

Neighborhood Socioeconomic Status: Operationalization of Indicators

To measure neighborhood socioeconomic status we explored all, to our knowledge, available data sources on social, economic and contextual factors in Madrid, Spain. We looked for readily available indicators (to ease replicability), that were measured at the neighborhood or census section level (to improve granularity) and that were available for several years (to allow for further studies looking at longitudinal changes). After a literature review and data exploration we used seven indicators in four domains.

Operationalization of indicators:

1. Education:
 - a. Primary education: people with primary studies or below / people aged 25 or above
 - b. University education: people with university education or above / people aged 25 or above
2. Wealth:
 - a. Property value: average sales price of housing properties in EUR per m²
3. Occupation:
 - a. Part-time employment: workers in part-time employment / all workers
 - b. Temporary employment: workers in temporary employment / all workers
 - c. Manual occupational class: workers in manual or unqualified jobs / all workers
4. Living conditions:
 - a. Unemployment rate: individuals registered as unemployed / all people aged 16 to 64

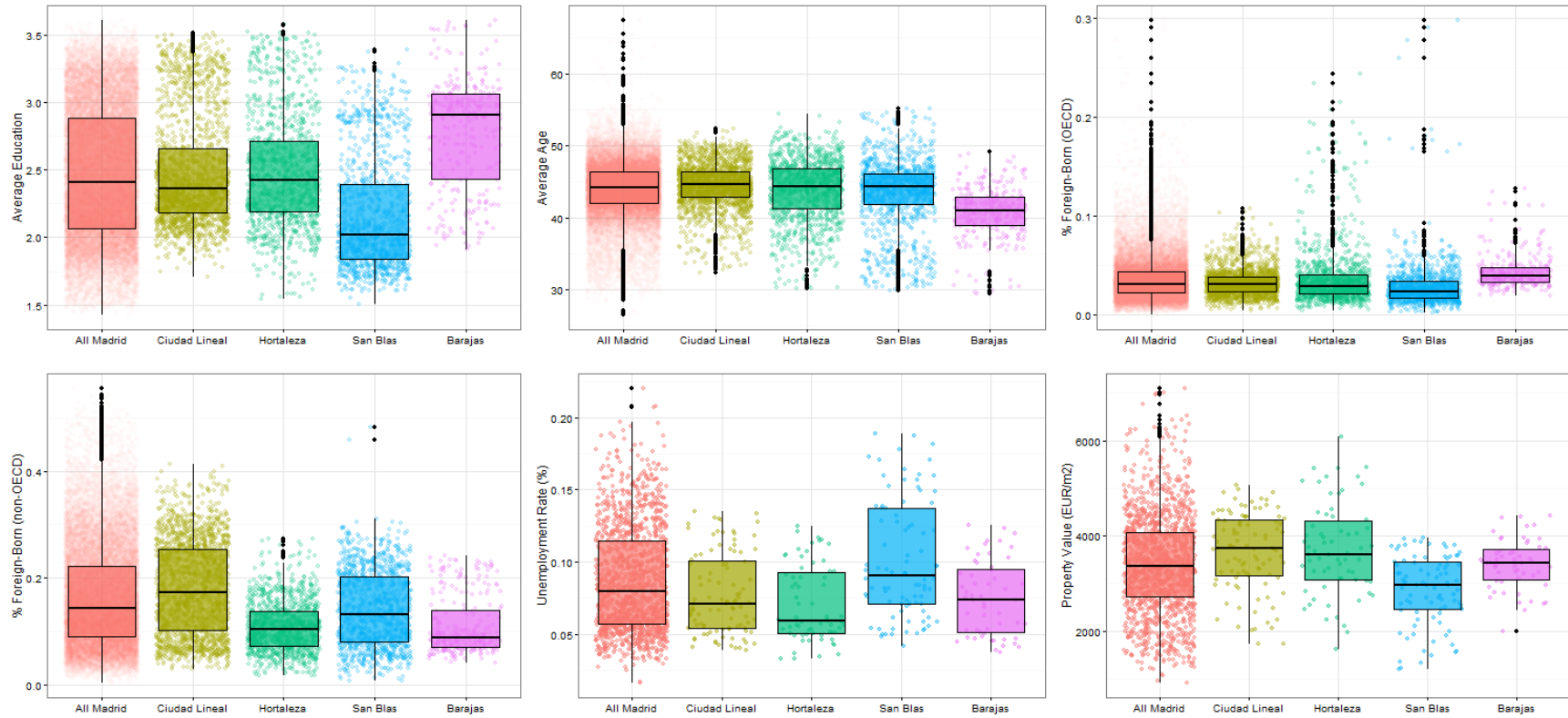
Data Sources:

1. Education: The education indicators were obtained from the Padron, a continuous universal census of the entire population used for administrative purposes. It includes data on education level which we recategorized into the four typically used levels in Spain: no formal studies, primary education, secondary education, and university education. We also obtained proportion of people above age 25 to use as the denominator.
2. Wealth: Property value was obtained from the Idealista Report, a yearly study of neighborhood-level sale prices of all housing sold through the biggest real state corporation in Spain (Idealista). All data was

downloaded from the statistics website of the City Government of Madrid. Property value data from the IDEALISTA Report contains data for all houses listed for sale in their website on the first day of each year. The report contains data at the neighborhood level (n=128 each year). To translate this to the census section level, we obtained data from the IDEALISTA API (<http://developers.idealista.com/access-request>) on April 18th 2016. We collected all housing units for sale on that day, including their price, size and geocoded location. We overlaid a census section polygon file and assigned each housing unit to a census section. With this, we constructed a measure of average property value per census section for 2016. We then used a weighted linear mixed model with property value at the census section as the dependent variable, and property value at the neighborhood level (from the IDEALISTA Report 2016 data) as a fixed and random coefficient (at the neighborhood level, with an unstructured covariance structure), and the following fixed effects for each census section: % primary education, % university education, % immigration from non-oecd countries, % people below age 25, % people above age 25, and a quadratic fixed term for each indicator. Each observation was weighted by the number of housing units on sale on each census section. We then predicted the property value in each census section in 2013 by replacing the data above with the respective data from 2014. To diagnose this imputation we correlated the predicted values for 2016 with the observed values in 2016, finding a pearson correlation coefficient of 0.93.

3. Occupation: The total number of workers, and the number of workers in part-time and temporary employment along with the occupational class were obtained from the Social Security registries. These were downloaded from the statistics website of the City Government of Madrid.
4. Living conditions: Registered unemployment was obtained from the statistics of the Employment Service (SEPE), downloaded from the statistics website of the City Government of Madrid. The denominator was, given the lack of a better measure for the active population at this geographical level, the amount of people between 16 and 64 years of age in the neighborhood, obtained from the Padron.

Appendix Figure 1. Distribution of key sociodemographic and socioeconomic variables in the four districts as compared to the entire city of Madrid



Appendix Figure 2: Comparison of the OR of Lack of Diabetes Control and the Change in Average HbA1c % in models using complete case analysis (ignoring missing data) and in models using conditional mean imputation of missing HbA1c %

