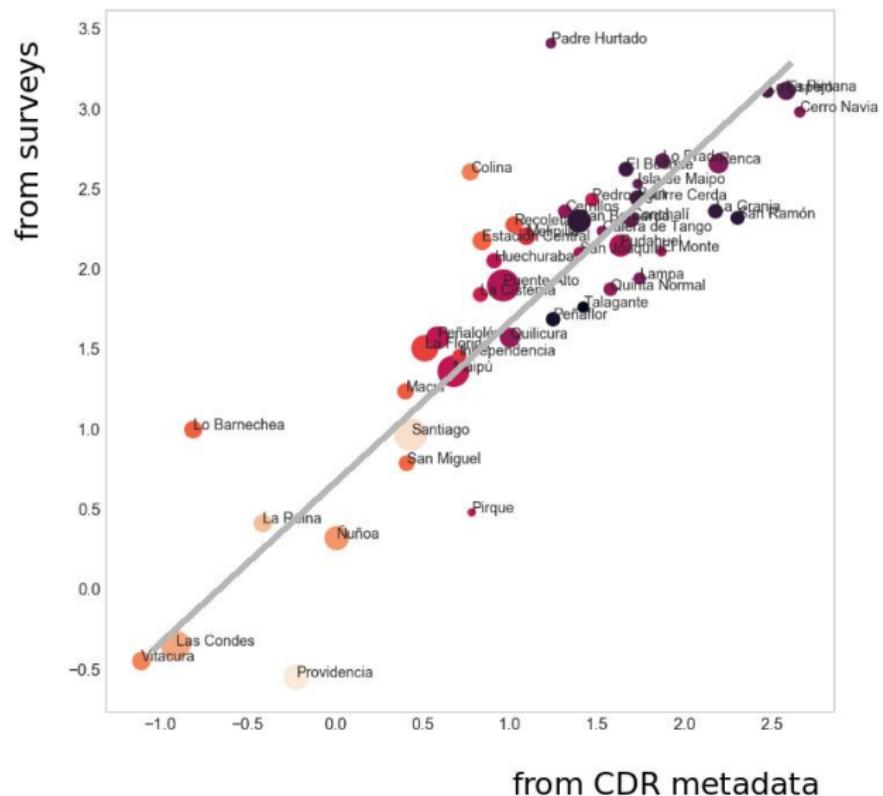
Population distribution, census vs CDRs is pearson .94

Representativeness (comuna level)



Gender-disaggregated population distribution, census vs CDRs is Pearson .85

Representativeness (comuna level)



GSE metadata vs census data, census vs CDRs is pearson .88