

# Ciencia de Redes (Humanas y Sociales)

## Clase #6b

Carlos Sarraute

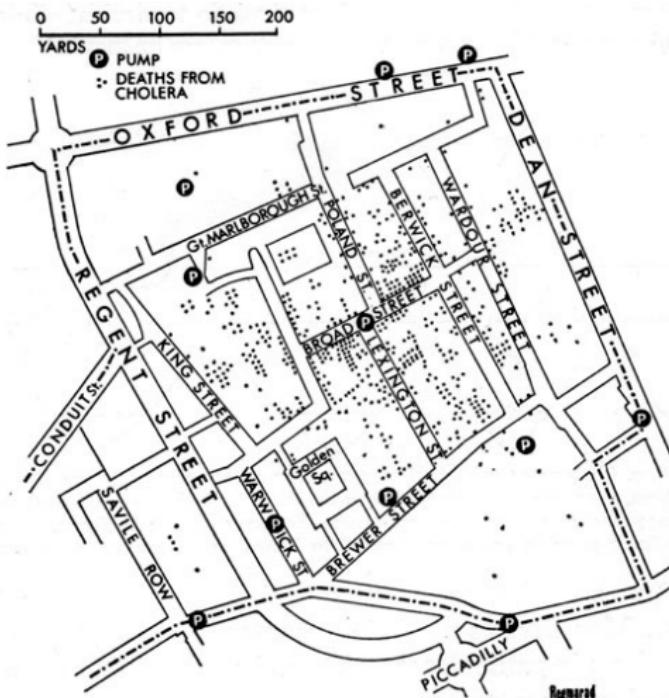
Grandata Labs

Abril - Junio 2019

# Agenda

- 1 Chagas Disease
- 2 Generation of Risk Maps
- 3 Prediction of Long-term Mobility
- 4 (New) Chagas Potential Prevalence Index
- 5 Conclusion

# Data Visualization & Epidemics



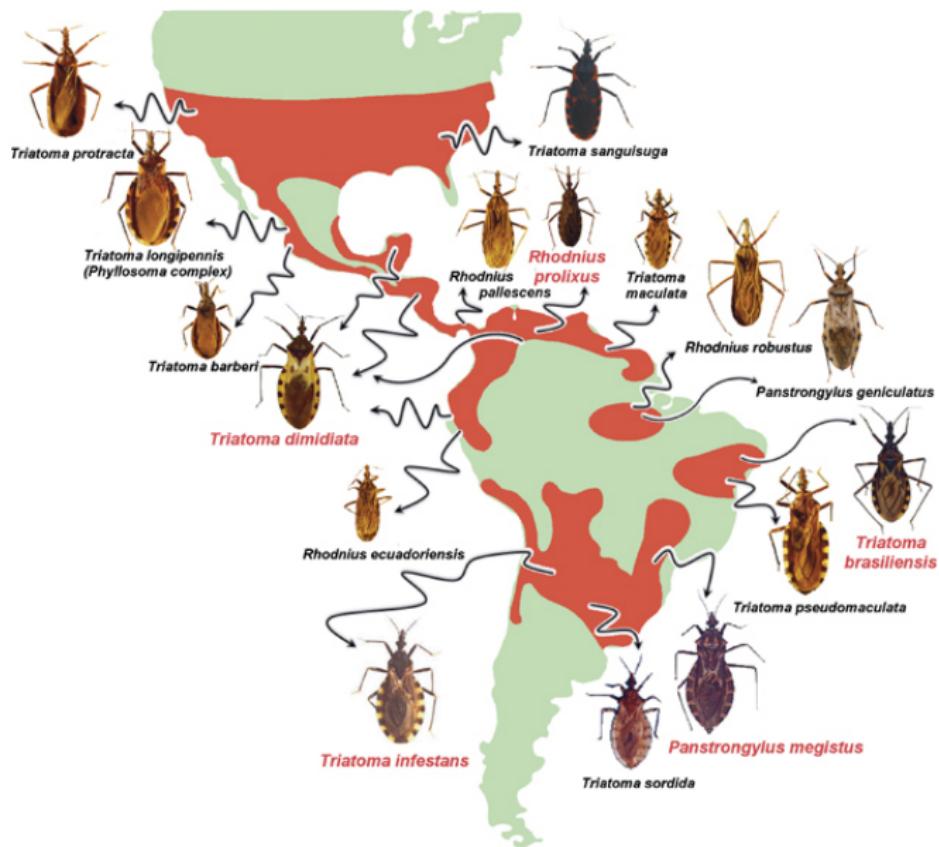
London cholera cases.

Map made in 1854 by Doctor John Snow.

# Chagas Disease

- Chagas is a **neglected tropical** disease.
- Chagas disease is caused by the *trypanozoma cruzi* parasite that extends through the American continent and, recently, some parts of Europe.
- With over **65 million people** exposed and endemic in more than 21 Latin American countries.
- **Asymptomatic phase:** infected individuals can carry the parasite during 10 to 20 years without presenting any symptom.

# Vector: the “kissing bug” or vinchuca

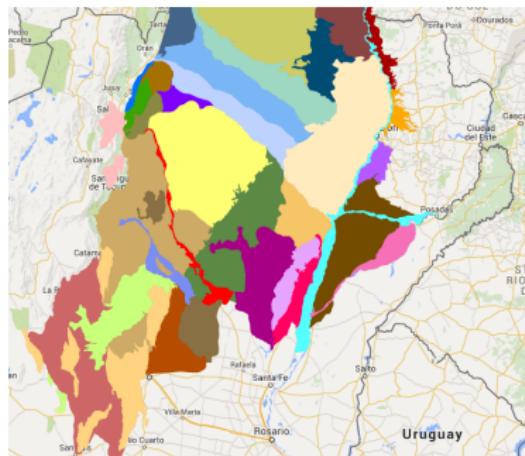


## Housing of the “kissing bug”



# Epidemiology of Chagas

- In Argentina, the disease is endemic in the *Gran Chaco* region.
- National estimates account for at least 1.5 million infected users.
- Only 1 to 2 thousand treatments done yearly.
- About 30% of all infected will develop severe heart and/or digestive disorders.



# Gran Chaco of Argentina



# Epidemiology of Chagas in Mexico

- In Mexico, the disease is endemic in the states shown in the map.
- Non official reports estimate 5.5 million affected persons.
- Less than 0.5% of infected have access to treatments.



# Epidemiology of Chagas

Chagas is a disease:

- Different means of transmission
- Epidemic and unattended.
- Lack of official statistics and public detection campaigns.
- Very low proportion of the infected population knows that they have the disease.
- Asymptomatic phase can last more than 10 years.
- Associated with the Gran Chaco in Argentina...
- ... where are the infected persons today?

# Agenda

- 1 Chagas Disease
- 2 Generation of Risk Maps
- 3 Prediction of Long-term Mobility
- 4 (New) Chagas Potential Prevalence Index
- 5 Conclusion

## Mobile Phone Data

The Argentinian and Mexican datasets consist of **anonymized** call detail records (CDRs), each from one national telco. The former ranges 5 months of data whilst the latter 24.

Each call record consists of:

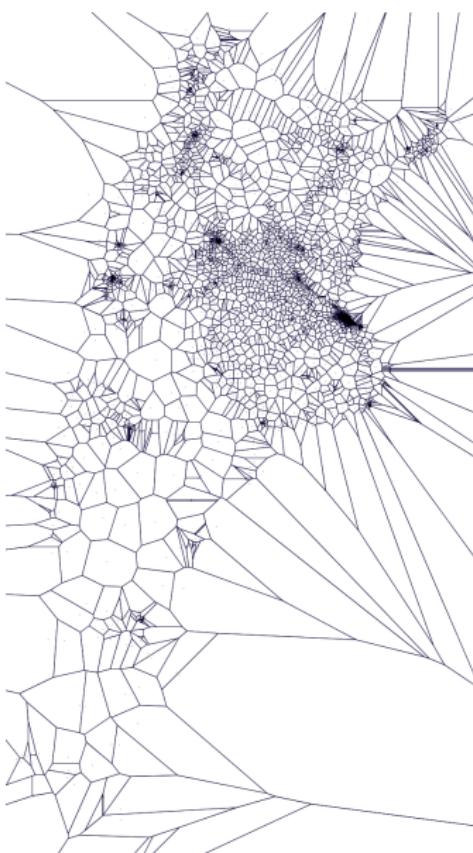
- Origin and destination users.
- ID of origin antenna.
- Start time and duration in seconds.

In all, each dataset has more than 9 billion geolocalized calls.

The population coverage of each telco is of 8 million and 2 million mobile lines for the Argentinean and Mexican datasets respectively.

The antenna datasets consists of more than 5 thousand geolocalized antennas.

# Voronoi Cells in Argentina



# Methodology (1)

## Home Prediction

- As a first step, we determined each user's residence antenna. This was chosen to be the most used frequently used antenna during week evenings.
- Users for which the inferred home antenna is located in an *endemic area* will be considered the set of *endemic users*.

# Methodology (1)

## Home Prediction

- As a first step, we determined each user's residence antenna. This was chosen to be the most used frequently used antenna during week evenings.
- Users for which the inferred home antenna is located in an *endemic area* will be considered the set of *endemic users*.

## Aggregation of vulnerable users

- If a given user communicated with the *endemic area* in the selected period of time, we tagged him as potentially *vulnerable*.
- We aggregated vulnerable users and total users (residents) per antenna.

## Methodology (2)

### Heat Maps

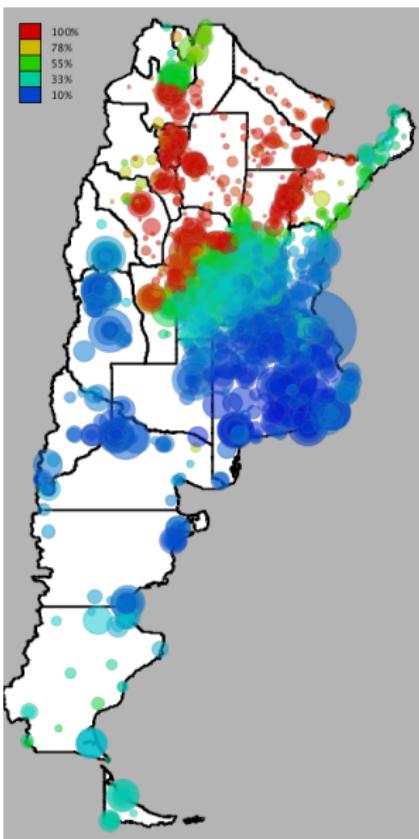
We generated heat maps from these results, plotting a colored circle around each antenna, which encodes its *vulnerable* communication patterns.

- the **area** depends on the volume of use.
- the **color** corresponds to the percentage of vulnerable use in that antenna (whether calls or users).

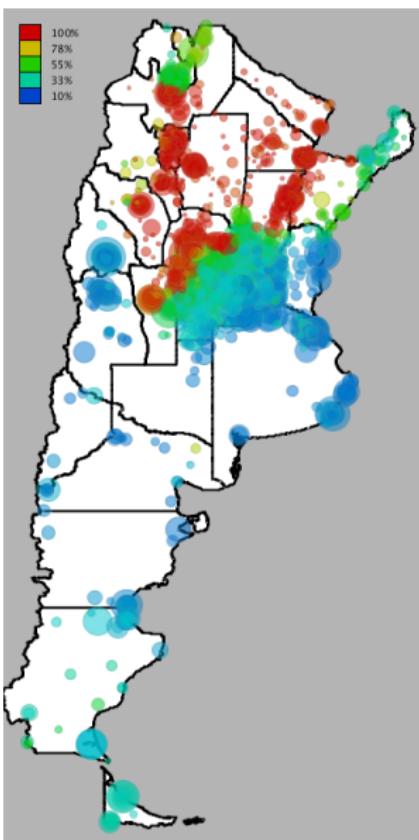
### Antenna Filter

$\beta$  will be our parameter to filter antennas by minimum vulnerable interaction rates.

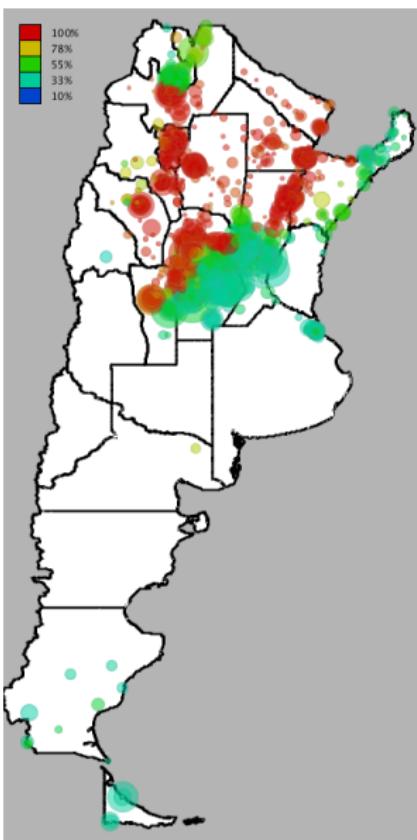
# Heat Map Argentina, $\beta = 1\%$



# Heat Map Argentina, $\beta = 15\%$



# Heat Map Argentina, $\beta = 30\%$



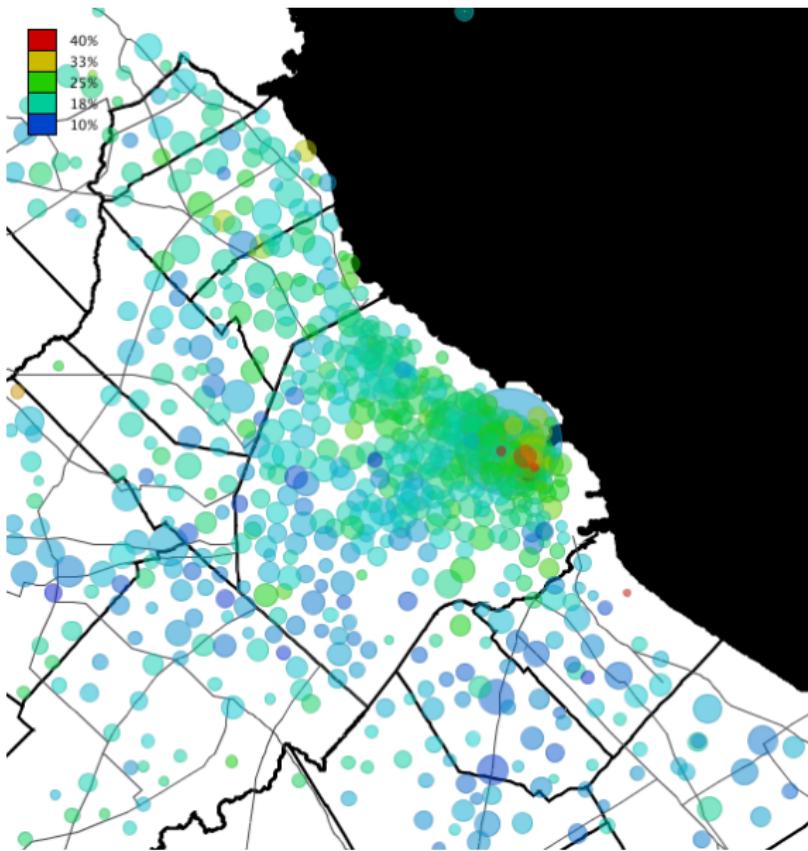
# Focused Areas

We focused visualizations in specific regions outside of Gran Chaco.

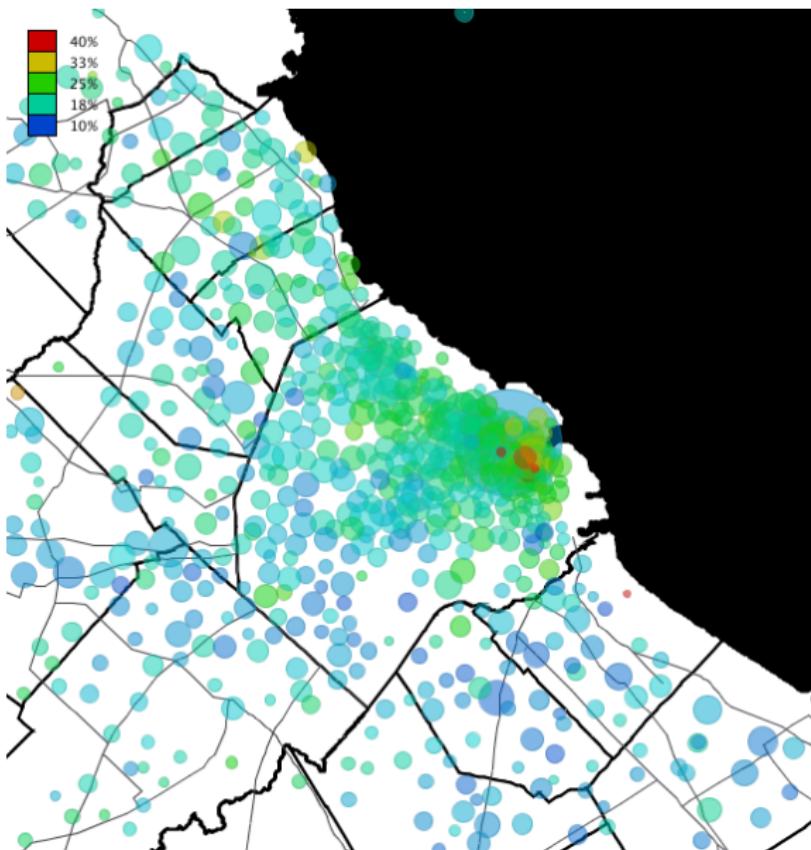
## Studied Areas

- Tierra del Fuego and South Santa Cruz
- Chubut.
- East Río Negro.
- Buenos Aires.
- Capital Federal, South and North CABA, and AMBA.
- Central Córdoba and Santa Fe.

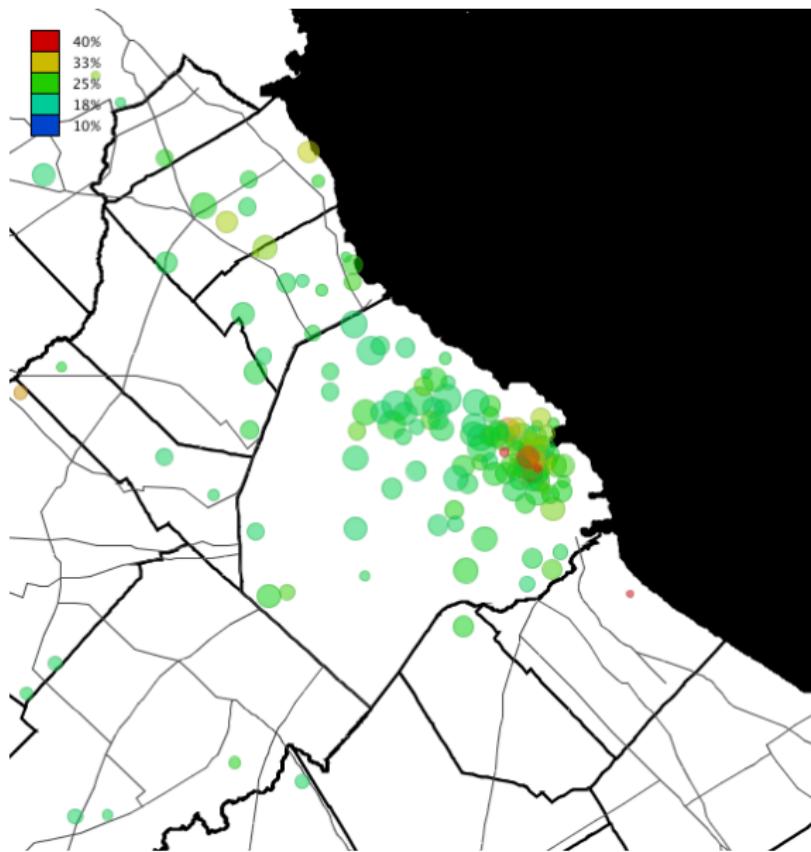
# Heat Map Buenos Aires area, $\beta = 2\%$



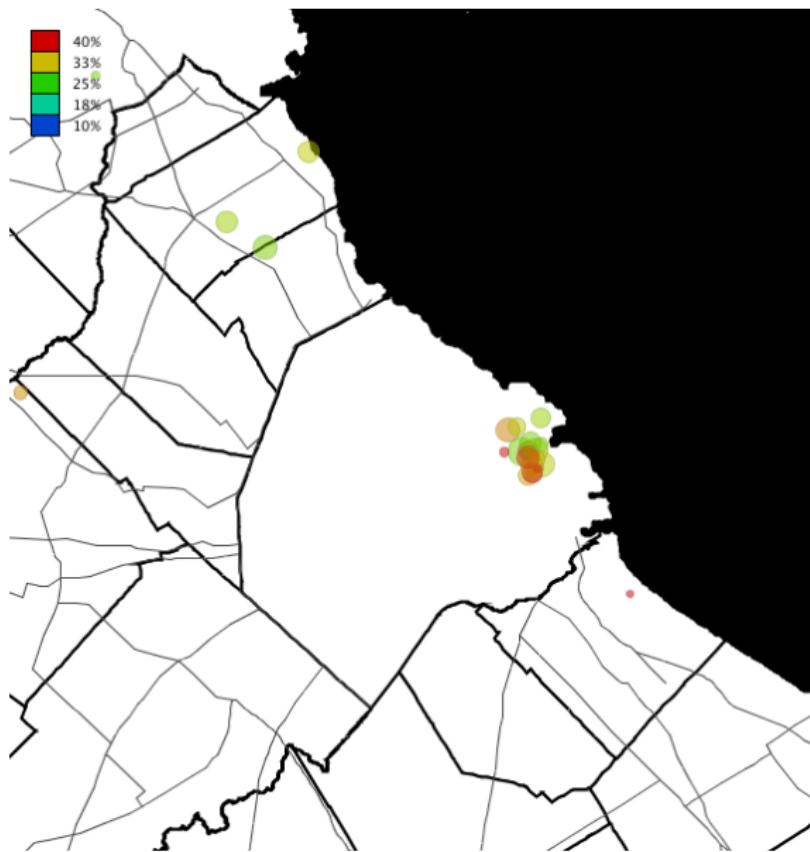
# Heat Map Buenos Aires area, $\beta = 10\%$



# Heat Map Buenos Aires area, $\beta = 20\%$



# Heat Map Buenos Aires area, $\beta = 28\%$



## Argentine Communities Detected

After filtering the maps, we are left with few antennas that stick out for their risk level. Using their location, we assign them to Argentinian communities or cities.

### Some communities outside the big cities

- Cordoba: Freyre, La Tordilla, Balnearia, ...
- AMBA: Avellaneda, Parque Patricios, San Isidro, ...
- Bs As Province: Lima, San Nicolas.
- La Rioja: Chamical and Malanzán.
- Salta: Tartagal.

# Agenda

- 1 Chagas Disease
- 2 Generation of Risk Maps
- 3 Prediction of Long-term Mobility
- 4 (New) Chagas Potential Prevalence Index
- 5 Conclusion

# Mobility Classifier

## Target Variable

Let  $T_0$  and  $T_1$  be time periods corresponding to (01/01/14, 31/07/15 ) and (01/08/15, 31/12/15) respectively.

Consider  $U_0$  to be the set of users that lived in the **mexican** endemic region  $E_Z$  during period  $T_0$ . Then our target variable  $Y$  for every user  $u$  is:

$$Y_u = \begin{cases} 1 & \text{if } u \in U_0 \\ 0 & \text{in other cases.} \end{cases}$$

where  $u \in U_0$  iff the user's home antenna is in  $E_Z$  during  $T_0$ .

## Feature Matrix Description

The features were constructed using data during the period  $T_0$  and amount to a total of 130 variables per user.

- Volume of ten most used antennas.
- Mobility diameter.
- Graph data and communications, segmented accordingly:
  - Month of interaction.
  - Time of the week.
  - Vulnerability of the interactions.
  - Duration, volume and direction of the interactions.

### Feature - Target Correlation

As expected, there was a high Pearson correlation between  $Y$  and calling patterns with the endemic region.

# Agenda

- 1 Chagas Disease
- 2 Generation of Risk Maps
- 3 Prediction of Long-term Mobility
- 4 (New) Chagas Potential Prevalence Index
- 5 Conclusion

# Housing Conditions

We used information from the 2010 National Population and Housing Census, and we constructed an index that quantifies for each house its viability conditions for lodging the *vinchuca*.



## Condiciones de la vivienda

- Material predominante de los pisos
  - Cerámica, baldosa, mosaico, mármol, madera, alfombrado
  - Cemento o ladrillo fijo
  - **Tierra o ladrillo suelto**
  - Otro
- Material predominante de la cubierta exterior del techo
  - Cubierta asfáltica o membrana
  - Baldosa o losa (sin cubierta)
  - Pizarra o teja
  - Chapa de metal (sin cubierta)
  - Chapa fibrocemento o plástico
  - Chapa de cartón
  - **Caña, palma, tabla o paja con o sin barro**
  - Otro
- Revestimiento interno o cielorraso del techo (si / **no**)

## Health Vulnerability

The proposed notion of health vulnerability is composed of the following associated factors:

**Access to health services from the state:** The proximity to health providers was used as the main indicator. The walking time was calculated from various points to the nearest Health Center.

**Socio-economic Index of the population (SEI):** To construct the SEI, we used census information. Relevant variables (educational level, indicators of unsatisfied basic needs, etc.) were selected and combined using *variational autoencoders*, a method for dimensionality reduction based on neural networks.

# Accessibility to Hospitals and Health Centers

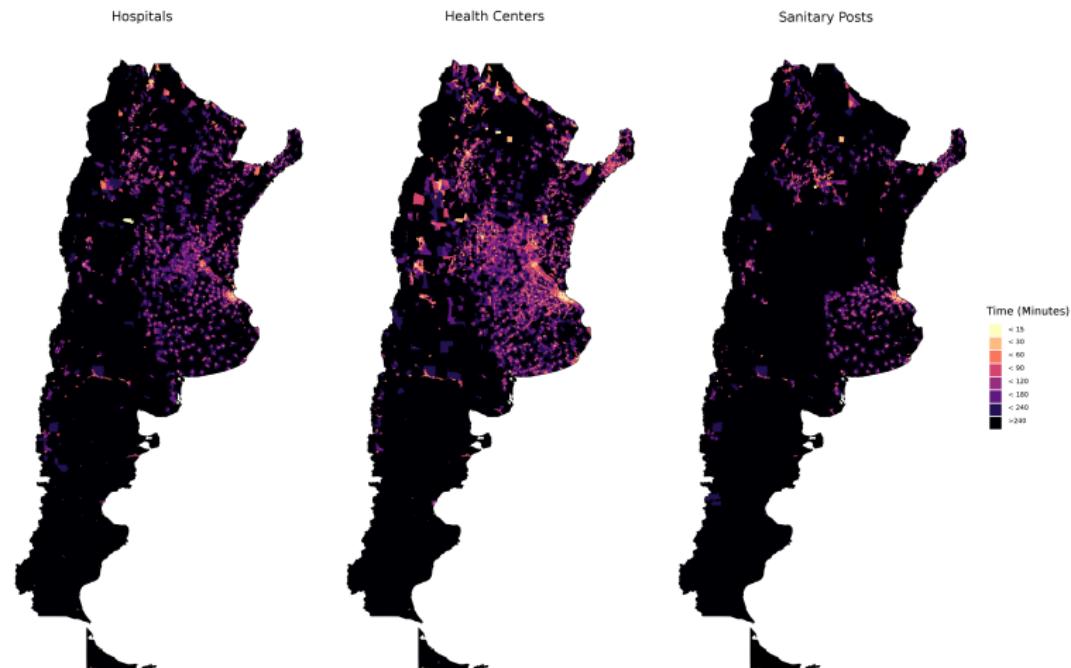


Figure: Walking times to health providers. Aggregated by census block.

## Socio-Economic Index (SEI)

We decided to calculate an index for each head of household in the Census 2010 dataset. Indicators used:

| Variable                                | Unit              |
|---|-------------------|
| Condition of home ownership             | Housing           |
| Quality of Materials                    | Housing           |
| Quality of Connection to Basic Services | Housing           |
| Quality of Construction                 | Housing           |
| Overcrowding                            | Household         |
| Unsatisfied Basic Needs (UBN) indicator | Household         |
| Educational level of the Household      | Household         |
| Number of Unemployed in Household       | Household         |
| Existence of domestic services          | Household         |
| Activity condition                      | Individual (head) |
| Educational level                       | Individual (head) |

# Resulting SEI by province

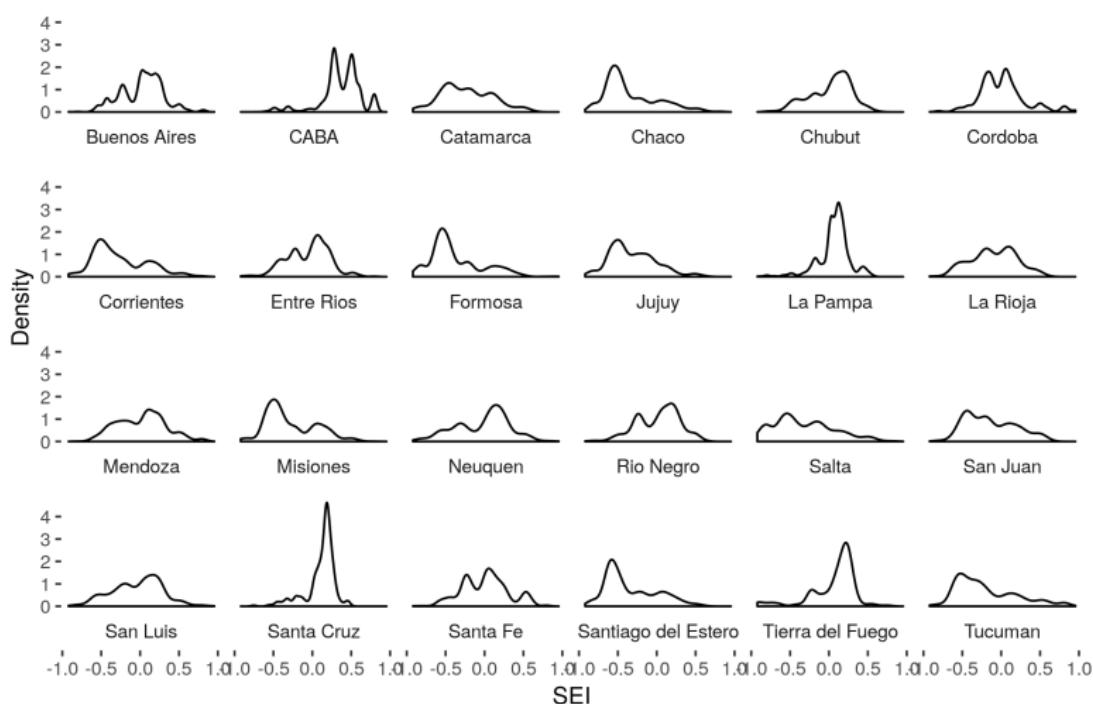


Figure: Density plot of the SEI by province.

## Chagas Potential Prevalence Index (ChPPI)

For the final construction of *ChPPI*, the following variables were taken into account:

- local affinity with endemic area of Chagas,
- health vulnerability, and
- population density.

From the point of view of public policies, with limited funds, it makes sense to prioritize areas with high population density and/or high “affinity”.

# Chagas Potential Prevalence Index (ChPPI)

The final index was composed as follows:

$$ChPPI_r = \frac{HV_r^\alpha d_r^\beta AI_r}{\frac{1}{R} \sum_{r=1}^R HV_r^\alpha d_r^\beta AI_r} \quad (1)$$

where

- $HV_r^\alpha$  moderates the effect of the Health Vulnerability component with  $\alpha$  being the parameter that determines the impact;
- $d_r^\beta$  penalizes depopulated areas and  $\beta$  functions as the regulator;
- $AI_r$  is the Affinity Index for block  $r$ . In the denominator,  $R$  is the total number of census blocks.

# Results

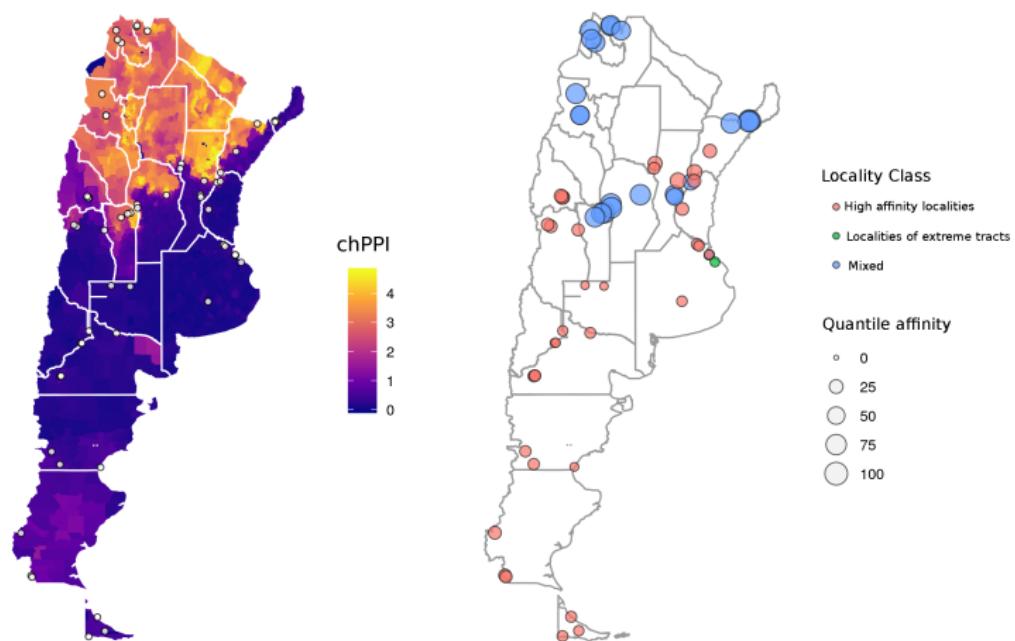


Figure: Localities selected as potential areas of intervention.

# Agenda

- 1 Chagas Disease
- 2 Generation of Risk Maps
- 3 Prediction of Long-term Mobility
- 4 (New) Chagas Potential Prevalence Index
- 5 Conclusion

# Conclusions

## Expected

Heat maps show *temperature* falling from **Gran Chaco** outwards

# Conclusions

## Expected

Heat maps show *temperature* falling from **Gran Chaco** outwards

## Unexpected

The affinity does not decrease in a continuous gradient as the localities move away from the endemic area (EA).

On the contrary, localities were detected in Province of Buenos Aires and Patagonia whose degree of affinity is much higher than population centers in provinces closer to the EA.

This suggests the existence of considerable migrations from endemic regions to the highlighted localities.

## Detection campaigns

The maps provide a tool to **prioritize detection campaigns for Chagas Disease.**

In 2019/2020, four diagnosis campaigns will be run by Fundación Mundo Sano and Fundación Bunge y Born, based on a detailed analysis of the maps and the field knowledge of the Foundation experts.

The **diagnosis campaigns** will be run in:

- Center region of Argentina
- Patagonia region
- Province of Buenos Aires
- Metropolitan region of Buenos Aires