

## Report

This project gathers data on Dutch Sinti and Roma WWII victims, retrieved from an online memorial site<sup>1</sup>. Data is structured into csv-files containing ID and names of the victims and their familiar relations, dates and places of birth and death, and their ages. Results are presented through a user interface.

### Structure:

The website was scraped and the individuals' data was stored locally as html files, which were then parsed to retrieve the aforementioned information into csv files.

Longitudes and latitudes were gathered by geocoding with timeouts to ensure the code is robust in case of external server/service failure.

We started by mapping the places of birth and death of the victims and creating a network graph to show relationships between victims and their family when interacted with. With the visuals ready, we added callback functions to be able to select victims and show their place of death/birth, and vice versa.

### Obstacles:

The scraping process revealed two obstacles: IDs were only present in individuals' personal URLs, and multiple people shared the same names. We were able to solve both of these problems by naming the html files using both the person's ID and name.

Some text blocks containing specific relationship types also included whether the person survived, as well as spaces and new lines. This was solved by splitting the string at the spaces and only taking the first index of the list created.

Many blocks were empty, resulting in error when trying to extract data. To avoid this we made if-statements; for empty blocks, we assigned the variable as “unknown”.

While geocoding, some locations weren't found, either because the data was missing from the website, or due to city names not being found by the geocoder. In such instances values were set as missing values (NaN) and not included in the map visualization.

We chose to use Plotly Express for the map and Cytoscape for the network graph, which wasn't straightforward to link. First we had to have a common key between the circles and the nodes, we chose to add the birth and deathplaces to the definition of the nodes. The map is colored based on a column in the dataset, therefore we had to create an additional column, which specified the colour. When clicking a node in the network graph, the value in this column was changed to red. The colour of the nodes are defined in a stylesheet, which was changed to red when clicking a place on the map.

### Division of labour:

All group members have collaborated on solving all the given tasks.

---

<sup>1</sup> <https://www.joodsmonument.nl/en/page/344139/sinti-en-roma-namenlijst>