# Introduction

In this report we describe how we find and discover interesting aspects in a given large dataset. Since it is hard for humans to interpreted large dataset, we use different visualizations techniques to visualize and discover the data in a more accessible format. We make use of the D3.js library for implementing these visualizations.

<Section explanations>

<Image of application>

# The data set

The dataset that is visualized originates from <http://www.opencbs.nl>. This dataset is in GEOJson format and contains shape information of Dutch municipalities. Also we make use of tab-separated file containing statistical information about these municipalities. An example of data found in this file is the age distribution within these municipalities. A description of this dataset and the information in it can be found in the assignment[[1]](#footnote-1) description.

## Inspecting the dataset

The dataset contains various information, we inspect the dataset to find information that could deliver interesting results when visualized. We did this by formulating questions that could be resolved by visualizing this dataset.

Some questions we came up with are:

* In what municipality are there proportionately the most young unmarried woman?
* In what municipality are the most widowers and is there a correlation between widowers and the average age in that municipalities?
* Is there a direct correlation between number of cars and number of residents?

For visualizing the data and answering these questions we needed the following data from the dataset:

* **P ONGEHUWD** Percentage of unmarried people.
* **P GEHUWD** Percentage of married people.
* **P GESCHEID** Percentage of divorced people.
* **P VERWEDUW** Percentage of widows and widowers.
* **AUTO TOT** Number of cars.
* **BEDR AUTO** Number of company cars (minivans, trucks, etc.).
* **P 00 04 JR / P 95 EO JR** Percentage of inhibations aged 0 to 95 years in categories of five years.

Using a good visualization of this data, should make it possible to answer the questions. Also since we want to make comparisons between municipalities we need our visualizations to support this.

# Design

In this section we explain the general design of the visualization.

## General design

Since the dataset contains a large quantity of data about municipalities, a decision has to been made what to visualize. And also what visualization technique suits the selected data, so that it can be easily interpreted. More important is a valuable visualization, with this we mean that we can find valuable and useful information that meets the requirements.

## User interface

There a many municipalities and there is a lot of data. Showing this at once for all municipalities makes the visualizations to complex and hard to inspect. Therefor we needed an interface that allows the user to scroll through the data and select some aspects of the data that needs to be visualized. To support these actions on the visualizations we needed an interface.

We started with a map of the Netherlands showing all municipalities. This map is used to scroll to the municipalities by clicking them. After clicking a municipalities the data selected is shown. Since we also want to make comparisons between municipalities, the ability to select two municipalities was added. When selecting two municipalities, the data of these two is added in one visualization. This approach makes it easy for the user to compare data between municipalities.

Since it is not useful to show all the data of a municipality at once, also an interface is needed to give the user to be ability to select some aspects about the data. In this case where there are not to many categories radio buttons is a good approach. Since they are small enough to place it direct in the visualization. This way the user gets a direct overview of the data and options available for the visualization without going through dropdown or list constructions. Selecting an aspect of the data would recolor the map to show an average overview of the data. This enables the user to find interesting municipalities and select them to get more detail of the data.

# Implementation

## Interface

## Scatter plot

1. <https://dlwpswbsp.tue.nl/120-2014/4d0982b89c664c579bd307b3c4ae82ca/Assignments/assignments2014.pdf> [↑](#footnote-ref-1)