“Big Data at the service of social inclusion”

Final Project Report
May 2020

Principal investigators:
Víctor Martínez de Albéniz
Cynthia Echave Martínez

Collaborators:
Spyridoula Chrysikopoulou
Joan Estévez
Jordi Balsach
Mihalis Markakis

CONTENT

PRESENTATION

The world faces significant social inclusion challenges that have a direct impact on our economic, political, and social systems. Europe is no exception and although the welfare state can be better compared to other parts of the world, poverty is present in all countries and, unfortunately, rising. In fact, on the list of 17 sustainable development goals of the United Nations, poverty is the first objective.

Lack of social inclusion is manifested in situations of vulnerability, which vary according to the characteristics of the people, the households to which they belong, and the territory. Urban vulnerability is becoming increasingly complex and requires dynamic tools to identify it, to convey appropriate preventive and mitigation measures. Big Data can be one of these prediction tools. If statistical data can be jointly and dynamically analyzed, one can improve the effectiveness of the social policies of cities and their neighborhoods.

This has been the main motivation of the research project "Big Data at the service of social inclusion" funded by the Recercaixa program from 2018 to 2020. The project has been led by IESE Business School. It has been fortunate to count with the collaboration of the Department of Innovation, within the group of social rights of the City Council of Barcelona, for providing the key data from the study and for supporting the organization of participatory workshops. We would like to thank Luís Torrens for his collaboration throughout the development process of this project.

This report summarizes the results obtained from the project and the final reflections to which we have come. Now more than ever before the current COVID-19 crisis, it is essential to better understand the factors determining vulnerability, to establish resilience mechanisms that allow us to prevent exclusion risks.

1. BACKGROUND

From a global perspective, all territories are vulnerable to the effects of climate change, or to unexpected pandemics such as Covid-19. Its inevitable social and economic impact affect people's lives. Either because of the depletion of available resources, or by the lack of equitable access to basic services.

From a local perspective, cities are especially vulnerable due to poverty and social exclusion. Although it is still linked to the conditions of global economy, urban vulnerability is characterized by aspects that stop the development of a full life by people. Vulnerability prevention represents one of the most important challenges, and proper management of the phenomenon increases the environmental and social resilience of cities, strengthening their ability to react to negative shocks such as economic recessions.

Since the economic crisis of 2008, the identification of the vulnerability of people and the corresponding action measures have played a relevant role in the definition of current urban policies. Its incorporation as one of the main criteria in decision-making allows the establishment of mechanisms to prevent vulnerability and its immediate effect, social exclusion. If we understand that social exclusion is the stage of segregation and disengagement of an individual from the surrounding community, thinking about vulnerability means anticipating extreme situations.

The frame of reference on urban vulnerability studies are very varied and extensive. The methodologies for assessing it agree that vulnerability occurs in different factors: physical factors, work factors, and psychological and social factors.

This first section aims to analyze the background and motivations of this project.

1.1 Vulnerability, poverty and social exclusion

The latest data published by Eurostat in 2017 show that more than 113 million people in Europe are at risk of poverty and social exclusion, representing 22.4% of the total EU population. The EU has set ambitious targets for 2020 in terms of job creation, innovation, education and social inclusion, climate, and energy. In terms of poverty, the EU's overall goal is to reduce at least 20 million people suffering from risk of poverty and exclusion, by 2020.

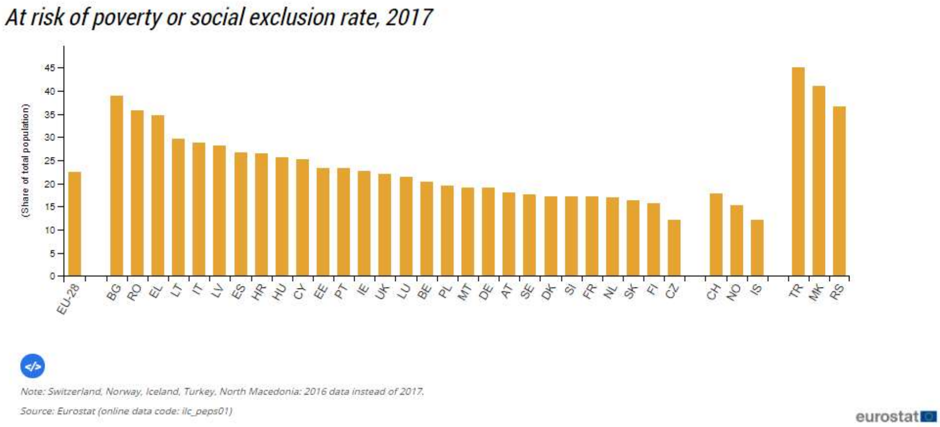


Figure 1: Index AROPE for 2017, by country. Source: Eurostat

The vulnerability with respect to poverty is measured by a multidimensional indicator that goes beyond economic incomes, called **AROPE (*At Risk of Poverty and Exclusion*)**. It is an indicator that considers the risk of social exclusion when people meet one of the following conditions:

- People living in households with a very low work intensity (9.5% of the EU population)
- People at risk of poverty after social transfers (16.9% of the EU population)
- People living with tangible privations (6.6% of the EU population)

In terms of age groups and gender, in 2018, 25.1% of women in Europe presented one of the AROPE conditions compared to 27.0% of men in the EU. The most vulnerable group by age is between 16 and 29 years old, mainly due to the vulnerability of young people in accessing the labor market.

In Spain, the recent publication of the "State of poverty and monitoring of the indicator Poverty and social exclusion in Spain 2008 - 2018" by EAPN (European Anti-Poverty Network) states that there are a total of **12.2 million people, representing 26.1%.** Of the total Spanish population at risk of poverty and / or social exclusion. The main target set by the EU 2020 Strategy is to reduce it by **1.4 million people.**

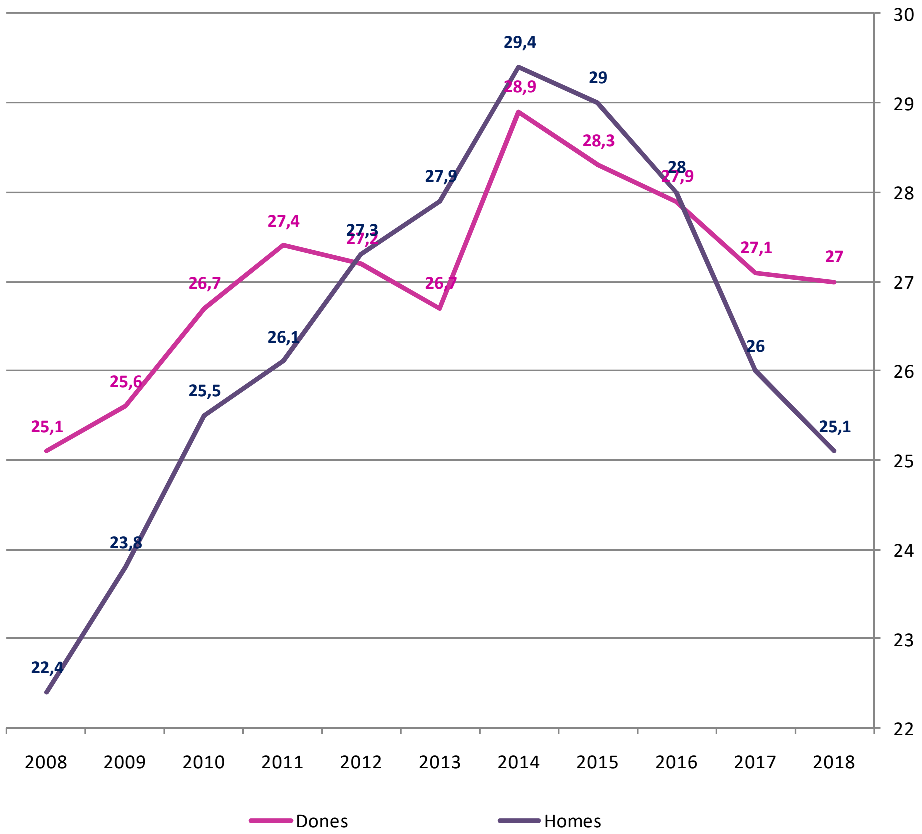


Figure 2: People at risk of poverty and social exclusion in Spain from 2008 to 2018 (% of population). Source: INE Spain

The AROPE rate is not homogeneous for all people and has significant differences according to gender, age, nationality, educational level and employment status, household type, disability, and territory. Since 2014 the AROPE has fallen sharply in men, a factor that has caused the rate of women to be two percentage points higher this year than that of men (27.0% and 25.1%, respectively). By age, it is worth highlighting the high AROPE rate among young people aged 16 to 25 (33.8%) and the increase for the fourth consecutive year among those over 65 (17.6%).

Women are at an even higher risk of poverty than men, and the following graphs show the evolution of gender differences over the last ten years.

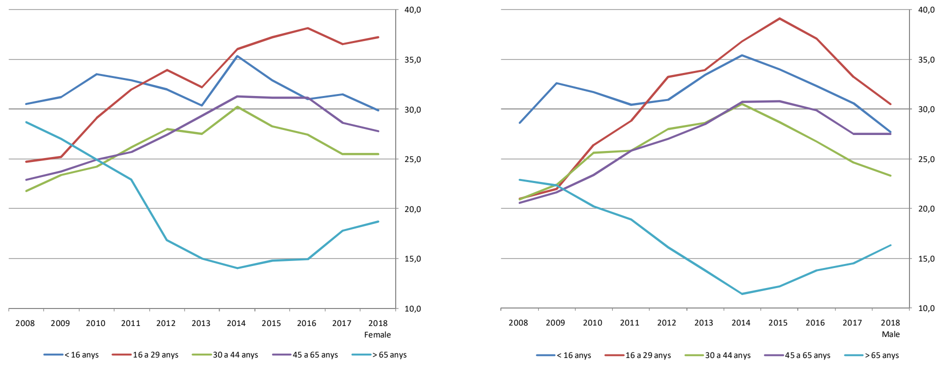


Figure 3: Risk of poverty by gender and age group in Spain between 2008 and 2018. Source: INE 2018

In Catalonia, **the AROPE rate is 21.9%,** while the city of **Barcelona has a rate of 15.3%, compared to 19.7% in the Metropolitan Area**, according to statistics from Barcelona City Council. Regarding material deprivation, the rates are 33.4% in Barcelona, 42.6% in the metropolitan area and 43.4% in Catalonia. It is surprising to see that almost **48.8% of the inhabitants of the city of Barcelona had difficulty covering monthly expenses**, while around 23.1% did not have the capacity to face unexpected expenses.

The Barcelona City Council has directed the activity of the city's social agents towards a city model that has made substantial progress in safeguarding social rights. In this line, the *"Strategy for inclusion and reduction of social exclusion in Barcelona 2017 - 2027"* was drafted, which establishes 41 targets, classified into five strategic lines.

Some of the main conclusions of the document point to the need to fight the stigmas and disqualify poverty. The document explains the three stages of the disqualification process that people experience: **1st phase of fragility; 2nd phase of dependency; and 3rd stage of rupture**:

During the first phase, people begin to be identified as people with social problems and have the impression that their environment points them as "poor". In the second phase, that of dependency, people gradually accept various forms of help and become aware of their own social vulnerability. Households that receive help from social services or welfare organizations considered dependent. Finally, the breaking phase is characterized by the disappearance of ties with most welfare *and social service organizations and also by the adoption of private lifestyles geared towards mere survival. People seek financial help or food through informal circuits and social organizations and develop mechanisms to reject institutional interventions in their own lives and decisions."*

Source: Barcelona Inclusion Strategy 2017 – 2027.

1.2 SOCIAL INNOVATION AND COLLECTIVE ACTION

A common way of managing urban vulnerability and the fight against poverty and social exclusion in cities is through the synergy of two components: 1) the social service structures offered by local institutions and governments and 2) the network of third sector entities. The initiatives resulting from this synergy make it possible to carry out measures tailored to the needs of citizens through a more flexible way of providing social services and gradual implementation in the territory (Garrone et al. 2017[1]). It is true that in many cases, the initiatives are based on an important volunteer base. Social innovation through collective action can drive progress in managing urban vulnerability to poverty and social exclusion. Initiatives formulated by society or entities can be promoted, through responsible consumption networks, associations that look after social rights, or business models that facilitate the transition to a more sustainable and egalitarian model.

In 2017, the Institut d'Estudis Regionals i Metropolitans de Barcelona published the study "Social and political innovation, institutional density and urban vulnerability in Metropolitan Barcelona" (Antón-Alonso et al. 2017[2]). The study is characterized by the elaboration of the index of urban vulnerability (IUV) applied to all the districts of the municipalities within the Metropolitan Region of Barcelona from the data of 1991, 2001 and 2011. The IUV considers 4 dimensions of vulnerability and assesses them through 8 indicators that integrate various phenomena of a socio-economic, labor, socio-demographic and residential nature.

The results of the study show that the rate of urban vulnerability is declining in some neighborhoods while growing in others. For example, in Barcelona in 1991,

[1] Paola Garrone; Angelamaria Groppi; Paolo Nardi (2017)*"Social innovation for urban liveability. Empirical evidence from the Italian third sector"*. Journal Industry and Innovation. Volume 25, 2018 - Issue 6: Social Innovation and Services.

[2] https://iermb.uab.cat/ca/estudi/densitat-institucional-barris-i-el-seu-rol-en-els-processos-de-vulnerabilitat-urbana-innovacio-i-metropoli-2a-fase/

20.5% of the city's neighborhoods had high urban vulnerability; later in 2001, they increased to 43.6% (17 neighborhoods); and finally, in 2011 the rate dropped to 30.8%, equivalent to 12 neighborhoods.

Given that vulnerability is a complex phenomenon and responds to several factors, the results of the study projected in the territory, show that in general terms, Barcelona in 20 years increased its degree of vulnerability, as well as other municipalities of the Metropolitan Region. In the specific case of Barcelona, on the one hand, there are neighborhoods that fail to reduce the value of vulnerability during that period of time, which indicates that there have not been enough measures to improve their poverty conditions. On the other hand, we see that during this period, the vulnerability has spread to neighborhoods traditionally considered middle class, such as the neighborhoods of the Eixample and Gràcia Districts.

Given the complexity and diversity of circumstances linked to vulnerability in various contexts, special attention is paid to the role of the scale of action in the effectiveness of established measures: the lessons learned clearly point out that it is unsuccessful to unify and homogenize solutions.

Therefore, it is important to define the scale of action, as it affects the proper sizing of the services offered to citizens and their corresponding public investment. Specifically, more local interventions require the reorganization of statistical data, in order to provide adequate monitoring of the actions implemented and the objectives achieved.

As a result, adjustments need to be made to the way socio-economic and demographic data are collected, and we have witnessed significant progress in recent years. Two factors have facilitated this improvement in the richness of the data: on the one hand, new technologies offer alternatives to move in this direction and improve the understanding of the behavior of the dynamics of urban ecosystems; on the other hand, public institutions have developed open-data dissemination platforms, which for the first time provide wide access and allow an in-depth study of the phenomenon of vulnerability.

1.3 Big Data and social inclusion

Statistics are undoubtedly how the demographic, economic and social dynamics of territories have been analyzed. With the arrival of Big Data, a new dimension of analysis of these dynamics has opened: it is now possible to approach an analysis of individual behaviors, where the characteristics of each are an integral part of analytical models. This level of detail allows for a better understanding of the integral, "metabolic" behavior of a city. Specifically, it is possible to give **descriptive perspectives**, where the magnitudes related to vulnerability can be made explicit and quantified; **predictive,** where these magnitudes can be predicted in the future, depending on variations in the indicators that determine them; and finally **prescriptive**, where social policies can be evaluated and optimized through the control of their predicted effects on vulnerability.

Currently, most applications of data processing and statistical techniques and methods with Big Data have been made within the business sector, in the study of consumption patterns to maximize profits. The use of these techniques in the public sector has been much more limited. It is true that the so-called "e-governance" has become increasingly important in the processes of participation and elections, or in telematics management at the service of the citizenship. However, these advances have usually been restricted to data management, and there has been a lack of statistical uses which would have allowed a deeper analysis of social dynamics. Recently, data journalism has become popular (Yárnoz 2019[3]), where public management scrutiny has played a relevant role. Platforms have also been launched to bring Big Data professionals closer to data and management "owners" (DataForGoodBCN 2020[4]). These initiatives explore the possibilities of creating prediction models based on existing statistical data, and why not, identify new variables needed to be collected to make the action of social services in the territories more effective.

Most studies in urban economics emphasize the need for small-scale analysis: the more "granular" the observed data, the better the quality of the models to explain the variations, and the more reliable the conclusions. In the case of vulnerability, the segregation of data on a small scale gives a more realistic and everyday x-ray of poverty in our immediate environments and the response capacity that the solidarity fabric provides, precisely at this scale of action.

This is therefore the great potential of Big Data: to discover and accurately measure, through small-scale data, the impact of different variables on people's vulnerability. This is precisely the goal of this project.

[3] https://elpais.com/elpais/2019/08/30/opinion/1567195966_013349.html
[4] https://www.meetup.com/DataForGoodBCN/

2. Project Description

This project *"Big Data at the service of social inclusion"* raises as a starting point that statistical data can affect the improvement of the understanding and management of social vulnerability and discuss measures that promote inclusion.

2.1 Structure

The phenomenon of vulnerability is complex and includes determinant objectives, such as indicators of material or health needs, and subjective, related to the psychology of individuals. In particular, people do not behave in the same way in circumstances of vulnerability and poverty. The perception and stigma of poverty impacts on people and territories, an aspect that ends up conditioning the effectiveness of the measures used by local authorities. The prejudices sown by our societies exacerbate inequality between territories, a factor that can stop the processes of risk mitigation and improvement.

In this sense, we believe that an approach at various geographical levels of analysis can help to improve the perception of urban realities. It is important to integrate local aspects, which include objective realities and relationships with the territory, and subjective aspects linked to neighborhood identities. The implication of this type of approach is to guide social cohesion policies at the individual or local level; within the neighborhood; and also between neighborhoods.

To develop these policies, this project sets out the following stages and objectives:

1. Identify from statistical models the patterns of social vulnerability at the urban scale of proximity, and the temporal variations of them.
2. Attribute variations in vulnerability to external causes, of different types.
3. Define a vulnerability predictive tool based on statistical models.
4. Validate the conclusions of the quantitative study with the social entities, active in the field, and detect the qualitative aspects to be integrated in future studies.

2.2 Methodology

The methodology applied in the project analysis process is based on the use of advanced statistical models to relate the vulnerability variables of an individual with variables related to the context of these individuals. Specifically, we want to relate how the probability of a person being in a state of dependence, both health-related and material, depends on their personal condition. This exercise seeks to identify the variables that affect urban vulnerability and the risk of social exclusion, as well as the quantitative measure of their impact.

The following data have been collected:

- Micro-data on Social Services provided by Barcelona City Council, to measure the incidence of vulnerability, in dependency benefits, as financial aid.
- Micro-data at the census tract level on demography, economy, housing, and territory, among others. For the validation of the results in the fieldwork, these main data have been supplemented with data on environmental quality in public space, where six variables of urban liveability have been applied (Echave and Rueda 2008[5]); two ergonomic variables: road accessibility and pedestrian space on the street; two psychological variables: green volume per street section and diversity of economic activities; and two physiological or comfort variables: air quality and street noise level. The set of vulnerability variables considered in this study are organized based on the different dimensions in which an individual establishes his relationship with his environment.

Specifically, we have classified the context variables into the different categories, linked to the mechanism of action on vulnerability, shown in Figure 4. Annex A.1 contains a formal description of the variables used and the corresponding sources.

[5] Echave, C. and Rueda, S., 2008. Liveability index in the public space. In *International conference Walk21 IX: a moving city, Barcelona, Walk21, Cheltenham, Gloucestershire (UK)*.

Autonomia (context personal)	Envelliment i Sobrenvelliment Prestació d'ajudes Serveis Socials Condicions físiques de l'habitatge
Economia (poder adquisitiu)	Renda per llar Preu de lloguer d'habitatge Preu de venda d'habitatge
Ocupació (context laboral)	Atur Ocupació per sector productiu
Entorn (context urbà)	Habitabilitat a l'espai públic Proximitat a equipaments i transport públic
Identitat (context social)	Teixit associatiu Participació eleccions % Població estrangera

Figure 4: Vulnerability variables considered. Source: Own elaboration.

1. **Autonomy Factor**. It captures vulnerability due to limitations in personal development, such as lack of physical or mental abilities, and need for help to lead an independent life. The age of the buildings and the size of the homes can be a limiting factor.
2. **Economy Factor**. It represents the vulnerability created by purchasing power constraints, based on household income and the cost of housing, as one of the main expenses of the household economy.
3. **Employment Factor.** Represents the vulnerability created by pressures from the work context. In this case, unemployment has been considered, however, it should be possible to add the quality of employment with the% of temporary employment.
4. **Environment Factor**. It represents the vulnerability due to limitations of the environmental quality that the city offers to the citizens. It includes livable public space and proximity to services such as public facilities, public transportation, and neighborhood commerce. Congestion and pollution are not included but could be added in the future.
5. **Identity Factor**. The construction of identity is done collectively, hence the aspects linked by social cohesion. The presence of a foreign population, participation in elections and also the density of the associative fabric have been considered in this group.

From these data, statistical models have been created to relate the response variable - vulnerability - to the context variables. These models show the association / correlation between variables. By having access to the annual evolution of the indicators during the period 2013-2019 and introducing baseline values per unit of territory (neighborhood or census tract), the coefficients related to the variables can be interpreted as causal, and used to make predictions. In other words, if a public policy can modify a context variable, this variation will lead to variations in the degree of vulnerability, determined by our model.

In order to validate the quality of the models, the data were separated into a training set (2013-2018), in which the different models were calibrated; and in a validation set (2019), in which the reliability of the predictions can be compared.

In summary, we develop a statistical (econometric) model fed by low-level geolocated data. Although research in urban economics has created similar models before, to detect links between public policies and economic, educational or health outcomes, for example, it is the first time they have been used at this scale, in Barcelona. We hope that the models developed in this work can resonate and be applied in other geographies.

3. Results

3.1 Description of vulnerability degree

We are interested in understanding two different phenomena:

- On the one hand, we want to understand the vulnerability observable through aids to dependency, i.e., the one that is due (mainly) for reasons of lack of autonomy as we described in section 2.2, although this dependence is probably influenced by other factors, for example economy or environment. This vulnerability is "binary", in the sense that it requires help or not, and therefore does not vary much in its severity or duration (since once a situation of dependence is reached, it does not stop - for example, once a disability is recognized it lasts forever). For this reason, we focus on predicting **the incidence of dependency vulnerability** only.
- On the other hand, we want to understand the material vulnerability, leading to requiring financial assistance from Social Services. This type of vulnerability has much more varied underlying reasons, and as a result we will be interested in three dimensions. First, we want to predict the incidence of material vulnerability in the population (by census tract, age, and gender), to determine a person's likelihood of being vulnerable in this regard. Second, we want to predict the **severity of the situation of material vulnerability**, measured as the amount in euros per person per year needed to alleviate the problematic situation. Third, we want to predict **the duration of the situation of vulnerability**, measured as the probability of a person continuing to need help for two years in a row.

Variations by age and gender

Age and gender are the main determinants of vulnerability. Specifically, the degree of vulnerability measured in dependency benefits increases greatly with age. Figure 5 shows the percentage of dependency-assisted residents, by gender and age.

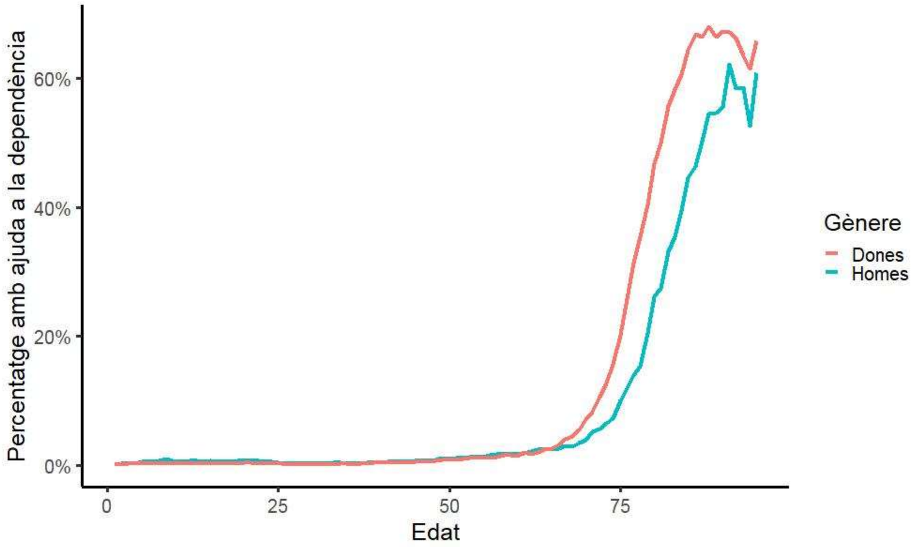


Figure 5: Percentage of residents with dependency aid, by gender and age, year 2019. Source: Own elaboration

The propensity to be vulnerable that is manifested in financial aid also has strong variations with age and gender, with children under 16 being more vulnerable, as well as women between 25 and 50 years old, usually in single-parent families with children. Figure 6 shows the distribution.

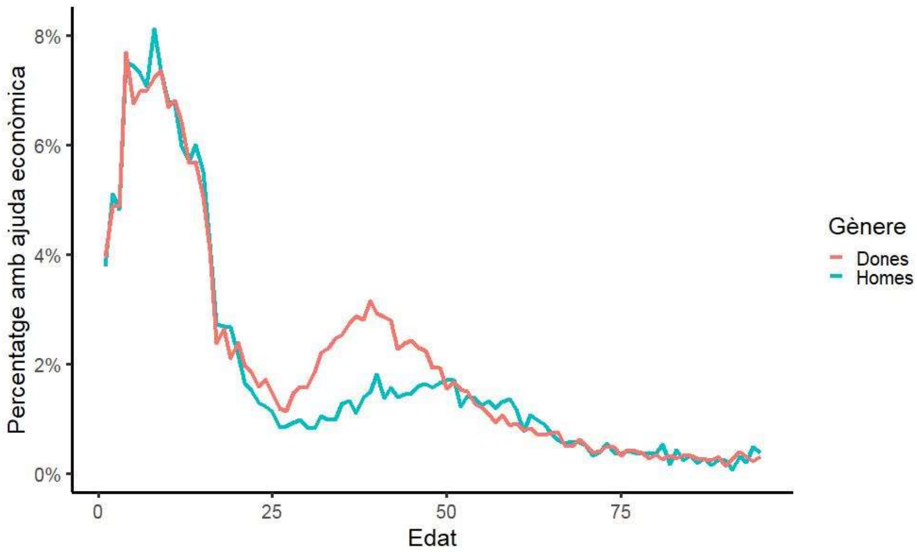


Figure 6: Percentage of residents with financial aid, by gender and age, year 2019. Source: Own elaboration

The economic amounts received do not vary as much, suggesting that the "severity" of the vulnerability does not vary throughout life. Figure 7 shows the average, by gender and age.

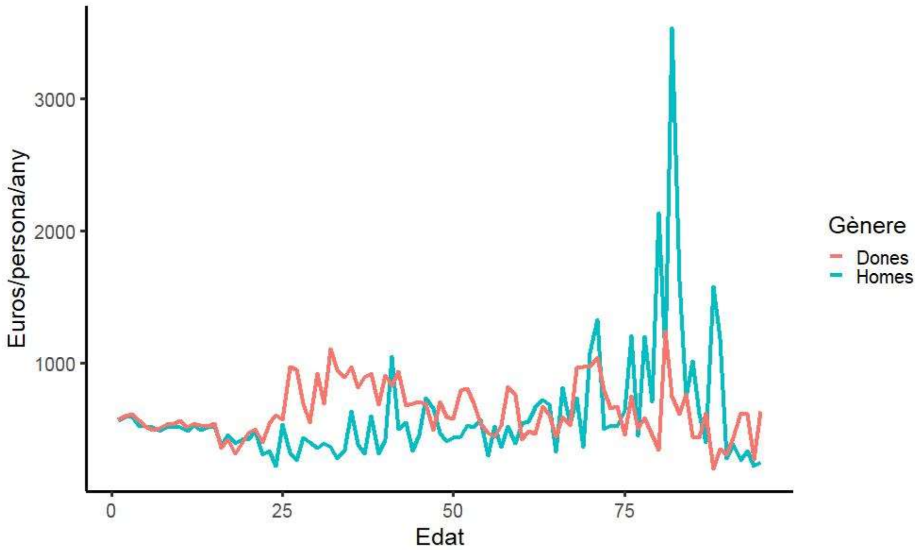


Figure 7: Average financial aid, by gender and age, year 2019. Source: Own elaboration

Finally, the duration of vulnerable people in the system, measured as the probability that a person with financial aid will continue in the system the following year, is high. As Figure 8 shows, around 60% of vulnerable people, who receive financial aid, stay in the system, that is only 40% are able to recover and reduce their degree of vulnerability. This percentage of "retention" in Social Services increases with age (before leaving from the age of 75 due to mortality): from the age of 50, vulnerability becomes chronic, possibly due to situations of permanent exclusion from the job market.

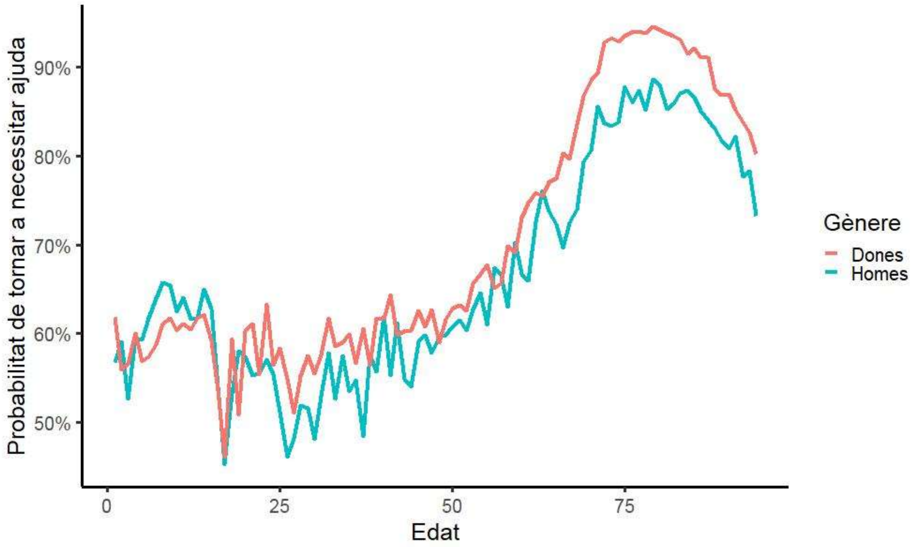


Figure 8: Percentage of repeat recipients out of the total number of recipients of financial aid, by gender and age, year 2019. Source: Own elaboration

<u>Variations in the territory</u>

The visualization of the variables in the territory, each one separately, shows the most common reading we have of the neighborhoods of Barcelona.

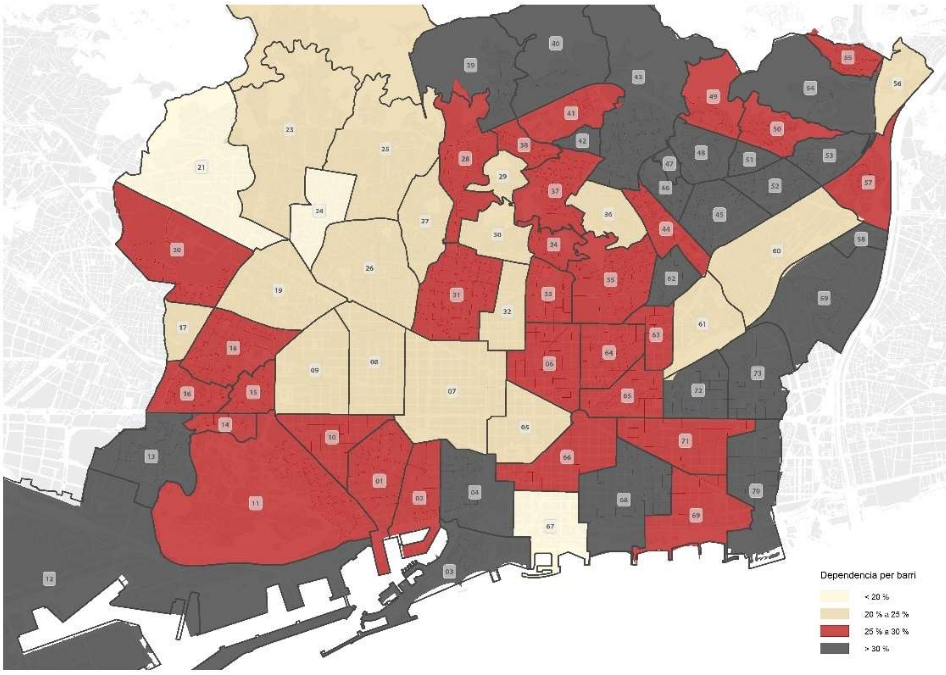


Figure 9: Percentage of residents aged 65 or over with dependency aid, by neighborhood, year 2019. Source: Own elaboration.

As shown in Figure 9, in the districts of Ciutat Vella, Horta-Guinardó, Nou Barris, Sant Andreu and Sant Martí the proportion of people with dependency assistance is the highest in the city, while in the Eixample, les Corts and Sarrià-Sant Gervasi is the lowest. This visualization, aligned with a stereotyped vision of a separate Barcelona in “Uptown” vs. Popular neighborhoods hides a more complex reality, as shown in Figure 10: within the same neighborhood we can see a high degree of variability in vulnerability.

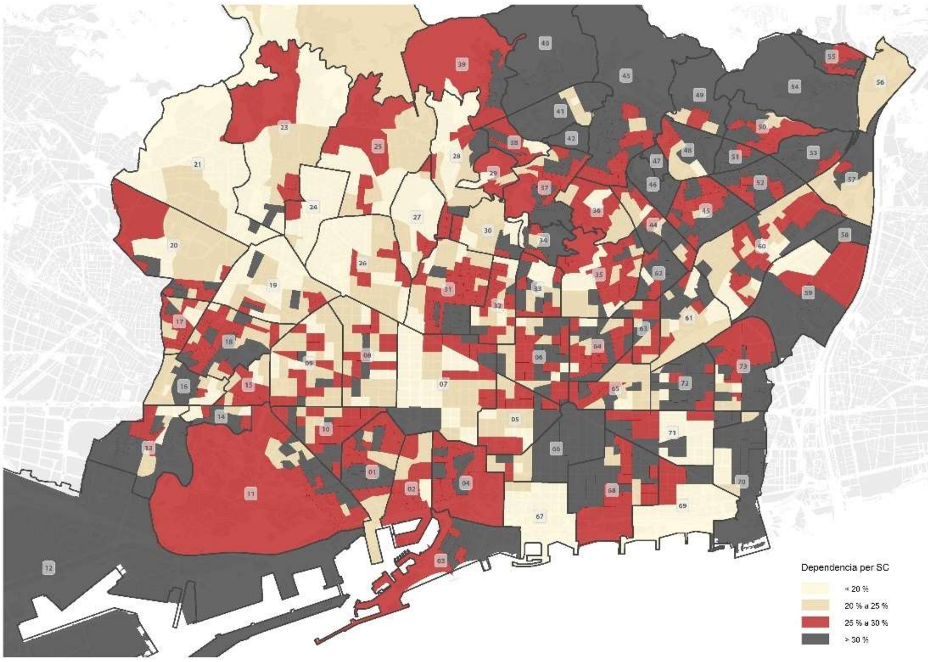


Figure 10: Percentage of residents aged 65 and over with dependency assistance, by census tranche, year 2019. Source: Own elaboration.

In terms of financial aid, the distribution in the territory shows even more marked differences between neighborhoods, as seen in Figure 11, with the same patterns as with dependency aid.

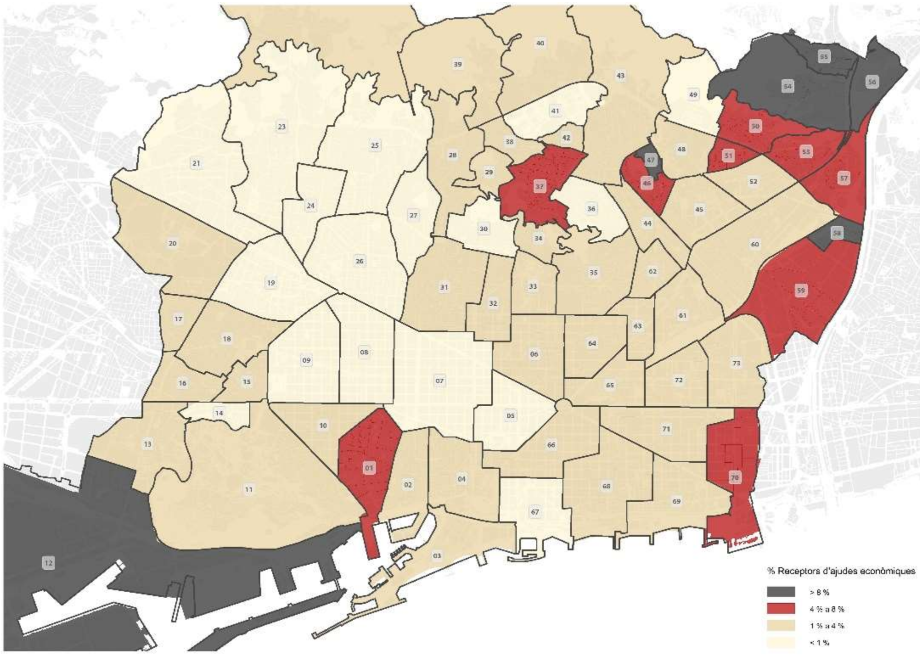


Figure 11: Percentage of residents with financial aid, by neighborhood, year 2019. Source: Own elaboration.

In addition to the percentage of people with financial aid, it is interesting to also measure the average aid, in € / person / year, as well as the percentage of people who stay in the system, those who receive help from those who already receive it the previous year. Figures 12 and 13 show these magnitudes. We can see that, although there is not much variation in the amounts, in the neighborhoods where there are more people receiving aid (for example in the district of Nou Barris), aid tends to be lower and the percentage of repeated recipients is less as well, suggesting a less serious but more widespread vulnerability.

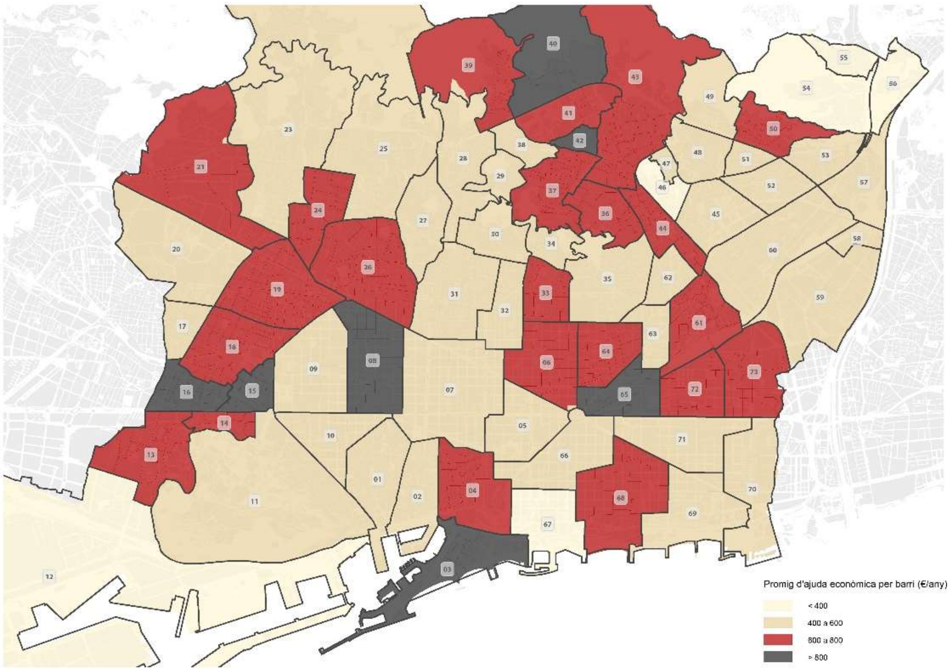


Figure 12: Average aid for recipients of financial aid, by neighborhood, year 2019. Source: Own elaboration.

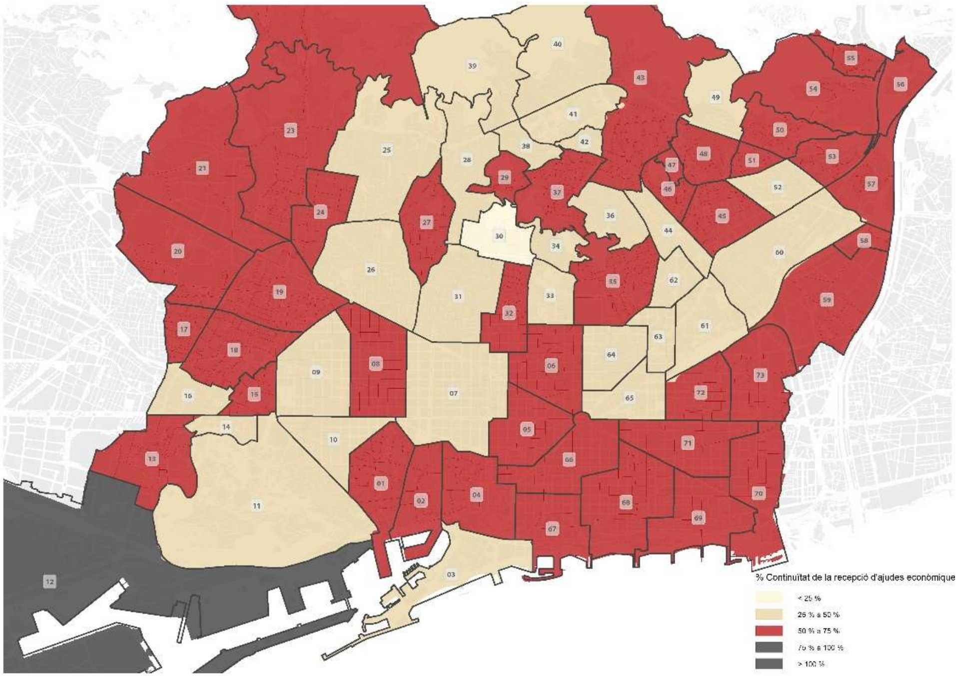


Figure 13: Percentage of repeat recipients out of the total number of recipients of financial aid, by neighborhood, year 2019. Source: Own elaboration.

These descriptions, although illustrative of the phenomenon, hide the deep reasons that cause the different degrees of vulnerability: if one takes into account that the population is unequally distributed in the city, and that socioeconomic conditions vary throughout the territory, it is reasonable to expect variations in the degree of vulnerability. The real question is what relationship exists between vulnerability and these variables.

3.2 Description of the socio-economic conditions

The factors which vary the most within the territory are the ones related to the economy and purchasing power: income, unemployment rate and the price of housing. Figure 14 shows that areas are more vulnerable to economic factors that look like those to be charged or the selling price of homes above the annual rent land. The 15% of the municipality meets this condition, especially in the neighborhoods of the districts of Nou Barris, Horta-Guinardó and Sant Andreu. The offer of rental prices in 53% of the use of the municipality representing annually between 33% and 50% of the total household income. 21% of the city's surface area can refer to libraries that exceed half of the annual income and 11% is around 75% of the income.

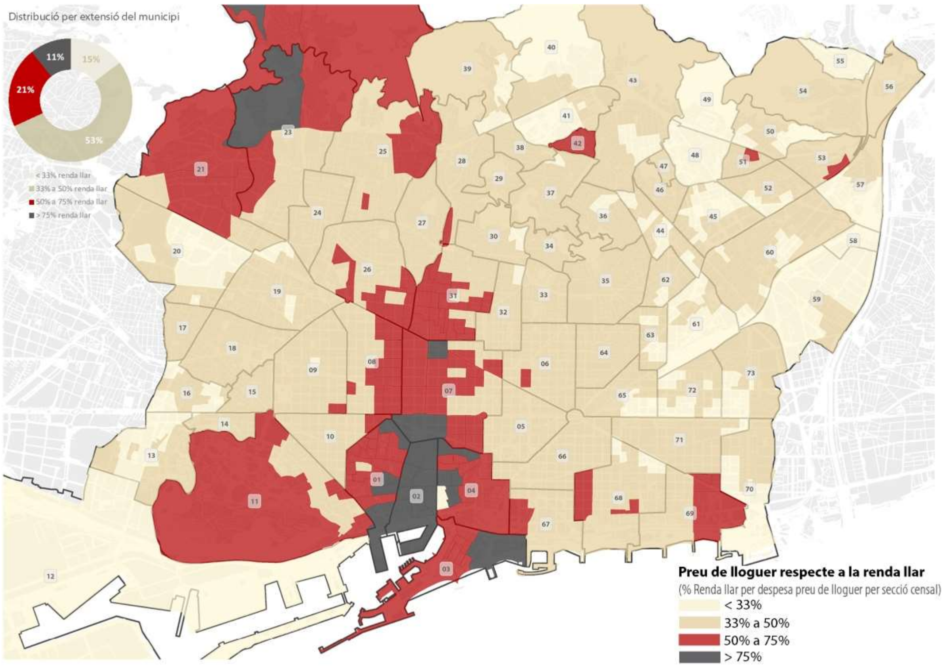
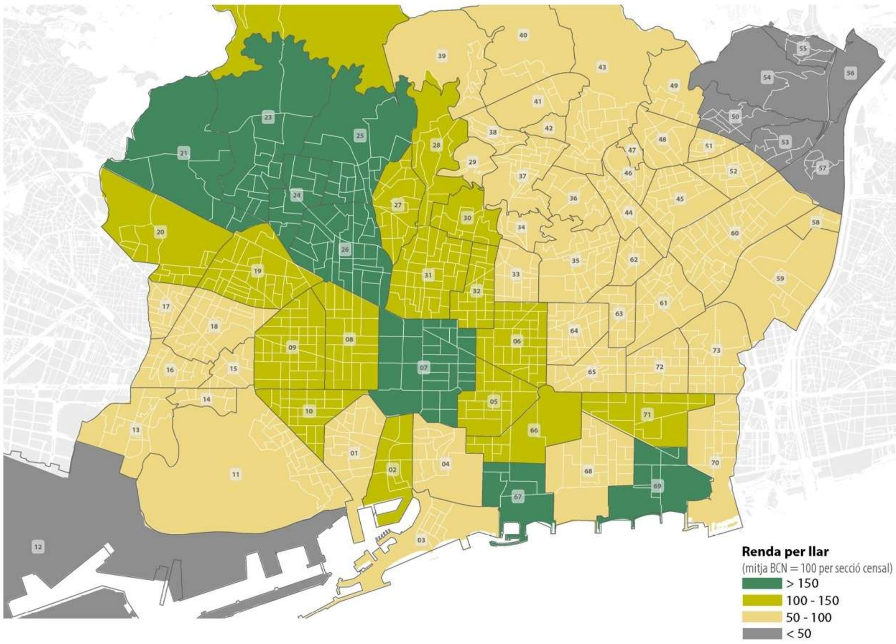


Figure 14: Ratio between rental price and average household income, year 2018. Source: Own elaboration.

Figures 15 and 16 show variations in income and housing cost variables.



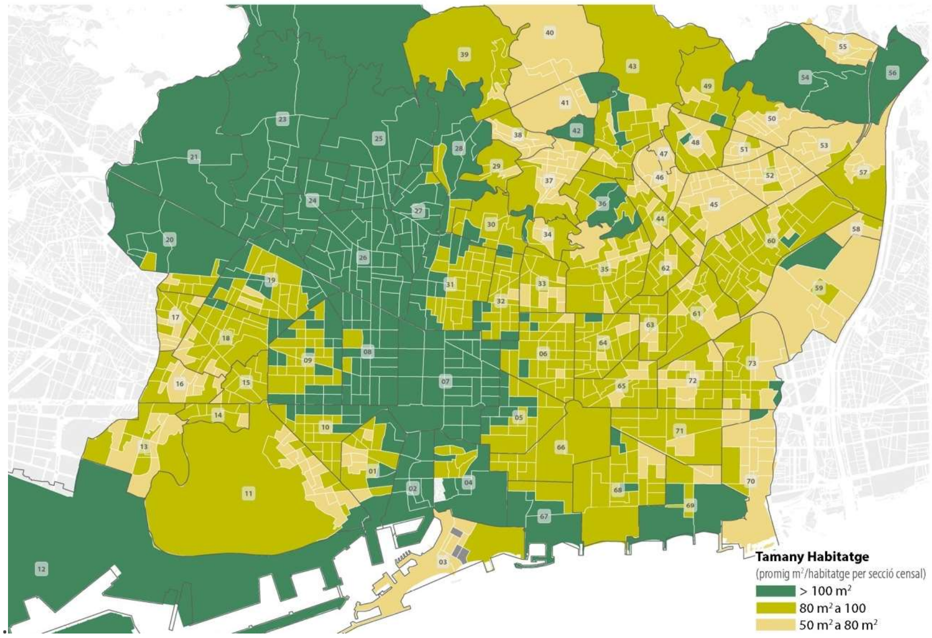
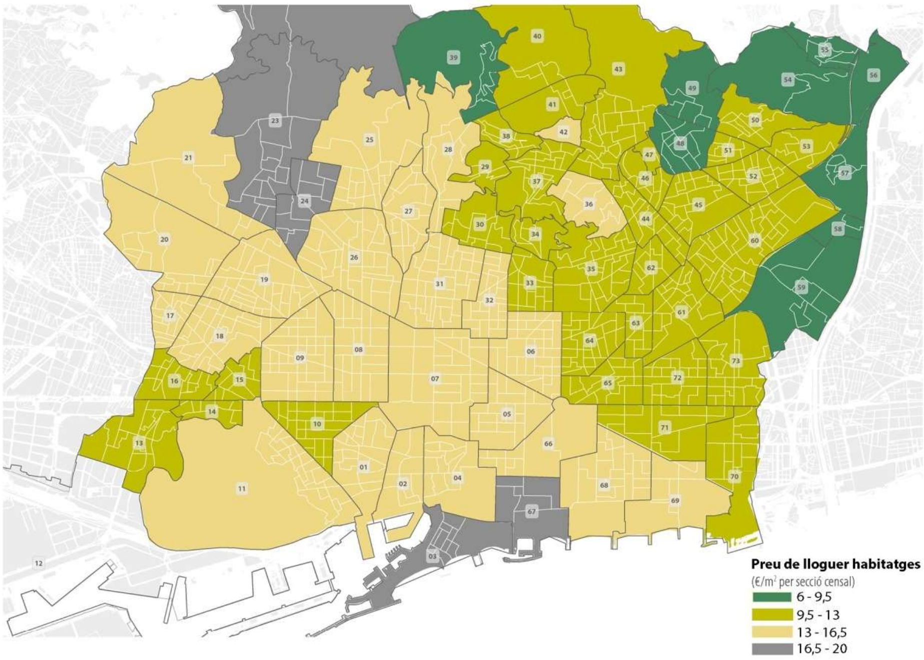


Figure 15: Household income and housing size, year 2018. Source: Own elaboration.



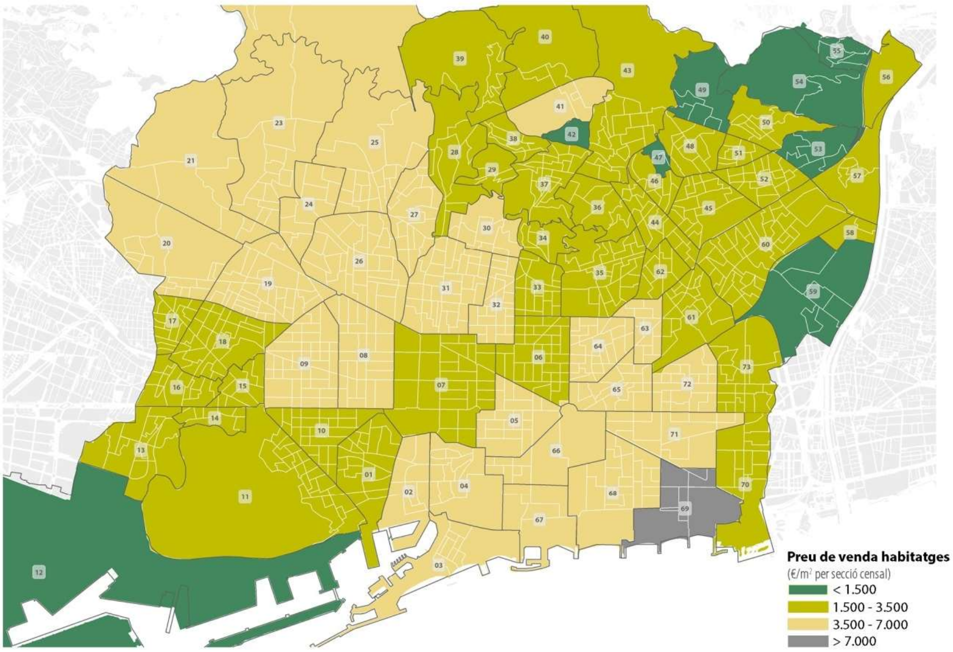
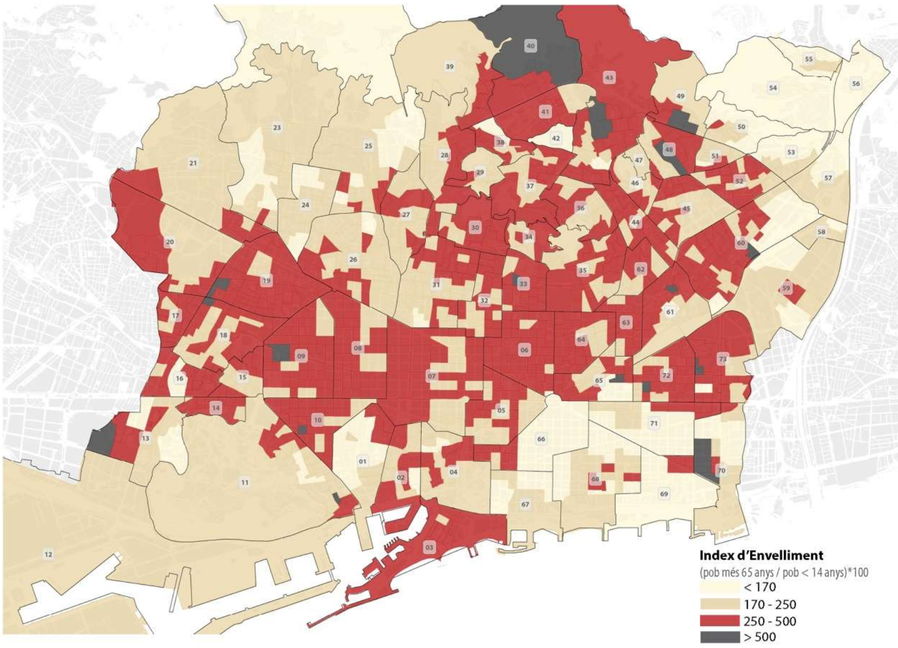


Figure 16: Rental and sale price, year 2018. Source: Own elaboration.

In the case of dependency as a factor of vulnerability, it is mainly conditioned by the age of the people. Therefore one of the critical factors in aid to dependency is the distribution of the elderly in the set of neighborhoods of the city differs when it comes to the Aging Index (people over 65 / people under 14 years) and the Over-Aging Index which considers people over 85 years of age with respect to people over 65 years of age. These magnitudes are shown in Figure 17.



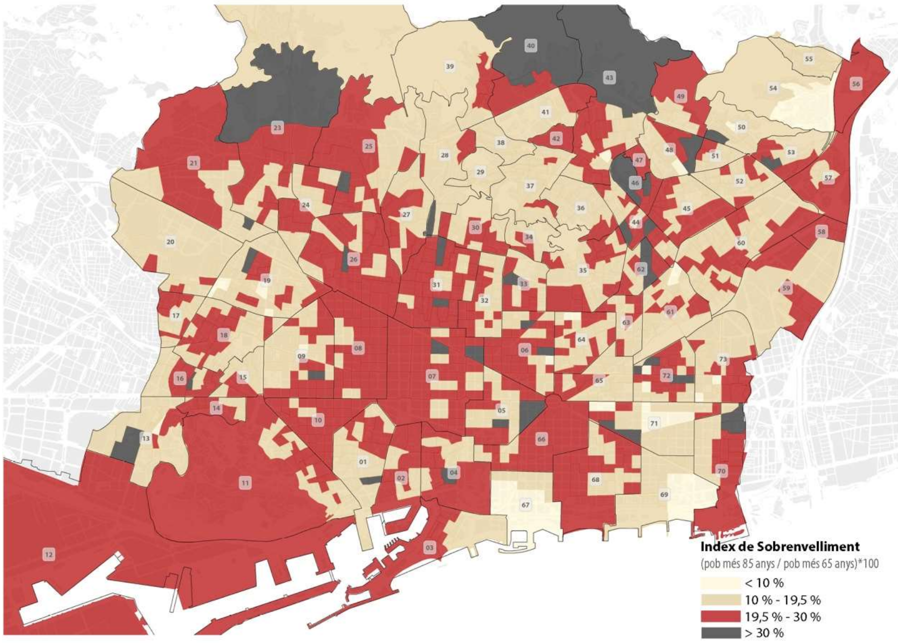


Figure 17: Aging and over-aging, year 2018. Source: Own elaboration.

45% of the total area of Barcelona presents an over-aging below-average in the city (19.5%). However, 56% of the area exceeds the average value of which 8% correspond to census tracts where people over the age of 85 represent more than 30% of the elderly. These cases are distributed in all districts, but where it is most pronounced is in low-income neighborhoods such as Barceloneta, Sant Martí de Provençals, El Turó de la Peira and Can Peguera, Vallbona and Trinitat Vella. In other neighborhoods with higher incomes, the rate of over-aging is also marked, for example in the Vila de Gràcia, Sagrada Família and Sant Antoni neighborhoods. Regarding the high-income neighborhoods, the neighborhoods of Sarrià-Sant Gervasi stand out for a high rate of over-aging: Sant Gervasi La Bonanova and Sant Gervasi Galvany.

3.3 Predictive models

After reviewing the variables of interest and the explanatory variables, we will focus on describing the different predictive models used. We create four predictive models, one for each of the four vulnerability variables described in section 3.1. Appendix A.2 describes the technical details and we report only a brief description here.

To predict the incidence of people receiving dependency benefits and financial aid, we are interested in modeling the likelihood of a resident being vulnerable and receiving such assistance. We adjust this probability with the information of the explanatory variables described in section 2.2 and annex A.1. A good approximation of this micro-behavior model is the Poisson model, with intensity proportional to the number of residents, and the explanatory variables.

To predict the amount of financial aid received (in euros per year per person), we can use a lognormal distribution, which reasonably limits the distribution of real aid. We use all relevant context variables again.

Finally, to predict the propensity to continue in the system, measured as the proportion of recipients of economic aid in year t-1 that follow economic aid in year t, we use a binomial model. This provides us with a prediction of the likelihood of staying in the system, which is directly linked to the duration of the vulnerability.

It is also possible to consider more advanced predictive models, based on Machine Learning (random forest in our case), described in Annex A.3. which have the particularity of improving the quality of predictions, although they may be less accurate when predictions have to be made outside the "usual" parameters, and especially when predictions are made about indicators in the future.

The final models are detailed in Annex A.2. In developing these models, it became evident that **it was necessary to introduce differential effects by age and gender**, which, like other variables, greatly conditions vulnerability (as shown in Figures 5 to 8). To a lesser extent, the introduction of structural **variations at the neighborhood level** (independent of other covariates such as income, housing characteristics, etc.) also improved the performance of the model, suggesting that there are intrinsic variations in neighborhoods, not connected with the quantitative variables included in our models. Indeed, as we contrasted in the fieldwork, there are neighborhoods that are reluctant to ask for help, on an equal footing; for example, a participant in Les Corts suggested to us that in his neighborhood, "I would rather die than ask for charity from Social Services".

Apart from these fixed effects, the models detected **strong relationships between the drivers of vulnerability, and their incidence, severity, and duration**. As shown in Figure 18, the models detect that each context variable may affect the characteristics of the vulnerability differently. As the models are implemented, these relationships are association / correlation. To establish causality, we must be able to ensure that variations are due to external causes, and for some of the variables this could be a reasonable hypothesis (e.g., unemployment, which when it appears is not of its own accord), while for others are not (for example, the percentage of new housing, after 1960, may vary because there is more vulnerability and therefore more willingness to change housing). In any case, even if the relationship is correlational, our models can be used to make predictions. This can be especially useful if you do not have vulnerability data. The inference process could be as described in Figure 18.

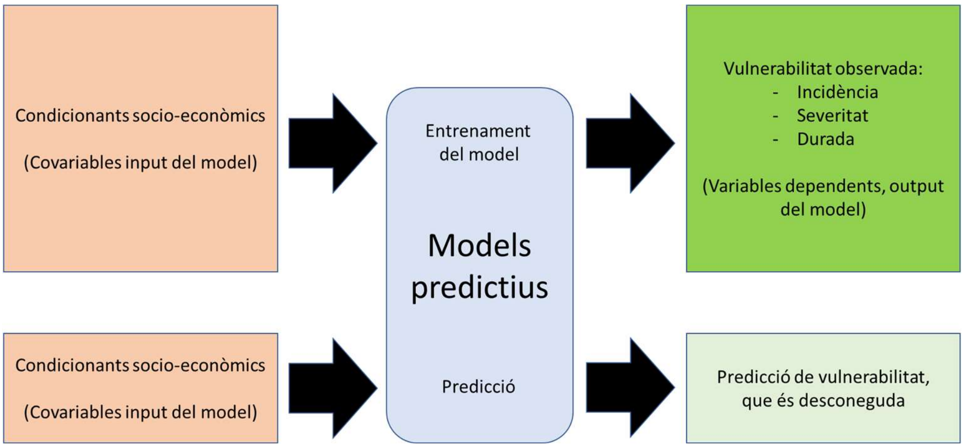


Figure 18: Use of models to create vulnerability indicators

.

3.4 Vulnerability drivers

Models can therefore create predictions of the vulnerability situation. Reliability is reasonably good (see model validation in Appendix A.4). While the models correctly describe the actual data, for example which characteristics of the population and territory are most vulnerable, the accuracy may be impaired due to ignorance of the specific effects of the year (which could not be estimated in the model as it is trained with data from 2013 to 2018, while validated with data from 2019). If one can include the year effect in the model, the accuracy improves significantly.

On the other hand, one of the main learnings of the models is to be able to discover the influence of the variables described in section 2.2 on the degree of vulnerability. We summarize here the main conclusions that we can draw from the analysis of the coefficients estimated by the model, in Figure 19:

- Over the years, both the incidence and the amount of financial aid and continuity in Social Services have increased rapidly (positive year coefficient), although growth has slowed (negative square year coefficient)
- The territory with the most immigration has a higher incidence, severity and duration of vulnerability related to financial aid (coefficient % of residents of Spanish nationality negative). However, this relationship does not exist in terms of dependency assistance, which is natural: dependency has to do with

physical and psychological problems, where they have no statistical relationship with the percentage of immigration.

- Having more stable populations (with less rotation in a place of residence) reduces the incidence of vulnerability, but increases its severity and duration (coefficient % of residents with> 5 years in the same address). This suggests that the territory with the most residential mobility is also the one with the highest incidence of vulnerability, although these are less severe and short-lived problems (perhaps because those affected move to other neighborhoods).
- Participation in citizen life (measured as electoral participation) is related to less vulnerability (coefficient % participation in elections negative). This seems to be a clear relationship of association (not causation), where lack of participation and vulnerability move in the same direction.
- The effect of income and unemployment rate strongly influences the incidence of the vulnerability, although it is not related to the amounts or durations of this vulnerability (coefficients log (annual income) and log (rate of employment).
- The level of education, apart from its influence on the unemployment rate, considerably increases the incidence and duration of the vulnerability (coefficient % of residents with a positive or lower level of ESO education), while reducing its severity.
- The cost of housing (both sale and rental) does not influence dependency but does increase the incidence of economic vulnerability (coefficients of log (sale price € / m2) and log (rents € / month)). An increase in rent increases the amounts of aid (severity of the vulnerability), but not the duration. A rise in buying and selling prices slightly decreases the severity of the vulnerability (perhaps due to an increase in the wealth of the household when it owns its home, linked to the well-known endowment effect in economics), but increases its duration.
- The constructive typology around the territory also has a distinctive effect, although it is a static effect with very little temporal variation. Census tracts with more numerous, newer or denser constructions (apartment towers instead of houses) have less vulnerability (negative coefficients log (number of buildings); % buildings after 1960; % dense buildings, with more than 20 appartments; and average age of buildings), although its severity may increase, as does the effect of rental prices which puts additional pressure on family finances.

- Finally, the environment has a relatively small effect, but worth noting. Areas with more trade and better connection to infrastructure (health, transport) tend to have less vulnerability (negative coefficients log (number of shops at 100m); log (distance to health center); metro stop at <100m; bus H or V at <100m). At the same time, when shops are mostly food-related, it is a sign of higher vulnerability (positive coefficient of% of food stores), perhaps due to lower material spending that make other less basic types of trade unviable.

Variable	Y_DEP	Y_ECO	Q	C
Year (2013 = 0, 2019 = 6)	0.277***	0.501***	0.597***	0.759***
Square Year (2013 = 0, 2019 = 36)	-0.009***	-0.032***	-0.084***	-0.100***
% Spanish nationality residents	-0.109	-1.541***	-0.560***	-0.498*
% residents with > 5 years at the same address	-0.488***	-0.375***	1.080***	1.130***
% participation in elections	-0.072**	-1.665***	-1.328***	0.142
log(annual income)	-0.106***	-0.416***	-0.014	0.087
log(unemployment)	0.098***	0.050*	0.005	-0.144*
% residents with an ESO level or lower	1.337***	3.934***	-0.551***	1.124***
log(sale price €/m2)	0.013	0.051**	-0.072**	0.444***
log(rentals €/month)	-0.034	0.195**	0.774***	-0.006
log(number of buildings)	-0.030***	0.012	0.054***	0.045*
% buildings after 1960	-0.134***	-0.149***	0.104***	-0.208***
% dense buildings, with more than 20 apartments	-0.153***	-0.306***	0.052	0.118
average age of buildings	-0.002***	-0.003***	-0.001***	-0.003***
log(number of stores in 100m)	-0.010***	-0.045***	-0.011*	-0.035***
% food stores	0.002	0.443***	0.209***	0.220***
log(distance to medical center)	-0.006***	-0.010***	-0.016***	-0.015***
metro stop at <100m	-0.004	-0.007*	-0.0002	0.020
bus station H o V a <100m	-0.002***	0.007***	-0.007***	-0.017***

*Figure 19: Coefficients of each variable within the different models (*p<0.1; **p<0.05; ***p<0.01).*

Figures 20 and 24 summarize the main relationships visually.

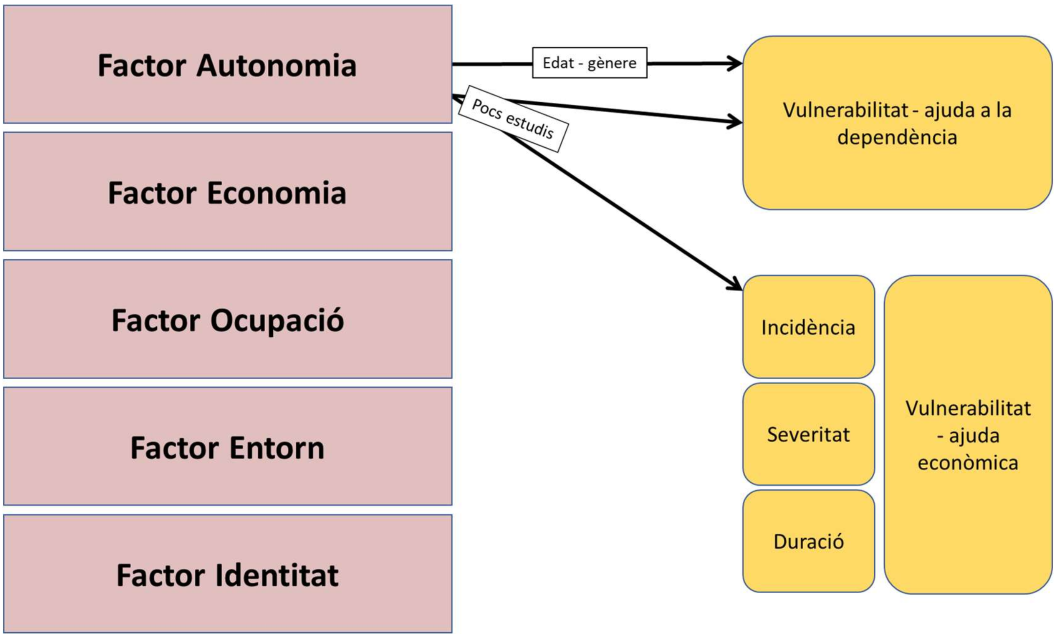


Figure 20: Summary of relationships between variables - autonomy factor.

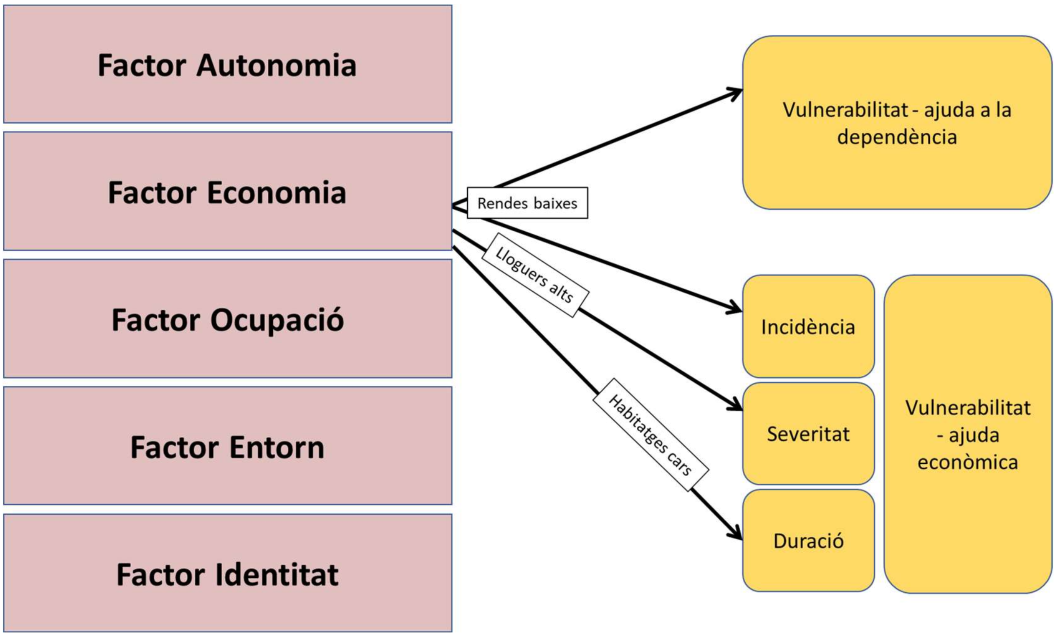


Figure 21: Summary of relationships between variables - economy factor.

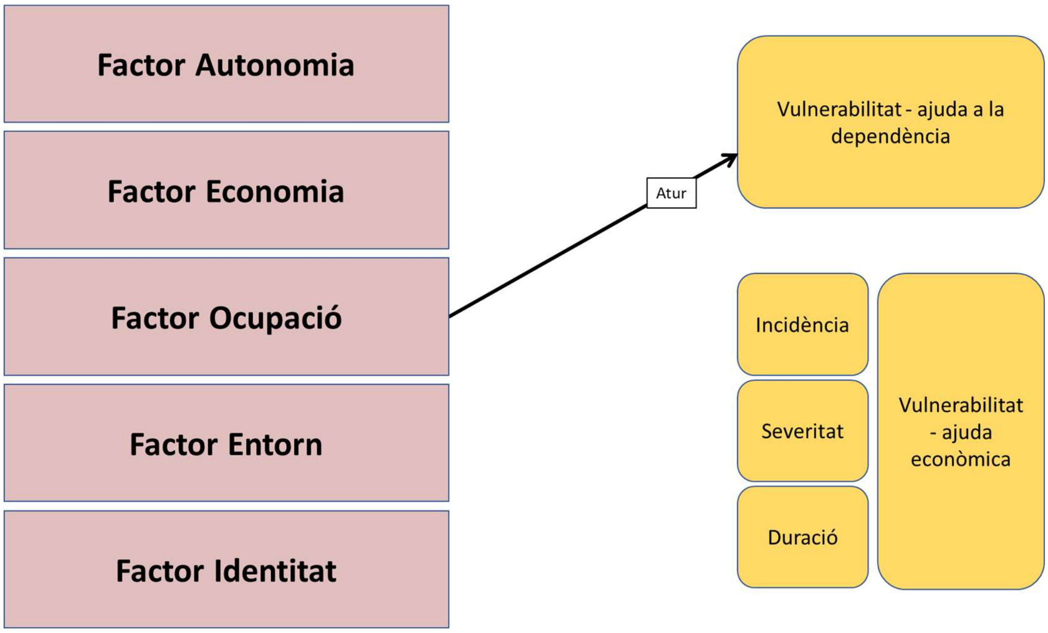


Figure 22: Summary of relationships between variables - employment factor.

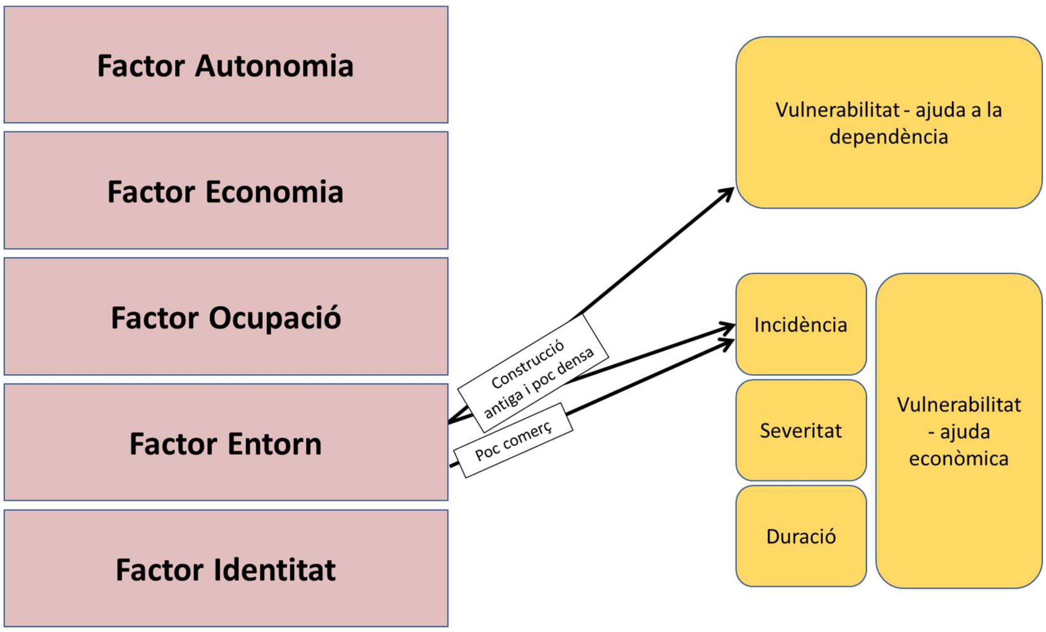


Figure 23: Summary of relationships between variables - environment factor.

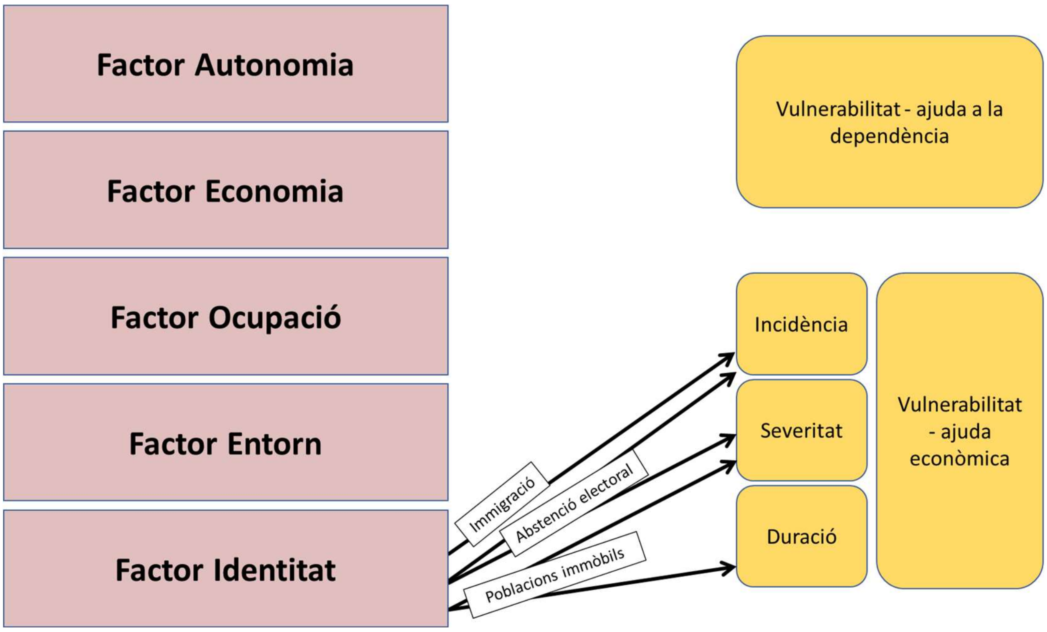


Figure 24: Summary of relationships between variables - identity factor.

3.5 DISPLAY OF RESULTS

To better visualize the data, the Shiny R package was used to develop an online application, which functions as a visual tool that can dynamically display the four variables of interest (incidence of dependency and economic aid; average aid; percentage of new users) on a map of all the census tracts of Barcelona, in age and gender ranges that can be selected. In other words, one can choose a segment of the population (for example women between 30 and 49 years old) and see the variation of the desired indicator in the territory. In addition to the real data, we also show the prediction of the model, and the error made by the model. This tool can help identify areas for model improvement and investigate the characteristics of SCs that have abnormally high or low behavior with respect to prediction.

The application is available on the project website https://social-inclusion.iese.edu/, requiring credentials to access it.

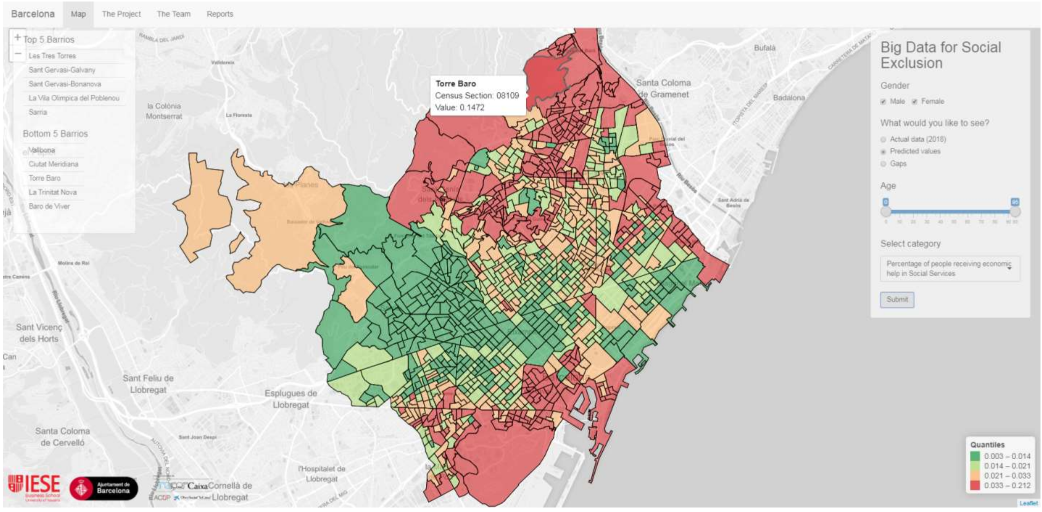


Figure 25: Display of results at https://social-inclusion.iese.edu/

4. Field Validation: Participatory Workshops

In a data-driven world, statistics are increasingly relevant. The way information is organized, synthesized, and interpreted, can influence our lives by shaping our decision-making, but also by influencing our perception and understanding of our contexts. To exchange opinions with people who are immersed in the management of social services, four participatory workshops were organized as part of the project in order to contrast the perception of poverty in the immediate environment.

That is why, following the work of statistical research, the project planned to make contact with the Territory through the organization of four workshops in different neighborhoods of the city of Barcelona. The object of these workshops was, first, to exhibit the results obtained with neighbor groups and neighbors of the districts in order to exchange with them, about their perception of urban poverty and vulnerability to social exclusion. We were particularly interested in contrasting their perception with the hard data of our study. **The material brought to the workshops was based on the assumption that a better understanding of the data can help improve social cohesion at the local level.**

The sessions of the workshops were organized around civic Centers with the support of the heads of social services of the involved districts and associations that are more linked to actions to support social vulnerability.

7

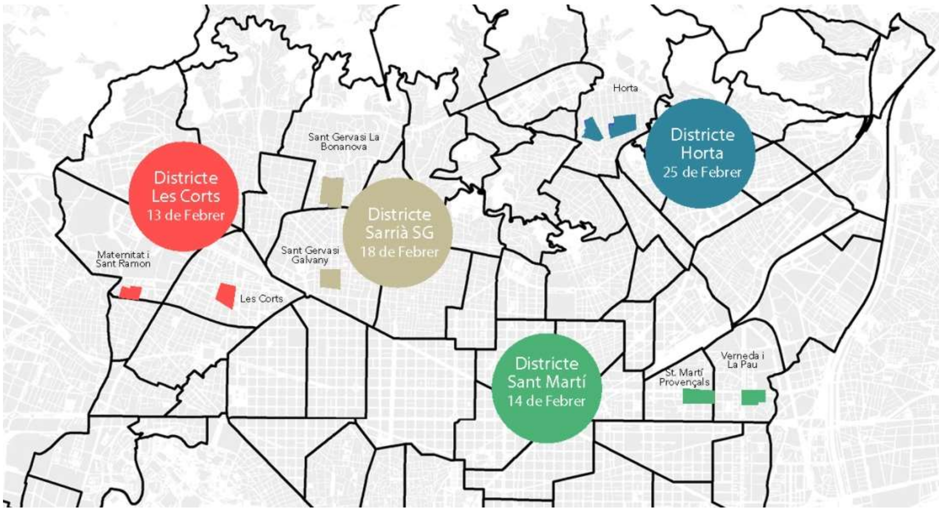


Figure 26: Participatory workshops, and analyzed census tracts. Source: Own elaboration.

Figure 27: Poster announcing the workshops

4.1 Dynamics of the workshops: Urban poverty and citizen empathy

1. Introductory Presentation

The sessions were opened with a presentation that introduced to the attendees the definition of urban poverty, vulnerability and the main aspects that define social exclusion. The objectives of the research study and the variables that were considered in the statistical analysis were also presented.

For each of the workshops, a couple of census tracts were chosen to exemplify the type of variables with which the analysis was looked at in the same neighborhood.

2. First discussion by groups: environment perception

Attendees were divided into groups and the discussion was encouraged in two phases. The first phase focused on the following questions:

- What do you understand by social vulnerability?
- Do you know how to identify social vulnerability of your neighborhood?

Attendees were provided with maps to mark the places where they perceived social vulnerability and poverty within their neighborhood were highest.

3. Results of the study throughout Barcelona

Once the discussion on the most vulnerable places in the neighborhood was over, the results of the variable-by-variable study were shown. The intention of this block was to show that in the rest of the city there are very heterogeneous situations and that there may be situations of vulnerability in a more distributed way than we normally think.

4. Second discussion by groups: improvement proposals

The second phase of the discussion consisted of discussing how actions could be improved for the detection and support of situations of social vulnerability through the involvement of all actors in the area: social services in the district and the city, associations and citizenship in general.

- Do you think statistics reflect reality?
- What do you suggest to improve cohesive actions in the neighborhood and in the rest of the city?

It should be noted that the group's dynamics varied depending on the neighborhood, as well as the attendance. However, the contributions in all the cases were beneficial and have enriched the study.

4.2 Workshop Les Corts District

DATE: 13 February 2020

PLACE: Cultural Centre Les Cristalleries

POVERTY PERCEPTION:

The participants pointed out the areas they thought were the most vulnerable to poverty and social exclusion within their neighbourhood and surroundings. In the case of Les Corts, the areas marked by people coincide precisely with the oldest areas of the neighborhood. One aspect that stood out from the workshop in Les Corts is that there is much loneliness in elderly people, but there is also a resistance to ask for help. The fact that they are traditionally high-income neighborhoods, makes it more difficult to accept a situation of vulnerability and need to ask for help to social services, unlike other neighbourhoods where poverty is not so stigmatised. But participants claimed that the neighborhood had seen a decline in purchasing power, and the very high prices of housing have led new generations to move to the district of Sants. In order to jointly analyze the variables of vulnerability, two census tracts were chosen, one from the district of Les Corts (19) and the other in the neighborhood of La Maternitat i Sant Ramon (20).

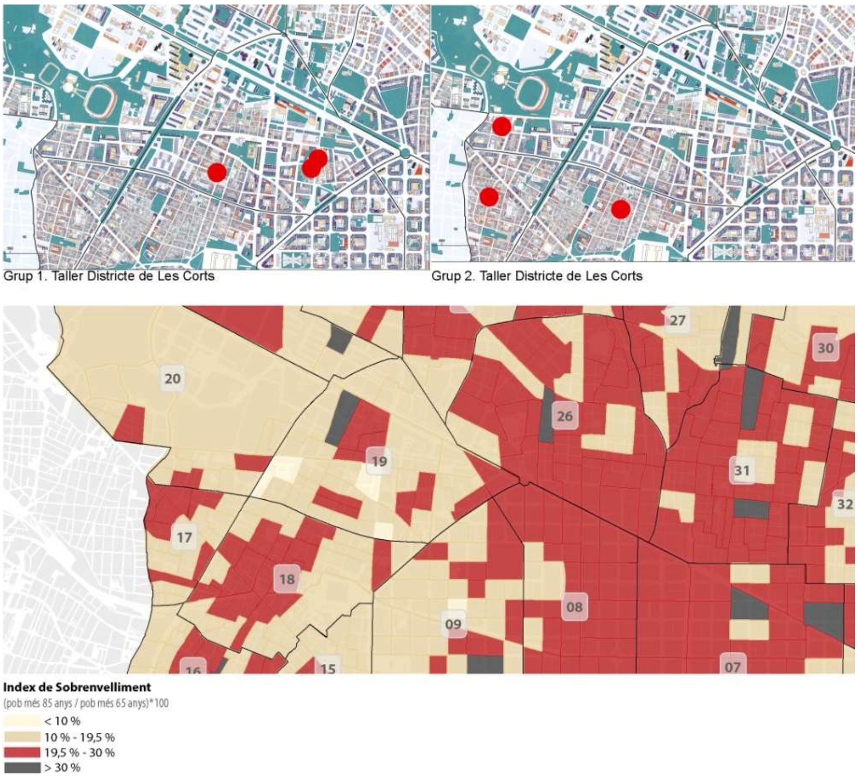


Figure 28: Les Corts District.

With regard to urban quality conditions, the values in both census tracts show different ratios of liveability, but exceed the ratios in the district area and at the scale of Barcelona. In the case of the census tract in the district of Les Corts, the resulting ratio is 5.3 (out of 10), the variables leaving low-score are air quality, green volume and the acoustic comfort. This is due to its proximity to Av. Numància.

In the case of census tract in the Maternitat and Sant Ramon districts, the values of each variable are satisfactory, giving as a global liveability ratio of 6.8 points.

SC 4011 Barri de Les Corts

	Barcelona	Barri Les Corts	SC 4011
Població	1.642.499	46.274	1.765
% Població estrangera	20,2	12,2	7,3
Index Sobrenvelliment	19,5	16,8	8,7
% Pob amb dependència SS	78,1	91,6	98,2
% Pob amb ajudes econòmiques SS	20,1	8,0	5,3
Ajudes SS (euros/ persona)	900	817	1.410
Renda per llar (€/m2)	100,00	120,0	120
Preu lloguer (€/m2)	12,90	14,3	14,3
Preu venda (€/m2)	4.344	4.650,0	5.605
Tamany habitatge (m2)	97,78	103,9	103,4
% Persones a l'atur	7,39	6,6	3,6
% Persones amb ocupació	96,00	93,4	96,4
% Edificis abans 1970	63,77	49,9	12,4
Distància mig a centre de salut (m)	216,91	161,6	190,6

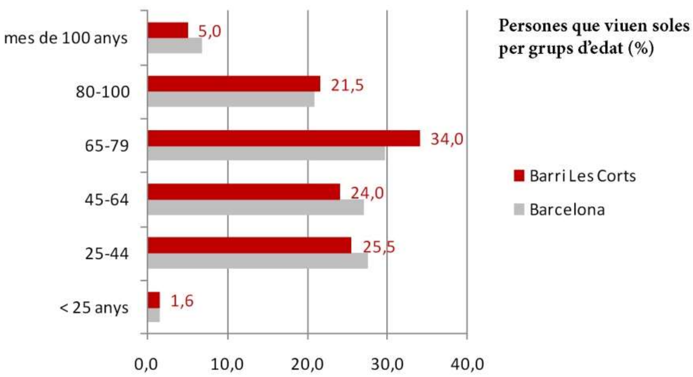
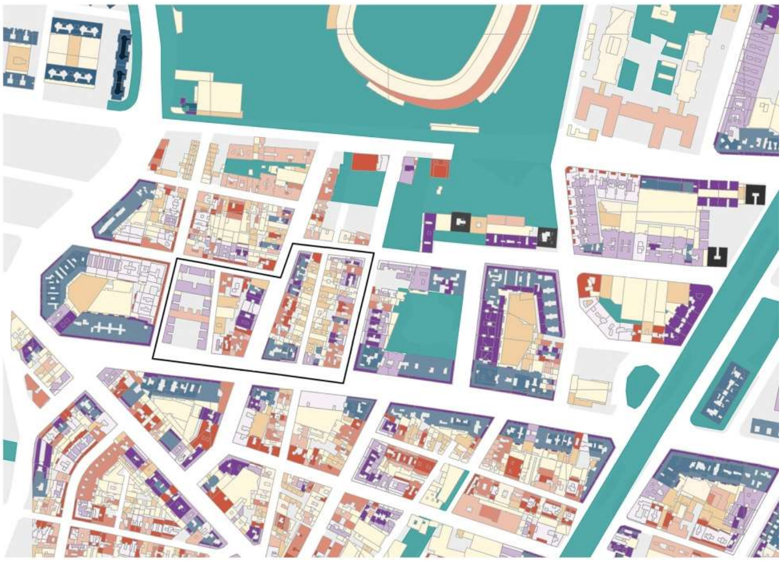


Figure 29: SC 4011

SC 4040 Barri de La Maternitat i Sant Ramon



	BCN	Barri Maternitat i St Ramon	SC 4040
Població	1.642.499	23.836	1.618
% Població estrangera	20,2	12,7	15,3
Index Sobrenvelliment	19,5	16,8	24,6
% Pob amb dependència SS	78,1	91,2	87,1
% Pob amb ajudes econòmiques SS	20,1	8,90	12,2
Ajudes SS (euros/ persona)	900	1.145	2.388
Renda per llar (€/m2)	100,00	114,2	114
Preu lloguer (€/m2)	12,90	13,4	13,4
Preu venda (€/m2)	4.344	4057	3.642
Tamany habitatge (m2)	97,78	100,1	86,2
% Persones a l'atur	7,39	3,6	3,6
% Persones amb ocupació	96,00	96,4	96,4
% Edificis abans 1970	63,77	39,8	76,0

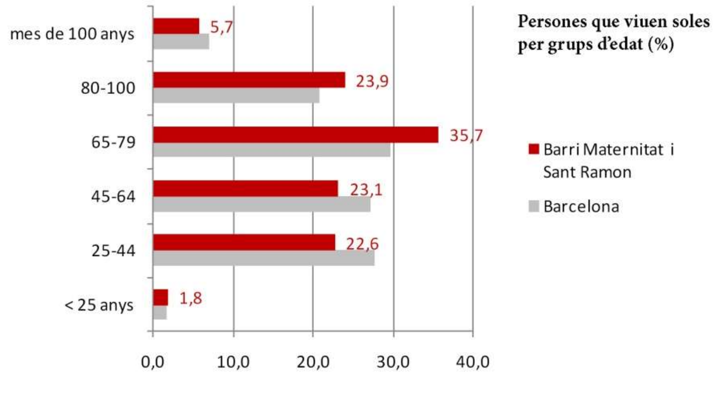


Figure 30: SC 4040

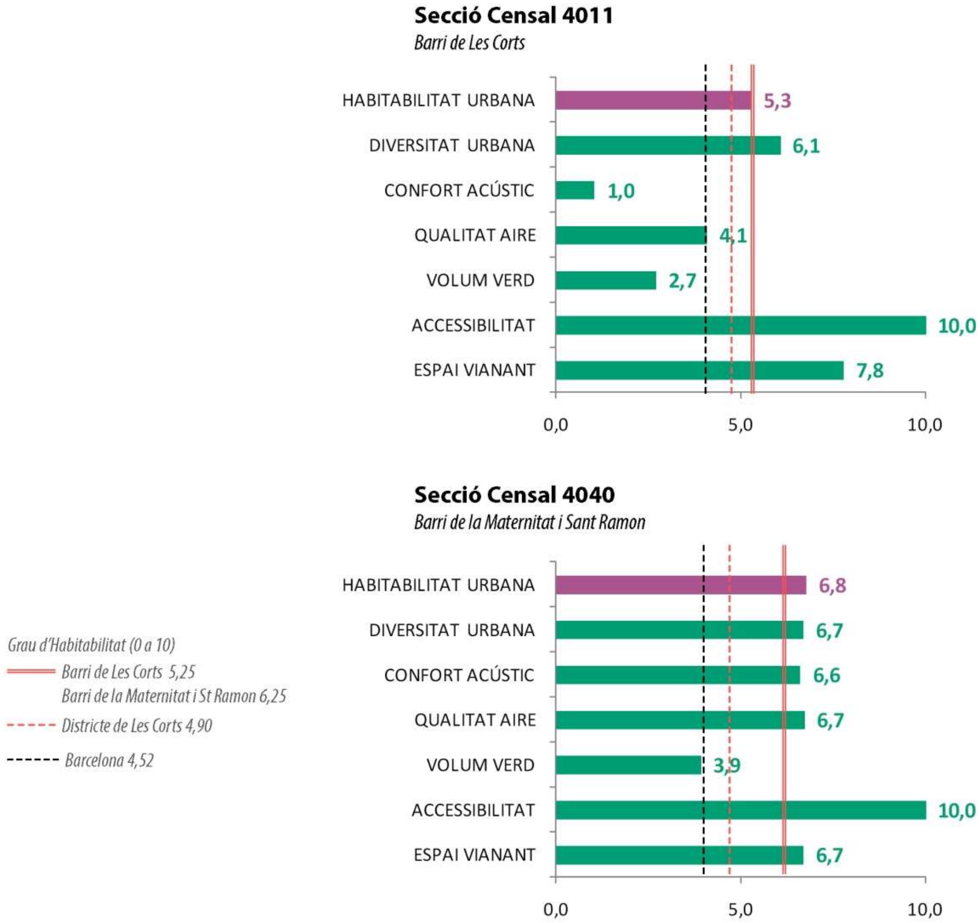


Figure 31: Comparison of SC 4011 and 4040

4.3 Workshop Sant Martí District

DATE: 14 February 2020

PLACE: La Verneda Civic Center

POVERTY PERCEPTION:

The second workshop in the Sant Martí de Provençals district was the busiest of all the workshops and three discussion groups were organized. The session was attended by representatives of the associations of the La Verneda Civic Centre, especially the School for Adults. Regarding the first question about identifying vulnerable areas within the neighborhoods of St. Martí de Provençals (72) and la Verneda and la Pau (73), attendees again identified the places that coincided with the highest levels of over-aging. Regarding the discussions around poverty and how to detect cases of vulnerability in the neighborhood, attendees insisted on the lack of sufficient care for the elderly. On the other hand, emphasis was placed on the fact that people with fewer resources tend to be more social cohesive and are therefore more active in the organization and call for collective action.

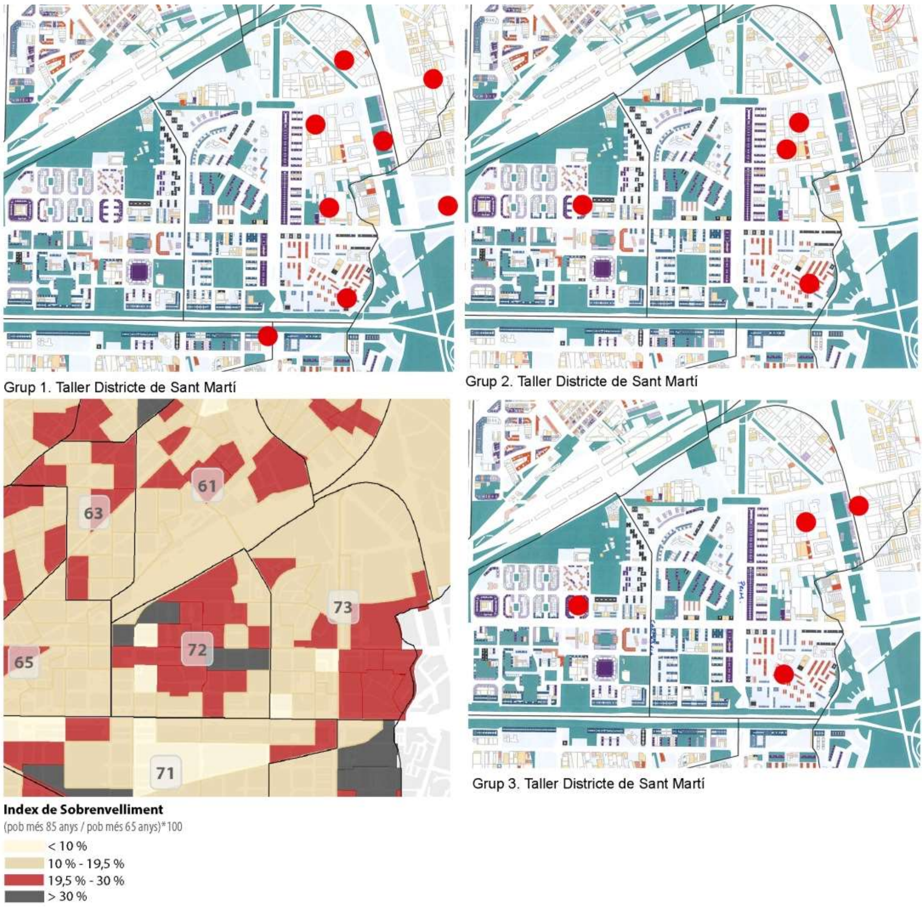


Figure 32: District Sant Martí.

SC 10120 Barri de Sant Martí Provençals

	Barcelona	Barri St Martí Provençals	SC 10120
Població	1.642.499	26.061	1.481
% Immigració	20,2	13,8	15,2
Index Sobrenvelliment	19,5	19,5	39,4
% Pob amb dependència SS	78,1	89,3	90,4
% Pob amb ajudes econòmiques SS	20,1	10,1	10,1
Ajudes SS (euros/ persona)	900	976	1.018
Renda per llar (€/m2)	100,00	67,4	67
Preu lloguer (€/m2)	12,90	10,9	10,9
Preu venda (€/m2)	4.344	2.926	3.785
Tamany habitatge (m2)	97,78	83,3	94,2
% Persones a l'atur	7,39	5,0	5,0
% Persones amb ocupació	96,00	95,0	95,0
% Edificis abans 1970	63,77	69,6	100,0
Distància mig a centre de salut (m)	216,91	249,9	300,2

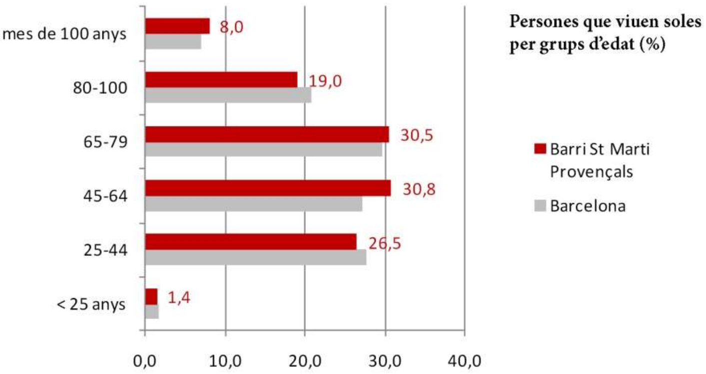


Figure 33: SC 10120

SC 10140 Barri de Verneda i la Pau

	Barcelona	Barri Verneda i La Pau	SC 10140
Població	1.642.499	28.791	1.648
% Immigració	20,2	13,8	18,0
Index Sobrenvelliment	**19,5**	**17,2**	**24,2**
% Pob amb dependència SS	78,1	81,9	83,2
% Pob amb ajudes econòmiques SS	20,1	16,2	13,9
Ajudes SS (euros/ persona)	900	960	1.053
Renda per llar (100 = mitja BCN)	100	57	57
Preu lloguer (€/m2)	12,90	10,2	10,2
Preu venda (€/m2)	4.344	2.672	2.056
Tamany habitatge (m2)	97,78	83,4	84,3
% Persones a l'atur	7,39	5,1	5,1
% Persones amb ocupació	96,00	94,9	94,9
% Edificis abans 1970	63,77	45,4	100,0
Distància mig a centre de salut (m)	216,9	341,2	47,7

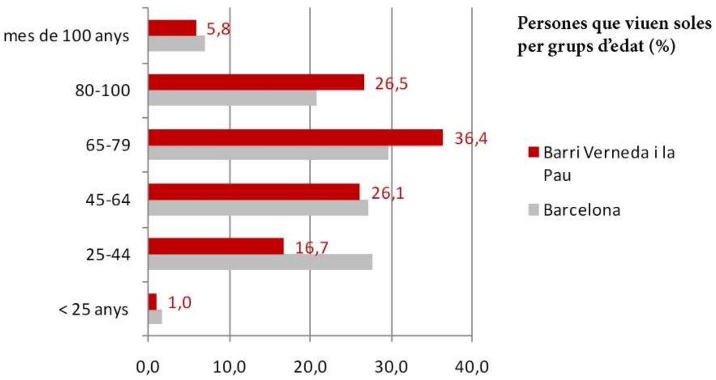


Figure 34: SC 10140

Regarding the census tracts analyzed in the workshop, we chose two located in the neighborhoods of St. Martí de Provençals and La Verneda i la Pau. Global liveability ratios are above the average value at the district level and also Barcelona.

The conditions of the public space in the case of Sant Martí de Provençals are better than those of La Verneda and La Pau. Despite its proximity to the Gran Via de les Corts Catalanes, these are very peaceful areas for vehicular traffic with good environmental conditions, in terms of air quality and noise. The least favored variable is urban diversity, due to low economic activity in both areas.

Figure 35: Comparison of SC 10120 and 10140

4.4 District Sarrià-Sant Gervasi Workshop

DATE: 18 February 2020

PLACE: Civic Centre La Vil·la Florida

POVERTY PERCEPTION:

In the case of the session at Vil·la Florida Civic Center, the dynamic was different from previous sessions, the discussion was focused on the methodology applied and the contrast of the data. Attendees agreed that the vulnerability in the neighborhoods of Sant Gervasi de La Bonanova (25) and Sant Gervasi Galvany (26) was characterized by elderly people living alone. Most of this profile is usually the case of women who often suffer from lack of care from relatives.

The census tracts chosen by these two districts are located in areas with above-average over-aging in Barcelona. These are high-income neighborhoods, and despite the number of people applying for social services, the value of financial aid per person is one of the highest compared to the other cases analyzed.

This indicates that in the case of neighborhoods where housing prices are high and in general the cost of living is significantly higher than in the rest of the city, the population that shows signs of vulnerability and asks for help has already moved into a more serious situation.

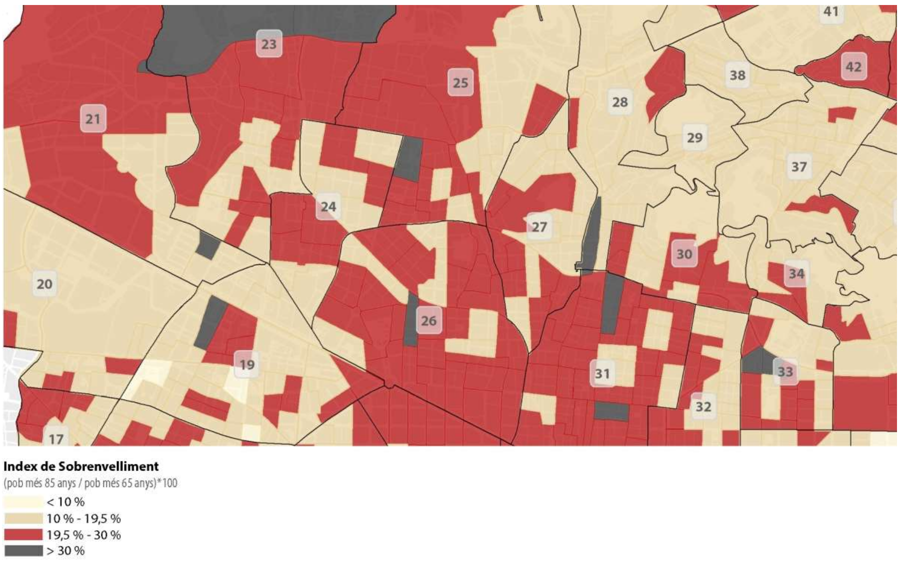
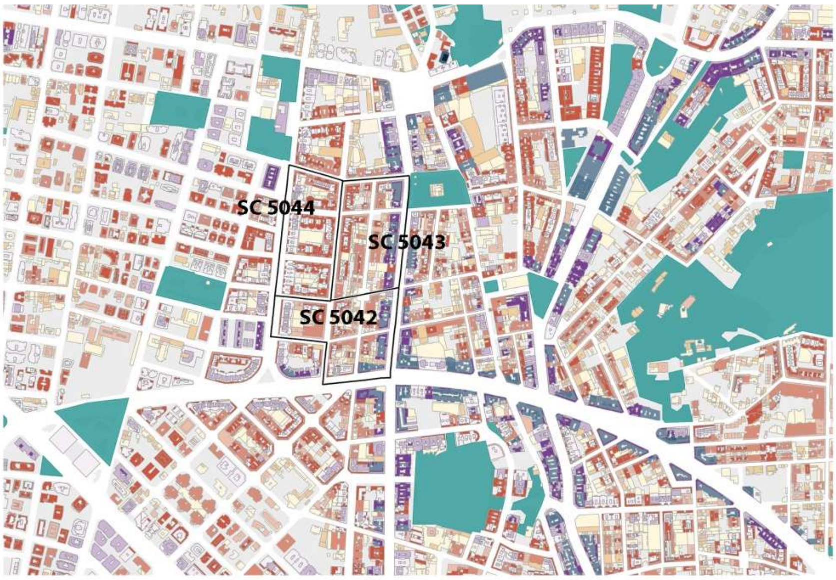


Figure 36: District Sarrià – Sant Gervasi



	Barcelona	Barri SG Bonanova	SC 5042	SC 5043	SC 5044
Població	1.642.499	25.919	1.311	1.466	1.295
% Població estrangera	20,2	17,1	10,3	9,6	5,6
Index Sobrenvelliment	19,5	21,3	24,4	16,2	29,4
% Pob amb dependència SS	78,1	92,8	93,9	92,2	97,8
% Pob amb ajudes econòmiques SS	20,1	6,24	7,1	9,1	2,2
Ajudes SS (euros/ persona)	900	871	895	2.176	302
Renda per llar (€/m2)	100,00	184,6	185	185	185
Preu lloguer (€/m2)	12,90	15,3	15,3	15,3	15,3
Preu venda (€/m2)	4.344	5.424	4.676	4.676	4.676
Tamany habitatge (m2)	97,78	155,2	121,6	115,6	122,2
% Persones a l'atur	7,39	2,1	2,1	2,1	2,1
% Persones amb ocupació	96,00	97,9	97,9	97,9	97,9
% Edificis abans 1970	63,77	65,2	67,9	51,2	97,1
Distància mig a centre de salut (m)	216,91	163,2	141,4	272,1	191,2

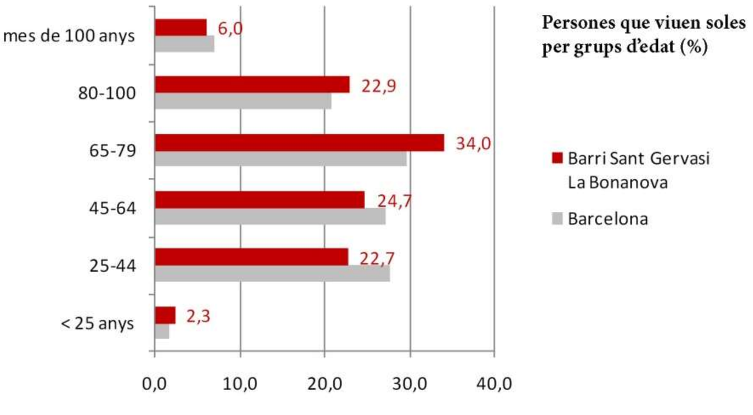


Figure 37: SC 5042/5043/5044 (Neighborhood of Sant Gervasi La Bonanova)

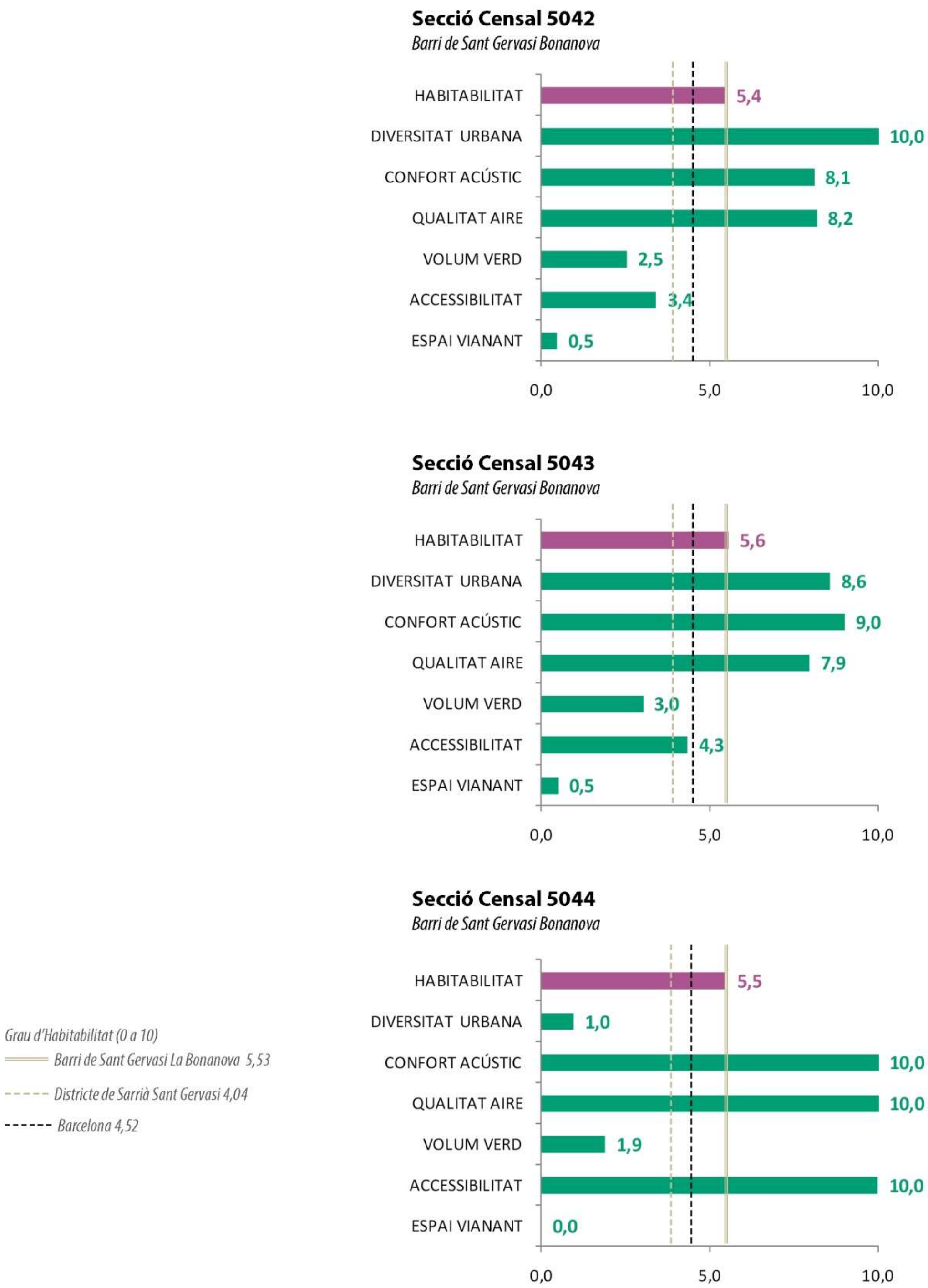
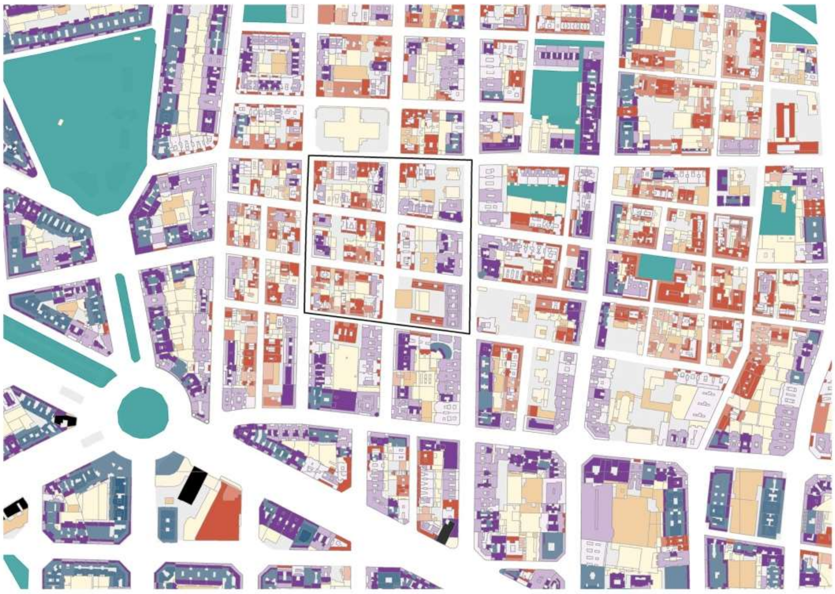


Figure 38: Comparison of SC 5042, 5043 and 5044

Figure 39: Photographs of SC 5042, 5043 and 5044. Source: Cynthia Echave.



	Barcelona	Barri SG Galvany	SC 5064
Població	1.642.499	47.588	1.588
% Població estrangera	20,2	18,7	7,8
Index Sobrenvelliment	19,5	22,4	29,5
% Pob amb dependència SS	78,1	95,3	92,7
% Pob amb ajudes econòmiques SS	20,1	4,65	6,4
Ajudes SS (euros/ persona)	900	742	1.260
Renda per llar (€/m2)	100,00	192,1	192
Preu lloguer (€/m2)	12,90	14,6	14,6
Preu venda (€/m2)	4.344	5.524	4.863
Tamany habitatge (m2)	97,78	171,1	136,6
% Persones a l'atur	7,39	2,2	2,2
% Persones amb ocupació	96,00	97,8	97,8
% Edificis abans 1970	63,77	74,5	72,8
Distància mig a centre de salut (m)	216,91	102,1	61,7

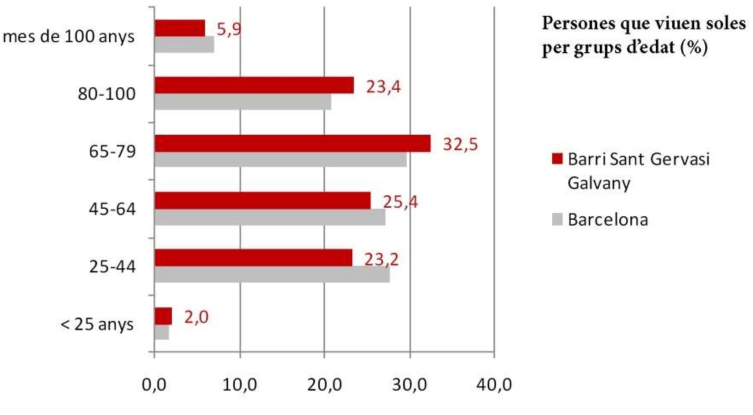


Figure 40: SC 5064 (Neighborhood of Sant Gervasi Galvany)

Figure 41: Photographs of SC 5064. Source: Cynthia Echave.

With regard to the conditions of the public space and the degree of liveability, the analyzed case in Sant Gervasi Galvany district presents good levels of accessibility and adequate compliance with sound levels and air quality. On the other hand, the level of urban diversity is very low and the space for pedestrians is limited due to the characteristics of the urban fabric and morphology. The streets are narrower, and the sidewalks are accessible but do not exceed 50% of the road space.

The same as in the analysis area of Sant Gervasi La Bonanova, the buildings are very old, 72.8% were built before 1970. However, it is an area with rents with average prices of 14.6 € / m2. Given its proximity to Av. Diagonal prices are higher than the city average.

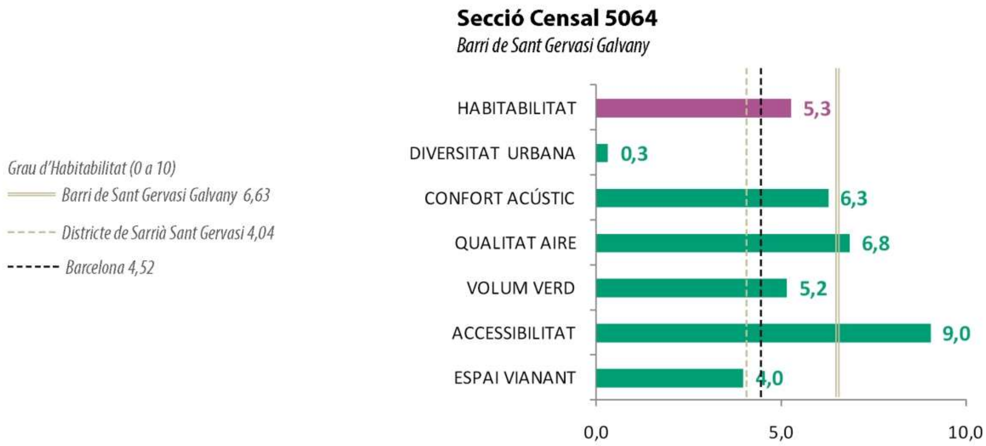


Figure 42: SC 5064

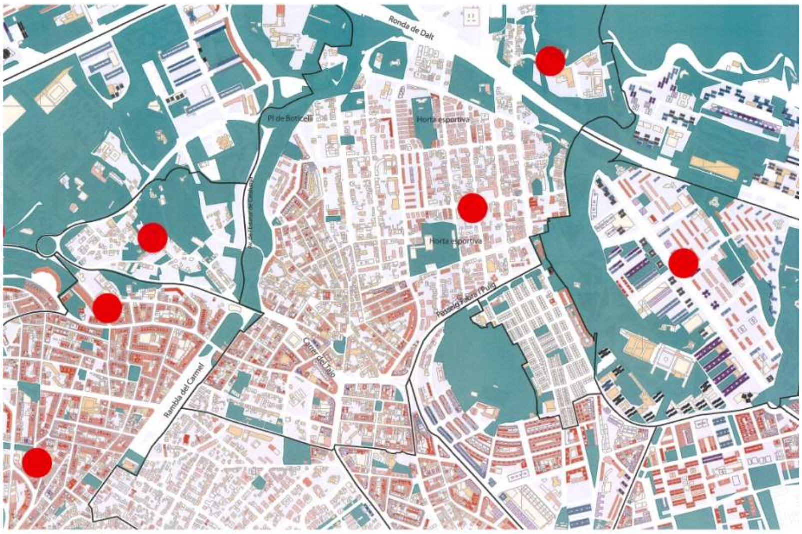
4.5 District d'Horta-Guinardó Workshop

DATE: 25 February 2020

PLACE: Cívic Centre Matas i Ramis

POVERTY PERCEPTION:

The fourth workshop was held at the Matas i Ramis Civic Centre. The session was characterized by the intervention of representatives of the social services of the Horta District and representatives of associations to help the elderly and neighborhood associations. In this case, the exercise of identifying vulnerable areas within the Barrio d'Horta was mostly consistent with the available statistical data. Attendees agreed that the neighborhood had experienced a deterioration in the atmosphere that characterized it. The growing influx of immigrants from other parts of the city as a result of gentrification was highlighted. They also referred to the replacement of traditional businesses and in the strict sense of vulnerability stressed the need for better synergy with nursing homes.



Taller Districte d'Horta

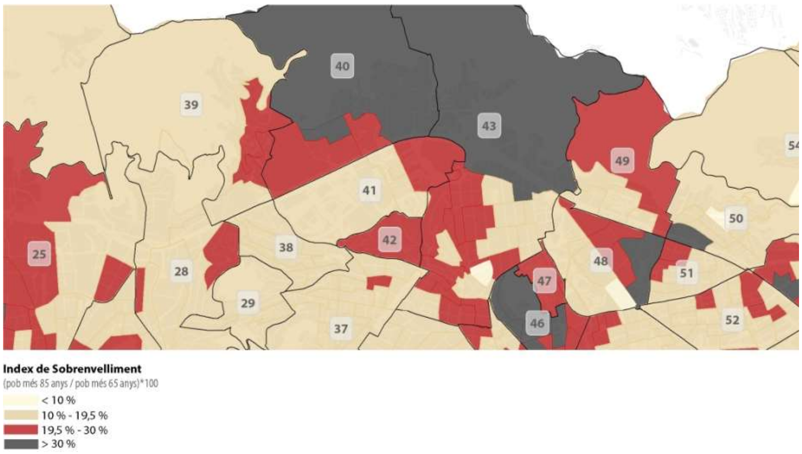
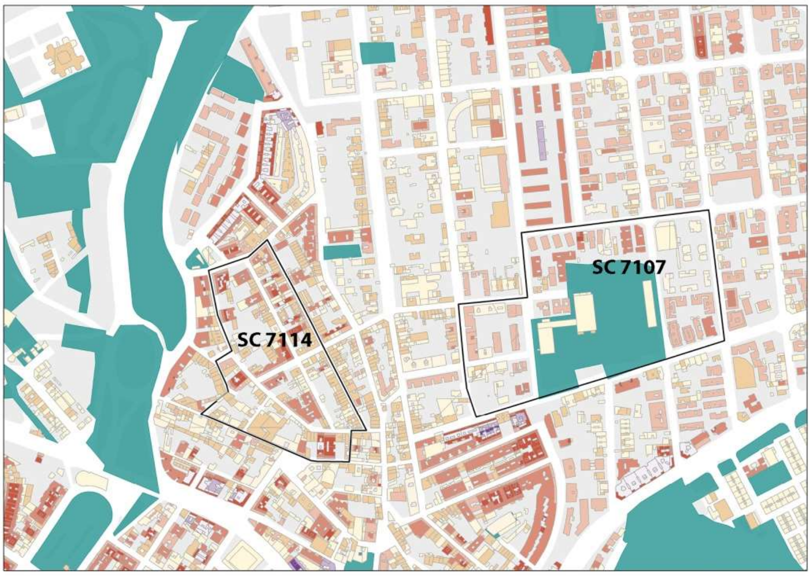


Figure 43: District of Horta - Guinardó

SC 7107 - SC 7114 Barri d'Horta



	Barcelona	Barri Horta	SC 7107	SC 7114
Població	1.642.499	27.340	1.302	1.474
% Població estrangera	20,2	13,8	4,4	3,6
Index Sobrenvelliment	19,5	19,6	12,3	18,4
% Pob amb dependència SS	78,1	82,5	85,0	89,6
% Pob amb ajudes econòmiques SS	20,1	16,72	14,0	12,3
Ajudes SS (euros/ persona)	900	1.332	2.790	1.110
Renda per llar (€/m2)	100	79,8	80	80
Preu lloguer (€/m2)	12,90	11,7	11,7	11,7
Preu venda (€/m2)	4.344	2.945	2.216	2.216
Tamany habitatge (m2)	97,78	86,4	84,9	89,0
% Persones a l'atur	7,39	4,9	5,0	5,0
% Persones amb ocupació	96,00	95,1	95,0	95,0
% Edificis abans 1970	63,77	61,5	35,2	46,4
Distància mig a centre de salut (m)	216,91	166,7	134,6	167,2

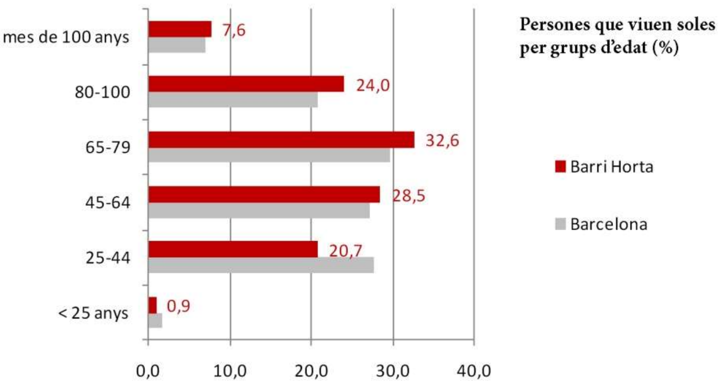


Figure 44: SC 7107 and 7114

The census tracts chosen for analysis in the workshop are characterized by good conditions in terms of sound levels and air quality. Its location is far from emissions from major sources of pollution from vehicular traffic. The weakest variables in these cases are again urban diversity, these are areas with a strong residential character, and activity is concentrated around public facilities in the area. Green spaces are not perceived from the street, it is very well linked to the private gardens of the surrounding houses.

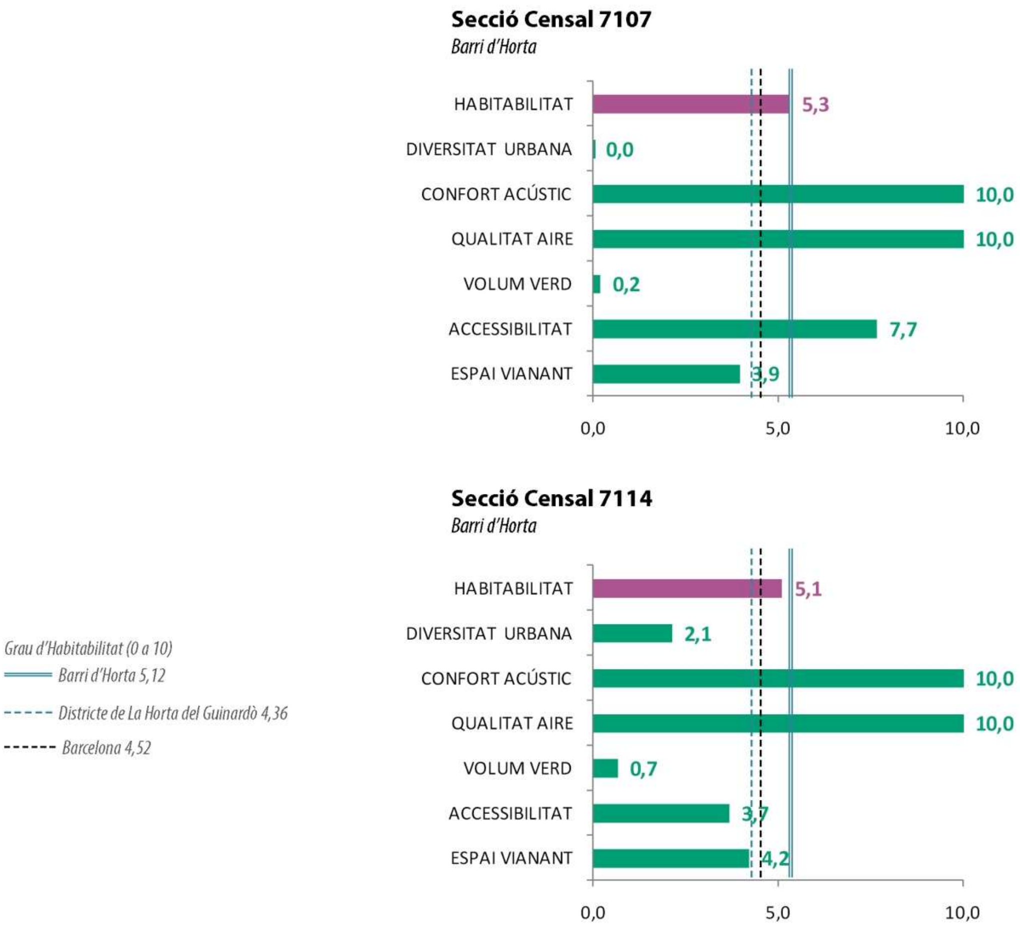


Figure 45: Comparison of SC 7107 and 7114

5. Conclusions

This project has produced a granular quantitative analysis of the vulnerability. Two scientific contributions stand out:

- On one hand, our study has identified the impact of different socioeconomic factors that influence vulnerability. Monitoring and controlling these conditions can be a useful way to prevent vulnerability. In fact, many of them are already actively managed by public policies (progressive tax system; minimum income; unemployment benefits, etc.). Others, on the other hand, are largely non-existent (rents and housing price controls, immigration, etc.), but are part of the current political debate. Finally, a third group of indicators does not even appear in the public debate, such as the predictive role of citizen participation or shop type landscape. It may be helpful to monitor the health of neighborhood businesses, as it may reveal an increase in vulnerability before this increase materializes in requests for help from Social Services.
- On the other hand, our models can be used to predict the evolution of vulnerability, and therefore be a tool for anticipating problems. One possible use would be, for example, to use the model to detect areas of the city where, due to their characteristics, there should be a high incidence of vulnerability, but not observed yet in the Social Services records. These areas could contain a significant vulnerable population that has not yet taken the "step" of contacting Social Services, although they would need to. This problem is a real fact, as we witnessed in the comments presented in the framework of the participatory workshops.

Finally, in the execution of the project we were also able to observe possibilities for improving this research. Although the scale of analysis, at the level of census tract, has been very satisfactory to be able to give a vision of proximity, we have observed that there can be important variations within a census tract: in the District of Les Corts, for example, the urban plot is a mixture of old buildings (prior to the 1960s), with relatively high vulnerability populations, and new buildings with higher-income populations. In this sense, a treatment at the cadastral reference level would be more accurate, as it would allow the use of changes in the building to predict vulnerability and other phenomena such as gentrification. This level of analysis would be much richer, but could create data privacy issues, and would require deeper collaboration by different administrations (e.g. giving access to the anonymized Census, to property records and rents, and tax data).

We are confident that in the future more studies like ours will be able to use the wealth of socioeconomic data to identify the root causes of vulnerability and suggest measures to correct them.

Appendices

A.1: Description of variables used

The aim of this project is to uncover the relationship between socio-economic micro-dynamics and vulnerability in the city of Barcelona, by constructing econometric models at a level of geographical detail as small as possible. A vital part of the project is the understanding of the current situation of administrative-territorial division of Barcelona. The city is divided into 10 districts, and since 2006 each district is divided into a different number of neighborhoods, totaling 73 neighborhoods. Each neighborhood is in turn divided into several census tracts (abbreviated SC), which in its current conception (from 2013 onwards) identifies a total of 1068. The SCs are the smallest administrative division with an electoral role and are regulated by law so that the population of each covers between 1,000 and 2,500 inhabitants.

<u>Dependent variables (response)</u>

The most relevant and most innovative source of data for research in urban economics was the Social Services Department of Barcelona City Council. For each year between 2013 and 2019, the anonymized list of people who received some kind of help from the City Council was obtained. For each individual, we knew:

- An anonymous identifier
- Age - or year of birth
- Gender - male or female
- The SC where s/he lives
- The type of aid received: TAD (telecare), DEP (dependency), SAD (home service), MEN (social dining room) and ECO (aid in €, as well as the total received in the corresponding calendar year)

Additionally, different individuals living in the same house could be identified, including individuals in the house who are not registered. From this original data, the data were processed to obtain two types of variables to study. For each possible micro-unit (combination of gender, year of birth and SC, a total of 2 x 96 x 1068, approximately 200,000 observations per year, between 2013 and 2019), four variables of interest are calculated, of two types:

1. The number of dependent people, those who receive TAD, DEP, SAD or MEN help. It is common for an individual to receive multiple of these concepts at once, for example TAD, DEP and SAD.
2. The number of people receiving financial aid (including those for whom a house member receives it), the total amount received in €, and the percentage of those who receive it and who did not receive it in the previous year.

Summarizing, we have four response variables, with panel structure, lasting 7 years (between 2013 and 2019) and cross section of 205,056 different levels.

<u>Independent variables</u>

We combined data from three main sources: the statistics department of Barcelona City Council (based on census data), the open data platform of Barcelona City Council, and the Spanish Ministry of Finance.

Figure A.1 describes the type of data obtained, as well as its level of granularity.

Variable	Source	Geographic level	Records by gender	Records by year of birth	Period
Population	Estadística Ajuntament de Barcelona https://www.bcn.cat/estadistica/catala/dades/tpob/pad/padro/a2013/edat/edata13.htm	SC	Yes	Yes	2013-2018
Immigration	Estadística Ajuntament de Barcelona https://www.bcn.cat/estadistica/catala/dades/tpob/pad/padro/a2013/nacio/nacio13.htm	SC	Yes	No	2013-2018
Household	Estadística Ajuntament de Barcelona	SC	No	No	2013-2018

	https://www.bcn.cat/estadistica/catala/dades/tpob/pad/padro/a2013/llars/ocu04.htm				
Educational level	Estadística Ajuntament de Barcelona https://www.bcn.cat/estadistica/catala/dades/tpob/pad/padro/a2013/nivi/nivi13.htm	SC	Yes	No	2013-2018
Duration in the same address	Estadística Ajuntament de Barcelona https://www.bcn.cat/estadistica/catala/dades/tpob/pad/padro/a2103/anys_alta_padro/ant13.htm	SC	Yes	No	2013-2018
Participation in elections	Estadística Ajuntament de Barcelona https://www.bcn.cat/estadistica/catala/dades/inf/ele/ele34/A19.htm	SC	No	No	2015, 2016, 2017, 2019
Age of buildings	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/data/ca/dataset/est-cadastre-	SC	No	No	2018-2019

	edificacions-edat-mitjana				
Housing surface	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/data/ca/dataset/est-cadastre-habitatges-superficie-mitjana	SC	No	No	2017-2019
Construction density	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/data/ca/dataset/est-cadastre-edificacions-nombre-locals	SC	No	No	2018-2019
Sale and rental prices	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/data/ca/dataset/est-mercat-immobiliari-compravenda-preu-total	Neiborhood	No	No	2013-2018
Unemployment	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/d	Neiborhood	Yes	No	2013-2019

	ata/ca/dataset/est-atur-sexe				
Commerce	Open Data Ajuntament de Barcelona https://opendata-ajuntament.barcelona.cat/data/es/dataset/cens-activitats-comercials	SC	No	No	2014, 2016, 2019
Income	Agencia Tributaria https://www.agenciatributaria.es/AEAT/Contenidos_Comunes/La_Agencia_Tributaria/Estadisticas/Publicaciones/sites/irpfCodPostal/2013/jrubik2f09e7b141531bf6a19f2f81d4fa1b08282fd32f.html	Postal code	No	No	2013-2017

Figure A.1: Independent variables used in the models.

As can be seen, not all data for the period 2013-2019 exist, years in which we have data on the response variables. In order to avoid losing observations, we replace the missing data with the data of the same indicator in the previous year in which it is available. For example, for a given zip code, we assume that the revenue for 2018 and 2019 is the same as for 2017, the nearest year available.

A.2: Description of econometric models

Incidence of aid - Poisson models

To predict the incidence of people with dependency benefits (Y_{DEP}) and financial aid (Y_{ECO}), we use the Poisson model. This model has the advantage of being discrete (the variable must be an integer) and additive, which is an important requirement in our context. Specifically, if $Y_{DEP,a}$ and $Y_{DEP,b}$ represent the dependents of groups a and b (e.g. women aged 30–39 years and women aged 40–49 years), and follow Poisson distributions with intensity λ_a i λ_b, then those dependent on the union of these two $Y_{DEP,a+b}$ groups follows a Poisson distribution with intensity $\lambda_a + \lambda_b$.

The specification chosen for our statistical model is as follows. For each census tract i, each age j and each gender k, with population $P_{i,j,k}$ and characteristics $X_{i,j,k}$, we asume that dependents $Y_{DEP,i,j,k}$ follow a Poisson distribution with intensity $\lambda_{i,j,k}$ on $\log(\lambda_{i,j,k}) = \log(P_{i,j,k}) + \beta X_{i,j,k}$. It can be seen that this specification "forces" the intensity $\lambda_{i,j,k}$ to be proportional to the number of residents $P_{i,j,k}$, in order to ensure that, when the model is added by age groups, genders or census tracts, maintain their meaning and validity. We use the same model for $Y_{ECO,i,j,k}$. We estimate this model with the glm package of R. The results are shown in Figures A.2 and A.3.

Incidence of dependency benefits

	Dependent variable: dependents
year_lin	0.277*** (0.005)
year_sq	-0.009*** (0.001)
residents_spanish	-0.109 (0.075)
log(households_in_cs)	0.645*** (0.022)
log(residents_by_gender_in_cs)	-0.484*** (0.021)
residents_long_tenure	-0.488*** (0.073)
election_participation	-0.072** (0.034)
log(income_gross)	-0.106*** (0.022)
log(unemployed)	0.098*** (0.025)
residents_education_low	1.337*** (0.033)
log(euros_per_m2)	0.013 (0.017)
log(rental_euros_per_month)	-0.034 (0.055)
log(number_of_buildings)	-0.030*** (0.005)
new_buildings	-0.134*** (0.013)
dense_buildings	-0.153*** (0.017)
housing_age_average	-0.002*** (0.0002)
log(shops)	-0.010*** (0.003)
shops_food	0.002 (0.015)
log(Distance_to_health_center)	-0.006*** (0.001)
MetroStopsLines100m	-0.004 (0.003)
HV_BusLines	-0.002*** (0.001)
Fixed effects	barri gender x age
Offset	log(residents)
Observations	1,191,561
Log Likelihood	-530,382
Akaike Inf. Crit.	1,061,324
Note:	*p<0.1; **p<0.05; ***p<0.01

Figura A.2: Model de Poisson per Y_{DEP}

Incidence of financial aid

	Dependent variable: receivers
year_lin	0.501*** (0.008)
year_sq	-0.032*** (0.002)
residents_spanish	-1.541*** (0.096)
log(households_in_cs)	1.798*** (0.029)
log(residents_by_gender_in_cs)	-1.687*** (0.027)
residents_long_tenure	-0.375*** (0.099)
election_participation	-1.665*** (0.042)
log(income_gross)	-0.416*** (0.046)
log(unemployed)	0.050* (0.026)
residents_education_low	3.934*** (0.049)
log(euros_per_m2)	0.051** (0.022)
log(rental_euros_per_month)	0.195** (0.076)
log(number_of_buildings)	0.012 (0.008)
new_buildings	-0.149*** (0.020)
dense_buildings	-0.306*** (0.026)
housing_age_average	-0.003*** (0.0003)
log(shops)	-0.045*** (0.005)
shops_food	0.443*** (0.020)
log(Distance_to_health_center)	-0.010*** (0.001)
MetroStopsLines100m	-0.000028
HV_BusLines	0.007*** (0.001)
Fixed effects	barri gender x age
Offset	log(residents)
Observations	1,191,561
Log Likelihood	-425,225
Akaike Inf. Crit.	851,009
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure A.3: Model de Poisson per Y_{ECO}

Amount of financial aid - lognormal model

Financial aid has different types, from support for children under 16 to a public transport subsidy. The allocation of these payments depends on the unique conditions of each case, and takes into account the situation of the household, including that of people not included in the list of social services, as they do not receive a direct payment. To alleviate this limitation, we add all the payments made to a household, and attribute the payment made to each person as the proportion of the total aid divided by the household members.

The distribution of this amount, paid to each person, has high variability. We assume that it follows a lognormal distribution (since it must be positive). Specifically the amount received by an individual p in a census tract i, age j and gender k, with characteristics $X_{i,j,k}$, is defined as $Q_{i,j,k,p} = \exp(\beta X_{i,j,k} + \sigma \varepsilon_{i,j,k,p})$ on $\varepsilon_{i,j,k,p}$ it is a normal residual. To estimate the model, we simply take the logarithm of this expression:

$$\log(Q_{i,j,k,p}) = \beta X_{i,j,k} + \sigma \varepsilon_{i,j,k,p}$$

We estimate this model with the lm package of R. The results are shown in Figure A.4.

Amount of financial aid

	Dependent variable:
	log(1 + helped_sum_euros)
year_lin	0.597*** (0.009)
year_sq	-0.084*** (0.002)
residents_spanish	-0.560*** (0.131)
log(households_in_cs)	0.086** (0.041)
log(residents_by_gender_in_cs)	0.025 (0.039)
residents_long_tenure	1.080*** (0.134)
election_participation	-1.328*** (0.064)
log(income_gross)	-0.014 (0.057)
log(unemployed)	0.005 (0.035)
residents_education_low	-0.551*** (0.064)
log(euros_per_m2)	-0.072** (0.030)
log(rental_euros_per_month)	0.774*** (0.100)
log(number_of_buildings)	0.054*** (0.011)
new_buildings	0.104*** (0.027)
dense_buildings	0.052 (0.034)
housing_age_average	-0.001*** (0.0004)
log(shops)	-0.000066
shops_food	0.209*** (0.027)
log(Distance_to_health_center)	-0.016*** (0.002)
MetroStopsLines100m	-0.0002 (0.006)
HV_BusLines	-0.007*** (0.001)
Fixed effects	barri
	gender x age
Observations	137,804
R2	0.319
Adjusted R2	0.318
Residual Std. Error	1.094 (df = 137525)
F Statistic	231.811*** (df = 278; 137525)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure A.4: Lognormal model for the amount received in financial aid, Q

Continuity in financial aid - binomial model

By having data during the period 2013-2019, we can track the duration of financial aid that a person receives. The aim of the support given by Social Services is precisely to give a "push" so that the person gains independence and quickly returns to an independent life, without the need for help.

In this sense, an important indicator to measure the chronicity of vulnerability is to follow the percentage of people who receive financial aid in year t-1, who also receive it in year t. In our incidence model (Poisson model), independence is assumed between individuals and years, and is therefore not suitable for capturing the duration of an individual within the system. Therefore, it is necessary to develop a new model. Therefore, we develop a binomial model, in which the probability that an individual receiving financial aid in year t-1 to a census tract and, age j and gender k, continues to the system in year t with probability with logistic structure,, $\alpha_{i,j,k} = \exp(\beta X_{i,j,k})/[1 + \exp(\beta X_{i,j,k})]$. Specifically, for an individual p (associated with a given year t-1), he continues to receive financial aid the following year if $C_{i,j,k,p} = 1$, where C=1 with probability $\alpha_{i,j,k}$ i C=0 amb probabilitat 1- $\alpha_{i,j,k}$. We estimate this model with the glm package of R. The results are shown in Figure A.5.

Continuity of financial aid

	Dependent variable:
	not_new/receivers_prev_year
year_lin	0.759*** (0.039)
year_sq	-0.100*** (0.006)
residents_spanish	-0.14442
log(households_in_cs)	0.504*** (0.087)
log(residents_by_gender_in_cs)	-0.293*** (0.082)
residents_long_tenure	1.130*** (0.297)
election_participation	0.142 (0.138)
log(income_gross)	0.087 (0.125)
log(unemployed)	-0.01152
residents_education_low	1.124*** (0.142)
log(euros_per_m2)	0.444*** (0.074)
log(rental_euros_per_month)	-0.006 (0.227)
log(number_of_buildings)	0.045* (0.024)
new_buildings	-0.208*** (0.061)
dense_buildings	0.118 (0.076)
housing_age_average	-0.003*** (0.001)
log(shops)	-0.035*** (0.013)
shops_food	0.220*** (0.065)
log(Distance_to_health_center)	-0.015*** (0.004)
MetroStopsLines100m	0.020 (0.012)
HV_BusLines	-0.017*** (0.003)
Fixed effects	barri
	gender x age
Offset	log(residents)
Observations	103,800
Log Likelihood	-68,782
Akaike Inf. Crit.	138,119
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure A.5: Binomial model for continuation in the system, C

A.3: Description of machine learning models

In parallel with statistical models based on econometrics, we have also developed models based on Machine Learning techniques. Specifically we use Random Forest to create predictive models for variables $Y_{DEP,i,j,k}$, $Y_{ECO,i,j,k}$, $Q_{i,j,k,}$i $C_{i,j,k}$.

We use the R ranger package to create the predictive models, which use all available variables, and create 500 trees per model.

A.4: Models validation

For each of the model used, we train the model with data from 2013 to 2018. With these data sets (training data), we can measure the quality of the prediction. When making predictions at a very low level (CS, age, and gender), the prediction is often very noisy, but we can track the error in% at higher levels. Specifically, we measure the percentage error

ERROR = |REAL DATA – PREDICTION| / PREDICTION

by CS (without differentiating age or gender), by age (without differentiating CS or gender) and by gender (without differentiating CS or age). For the different types of aggregation, we calculate the MedAPE (median absolute percentage error) which gives us an indication of the accuracy of the predictive model (as the prediction is made at a very granular level, using the MAPE is inappropriate because atypical predictions, with very high error (e.g., when the prediction is close to zero and the reality is 1 or more), and they do not reflect the overall prediction quality of the model. We generally see that the predictions within the training data are very satisfactory, for all models.

To validate the models, we test the predictions for 2019 by comparing them to the actual 2019 data, which were not available to the model at its creation. In this sense, it is a "bet" that can be contrasted with reality. We calculate errors in the same way in the validation data as in the training data. As expected, MedAPE is higher in the validation data than in the training data.

Figure A.6 shows the MedAPE of the different models, for the different variables, where we calculate the error at the census tract level (aggregated by age and gender the models are much more accurate). We see that the accuracy of the quantitative models (econometric and random forest) is quite good with the training data, although the actual data intrinsically has a lot of variability. With the validation data, the accuracy deteriorates, mainly because, not having the year 2019 contemplated in the training data, all predictions are skewed for 2019.

It is possible to improve a lot the performance of the econometric model by using the year data as a factor, rather than as a numerical variable. In order to use this technique, it is also necessary to include some 2019 observations in the model. By adding 1,000 2019 observations to the training data (out of a total of 20,000 in the model with fewer observations), the results improve a lot. Figure A.7 shows the same result. With this added help, models eliminate their bias, and improve prediction quality significantly.

	Tranining data 2013-2018		Validation data 2019	
Variable	Econometric model	Random Forest	Econometric model	Random Forest
Y_DEP	10.5%	1.4%	20.6%	12.2%
Y_ECO	37.4%	0.9%	53.6%	27.4%
Q	29.4%	23.2%	32.3%	32.3%
C	25.6%	4.4%	57.8%	48.1%

Figure A.6: Quality of the predictions of the different models, in MedAPE.

	Tranining data 2013-2018 +1000 observations of 2019		Validation data 2019 -1000 observations	
Variable	Econometric model	Random Forest	Econometric model	Random Forest
Y_DEP	11.5%	1.4%	9.1%	2.0%
Y_ECO	41.3%	0.9%	39.0%	31.0%
Q	28.9%	21.5%	34.3%	32.0%
C	26.5%	4.5%	54.4%	127.4%

Figure A.7: Quality of the predictions of the different models, in MedAPE, adding 1,000 observations of the year 2019 to the training data

A.5 Comparison of indicators by neighborhood and census tracts studied in the workshops

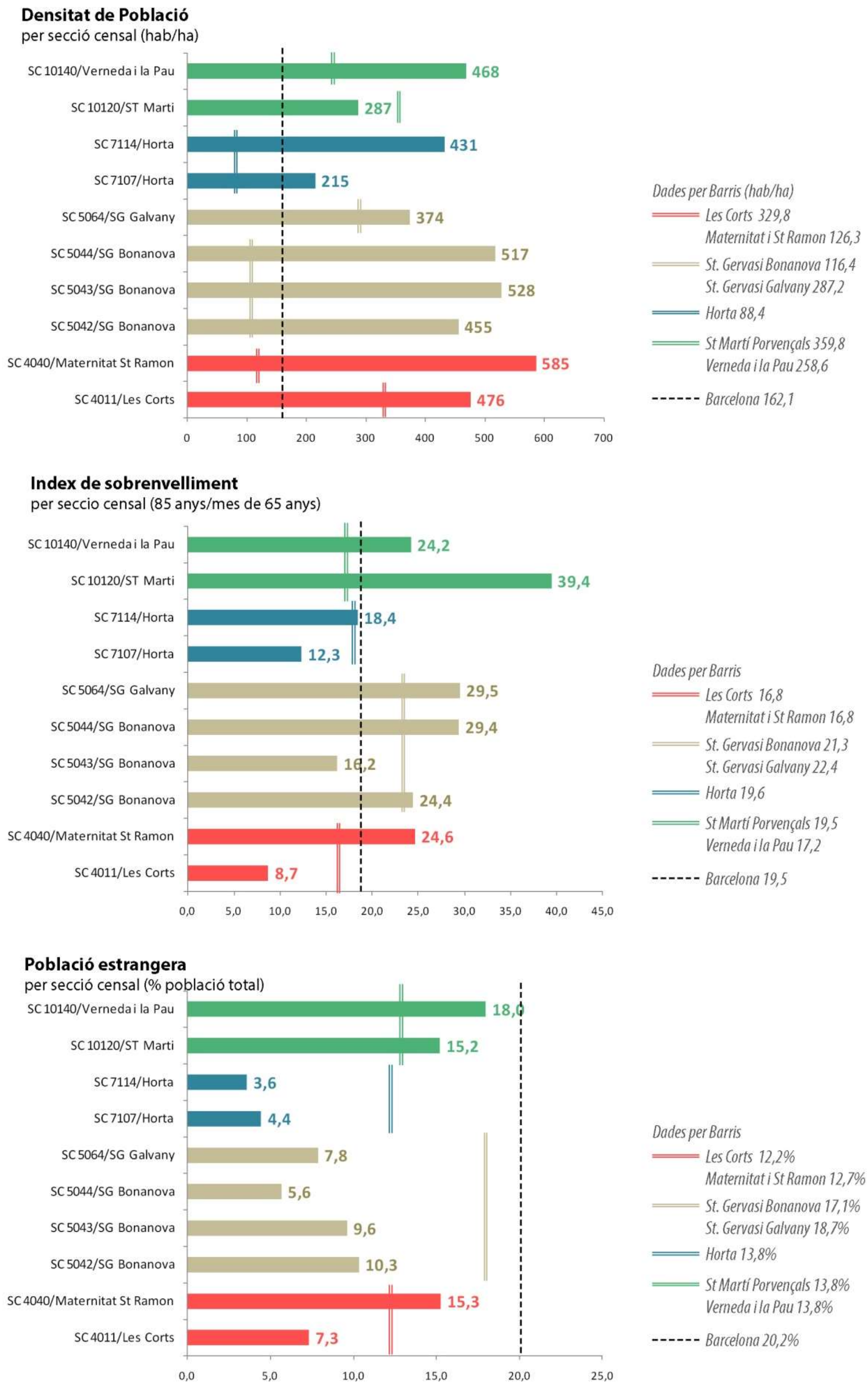


Figure A.8: Comparison of selected SC.

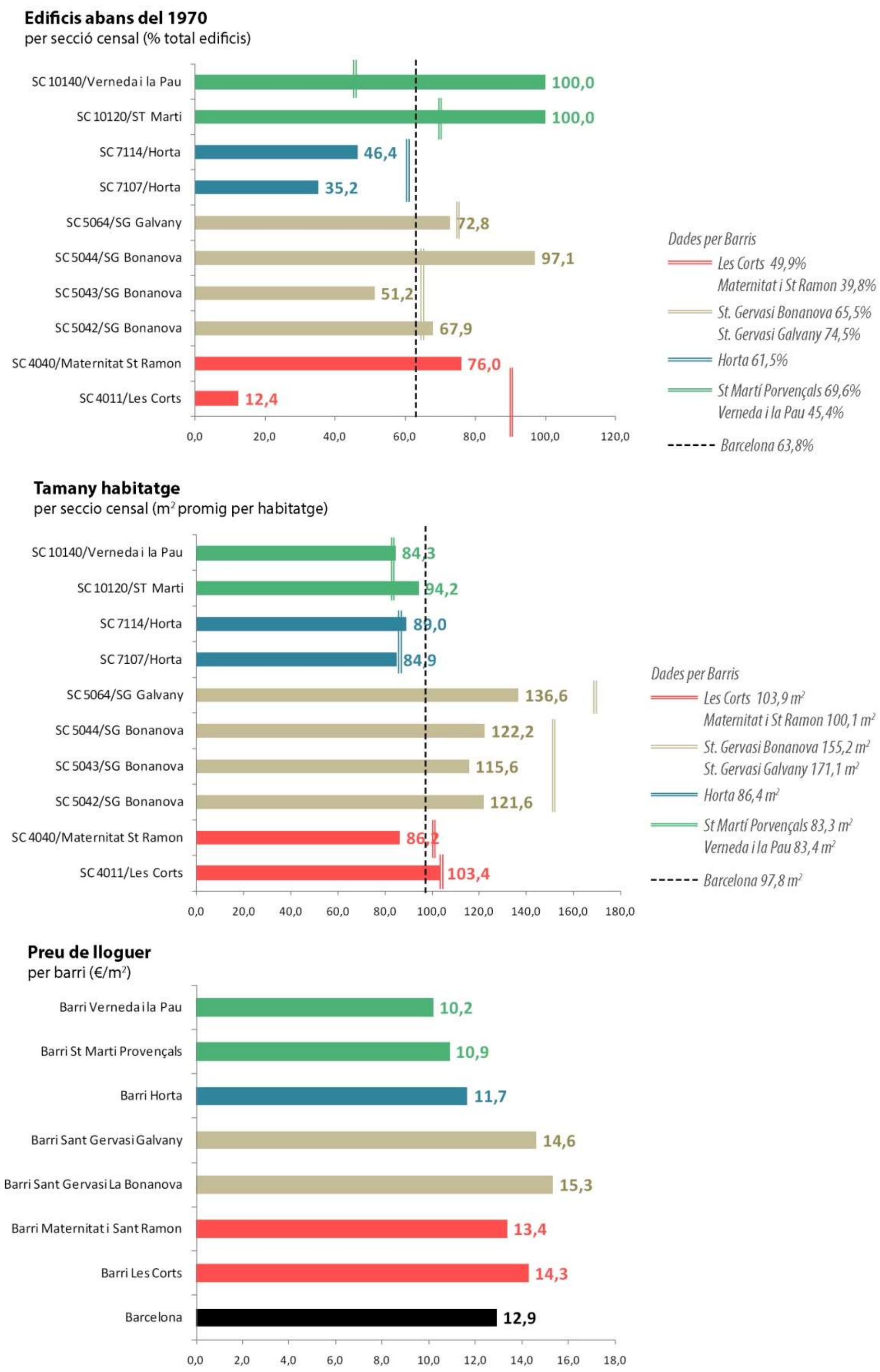


Figure A.9: Comparison of selected SC.

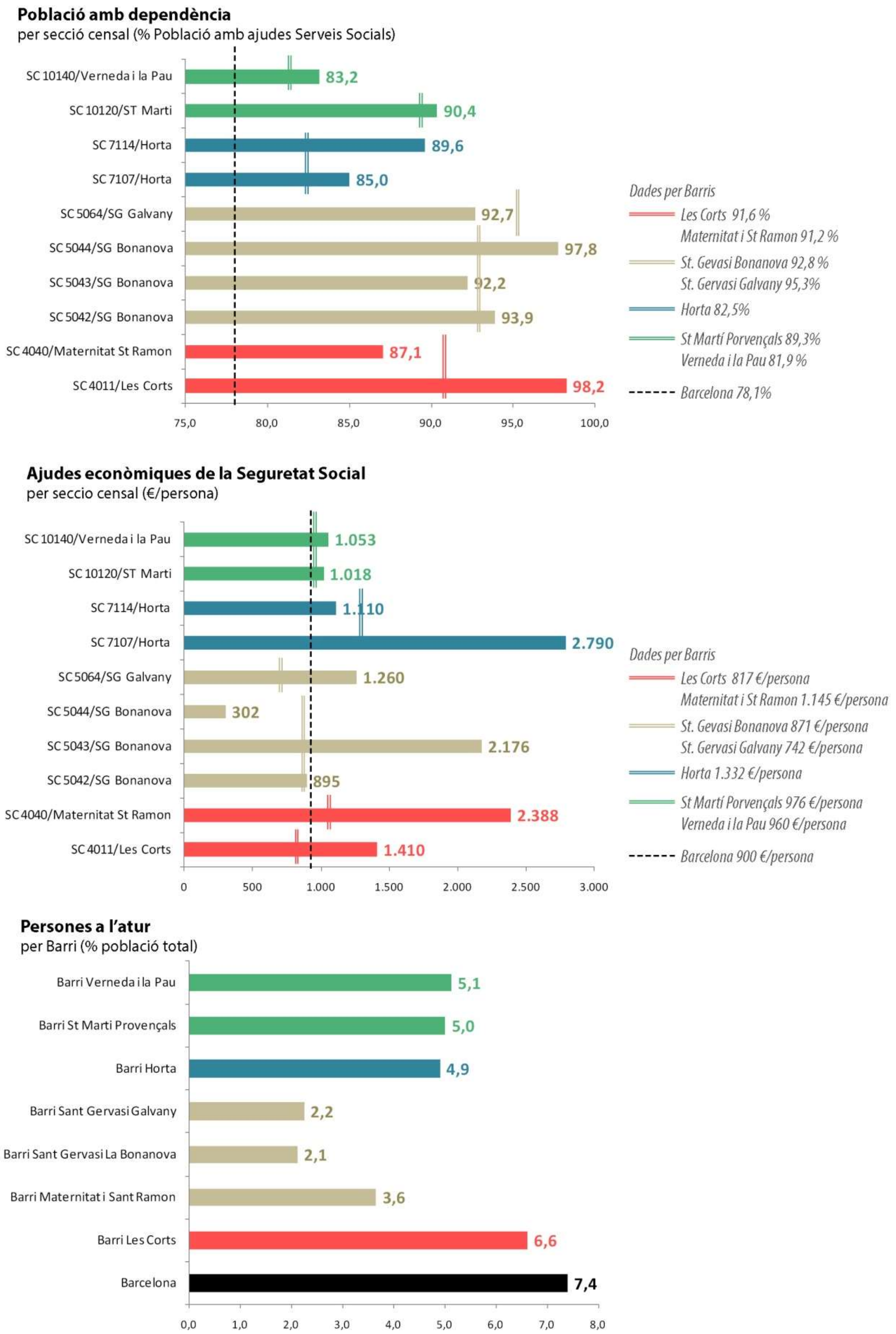


Figure A.10: Comparison of selected SC.