The pharmacies revenue
Links between sale, geography and demography

DTU - Course 02582
Victor SALLING s155453
Joseph SHVEDSKY s155250
Yannis FLET-BERLIAC s151399

OVERVIEW

The study has been conducted on the basis of two datasets, one giving the revenue of 371 Danish pharmacies and the other giving demographics insight over the Danish population for every region.
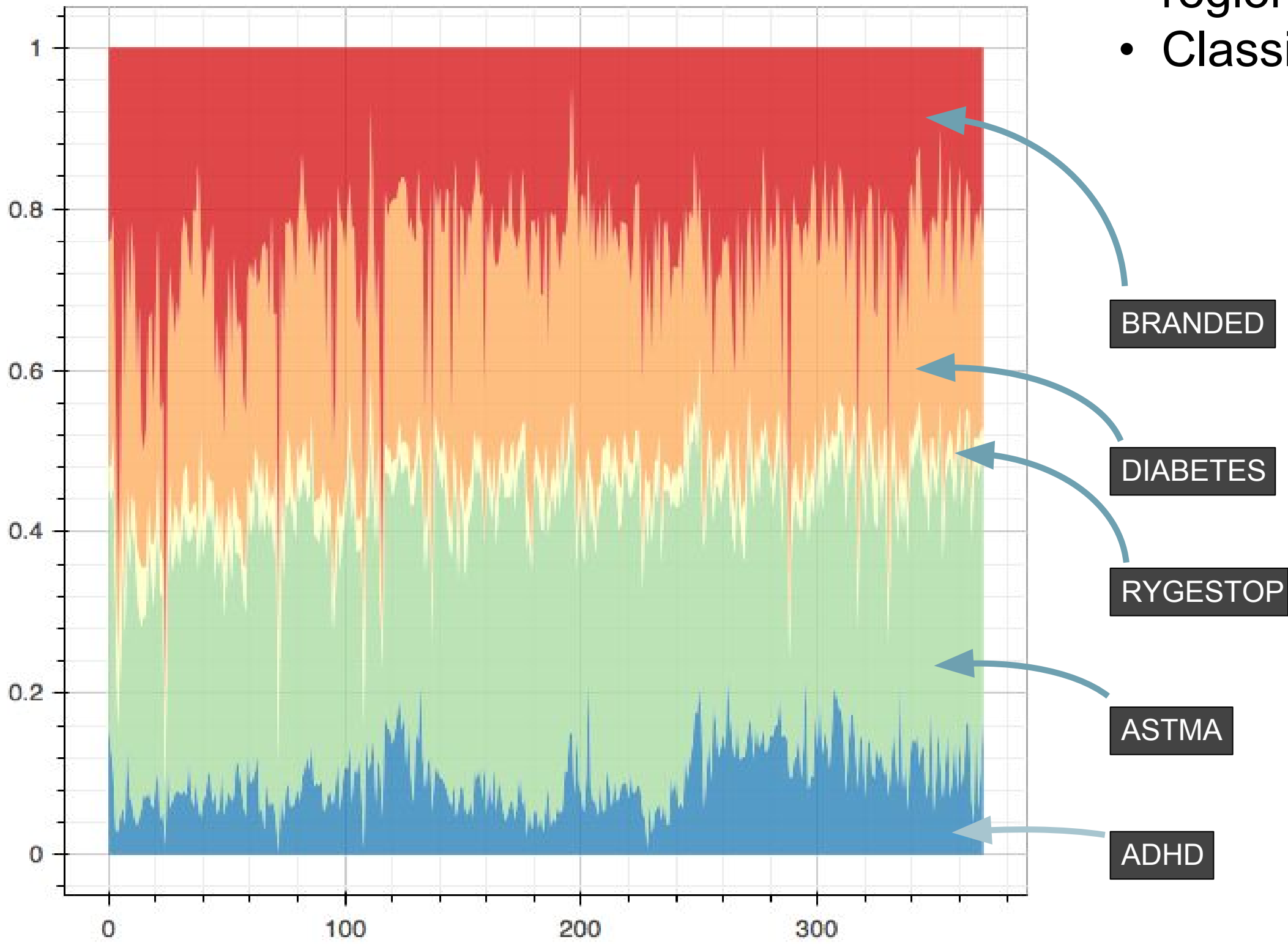
The purpose of this project was to find relations between sales, chains, catchment and geography by :
- Finding patterns in the relations between the revenues of the various categories of products sold at the Danish pharmacies;
- Combining those results with the population demographics related to the bricks, regions, chains and pharmacies;
- Classifying the pharmacies and predicting the surroundings of one given pharmacy.

Y-axis : Proportion of revenues

When plotting the proportion of each revenues for each brick one can have some insight over the share capital for the given pharmacies. For instance considering the outlier pharmacies (we can estimate around 10 of them), it is striking how the revenues from *Branded* medicine are predominant incomes compared to medicine from *Rygestop*, *Diabetes*, *Astma* or *ADHD*.
Moreover, as an overall note, one can easily appreciate that most of the revenues come from *Branded*, diabetes and *Astma* medicine more than those that concern *Rygestop*, or *ADHD*.

X-axis : Pharmacies in Denmark

Denmark is divided into 60 bricks in 5 regions : Capital Region of Denmark (Copenhagen), Central Denmark Region (Aarhus), North Denmark Region (Aalborg), Region Zealand (Roskilde) and Region of Southern Denmark (Odense). Each of these regions present different surroundings, which we will discuss in the following.
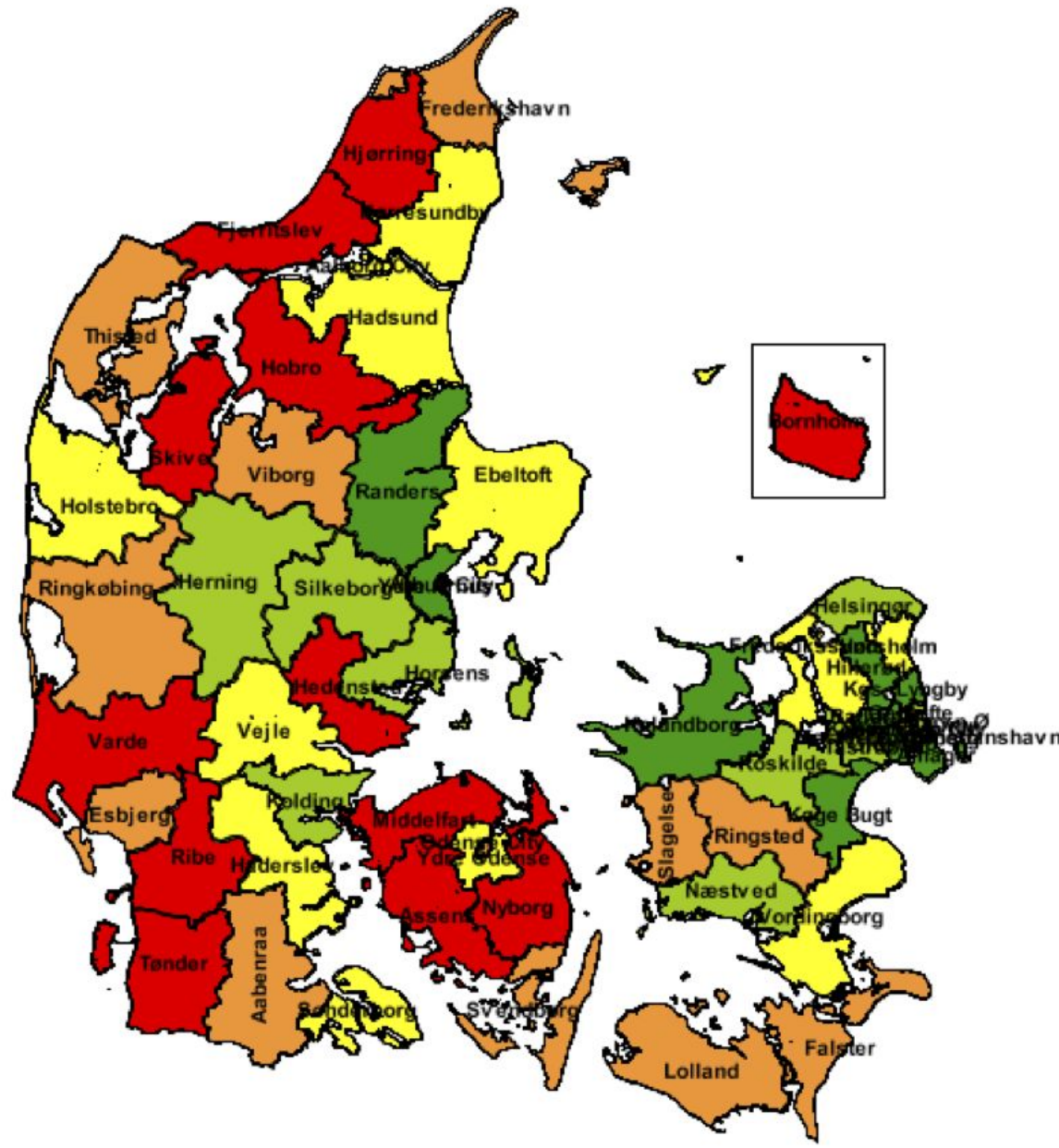
When we look at the graph after plotting, it seems that for the pharmacies in the blocks 23, 30, 31 and 32 there is a fewer need in *ADHD* medicine - those corresponds to countryside areas. It is also recognizable that pharmacies where *Branded* products are the most purchased are also situated in metropolis or nearby metropolis.

WHEN FEATURES SHOW PREVALENCE

In order to have a better understanding for example of the influence of the location of the pharmacies on their sales distribution, 2D plottings can be interesting.

Here are two examples that compare the sales of some pharmacies in proportion to their total sales revenues. The first one suggests more *Astma* problems on the countryside than in metropolis, and the second more *ADHD* sales in Aarhus than in Copenhagen
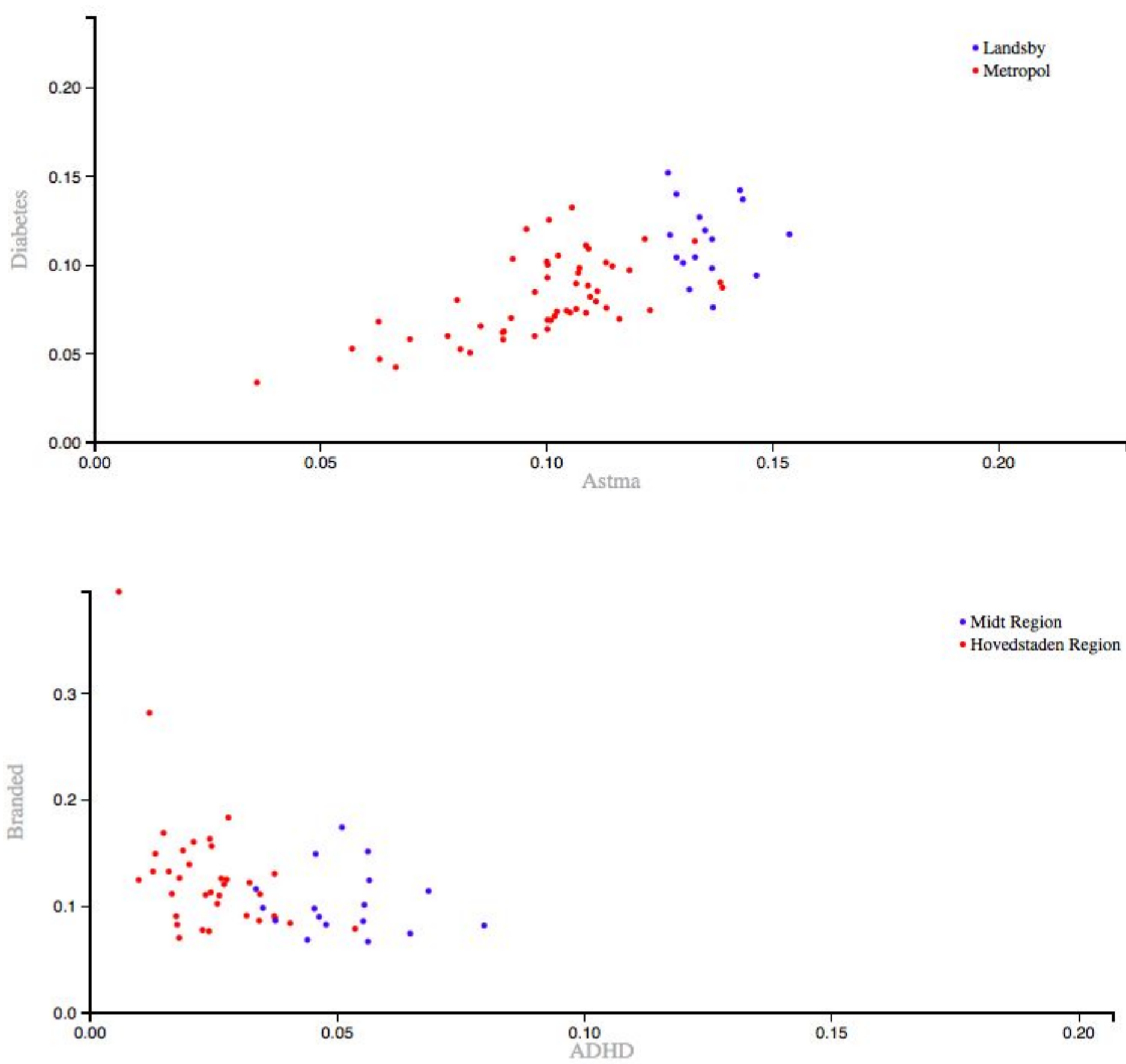
For the project we also decided to retrieve the following plot which gives insight into the importance within the variables made available to us.
The results are based on the Random Forest Classifier algorithm. The data has been normalized after concatenation (demographics together with all pharmacies' revenues). We used the surroundings as targets.
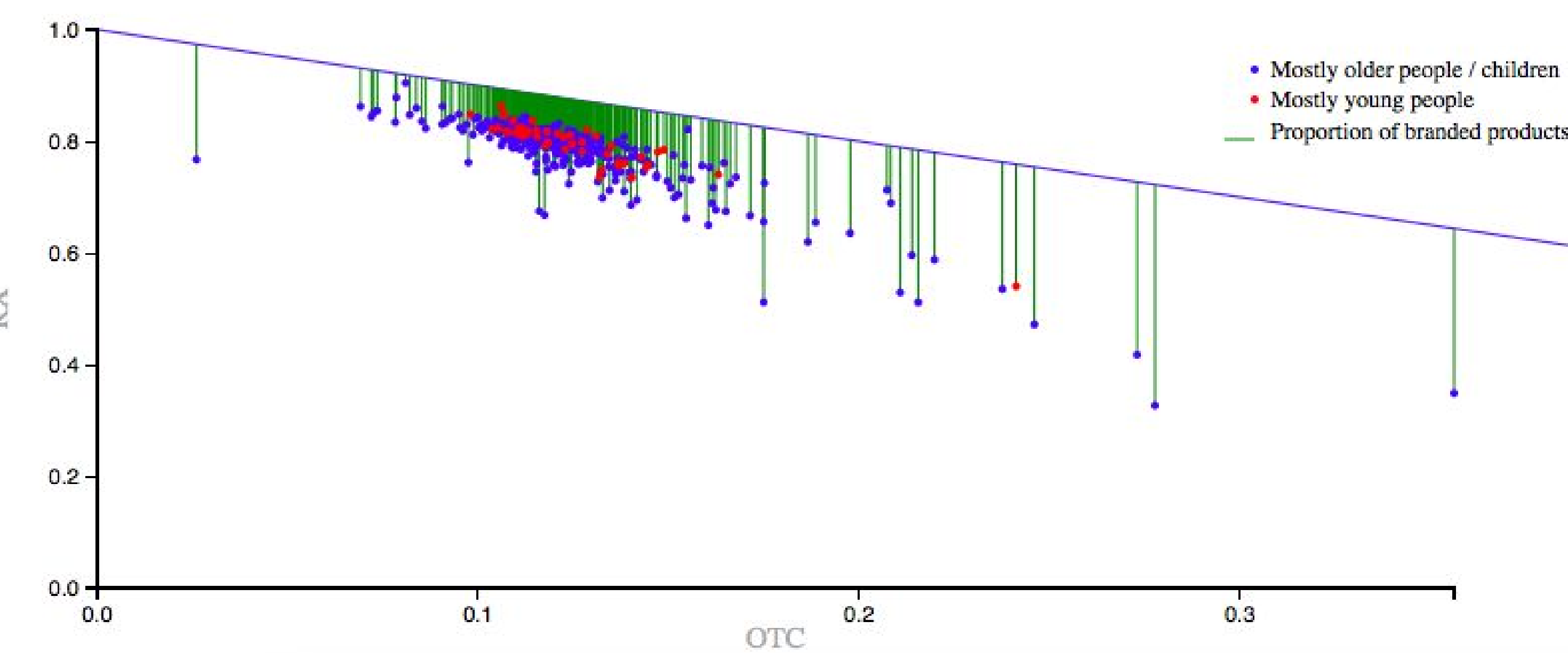It appears that the most important features when it comes to find the surroundings of a sample are : demographics of 1 year old individuals, proportion of *Rygestop*, proportion of *Branded*. Then it comes the demographics of 66 years old individuals, *Astma*, and the least informative medicine is *Diabetes* (probably because its proportion is low and quite stable).
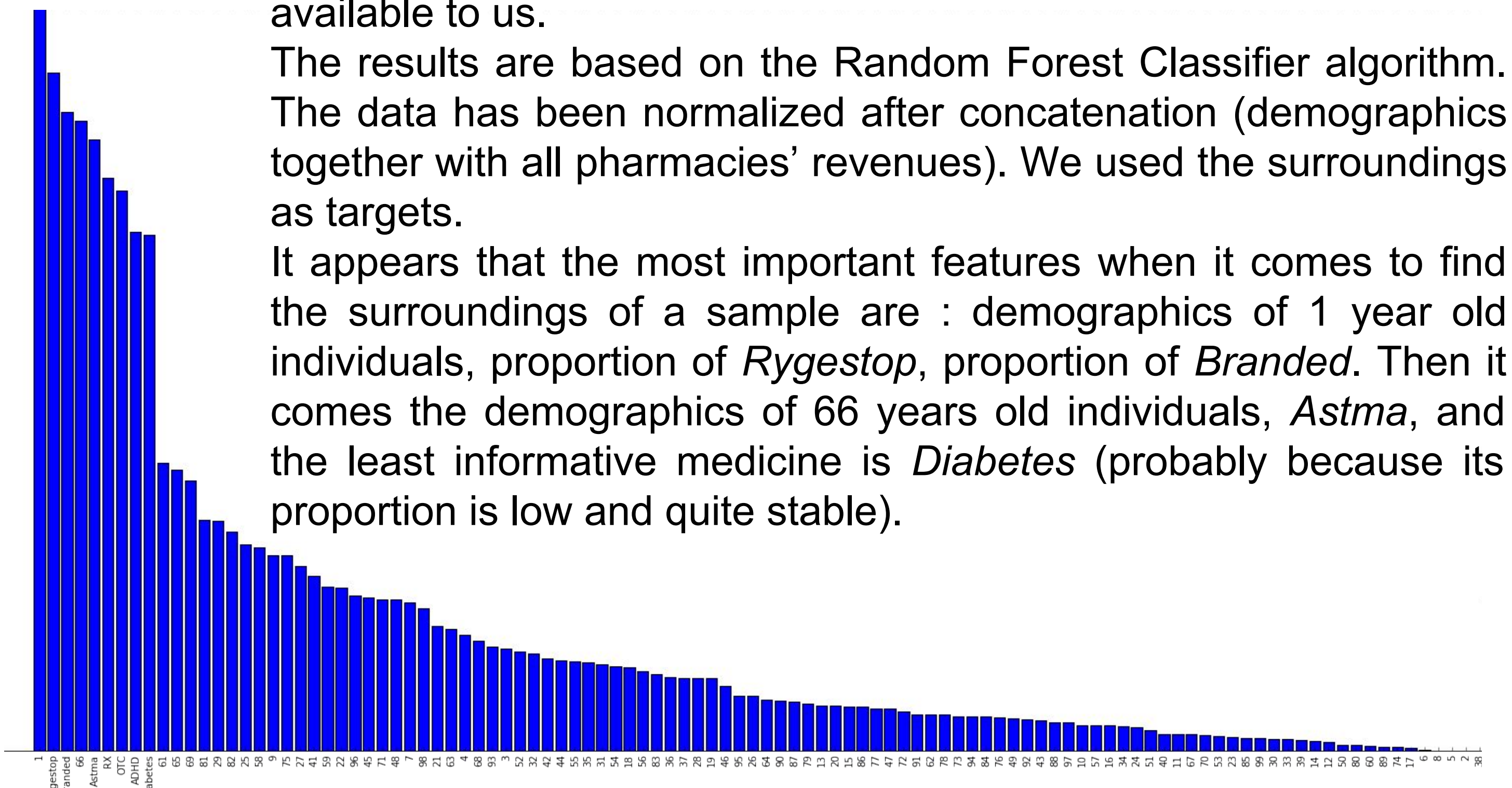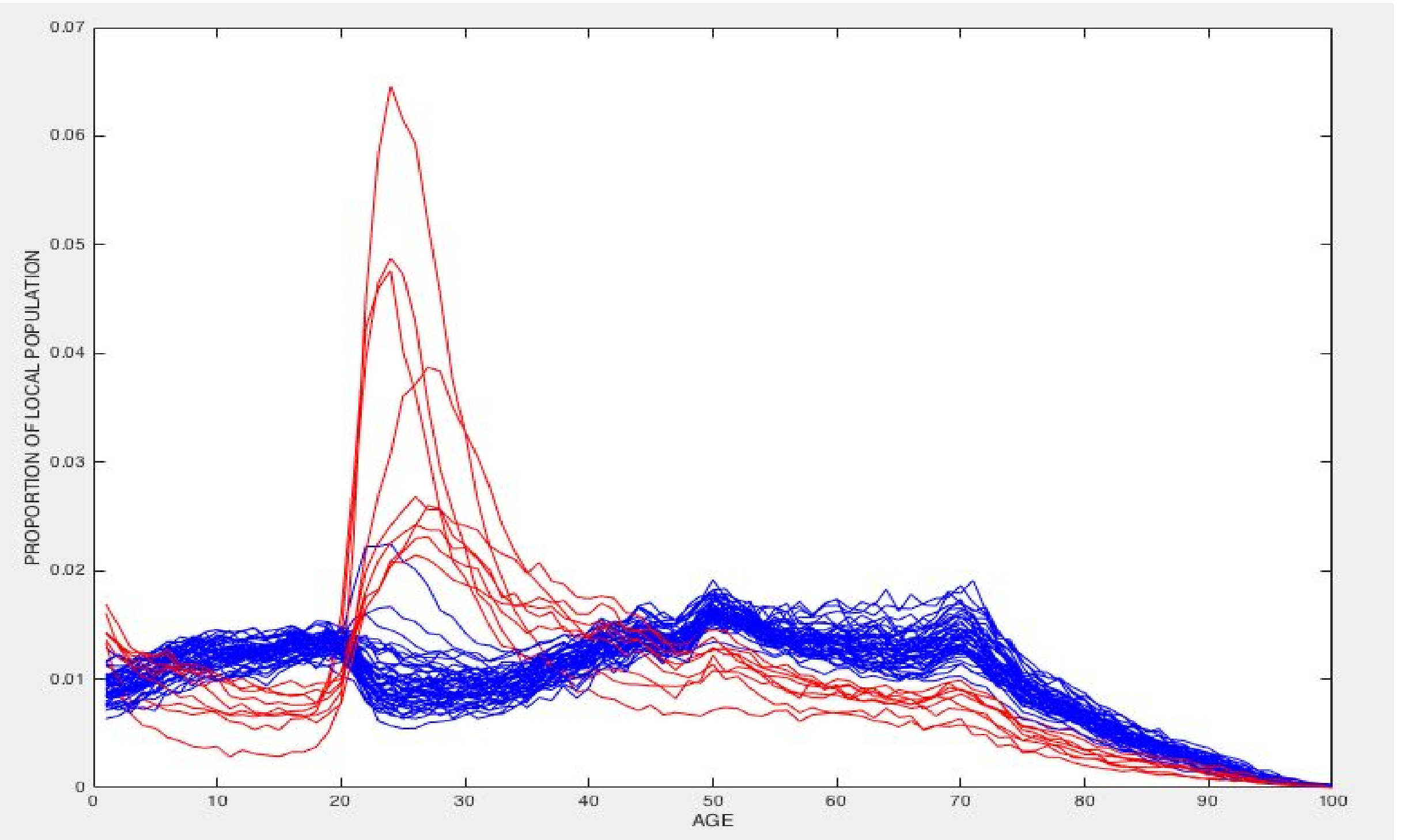

DEMOGRAPHIC STUDY

People of different ages have different health problems, so there is a reason to assume a correlation between the demographic distribution of a region and its medicine consumption.

We have plotted here the demographic distribution of all regions in Denmark, men and women combined, and clustered them into two groups using a K-means algorithm. We could have clustered them without machine learning using that visualisation. We end up with two groups: one with the regions where mostly 20-40 years old people are represented (red), and one with mostly older people and children (blue).

The 371 pharmacies of the dataset have much different total sales revenues. As a consequence, to compare them, it seems more relevant to divide their sales of different types of medicines by their total sales revenues (*OTC* + *RX* + *Branded*).

The proportion of *OTC* and *RX* can then be plotted for each pharmacy, with a red dot if the pharmacy is in a "young" area, and blue otherwise. This gives useful insight regarding the outlier pharmacies.

Additional information can be read on the graphic: the vertical distance from a dot to the line with equation y = 1-x is the proportion of *Branded* products in the total turnover.

This visualisation gives the intuition that the proportion of branded medicines sold in "young" area is smaller. We will explore this idea with more developed tools in the coming weeks.