Study of Accessibility of Waste Collection Facilities and Recycling Stations to the Residents of the City of València

Ondřej Hrdlička
Barbora Horňáková
Emma Hasund Vareberg
Sean Fraser

2025/2026 Data visualisation (course 36425)

Index

Introduction

This data visualization project develops a spatial data visualization dashboard to evaluate access to urban waste containers in the city of València. The project is linked to the municipal campaign **"València, más limpia, más verde, hagámoslo bien"** (Ajuntament de València, 2023) (**"València, cleaner, greener, let's do it right"** in English), which promotes waste reduction, separation of waste at origin, correct use of recycling containers, and improved selective collection infrastructure across the city. The key goal that the campaign mentions in regards to this project is increasing selective collection points moving toward zero waste, which can be aided by the visualizations provided.

The central question of this project then becomes: **How accessible are different types of waste containers from building entrances in València?**

Access to waste collection infrastructure is a well-documented factor in recycling behaviour, with proximity consistently shown to predict whether residents will use separate collection containers. Timlett and Williams (2008) found that convenience, particularly distance to collection points, was one of the most significant barriers to household recycling, while Bernstad (2014) demonstrated that the physical accessibility of organic waste facilities directly influenced source-separation rates, concluding that container placement strategies have measurable effects on waste diversion outcomes.

To answer this question, the project preprocesses official open data from the València open data portal in QGIS, utilizing buffering and spatial overlays, and a Shiny dashboard implemented in R. The final dashboard uses Leaflet maps, reactive filters, and scientific graphs to explore accessibility patterns by container type, distance threshold, city block, and neighborhood.

The main waste categories considered in this project are:

- **Glass**
- **Organic**
- **Paper / cardboard**
- **Plastic / packaging**
- **Residual waste**

The final application is intended to support interpretation of how evenly waste-separation infrastructure is distributed across the city of València and where underserved addresses (i.e., building entrances) or areas may exist.

Methodology

Firstly, official open datasets were collected from the València open data portal. The core dataset was **“Contenidors residus sòlids” / “Contenedores residuos sólidos”**, translated as **“Solid waste containers.”** The portal describes this dataset as geographic information related to solid-waste containers, including container type, loading type, and responsible company. This dataset was extended (joined) with the **“Contenidors Vidre” / “Contenedores Vidrio sólidos”,** which has data about glass containers, in order to contain all the common residential waste types.

The second key dataset was **“Portals dels carrers” / “Portales de las calles”**, translated as **“Building entrances.”** This dataset provides administrative geographic information about building doors in the city of València and was used as the origin-point layer for waste container accessibility calculations. This was later joined with the **“Vias”**, or **“Roads”** tabular dataset, to get full addresses for the building doors.

Population information was obtained from **“Illa de cases cadastrals amb dades de població” / “Manzanas catastrales con datos de población”**, translated as **“Cadastral blocks with population data,”** which generally matches with València’s city blocks. These population values were used to estimate the population residing at each individual building entrance of the given block.

The overall workflow then consisted of:

1. Downloading and cleaning the spatial layers.
2. Matching coordinate reference systems.
3. Joining entrances to population blocks.
4. Calculating nearest-container distances by type.
5. Creating service-area buffers.
6. Aggregating accessibility measures to population blocks and neighborhoods.
7. Exporting processed layers to GeoPackage.
8. Loading the GeoPackage layers into R with the sf package and transforming them.
9. Building a Shiny dashboard with Leaflet maps and charts.

Exploratory data analysis

Exploratory data analysis was performed in both QGIS and outside of it. Most of the datasets came from ArcGIS REST endpoints, although the population dataset needed to be sourced from a CSV table. The CSV geometry was stored as WKT and required cleaning because some long polygon definitions contained line breaks that interfered with QGIS CSV loading. The ArcGIS REST endpoints were limited by server transfer limits, so large layers were downloaded using paginated ArcGIS REST requests, two thousand features at a time, by QGIS.

The door layer contained limited useful attribute information, so it was later used primarily as an origin layer: each door was treated as a representative access point for residents leaving buildings.

Data Preprocessing

Data preprocessing began with conversion and cleaning.

All layers were then standardized in QGIS. To make sense of the data, container types were translated into English with values such as Glass, Organic, Paper / cardboard, Plastic / packaging, and Residual waste.

The glass container layer and the general solid-waste container layer were merged after field refactoring, which made the two source layers share the same schema before merging.

Preprocessing of geographic information

Spatial preprocessing was performed in QGIS using a projected CRS suitable for distance calculations. Since the analysis involved meter-based distances and buffers, the working CRS was **ETRS89 / UTM zone 30N - EPSG:25830**. Final web-map outputs were exported for use in Leaflet, which works with **EPSG:4326**.

The first spatial operation was the assignment of building doors to city blocks. A strict join with the *within* operation was initially problematic because some door points lay just outside the population-block polygons, especially along south and west block edges. This was interpreted as a positional mismatch between address-point placement and cadastral-block geometry. Doors were thus assigned to the nearest population block within a limited tolerance.

Next, the number of doors assigned to each population block was calculated using an attribute-based aggregation by block ID. This was again necessary because some doors lay slightly outside the block polygon. Population was then distributed across assigned doors within each block with the simple estimative computation of:

population per door = block population / number of assigned doors in block

This is obviously an approximation, but it allows the dashboard to estimate how many residents may be affected by poor access to waste containers.

Nearest-container distances were calculated from each door to the closest container of each type and to the closest container of any type. This information can then be used with thresholds.

Finally, service-area buffers were created around containers at predetermined distances of 50, 100, 150, and 200 metres. These buffers were dissolved by container type and buffer size, producing one polygon or multipolygon per type-distance combination.

Choosing graphs for different types of data

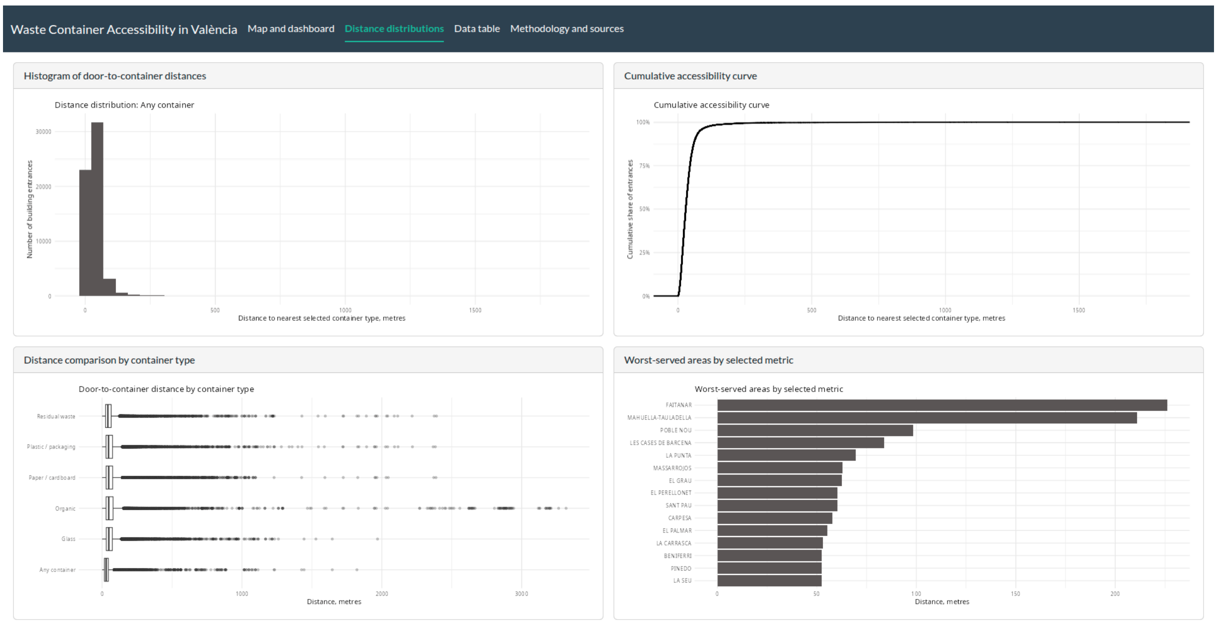
Several graph types for the dashboard were selected based on the structure of the data and the analytical question:

A histogram was chosen for door-to-container distances because the distance variable is continuous. The histogram shows whether most entrances are close to containers or whether there is a long tail of poorly served entrances.

An cumulative distribution function was chosen to show the share of entrances within a given distance. This graph is particularly useful for accessibility analysis because it directly answers questions such as **“What percentage of entrances are within 100 metres of a glass waste container?“**

A boxplot was chosen to compare distances across container types. This allows quick comparison of median distance, spread, and outliers for each of the types: glass, organic, paper, plastic, and residual-waste containers.

A **ranked bar chart** was chosen for neighborhood or population-block summaries. This graph identifies the worst-served areas by selected statistics, such as mean or median distance.



Map design and interactivity

The dashboard map was designed around multiple spatial layers exported from the preprocessing in QGIS:

- Container points.
- Building door points.
- Service-area buffer polygons.
- Population-block polygons.
- Neighborhood polygons.

The map uses a light, gray-scale basemap by default so that thematic layers remain visually prominent, although it also allows changing the base layer. All polygonal layers are displayed semi-transparently so that the basemap and point data remain visible underneath.

The Leaflet map includes layer toggles for:

- Summary polygons
- Service areas
- Containers
- Underserved doors

The map also includes controls for:

- Container type
- Buffer threshold
- Geographic aggregation
- Summary statistic

Clicking a neighborhood or population block filters the scientific graphs. This creates a linked map-graph interaction: the map is not only a display object but also a filter for the dashboard.

Dashboard design

The dashboard was implemented in Shiny using a multi-page layout. The implemented dashboard contains four main pages:

1. **Map and dashboard**
    - Leaflet map.
    - Container-type selector.
    - Distance-threshold selector.
    - Geography selector: neighborhoods or population blocks.
    - Layer visibility controls.
    - Summary value boxes.
2. **Distance distributions**
    - Histogram of selected door-to-container distances.
    - ECDF curve.
    - Boxplot by container type.

    - Ranked chart of worst-served areas.
3. **Data table**
    - Tabular summary of selected areas and metrics.
4. **Methodology**
    - Description of data sources, preprocessing, assumptions, CRS handling, and limitations.

Implementation

The implementation was written in R. Spatial data was loaded using the *sf* package, and transformed to the EPSG:4326 CRS for use with Leaflet.

The main map page contains the Leaflet map and the primary user controls. Users can select

- the summarization level, either neighborhoods or city blocks;
- the container types, including an option for allowing any container;
- the summary statistic used for polygon coloring, i.e. min, max, mean, or median;
- the service-area threshold, in values of 50, 100, 150, 200 m;
- and which map layers should be visible.

The main analytical layers are summary polygons, service areas, containers, and underserved doors, which can all be toggled independently. The map is linked to summary value boxes that report the selected area name, the number of doors in the current selection, and the selected distance statistic, allowing to view the mean, median, minimum, or maximum distance at a glance.

The code updates the main layers via *leafletProxy()*. Firstly, this means that changing any of the inputs does not reset the map pan and zoom, which improves usability, e.g., when the user is exploring a specific local area. Multiple observers of this *leafletProxy()*, one per main layer, mean that bi-directionally dependent reactive values don’t cause interference with one another, such as when clicking a neighborhood or population block recalculation in an earlier version of the project.

The application uses reactive expressions for filtering/specification:

- Containers by type.
- Service areas by type and threshold.
- Summary polygons by selected statistic.
- Doors by selected map polygon.
- Underserved doors by distance threshold.

A loading indicator thanks to the *shinybusy* library is also present for longer updates, especially when switching to city blocks, which can involve many geometries.

The distance-distribution tab contains four previously mentioned graphs generated with *ggplot2*. The same reactive filtered dataset is used for both the map and the graphs, ensuring consistency between spatial and statistical views.

A simple table is present in the Data table tab that shows tabular data for the current input configuration and the selected summarization layer.

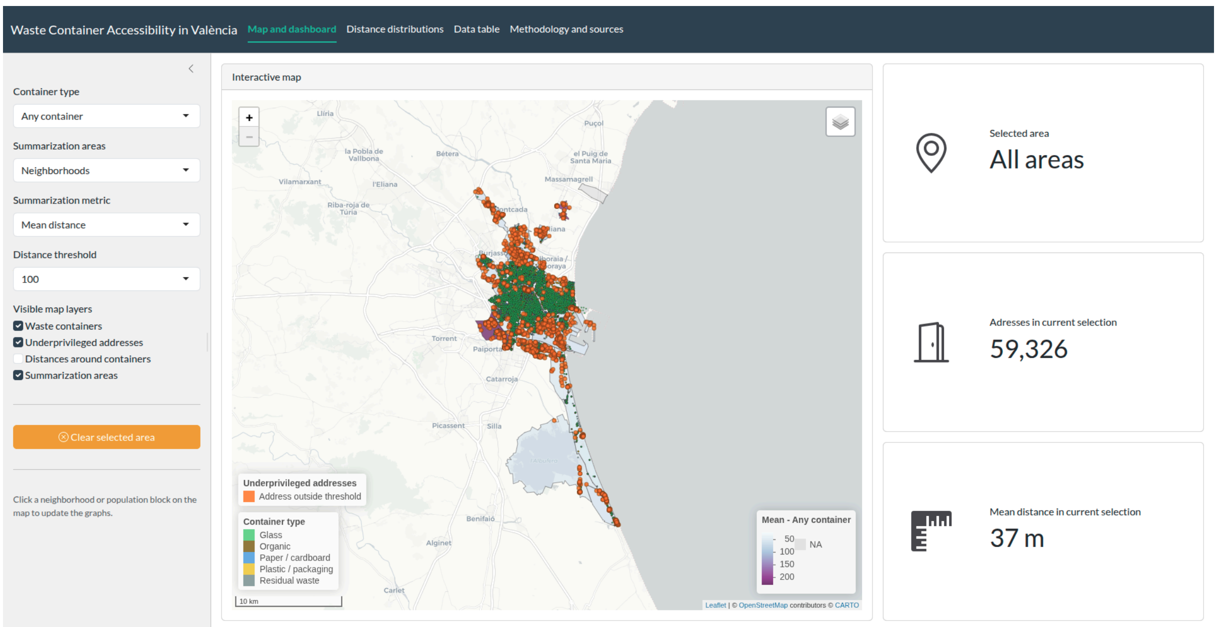
Results

The project produced a working geospatial dashboard structure capable of visualizing waste-container accessibility across València. The processed data layers support analysis at three spatial levels:

1. **Door/address level:** each entrance has distances to the nearest container by type and threshold indicators.
2. **Population-block level:** each block has population counts and waste container accessibility summary statistics.
3. **Neighborhood level:** each neighborhood has minimum, maximum, mean, and median nearest-container distances by type.

The service-area layer enables visualization of the areas within 50, 100, 150, and 200 metres of each container category. The door-accessibility layer enables identification of entrances outside selected thresholds, such as doors more than 200 metres from a glass or organic container.

The dashboard allows users to compare accessibility by container type. For example, a user can select **Glass** and **100 m** to view glass-container service areas and examine the distribution of entrance-to-glass distances. A user can then switch to **Organic** or **Paper / cardboard** to compare whether different recycling facilities are distributed with similar accessibility.



Discussion

The project shows how open municipal data can be transformed into a data visualization tool. The project allows for connecting point-level infrastructure data, building-entrance locations, population-block data, and neighborhood-level summaries in a single dashboard.

Browsing the map and focusing on the main bulk of the city of València, one can already see some patterns emerging. For example, the vast majority of addresses have a waste container less than 100 m away, but there is a cluster of addresses in the neighborhoods of El Carme and La Seu in the city center that don't. But further analysis is needed, for which this technical report ran out of time.

Given the current state of the implementation, several limitations with the result must be considered:

First, the distance calculations are based on Euclidean distance rather than walking-network distance. This means that barriers such as wide roads, gardens, walls, or restricted crossings may not be fully represented. A future version could use pedestrian-network distance for a more realistic measure.

Second, population was assigned evenly across building doors within each population block. This is a necessary approximation because the door layer does not contain household or resident counts. The method is useful for estimation, but it should not be interpreted as exact population at each entrance.

Third, the population-block dataset was last updated in 2020, so population values may not fully represent current residential distribution.

Conclusions

This project created a complete spatial data visualization for evaluating access to waste containers in València. It used official open data, *QGIS* spatial preprocessing, and an R *Shiny* dashboard with *Leaflet* and *ggplot2* graphs.

The final dashboard provides a useful way to explore whether residents have nearby access to different waste-container types and whether accessibility varies by neighborhood or population block.

References

Bernstad, A. (2014). Household food waste separation behavior and the importance of convenience. *Waste Management*, 34(7), 1317–1323. https://doi.org/10.1016/j.wasman.2014.03.013

Timlett, R. E., & Williams, I. D. (2008). Public participation and recycling performance in England: A comparison of tools for behaviour change. *Resources, Conservation and Recycling*, 52(4), 622–634. https://doi.org/10.1016/j.resconrec.2007.08.003

Ajuntament de València. (2025). *Contenidors residus sòlids* [Dataset]. València Open Data portal. https://valencia.opendatasoft.com

Ajuntament de València. (2024). *Portals dels carrers* [Dataset]. València Open Data portal. https://valencia.opendatasoft.com

Ajuntament de València. (2020). *Illa de cases cadastrals amb dades de població* [Dataset]. València Open Data portal. https://valencia.opendatasoft.com

Ajuntament de València. (2023). *València, más limpia, más verde, hagámoslo bien* [Municipal campaign]. https://www.valencia.es/cas/campañas-municipales