3b. Feature Matching

The feature matching takes the reference data and attempts to identify corresponding features in the OSM data set. Feature matching is a necessary precondition to compare single features rather than feature characteristics on study area a grid cell level.

Method

Matching features in two road data sets with each their way of digitizing features and a potential one-to-many relationship between edges (for example in the case where one data set only maps road center lines, while the other map the geometries of each bike lane) is not a trivial task.

The method used here converts all network edges to smaller segments of a uniform length before looking for a potential match between the reference and the OSM data. The matching is done on the basis of the buffered distance between objects, the angle, and the undirected Hausdorff distance, and is based on the works of Koukoletsos et al. (2012) and Will (2014).

Based on the matching results, the following values are computed:

- The number and length of matched and unmatched edges, in total and per grid cell
- A comparison of the attributes of the matched edges: Is their classification of cycling infrastructure as protected or unprotected the same?

Interpretation

It is important to visually explore the feature matching results, since the success rate of the matching influences how the analysis of number of matches should be interpreted.

If the features in the two data sets have been digitized differently - e.g. if one data set has digitized bike tracks as mostly straight lines, while the other includes more winding tracks, the matching will fail. This is also the case if they are placed too far from each other. If it can be confirmed visually that the same features do exist in both data sets, a lack of matches indicates that the geometries in the two data sets are too different. If however it can be confirmed that most real corresponding features have been identified, a lack of matches in an area indicates errors of commission or omission.

Sections

- Match features
    - Run and plot feature matching
    - Matched and unmatched features
    - Feature matching summary
- Analyze feature matching results
    - Matched features by infrastructure type
    - Feature matching success
- Summary

Match features

Run and plot feature matching

```
Interactive map saved at results/COMPARE/Rosario/maps_interactive/segment_matches_15_17_30_compa
re.html
```
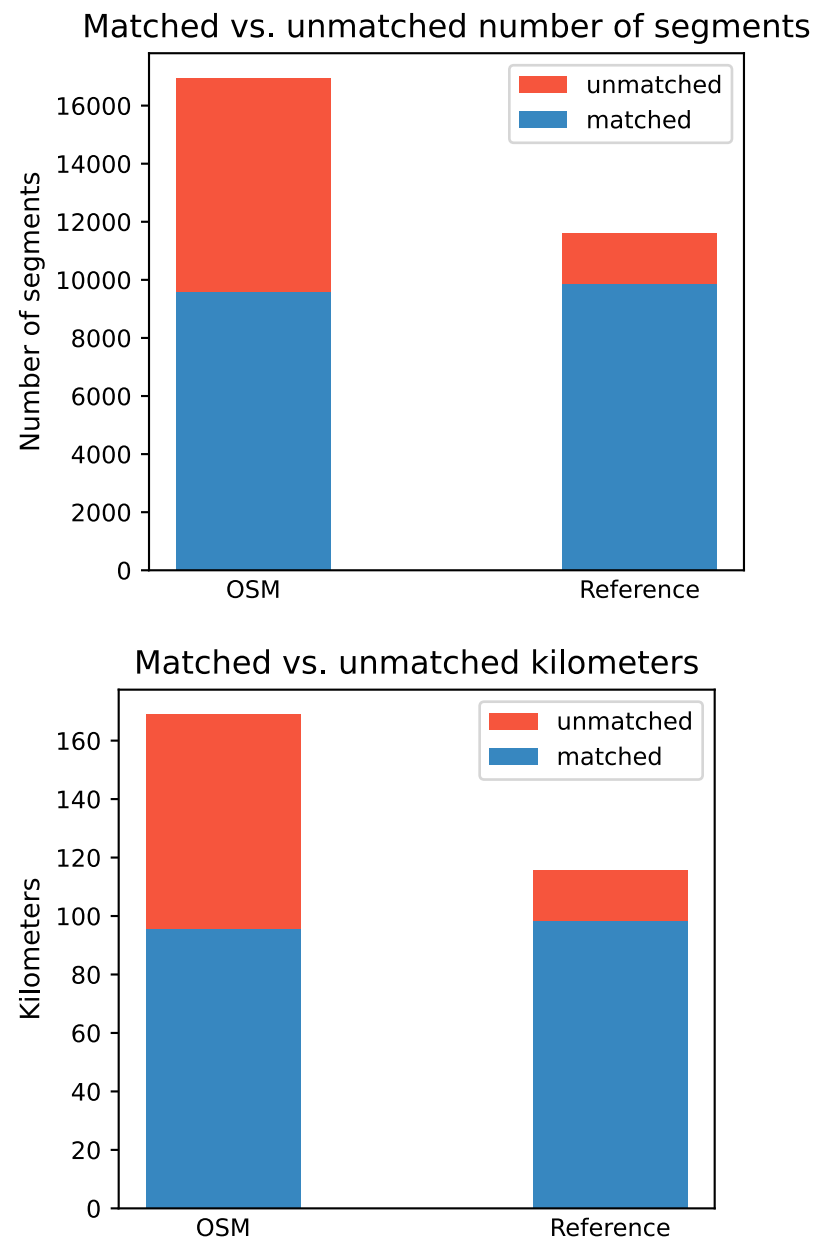

Rosario: OSM matched & unmatched features

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

BIKEDNA

Rosario: EMR matched & unmatched features

Map tiles by Stamen Design, CC BY 3.0 — Map data (C)
OpenStreetMap contributors

Feature matching summary

```
Edge count: 9565 of 16956 OSM segments (56.41%) were matched with a reference segment.
Edge count: 9859 out of 11603 EMR segments (84.97%) were matched with an OSM segment.
Length: 95.52 km out of 168.98 km of OSM segment (56.53%) were matched with a reference segment.
Length: 98.28 km out of 115.72 km of EMR segments (84.93%) were matched with an OSM segment.
```

BIKEDNA

Matched vs. unmatched number of segments

Matched vs. unmatched kilometers

Analyze feature matching results

Matched features by infrastructure type

Infrastructure type of matched features

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

Feature matching success

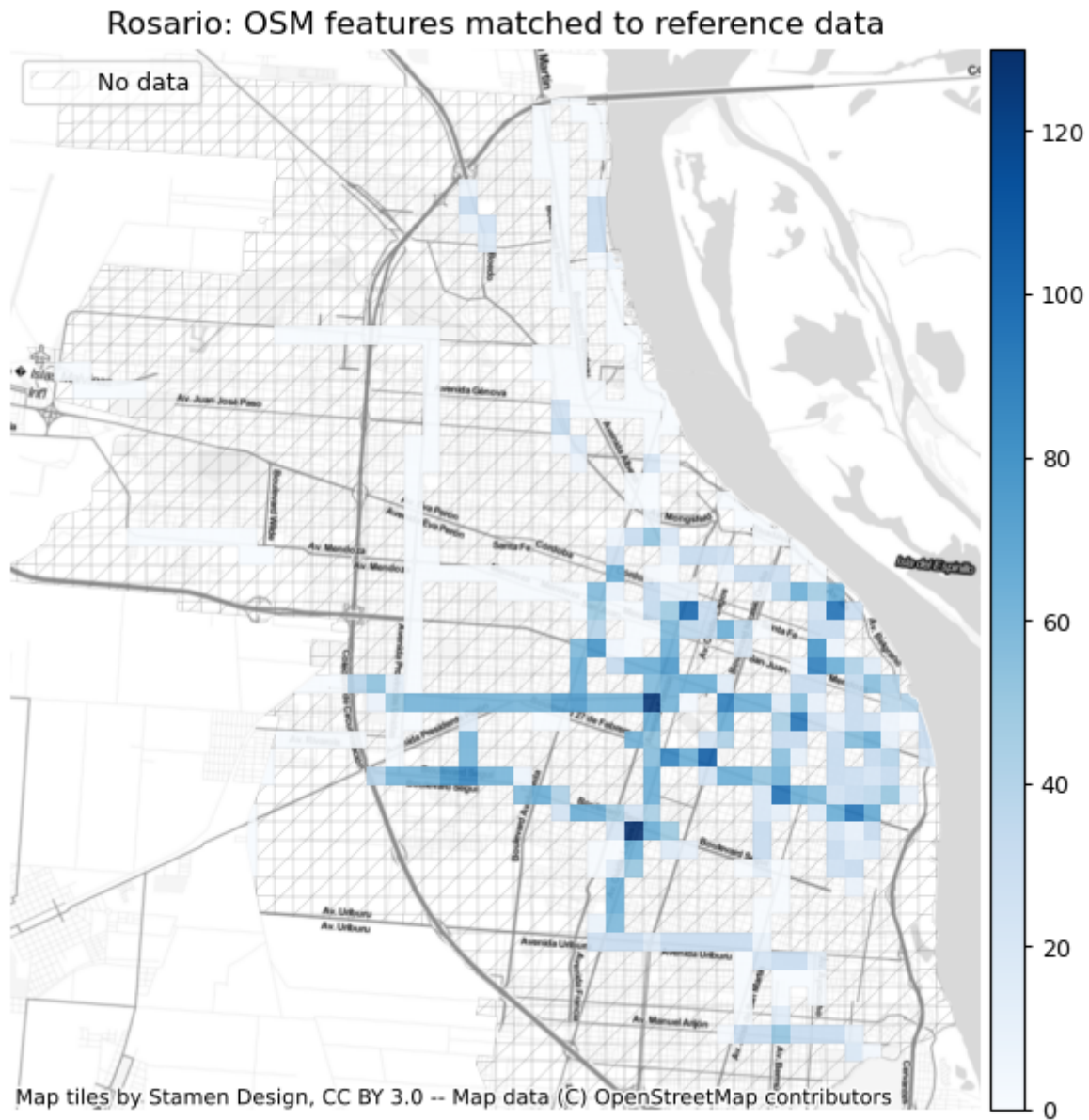In the plots below, the count, percent, and length of matched and unmatched segments in each data set are summarized.

> **Warning**
> The number of matched segments in one data set in a grid cell does not necessarily reflect the number of matched segments in the other data set, since a segment can be matched to a corresponding segment in another cell. Moreover, the local count refers to segments intersected with the grid cell. For example, a segment crossing 2 cells will thus be counted as matched in 2 different cells. This does not change the relative distribution of matched/unmatched segments, but it does entail that the overall summary of matched/unmatched segments above uses a different total count of segments than the plots below.
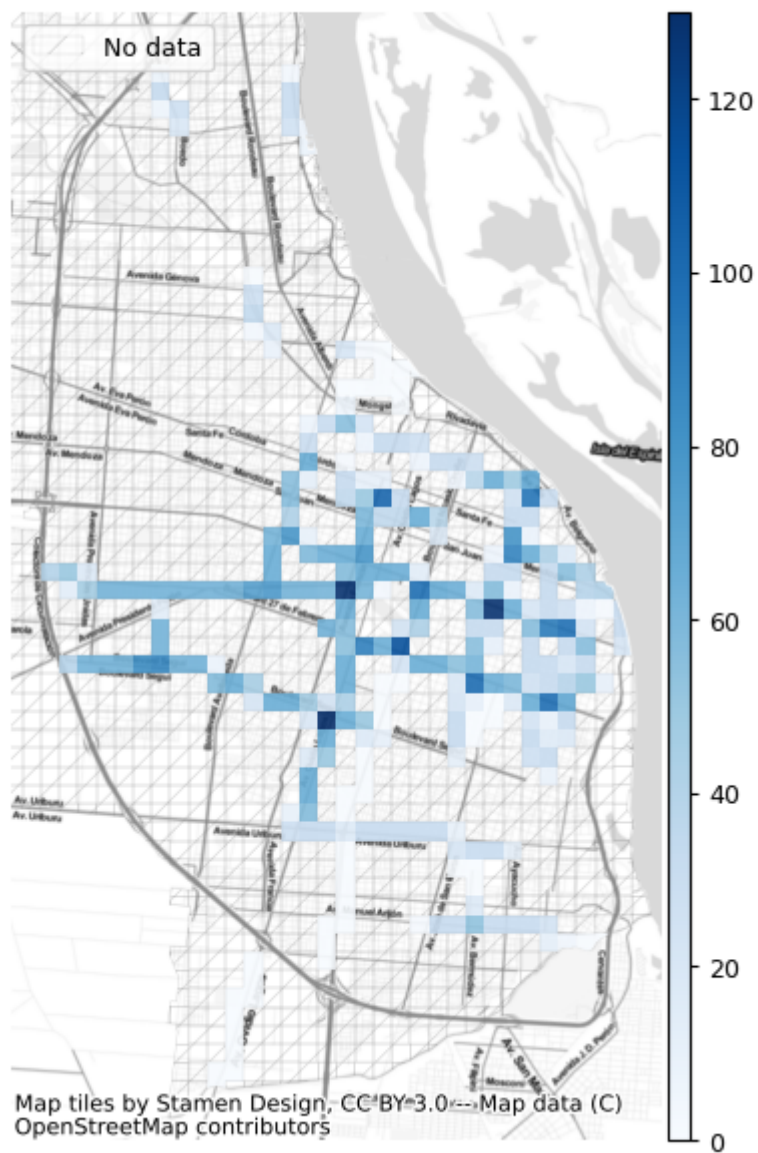
○BIKEDNA○

```
/var/folders/b0/lkvf88hn0673f5dlj9z0_2dr0000gn/T/ipykernel_37378/3687147659.py:4: UserWarning: `
keep_geom_type=True` in overlay resulted in 17 dropped geometries of different geometry types th
an df1 has. Set `keep_geom_type=False` to retain all geometries
  osm_segments_joined = gpd.overlay(osm_segments[['geometry','seg_id','edge_id']], grid, how="in
tersection")
```
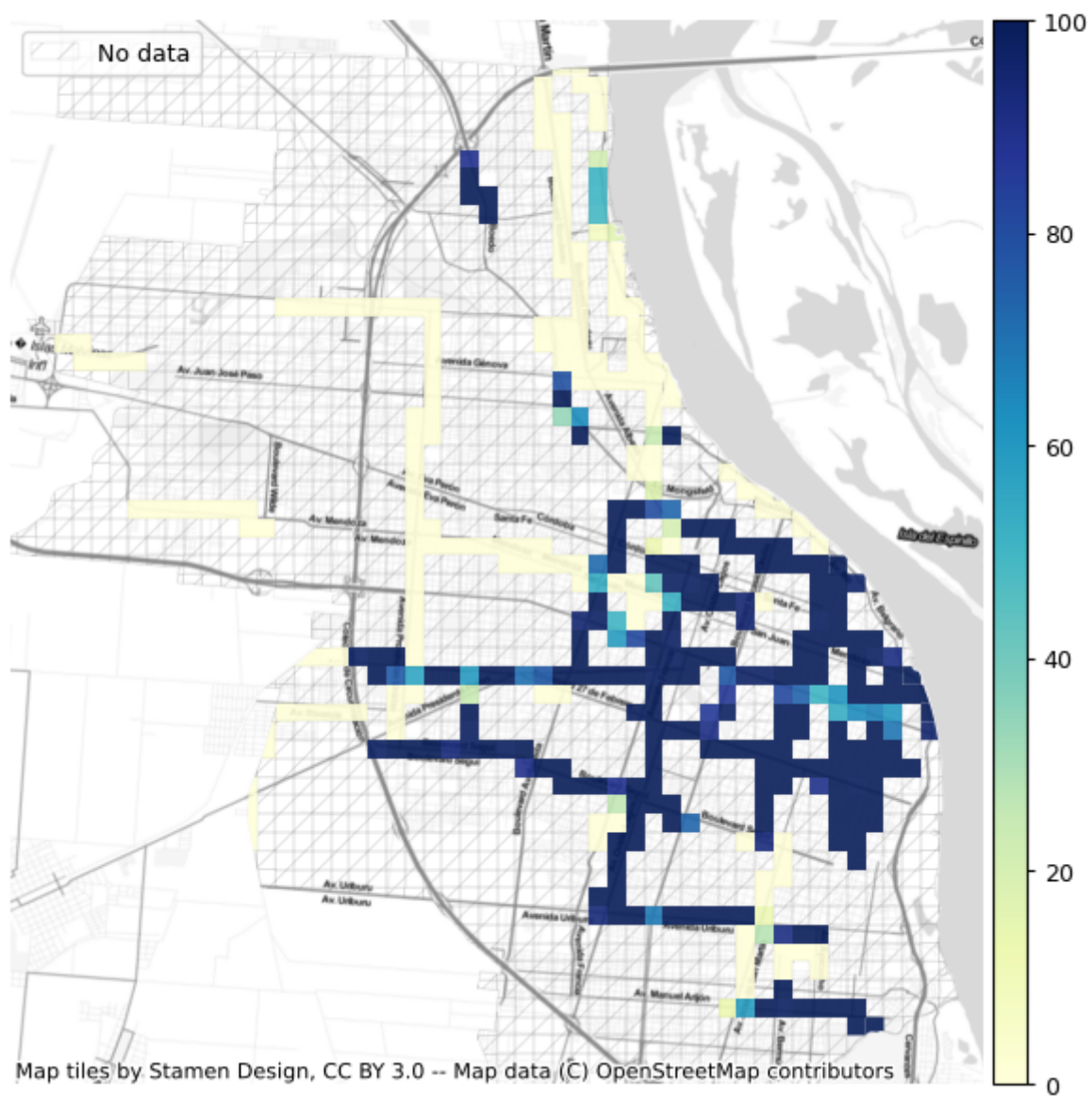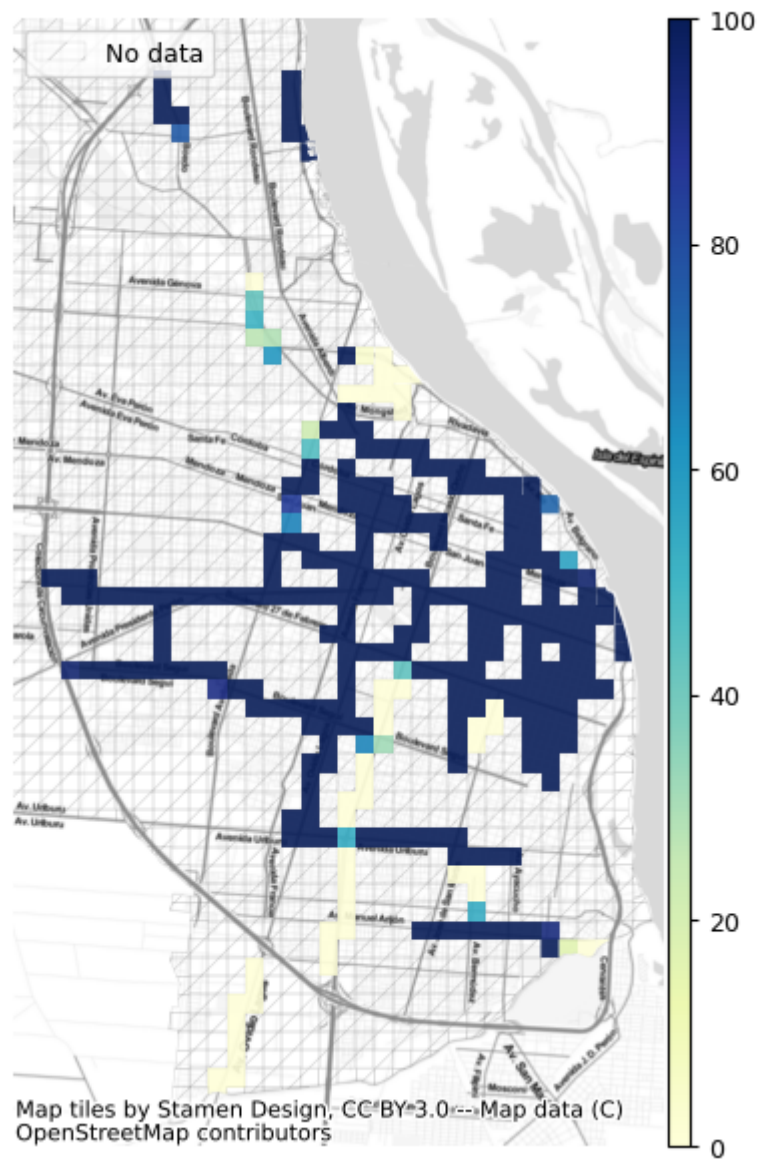

Rosario: OSM features matched to reference data

Rosario: EMR features matched to OSM data

Map tiles by Stamen Design, CC BY 3.0 — Map data (C)
OpenStreetMap contributors

BIKEDNA

Rosario: percent of OSM features matched to reference data2

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

BIKEDNA

Rosario: percent of EMR features matched to OSM data2

Map tiles by Stamen Design, CC BY 3.0 — Map data (C)
OpenStreetMap contributors

BIKEDNA

Rosario: length of OSM features matched to reference data (m)

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

BIKEDNA
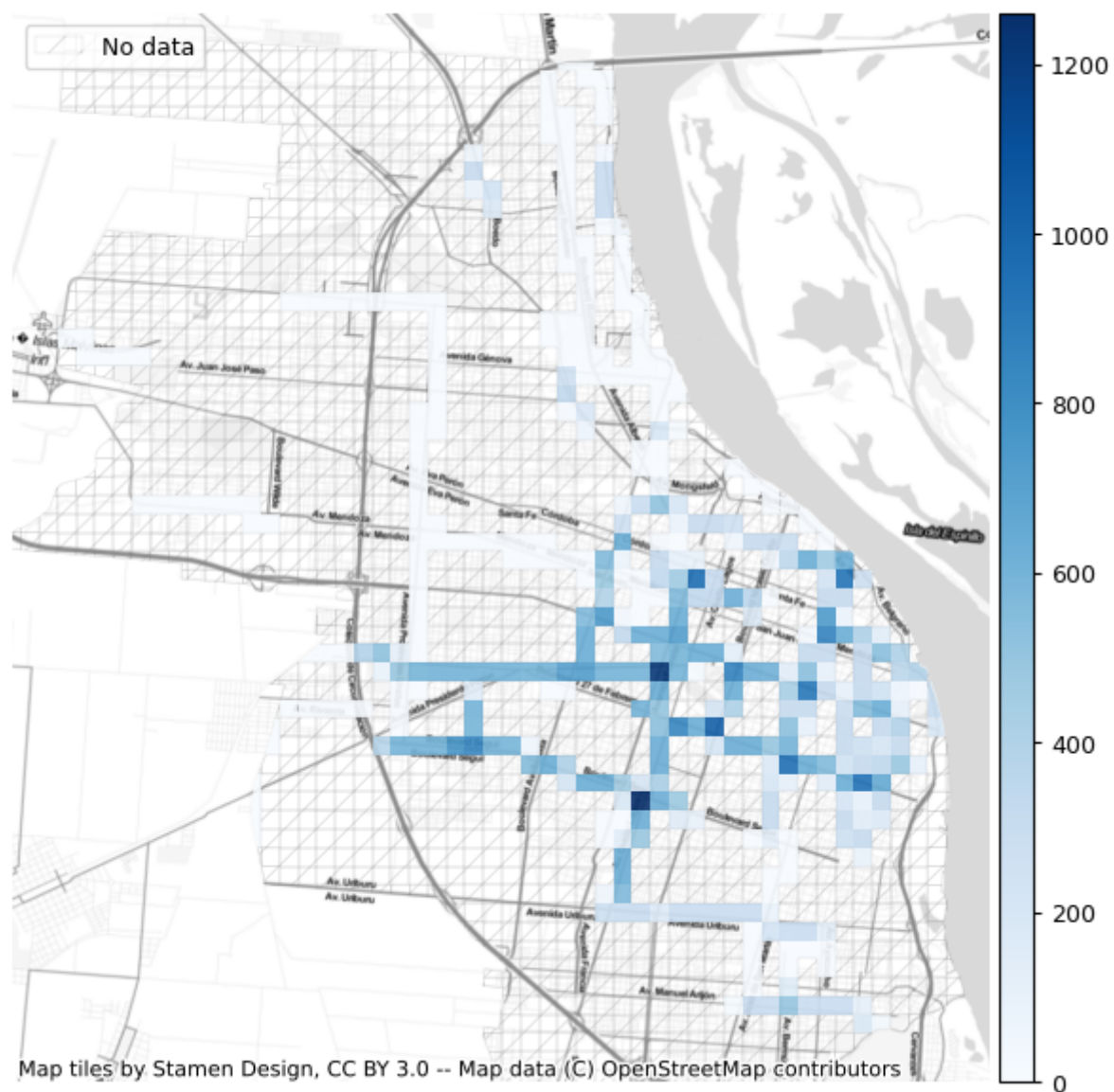
Rosario: length of EMR features matched to OSM data (m)

BIKEDNA
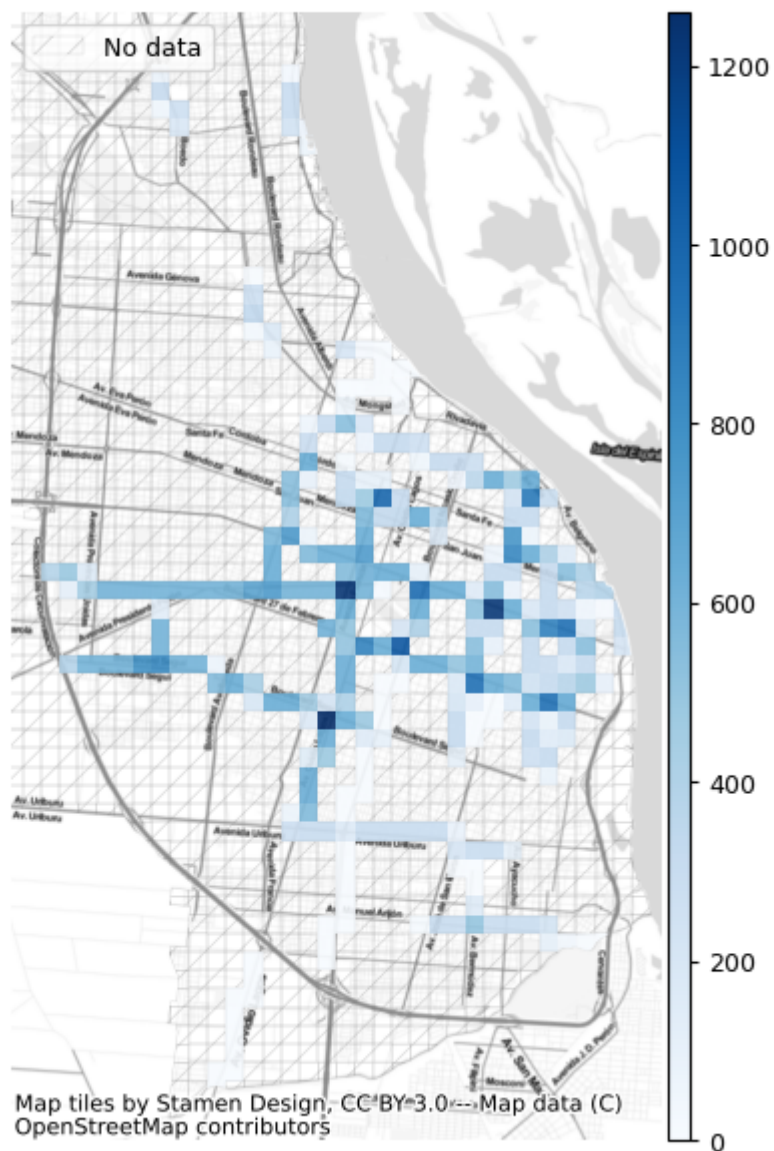
Rosario: OSM segments not matched to reference data

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors
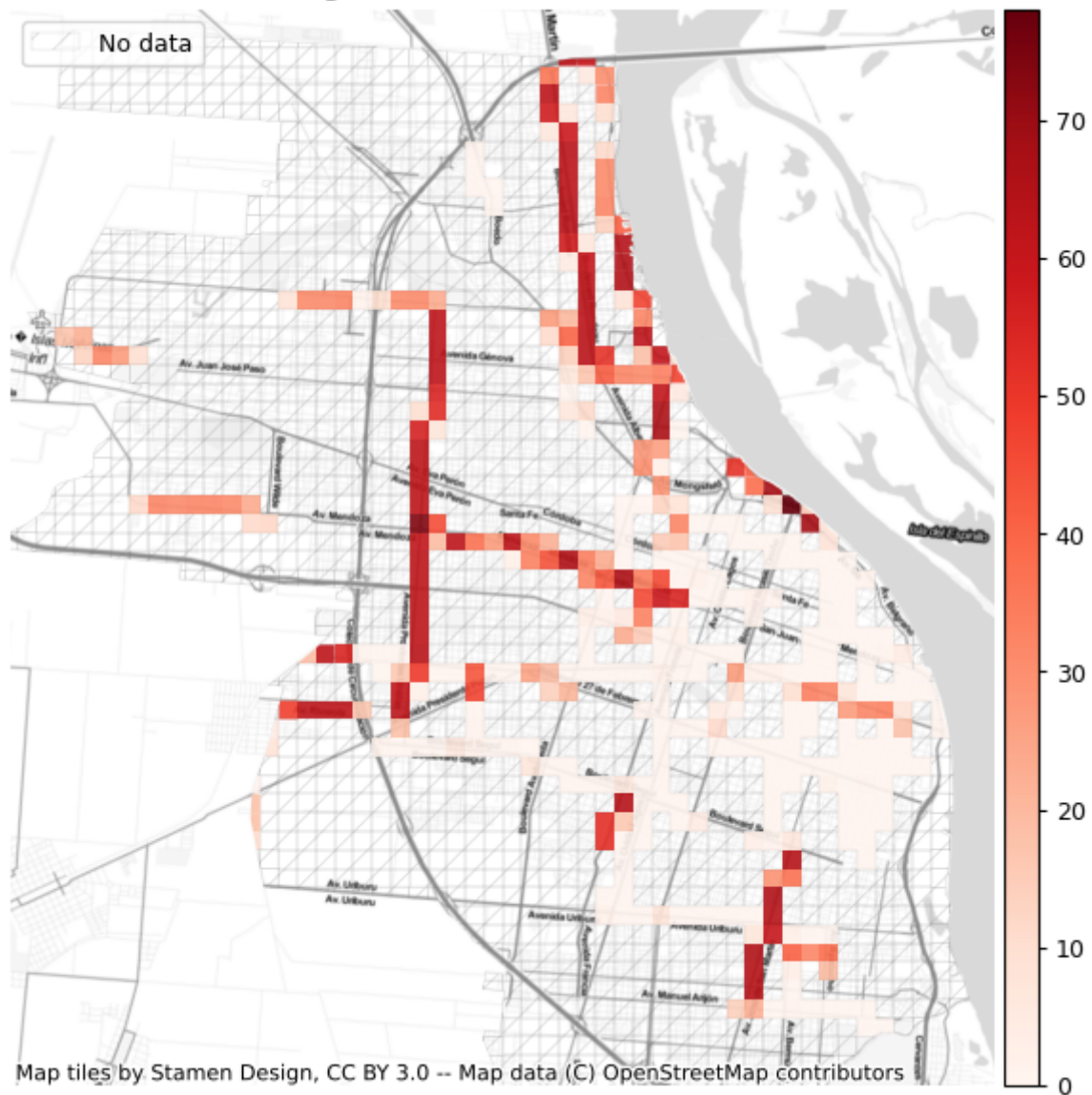
Rosario: EMR segments not matched to OSM data

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C)
OpenStreetMap contributors
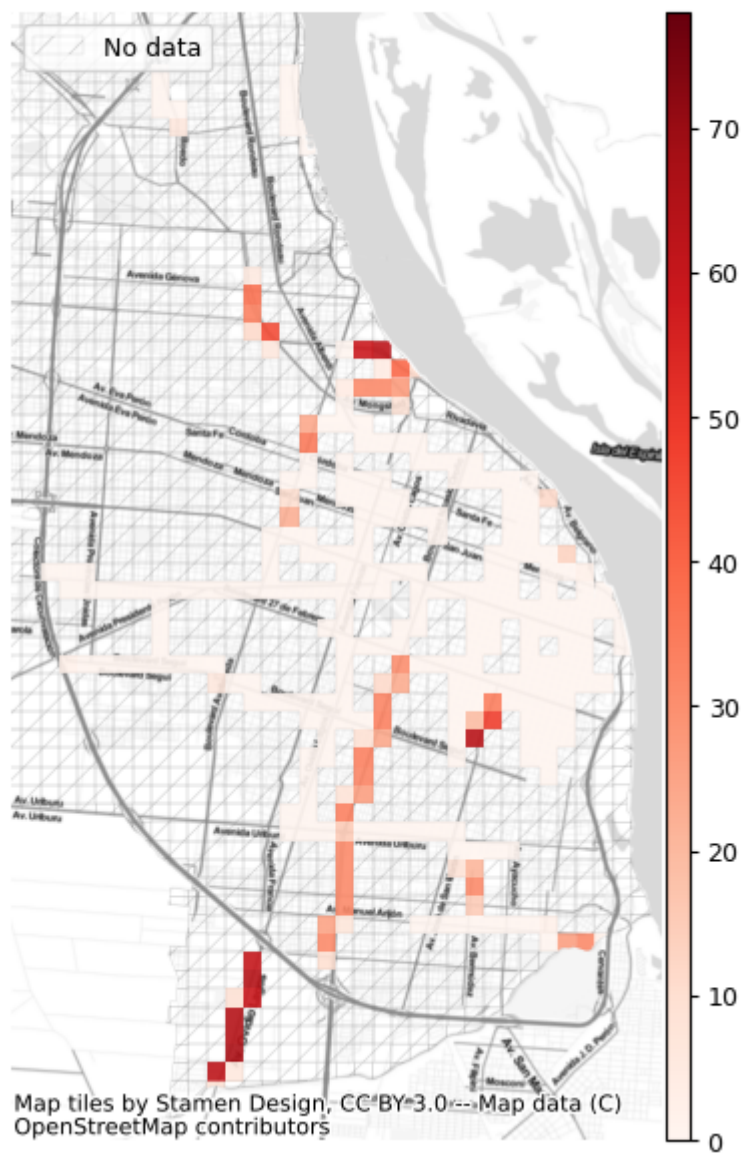
Rosario: percent of OSM segments not matched to reference data

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors
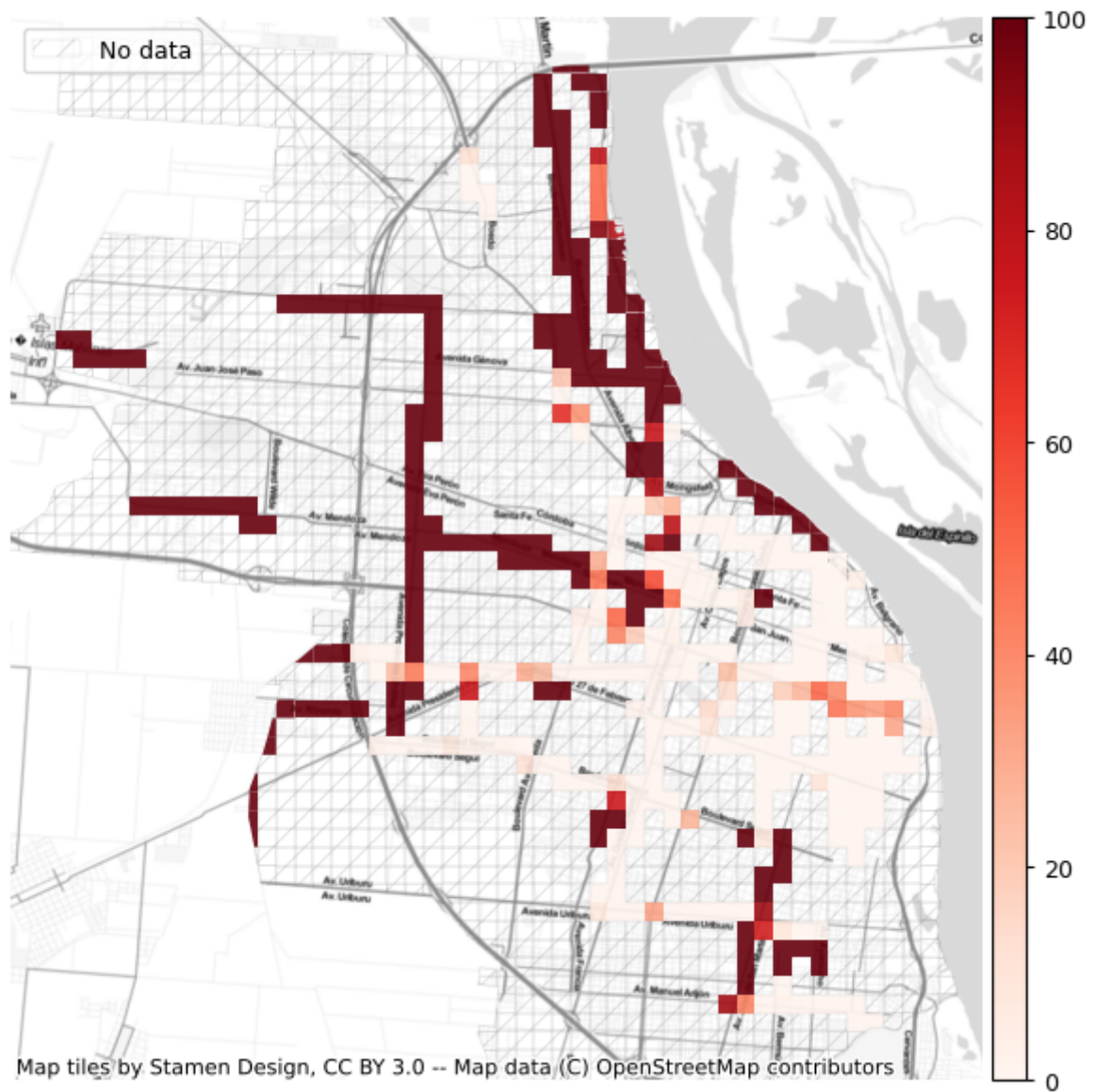
Rosario: percent of EMR segments not matched to OSM data

Map tiles by Stamen Design, CC BY 3.0 -- Map data (C)
OpenStreetMap contributors
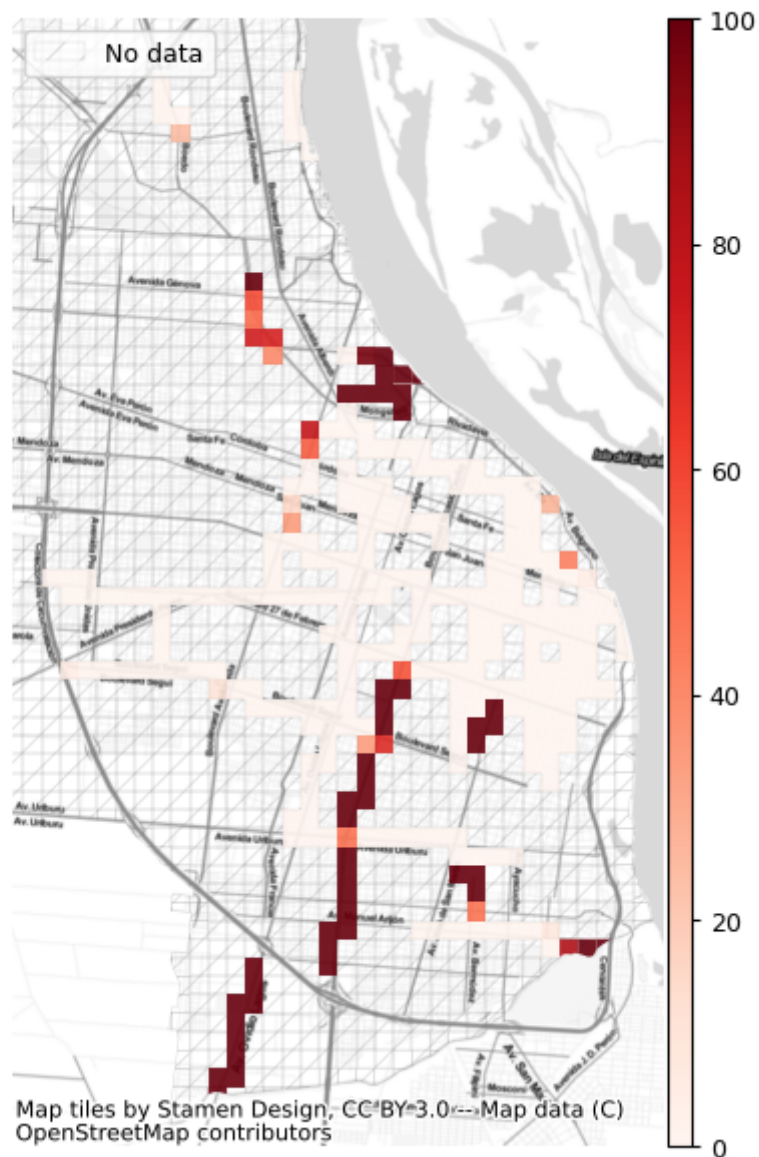
Rosario: length of OSM segments not matched to reference data

BIKEDNA
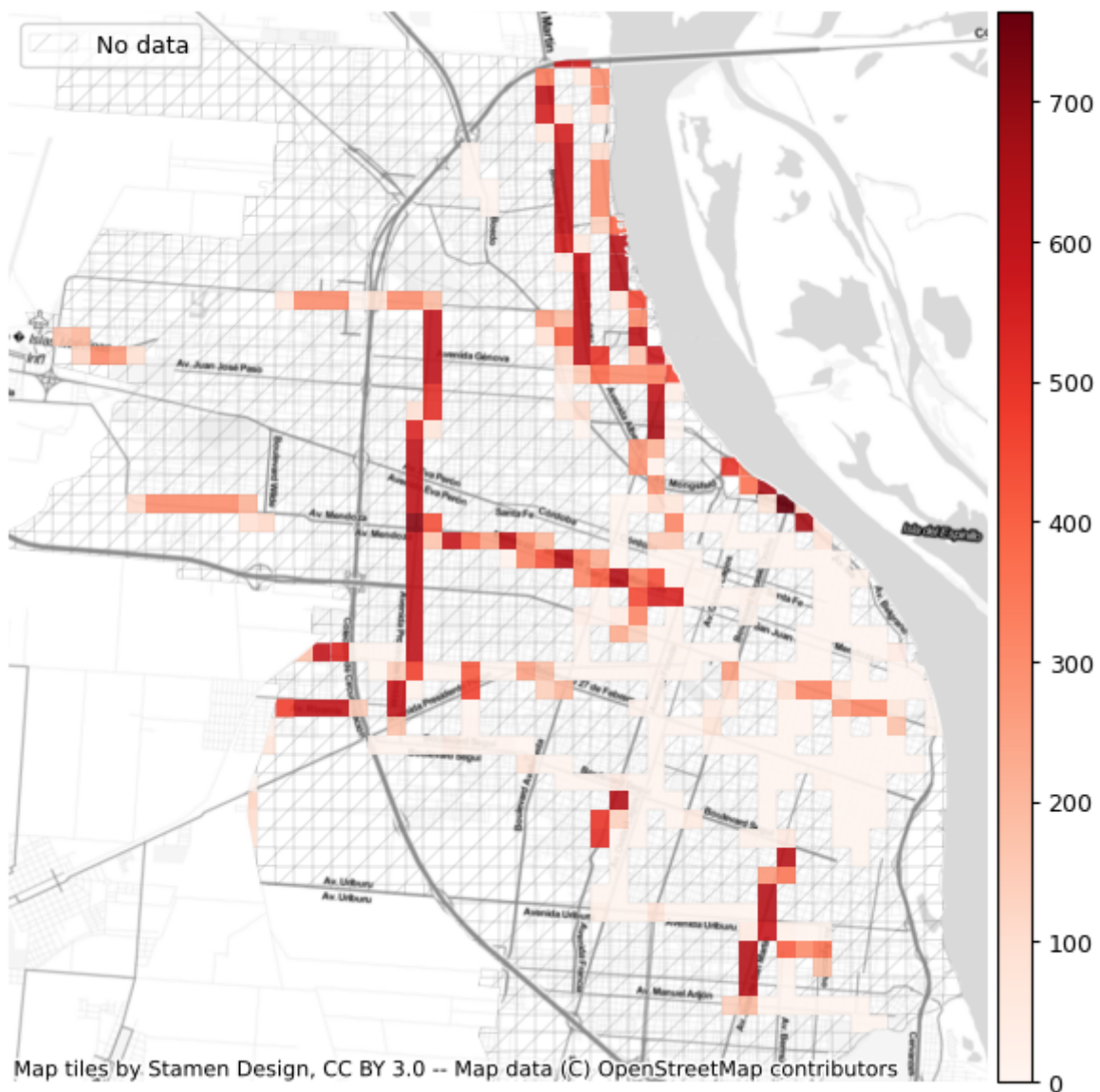
Rosario: length of EMR segments not matched to OSM data

Map tiles by Stamen Design, CC BY 3.0 — Map data (C) OpenStreetMap contributors
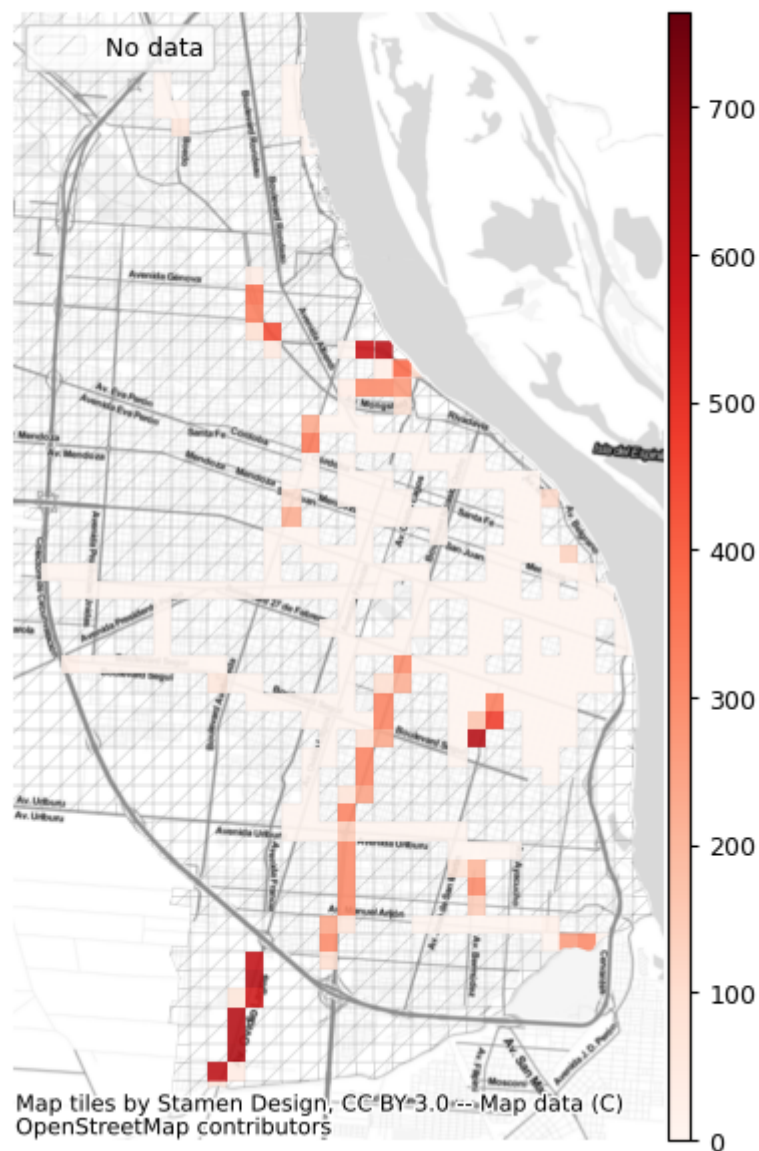
Summary

Feature Matching Results		
	OSM	**EMR**
Count of matched segments	9,565	9,859
Percent matched segments	56%	85%
Length of matched segments (km)	96	98
Percent of matched network length	57%	85%
Local min of % matched segments	7%	3%

BIKEDNA

| **Local max of % matched segments** | 100% | 100% |
| **Local average of % matched segments** | 90% | 96% |