The City Pulse of Buenos Aires

Carlos Sarraute[1], Carolina Lang[1], Nicolas B. Ponieman[2], and Sebastian Anapolsky[3]

[1]Grandata Labs – Buenos Aires, Argentina

[2]Physics Department, UBA

[3]Mobility and transport specialist

Workshop Big Data & Environment
November 12, 2015

Agenda

Presentation

Grandata

- Founded in 2012.
- Research team: 5 persons, based in Vicente Lopez.
- We investigate "Human Dynamics"
  - applying "Big Data" analysis to social relations and human behavior
  - integrating data from mobile phone operators and banks
  - to characterize and predict customer actions.

Presentation

Grandata

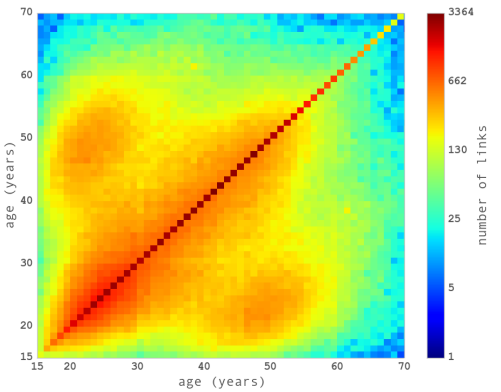
- Founded in 2012.
- Research team: 5 persons, based in Vicente Lopez.
- We investigate "Human Dynamics"
  - applying "Big Data" analysis to social relations and human behavior
  - integrating data from mobile phone operators and banks
  - to characterize and predict customer actions.

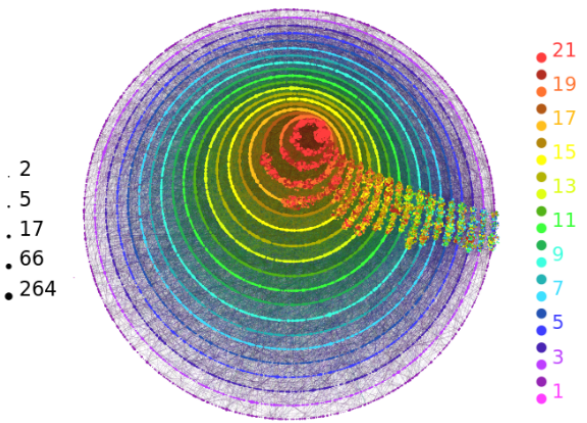
Scientific Collaborations

- Aline Viana from INRIA (Paris)
- Eric Fleury, Marton Karsai from ENS (Lyon)
- Sandy Pentland and his "Human Dynamics" team at MIT
- UBA and Fundación Mundo Sano

Some Research Lines

- Infer age and gender for all nodes (in Mexico)
- Study mobility patterns and mobility flows in the city (in Buenos Aires)
- Infer Home and Work location for all nodes.
- Integrating categorically different social networks to enhance our understanding of users social behavior and interactions
- Infer bank variables (total spending, total debt) for all nodes





Research Objective

In this project, we analyze the viability of using cell phone records to lower the cost of urban and transportation planning.

We generated an estimation of the distribution of people among different parts of Buenos Aires city, based on CDRs (call detail records).

Our objectives are to prove that these methods are reliable and to extend the analyses to other situations, taking advantage of the massiveness of the collected data.

Mobile Data Source

- We applied our methodology to the city of Buenos Aires, capital of Argentina.
  - Main political, financial and cultural center of the country.
  - It has **2,890,151 inhabitants**, divided in **48 neighborhoods** which are grouped in **15 communes**.
- Our dataset of call detail records (CDRs)...
  - ...has **200M calls**, corresponding to 4.95M phone users, in a period of three months.
  - ...contains, for every call: origin (caller), destination (callee), timestamp, duration and (geolocated) antenna used to connect.
- We used that information to associate each call to a particular neighborhood or commune.

Agenda

Location Distribution Matrix

We separated a typical week into 4 day groups and 4 hour groups:

Day groups	**Hour groups**	
Monday to Thursday	Morning	5am - 11am
Friday	Noon	11am - 3pm
Saturday	Afternoon	3pm - 8pm
Sunday	Night	8pm - 5am (next day)

The hour group selection corresponds to an analysis realized with the data of the Origin-Destination survey ENMODO 2010, from which we determined the peaks and valleys of mobility, for a typical working day in the city.

Location Distribution Matrix

- Let $\mathcal{C}$ be the set of communes and $R_{u,d,h,c}$ the number of calls made by user u on day group d, hour group h, in commune c.
- The proportion of calls that user u made in commune c during a combination of day group d and hour group h is

$$P_{u,d,h,c} = \frac{R_{u,d,h,c}}{\sum_{c' \in \mathcal{C}} R_{u,d,h,c'}}$$

  or 0 if the denominator is zero.
- The matrix P_u is the *Location Distribution Matrix* of user u.

Criteria for Filtering

We filter the users that don't provide enough information on their location; more precisely we only take into account the users that have enough calls in every one of the 16 day/hour groups. The user u is kept if

$$\sum_{c' \in \mathcal{C}} R_{u,d,h,c'} \geq \tau$$

for any combination of d and h, given a threshold τ (in our study $\tau = 1$). After filtering, we obtain a set of 73,000 users which we denote $\mathcal{U}$.

Scaling up to Census Population

We determine the home commune H_u for every user $u \in \mathcal{U}$. We consider that a user is at home on weekdays, at night:

$$H_u = \arg\max_{c \in \mathcal{C}} R_{u,\text{weekday,night},c}$$

With that information, we extend our predictions using the census data from INDEC (2010). The scaling factor F_c for commune c is:

$$F_c = \frac{\text{pop}_c}{\#\{u \in \mathcal{U} | H_u = c\}}$$

where pop_c is the population of commune c according to the census. The range of scaling factors goes from $\mathbf{17.26}$ in commune 2 to $\mathbf{93.29}$ in commune 8.

Summing up

We now define the expected quantity of people in a commune $\boldsymbol{c}$, during a combination of day group $\boldsymbol{d}$ and hour group $\boldsymbol{h}$ as:

$$\boldsymbol{EP_{d,h}[c]} = \sum_{\boldsymbol{u \in \mathcal{U}}} (\boldsymbol{P_{u,d,h,c}} \cdot \boldsymbol{F_{H_u}})$$

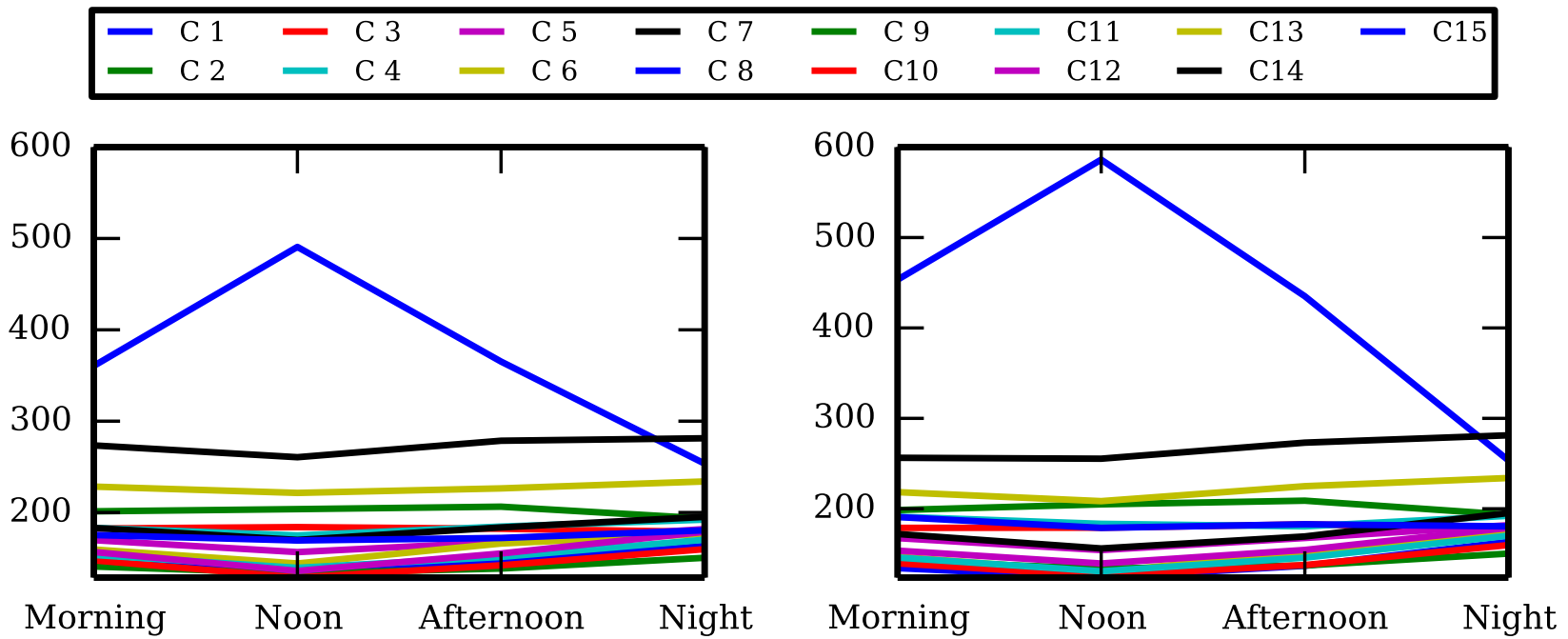
Additionally, the expected quantity of people found in commune $\boldsymbol{c}$, during a day group $\boldsymbol{d}$ and hour group $\boldsymbol{h}$, and that live in commune $\boldsymbol{c'}$ is given by:

$$\boldsymbol{EP_{d,h}[c][c']} = \sum_{\boldsymbol{u \in \mathcal{U} | H_u = c'}} (\boldsymbol{P_{u,d,h,c}} \cdot \boldsymbol{F_{c'}}) \,.$$

Note that $\boldsymbol{EP_{d,h}[c]} = \sum_{\boldsymbol{c' \in \mathcal{C}}} \boldsymbol{EP_{d,h}[c][c']}$.

Agenda

Method Validation



Comparison between the ENMODO survey (left) and the proposed methodology (right).

Method Validation

We validated the methodology against the ENMODO 2010 origin-destination survey, a widely used method in urban planning.

The obtained results are similar for both methods (the average difference between the estimations is $\mathbf{5\%}$).

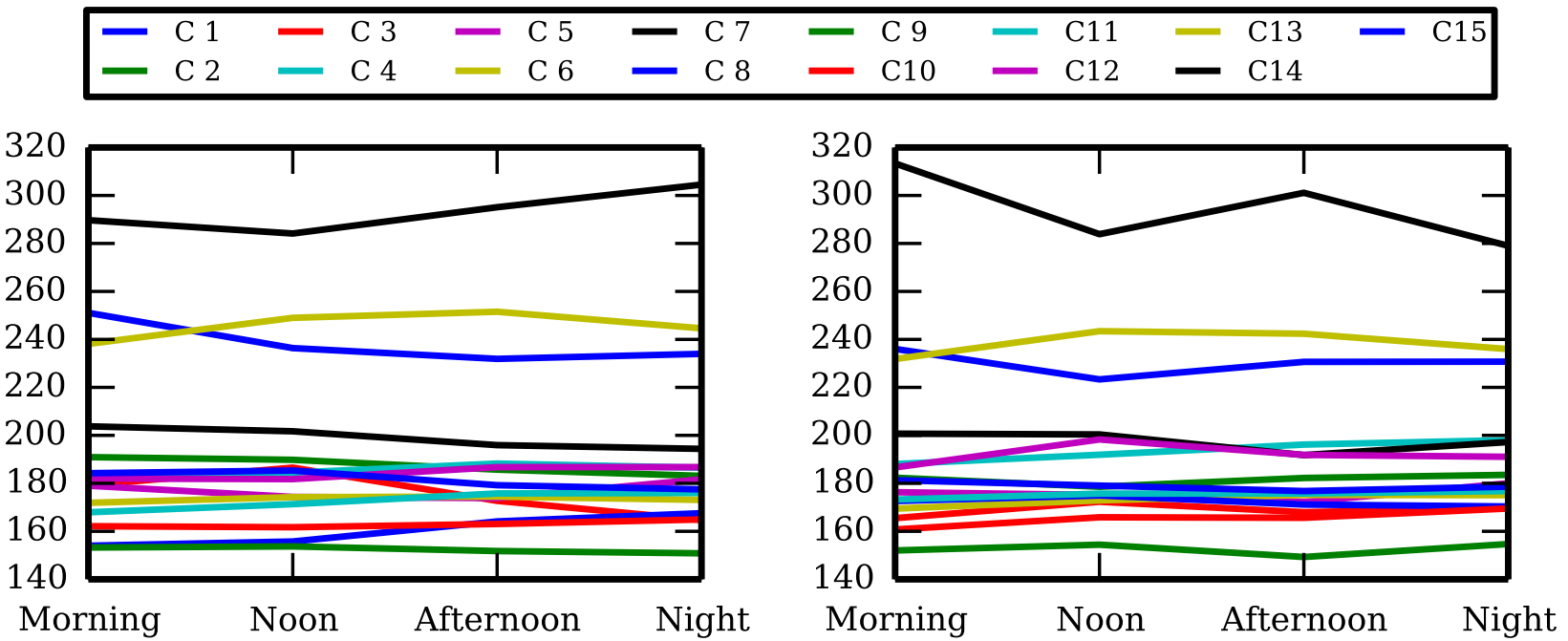
The highest variation appears in Commune 1, in the morning ($\mathbf{20\%}$) and the second highest in Commune 6, at noon ($\mathbf{11\%}$).

Extension to Weekends

We extended our analysis, as a next step, to weekends (regular origin-destination surveys are only held on weekdays). We are thus presenting **new results** on the mobility of the citizens in Buenos Aires.

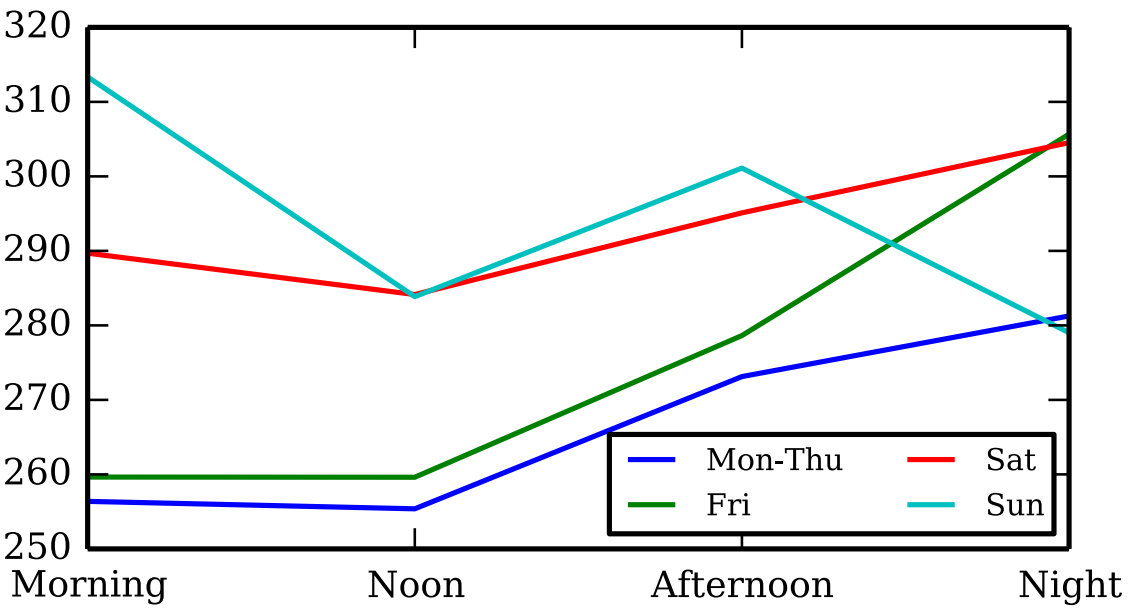
Weekend patterns differ widely from weekday patterns: the major pole of attraction isn't Commune 1 (central business district) anymore, whereas other communes (more famous because of their bars and night clubs) start attracting more people, mainly Commune 14.

Extension to Weekends



Our results for Saturday (left) and Sunday (right).

Analysis of Commune 14 (Palermo)



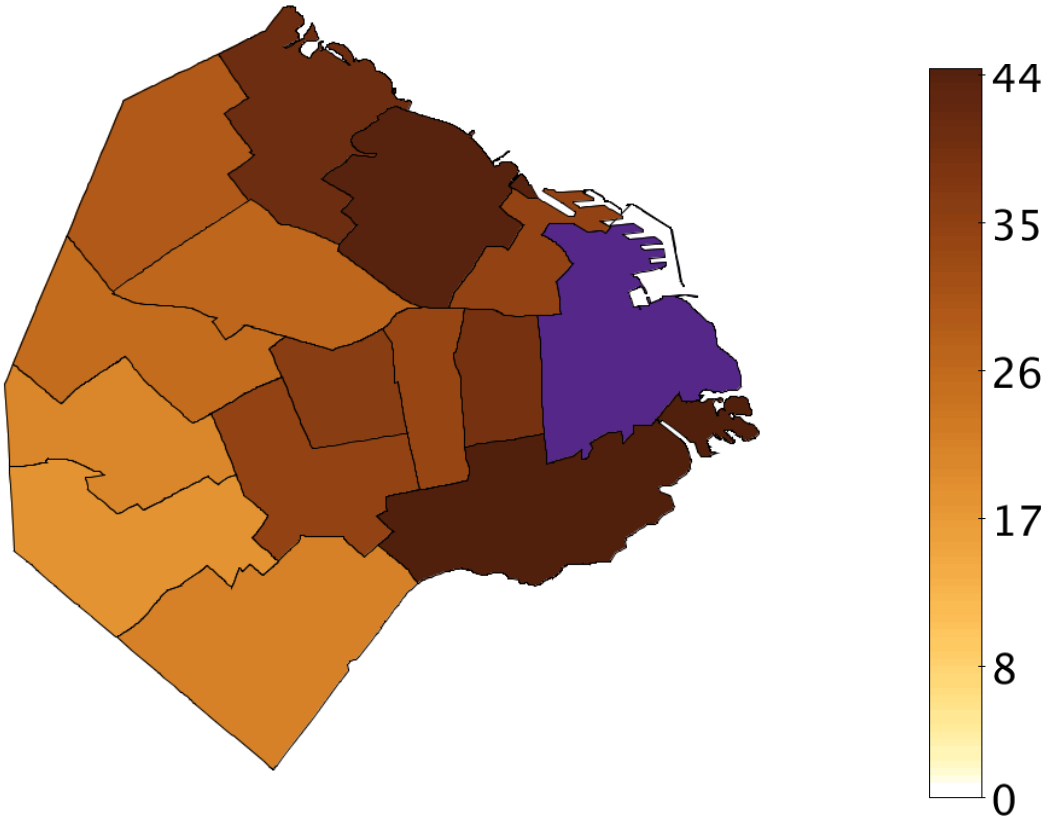
- Residential pattern on weekdays.
- Pole of attraction during Saturday
- Higher peak on Friday and Saturday nights.
- Sunday nights show less people than Fri-Sat nights (more similar to weekday nights).

Visualizing the City Pulse

We present here a comparison between Commune 1 and Commune 6, for Weekday Noon timeslot.

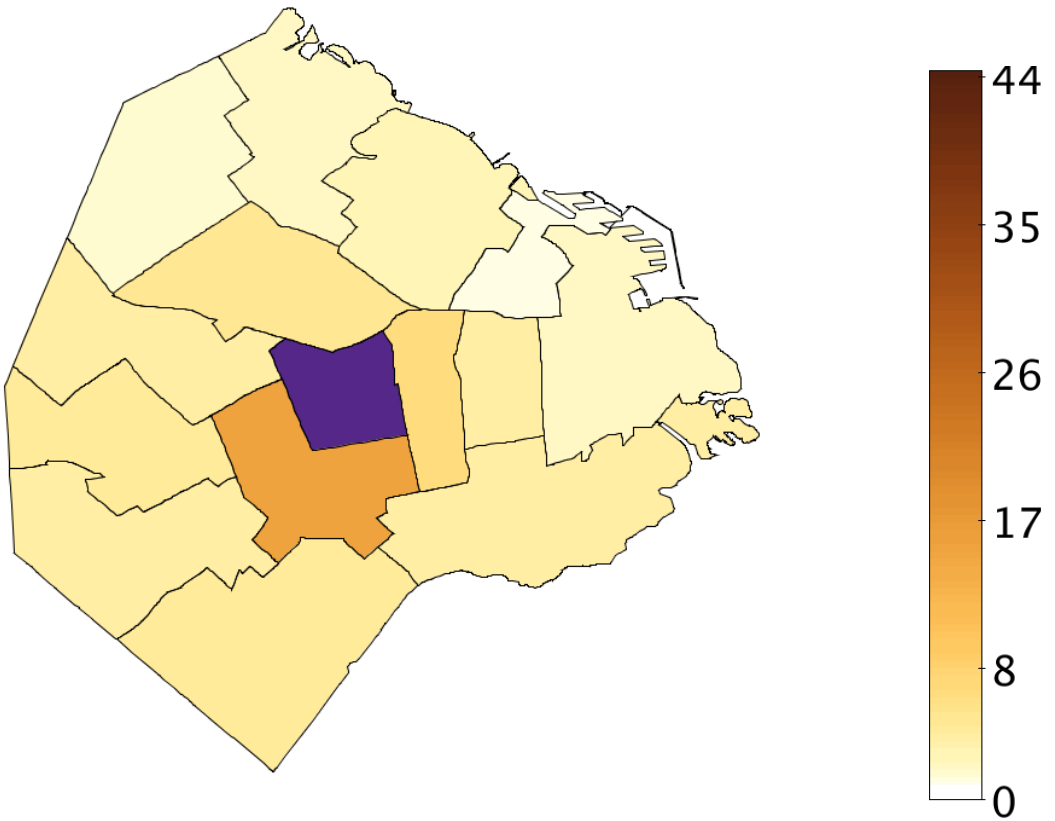
The plots show the home location of people located in the violet commune at the given timeslot. The scale used is in thousands of people and increases with color darkness.

City Pulse - Commune 1 (Microcentro)



- With moderate to high population.
- Central business district.
- Many people coming from all the city, even distant zones.

City Pulse - Commune 6 (Caballito)

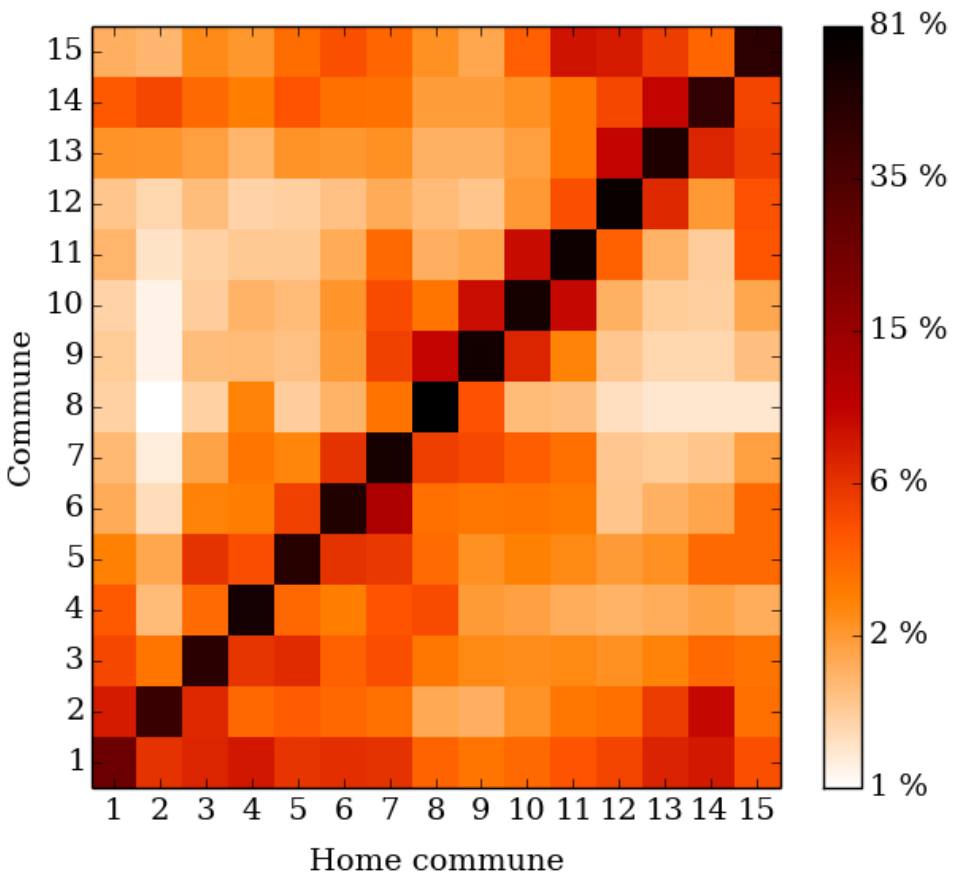


- One of the most populated and dense areas.
- Mainly residential district.
- Less people coming, and mostly from nearby neighborhoods.

The City Pulse Matrix

The urban mobility information can be used to generate the *City Pulse Matrix* (here for a typical weekday at noon):

$$CPM[i][j] = EP_{d,h}[i][j].$$

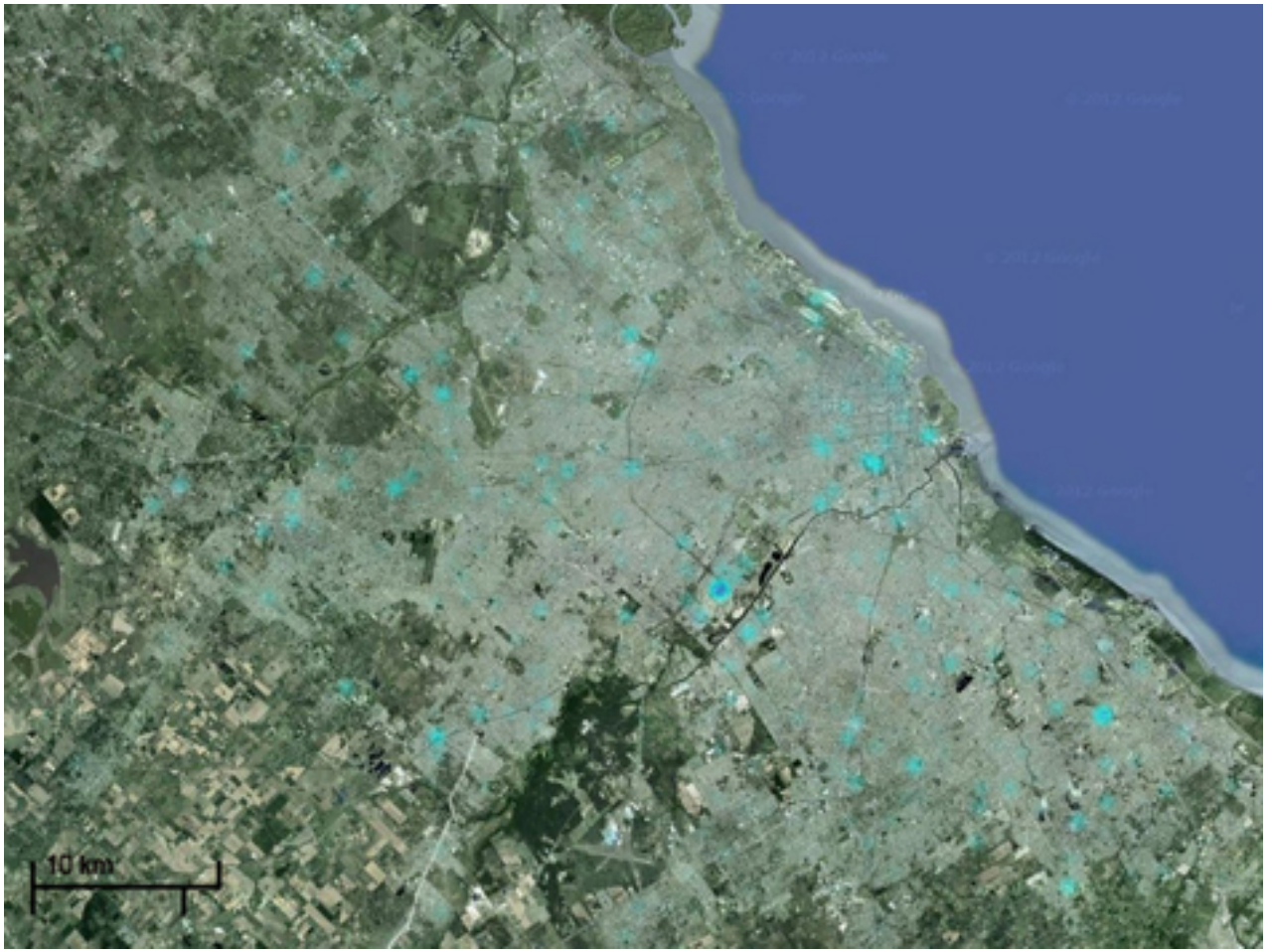


Agenda

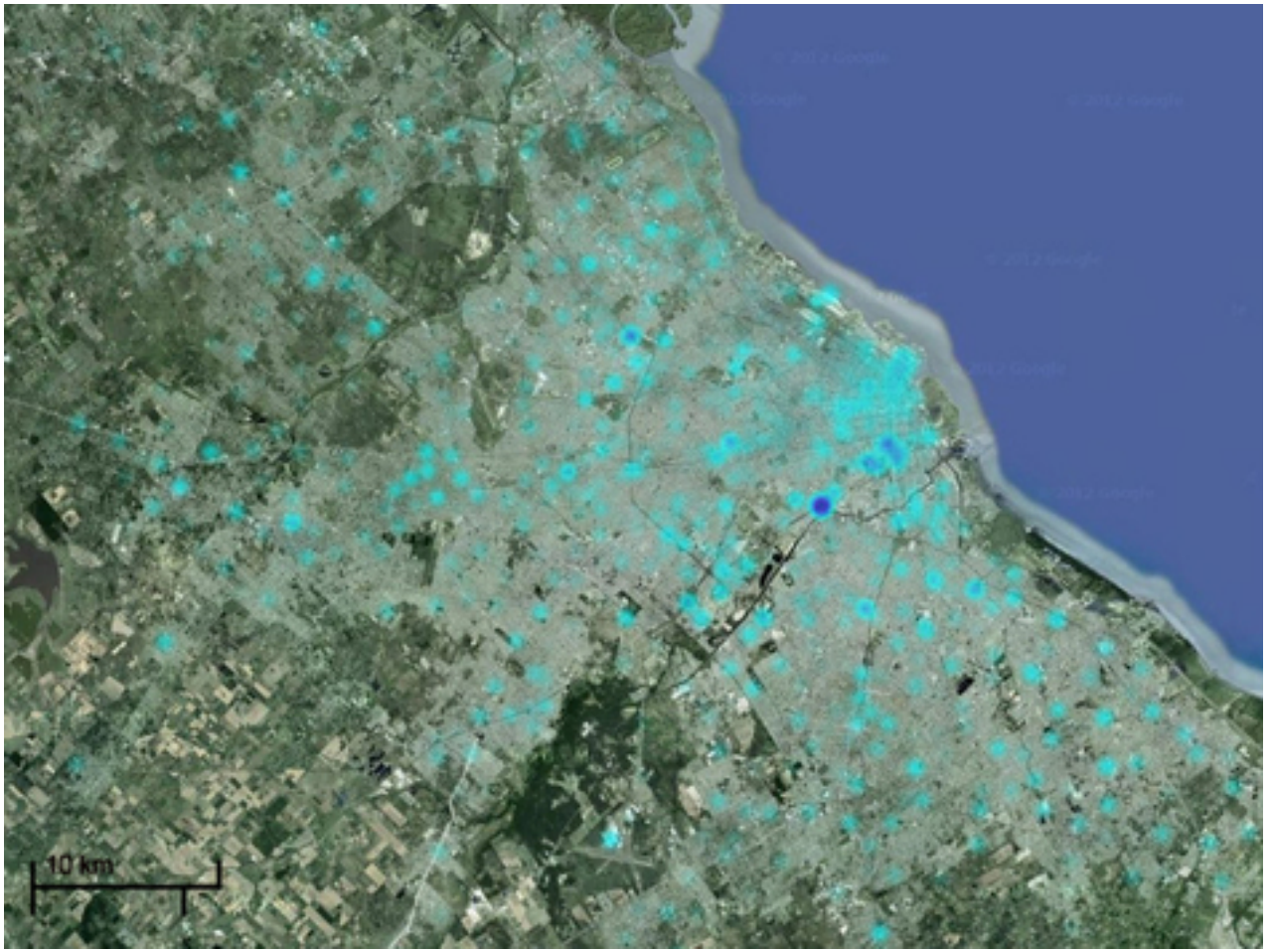
Commute from GBA to CABA

- The phenomenon of commuting is prevalent in large metropolitan areas (often provoking upsetting traffic jams and incidents), and naturally appears in mobile phone data.
- Who do we consider **commuters**?
  - Found in GBA (Greater Buenos Aires) during the night.
  - Found in CABA (Capital Federal) during the day.
- From the data, we can estimate the **radius of the commute** (the average distance traveled by commuters).
  - Considering the two most frequently used antennas as home and work, see Csáji et al. (2012)
  - We get an average commute radius of 7.8 km
  - As a comparison, the diameter of the city is about 14 km, and the diameter of the considered metropolitan area is 90 km.

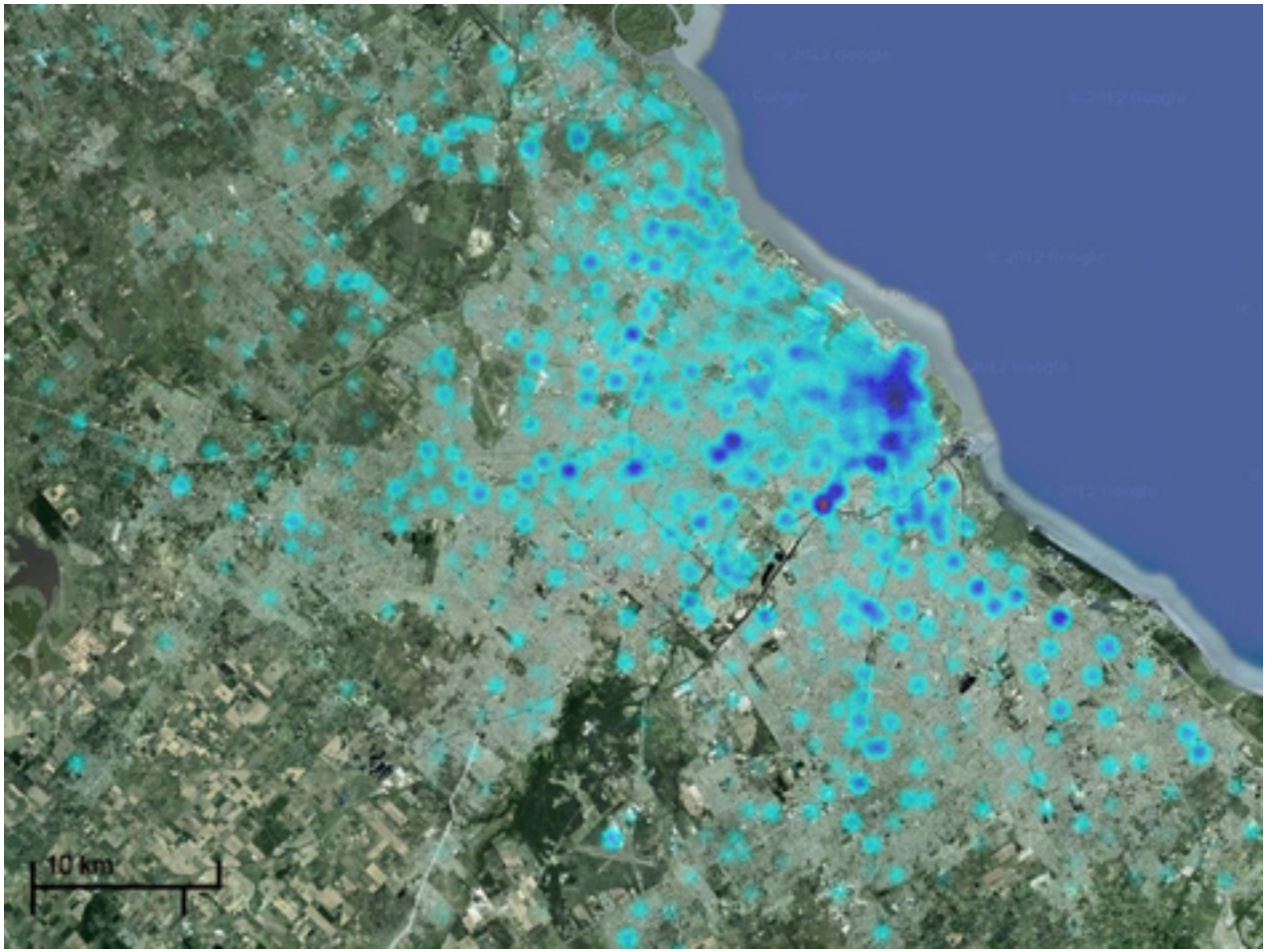
Commute to CABA - 05:00



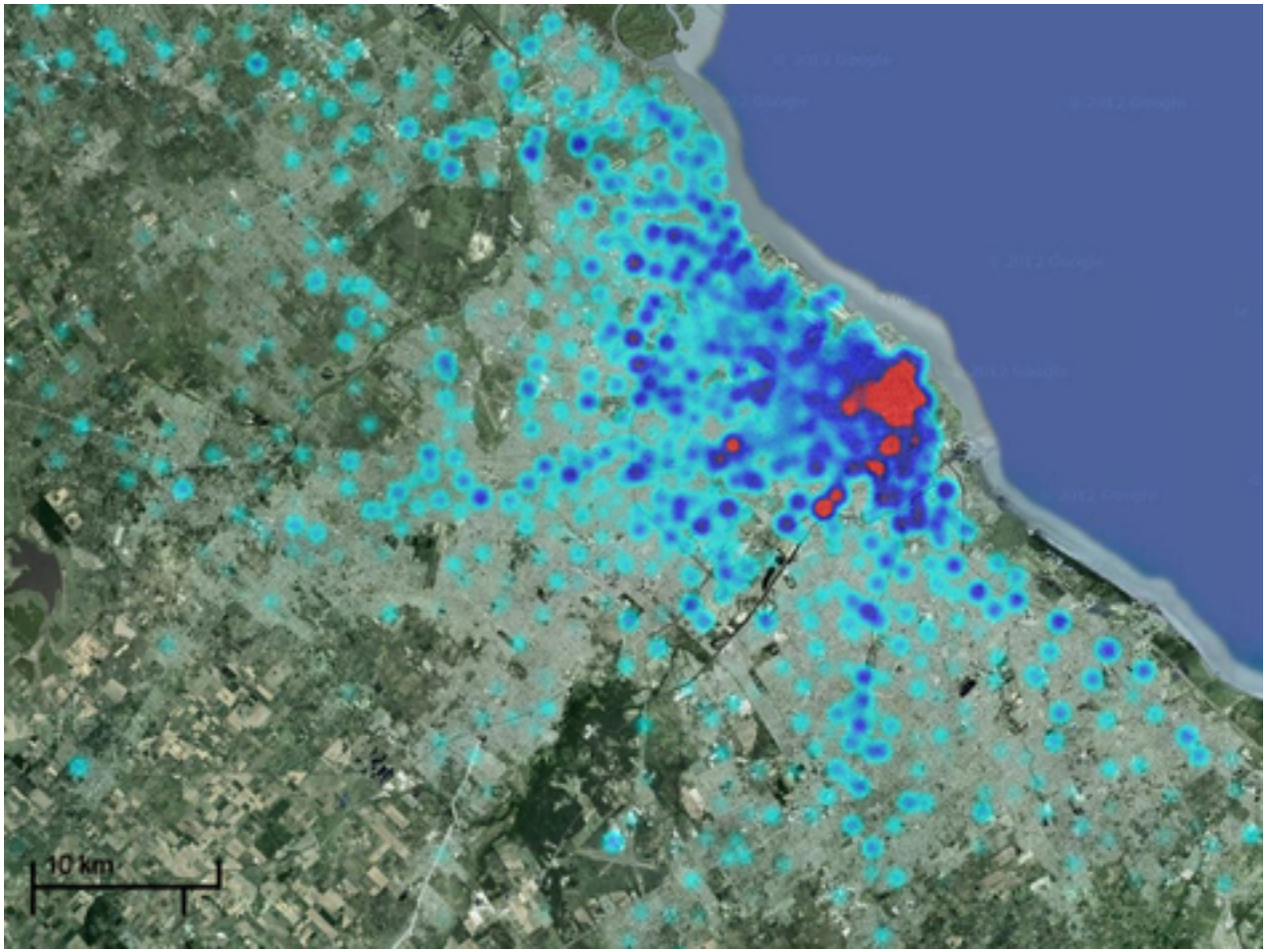
Commute to CABA - 06:00



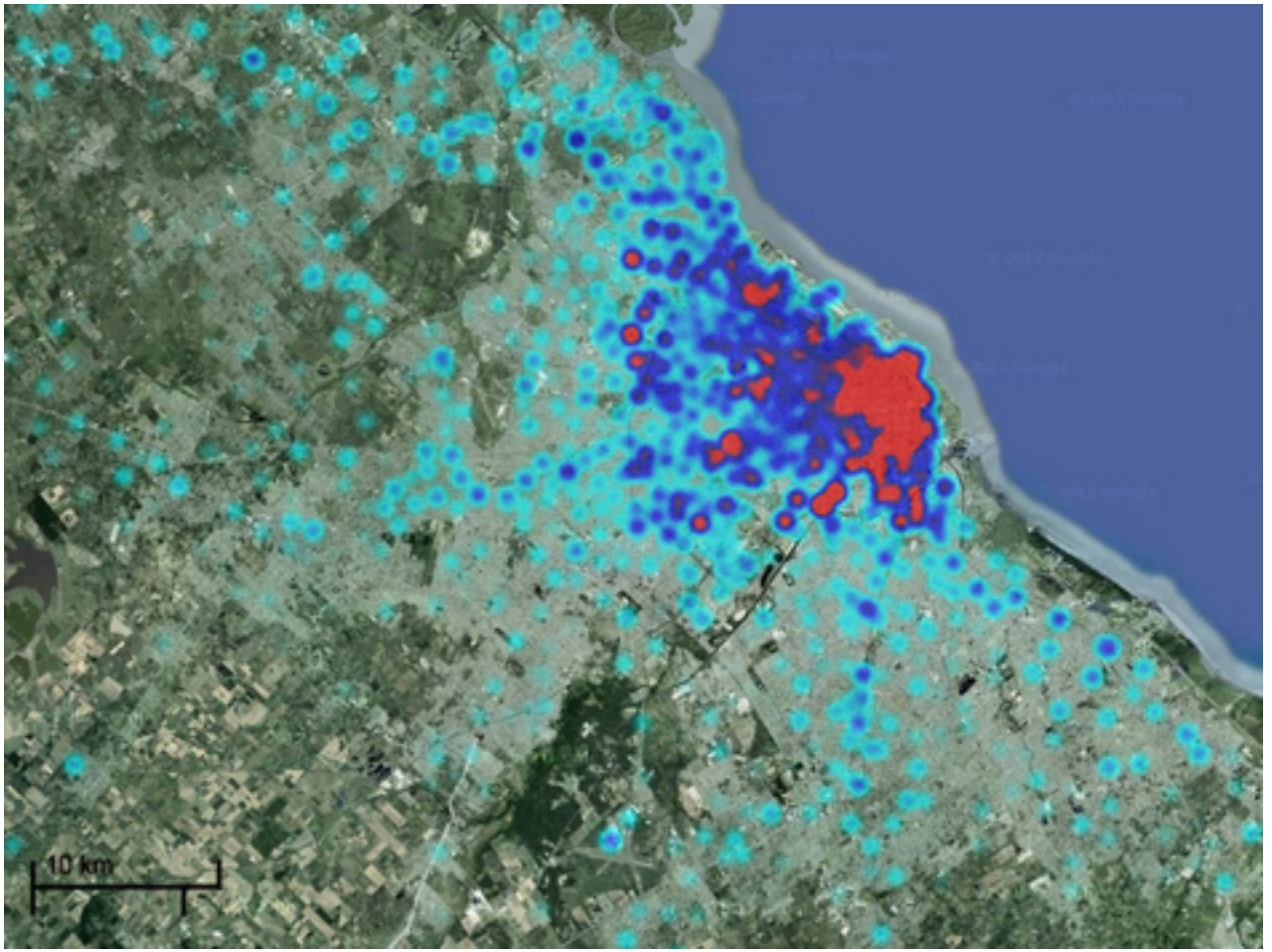
Commute to CABA - 07:00



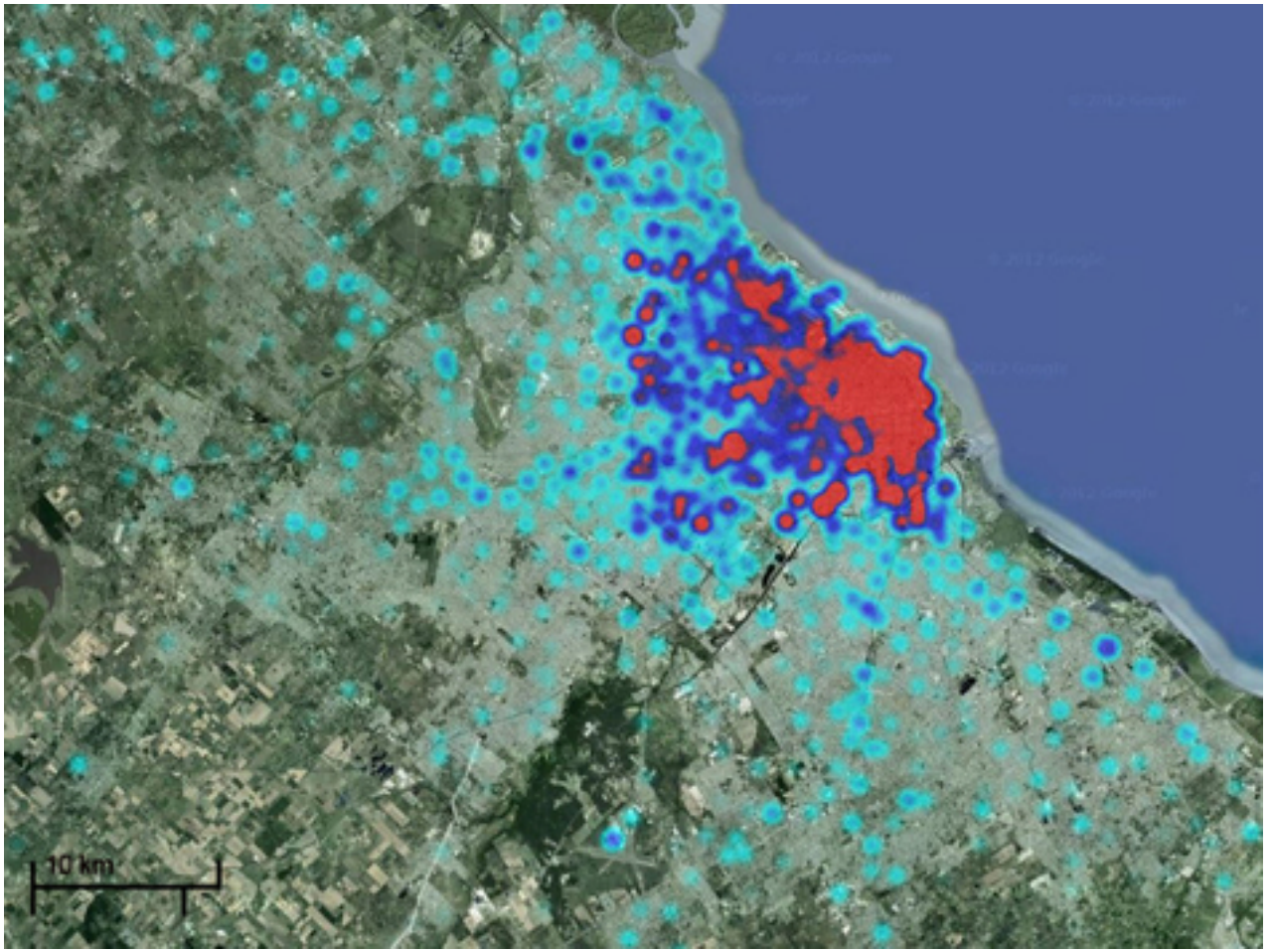
Commute to CABA - 08:00



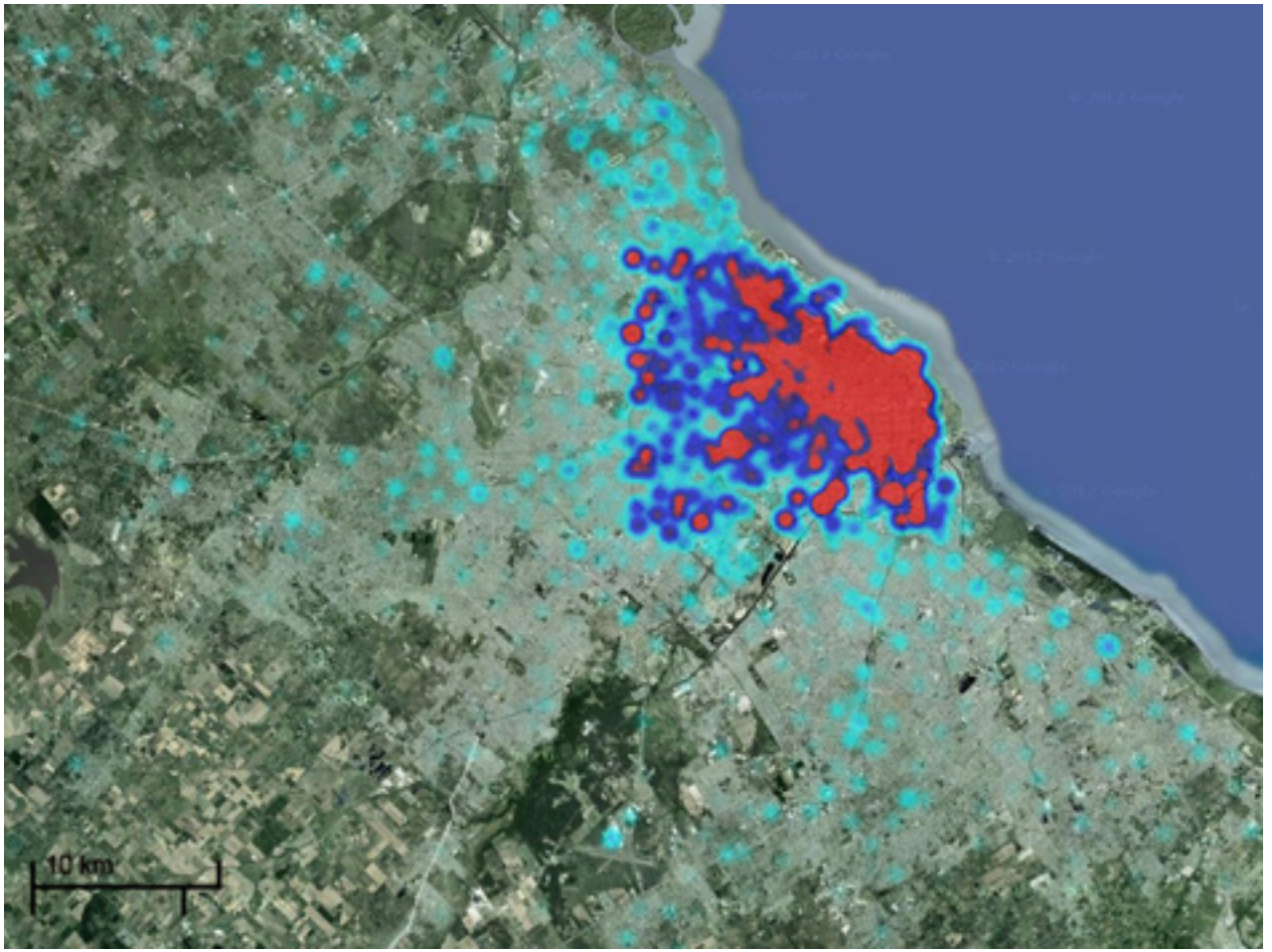
Commute to CABA - 09:00



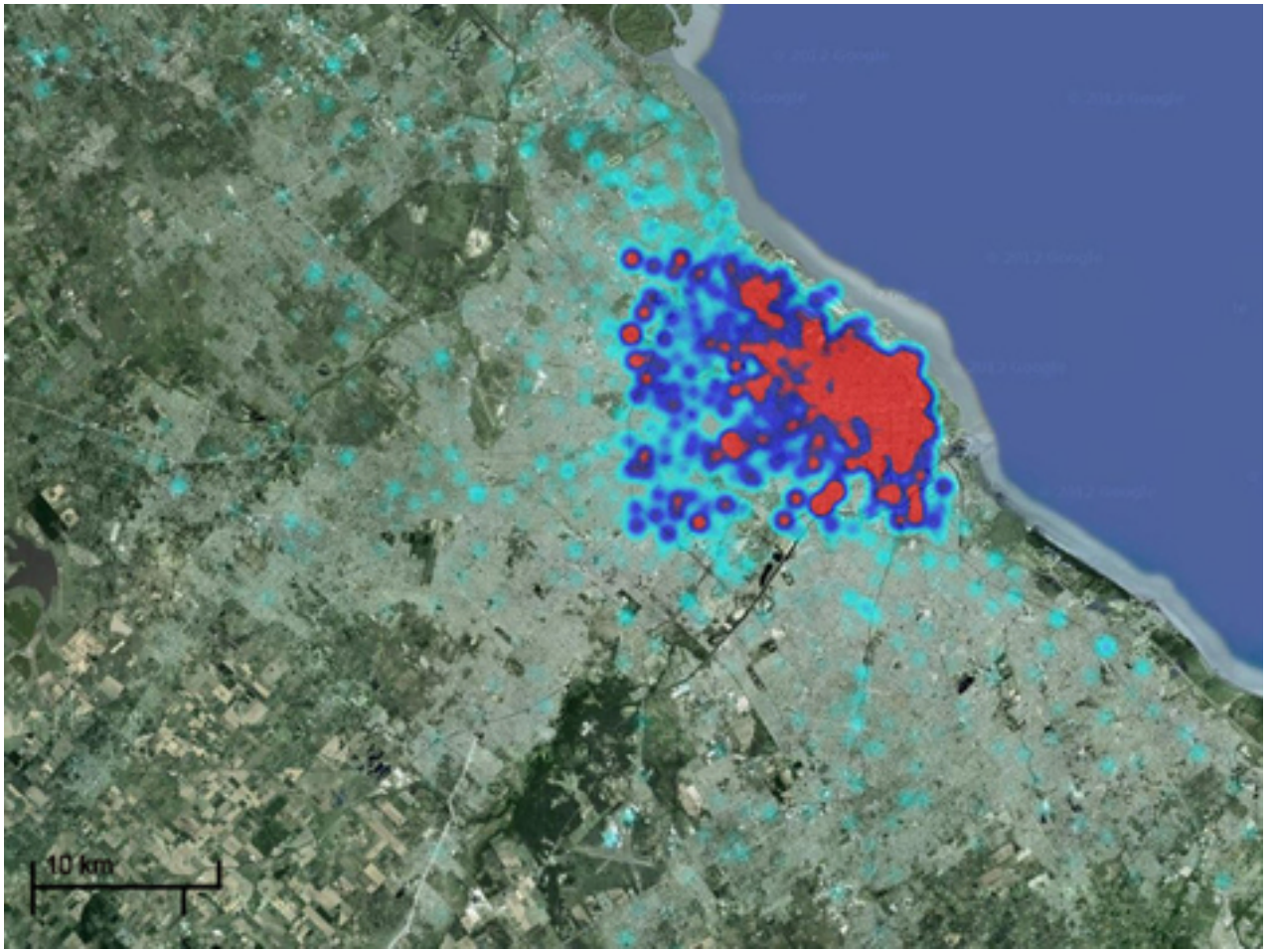
Commute to CABA - 10:00



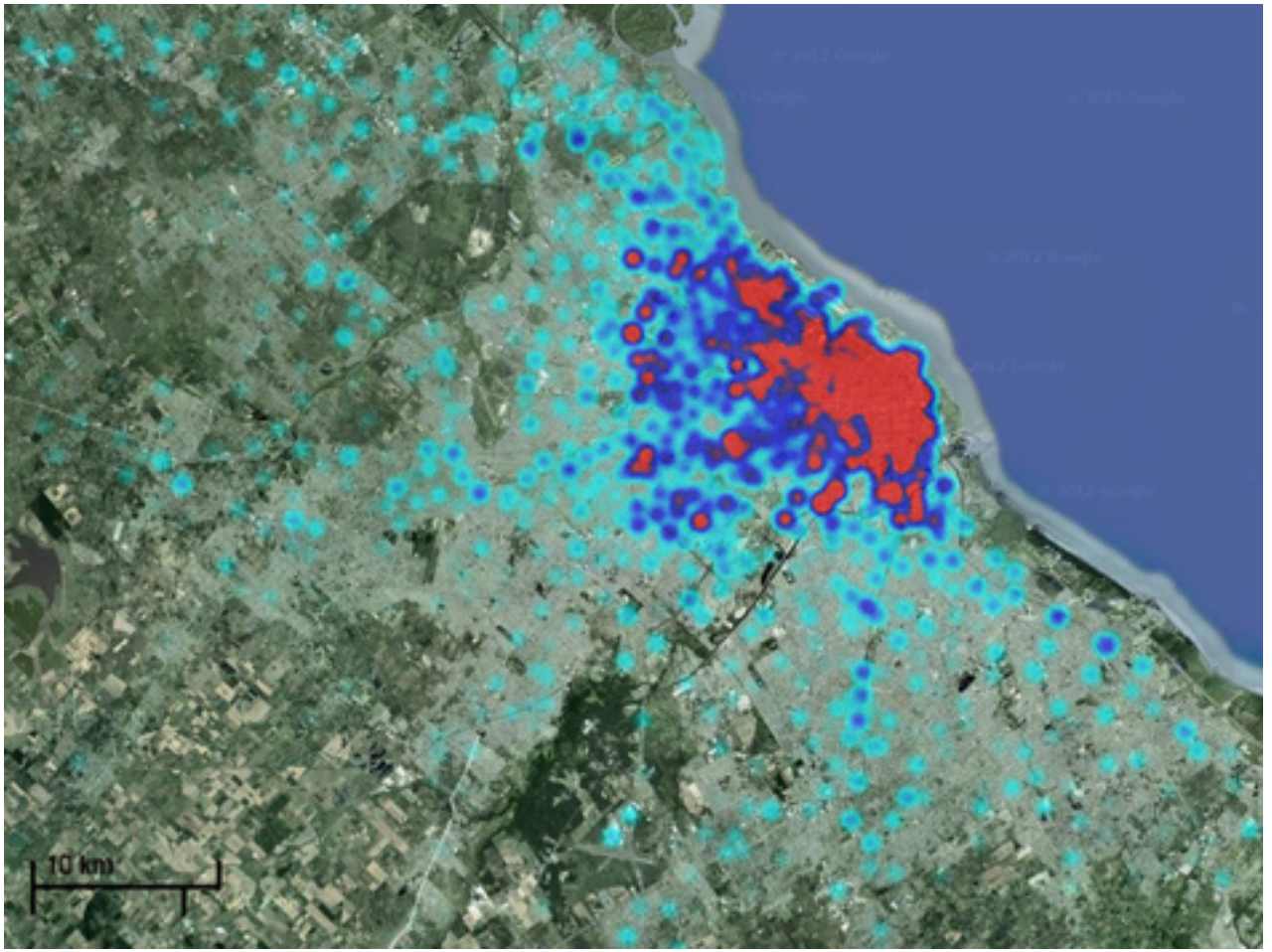
Commute to CABA - 12:00



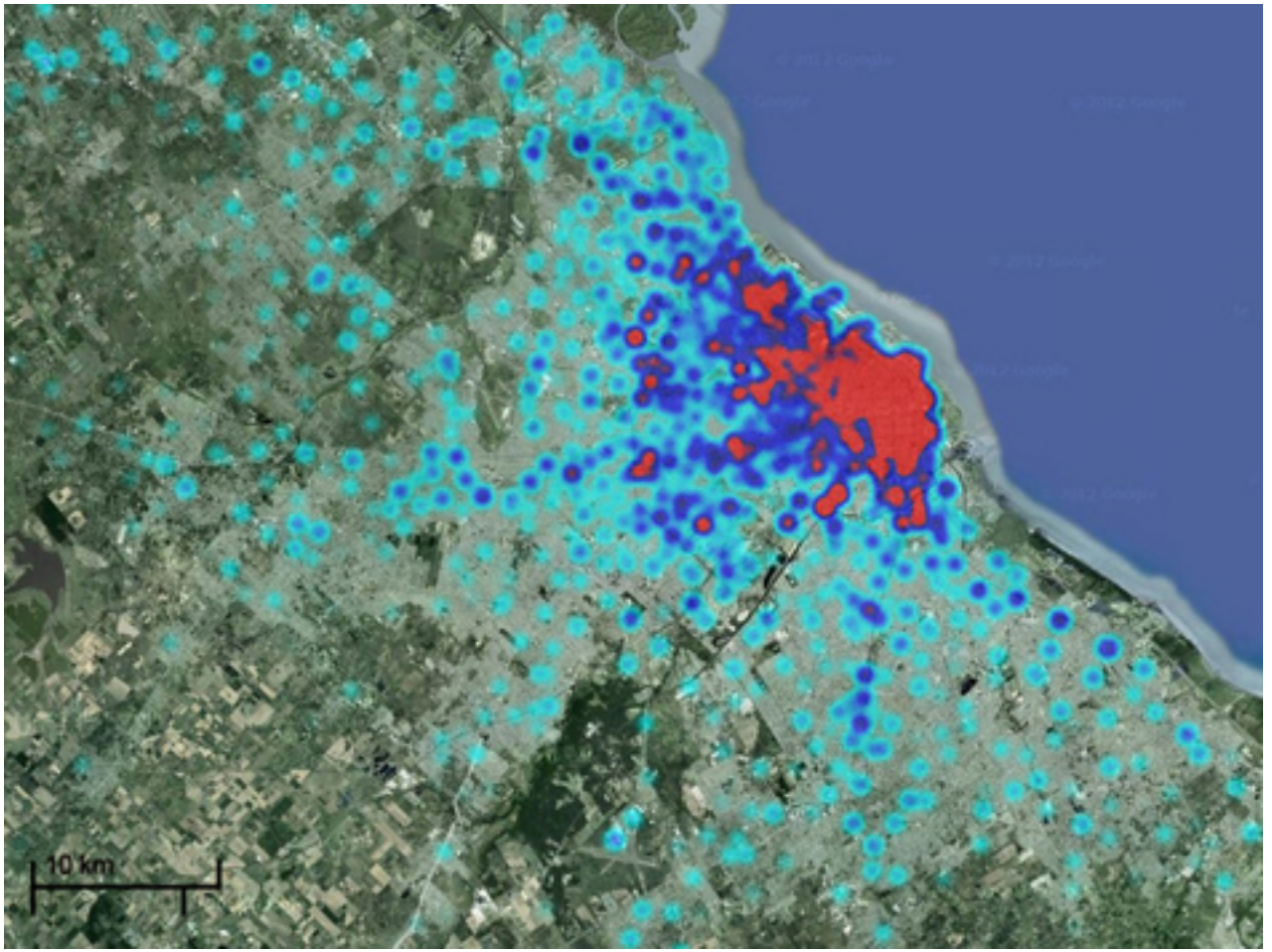
Commute to CABA - 14:00



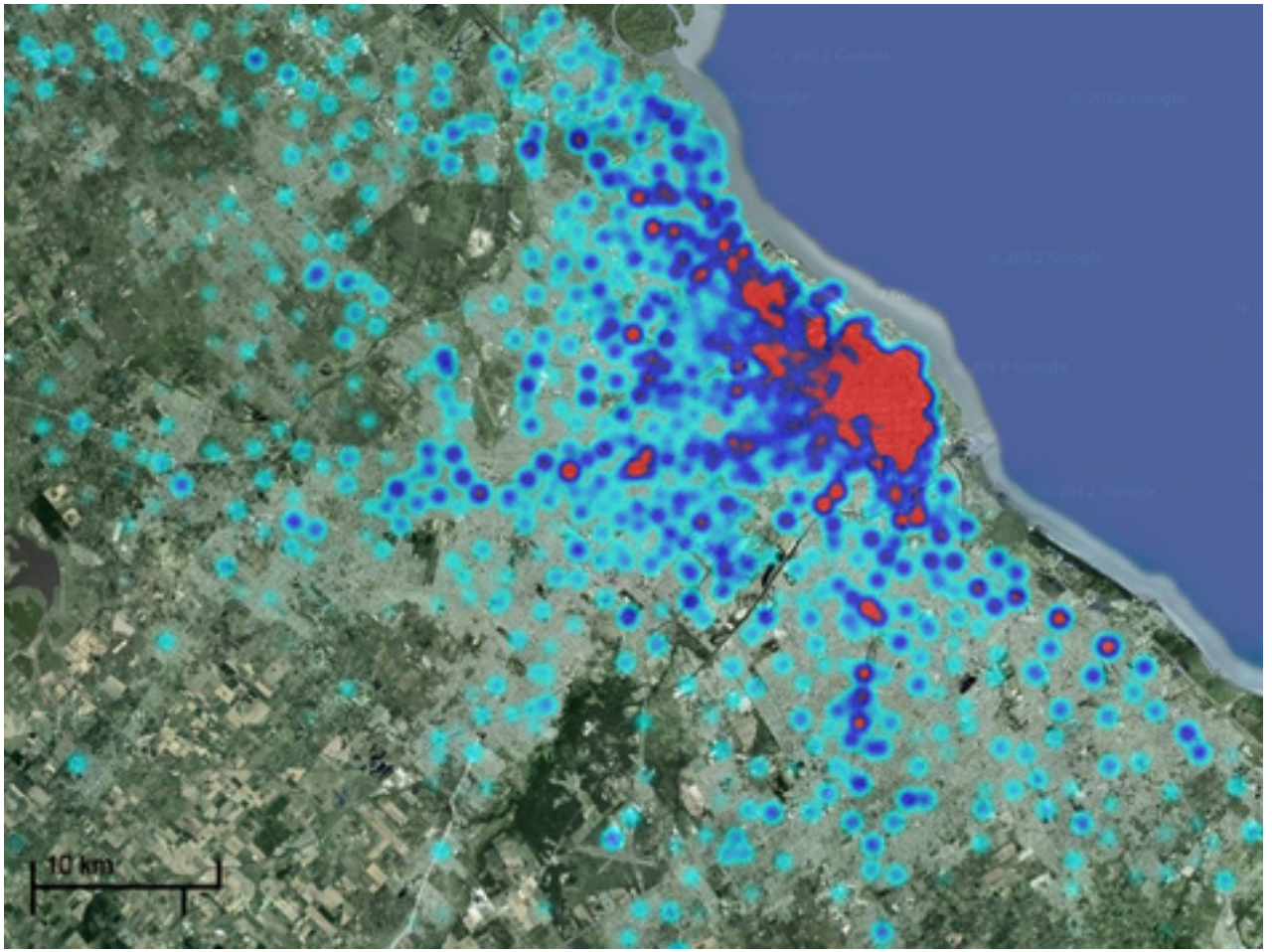
Commute to CABA - 16:00



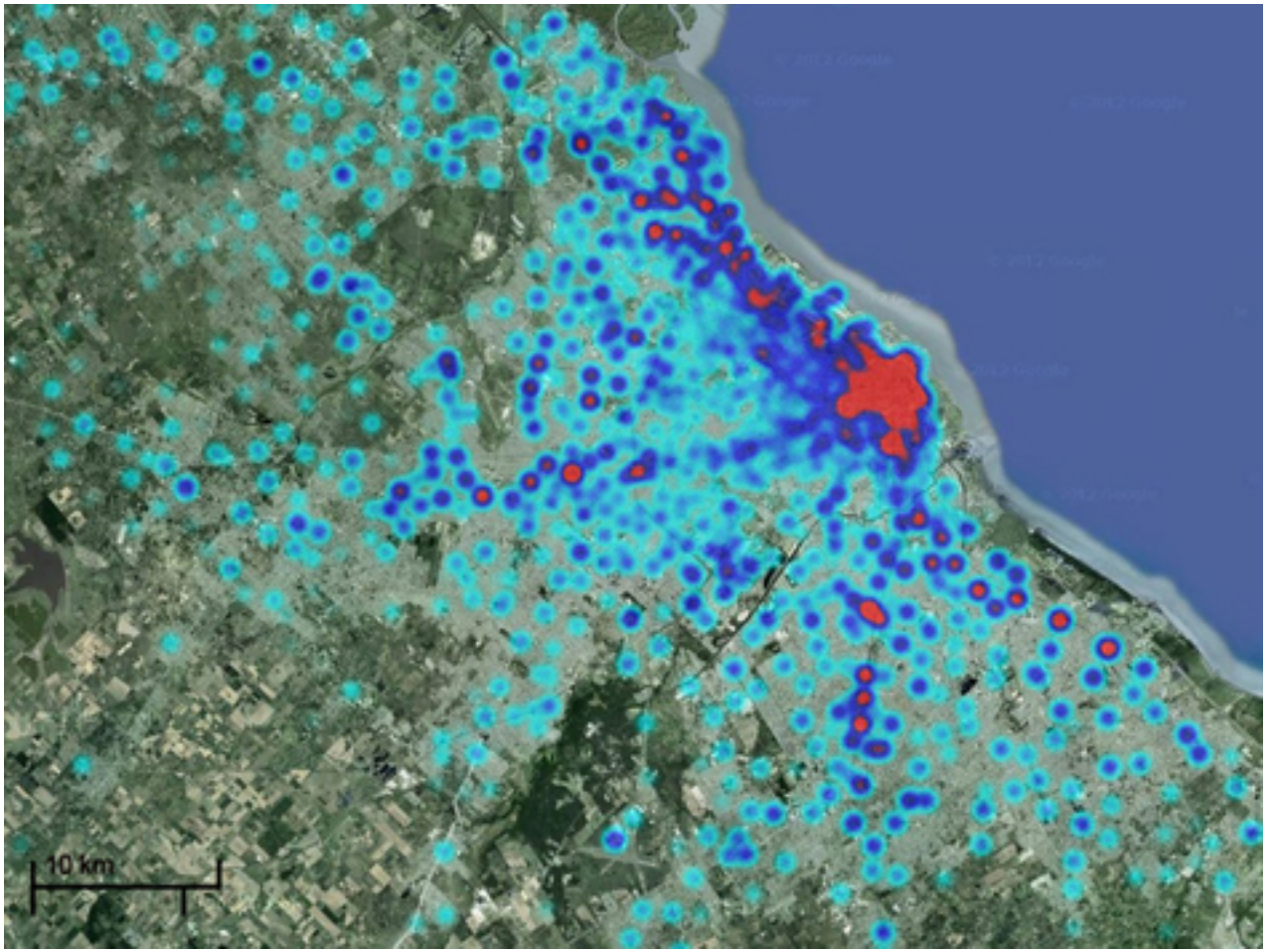
Commute to CABA - 17:00



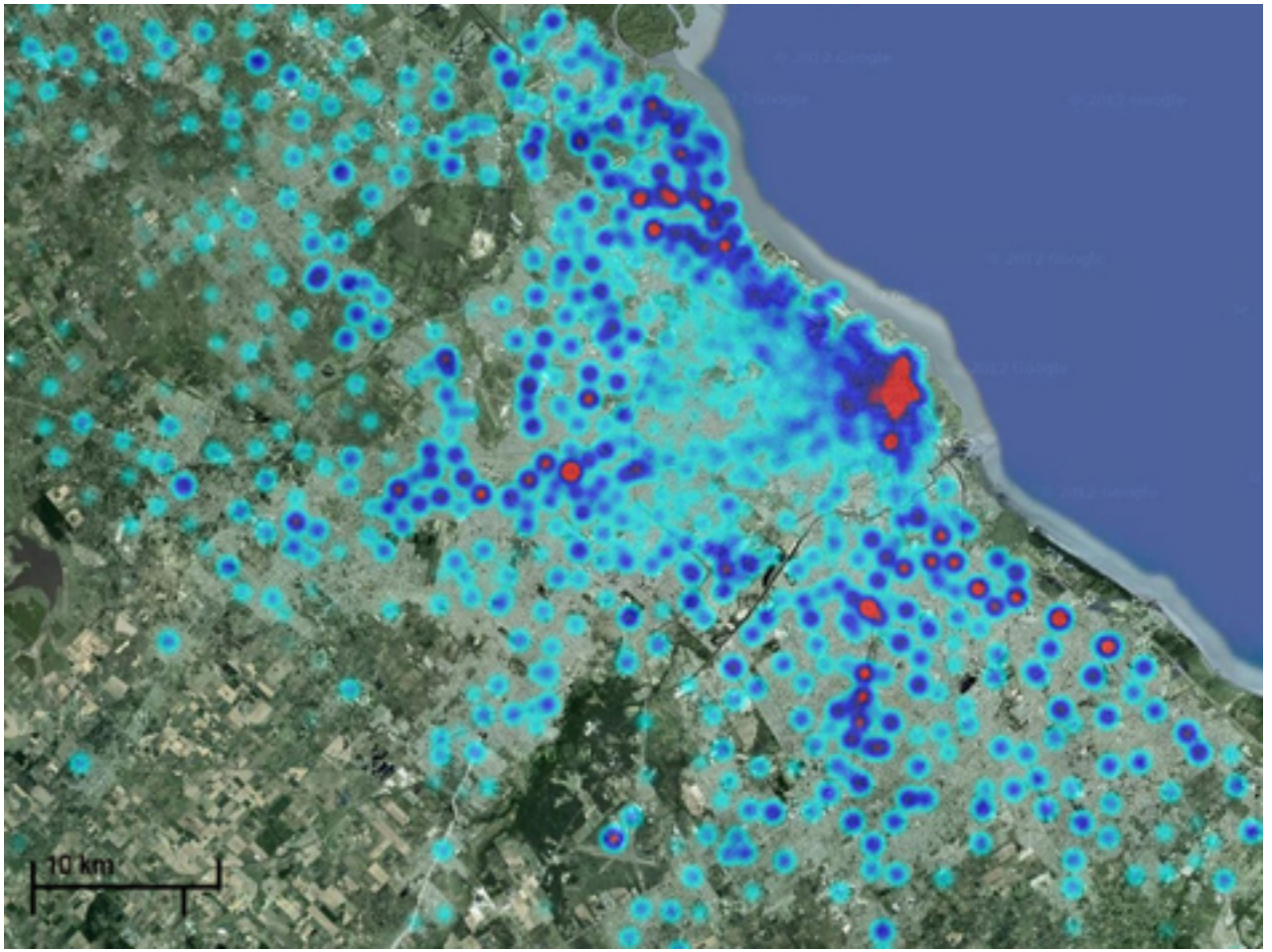
Commute to CABA - 18:00



Commute to CABA - 19:00

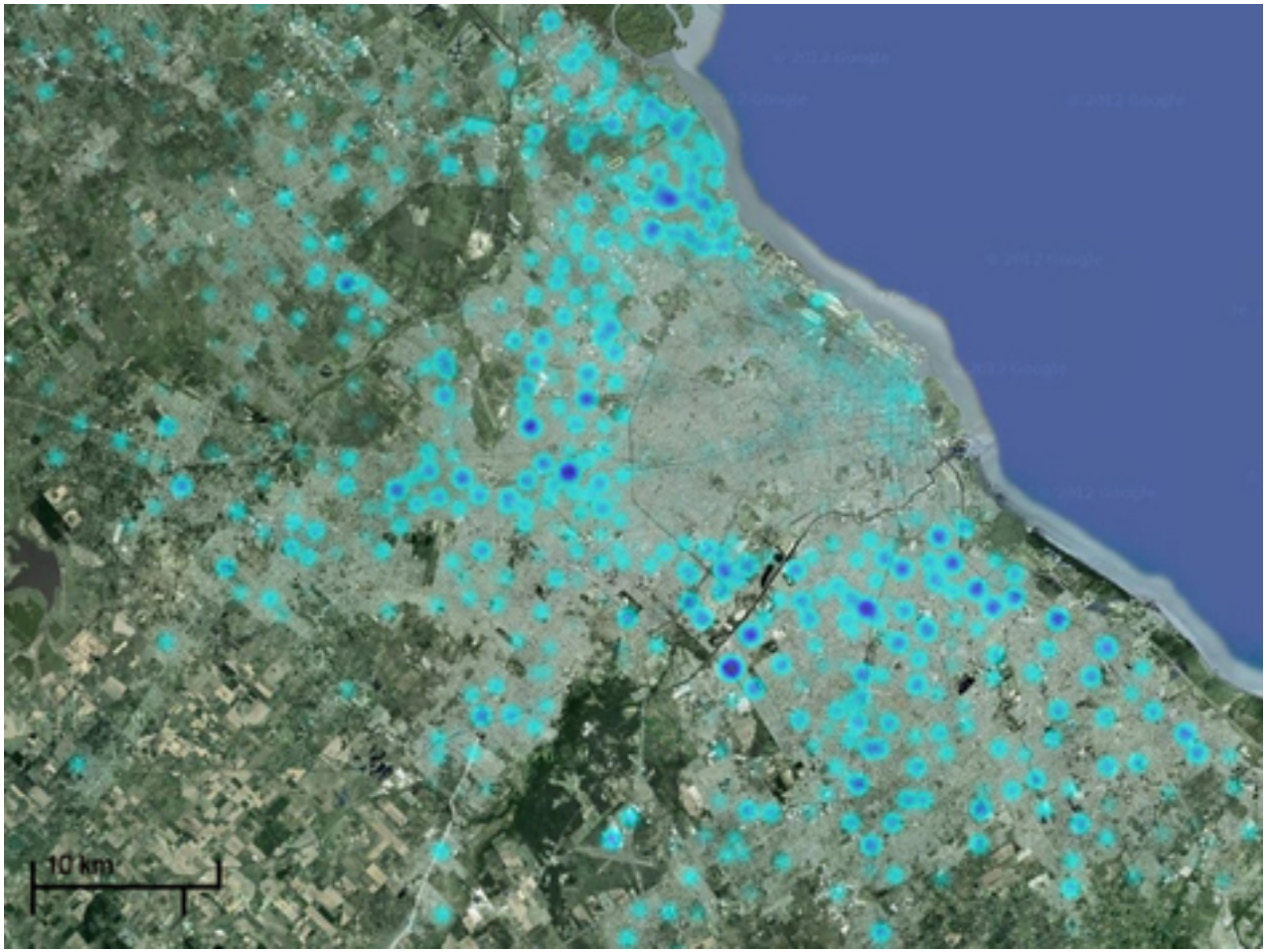


Commute to CABA - 20:00



Commute to CABA - 21:00

Commute to CABA - 23:00



Agenda

Conclusions

- We presented a methodology to **estimate the flow of people** between different parts of the city using mobile phone records.
- According to our validation, our prediction is reliable, presenting an average difference of 5% with the **origin-destination survey**.
  - Our method is less expensive and results can be updated much more frequently than traditional origin-destination surveys.
- By extending our prediction to weekends, we obtained **new results** that were consistent with our knowledge of the city.
- We also looked at the phenomenon of **urban commute** between the city and its surroundings.

Future Work (1)

- Achieve a finer spatial granularity with a richer dataset (e.g. GPS based).
- Analyze the mobility of citizens during a particular situation or event (such as a holiday or an flooding).
- **Model vehicle traffic** in streets, avenues and highways
  - Use the methodology of Wang et al. (2012)

Future Work (2)

- Urban air pollution represents one of the largest environmental health risks globally
- Air pollution is a leading environmental cause of premature mortality and growing concern in cities and in developing countries
- Vehicles constitute 47% of cities' total carbon dioxide emissions.
- Model of urban traffic can be used to:
  - Reduce traffic congestions (and thus carbon dioxide emissions).
  - Overlap traffic conditions with data from Air Quality sensing stations to improve Air Quality analysis

References

Secretaría de Transporte. Ministerio del Interior y Transporte. ENMODO (2009-2010). Resultados de la encuesta origen destino. Movilidad en el area metropolitana de Buenos Aires, 2010.

Instituto Nacional de Estadística y Censos (INDEC). *Censo Nacional de Población, Hogares y Viviendas 2010*, volume 1. INDEC, October 2010.

Balázs Cs Csáji, Arnaud Browet, VA Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D Blondel. Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, 2012.

Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2, 2012.

Thank you!

GRANDATA

Carlos Sarraute

charles@grandata.com