Patterns of People and Places

Is there any geographical pattern to crime?
Are we defined by where we live?

University of BRISTOL

Author: Kira Kowalska

Supervisor: Nello Cristianini

Overview

We develop a complete **pipeline** for knowledge discovery from geo-referenced* data. We then apply the data mining pipeline to data describing Bristol. The results are analyzed in order to determine whether geography plays a role in shaping social features of an area.

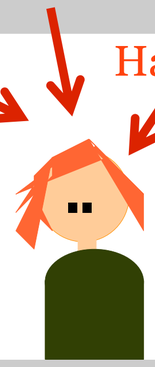
*data describing a location, e.g. average birth rates in British counties.

1 "Data Miner"

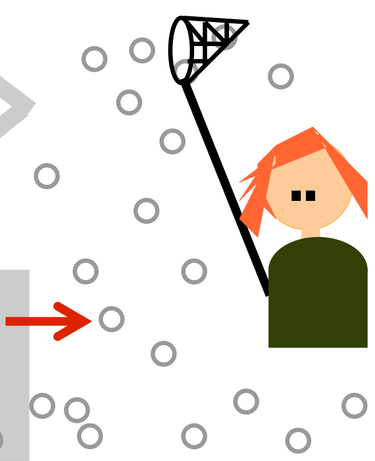
Mathematician

Code Writer

Hacker



Dataset



2 Finding Data

...from **data.gov.uk** initiative releasing governmental data to the public

...form **Census** records

...from **Police** data site police.uk/data

...from data mining **industry**

...form **non-profit** organizations

Each dataset contained information on a different social feature of areas within Bristol, e.g. school quality for each postcode unit.

3 Storing Data

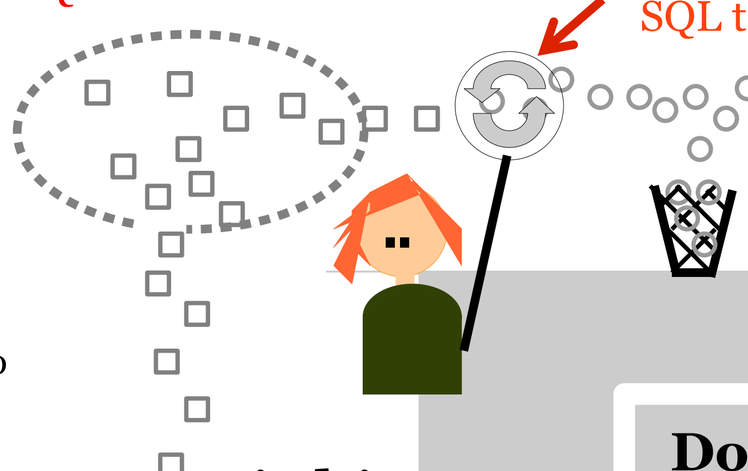
Different datasets used different geographical units as references.

Geography provided a way to link the datasets.

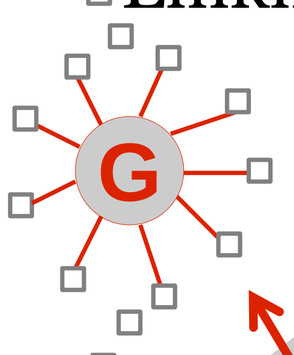
Geo-convertion tables were introduced into the database to convert between different geographical units and hence to make the datasets comparable.

SQL Database

Converting datasets to SQL tables

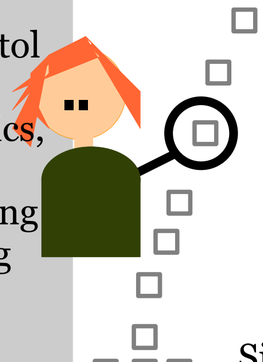


4 Linking Data



Linked Database

The linked datasets formed a rich information base on Bristol social characteristics, which was explored using the following techniques:



5 Data Mining

Correlations

Significant correlations between features of the data were found.

Clustering

Natural groupings in the data were examined using three clustering techniques: k-means clustering, hierarchical clustering, density clustering. The results were compared using average silhouette values. Best results were mapped.

Principal Component Analysis

The main driving forces that generated the data were found using principal component analysis. Principal component maps depicted geographical features influencing social characteristics in Bristol neighbourhoods.

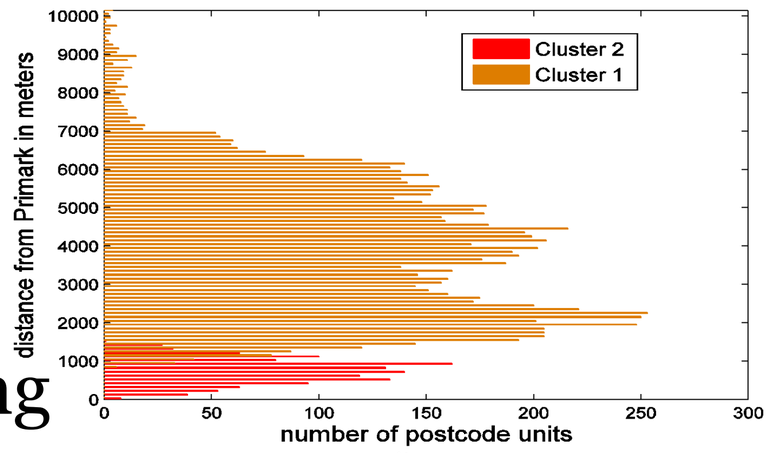
Multiple Linear Regression

Data outliers and geographical factors influencing the data were quantified using linear regression.

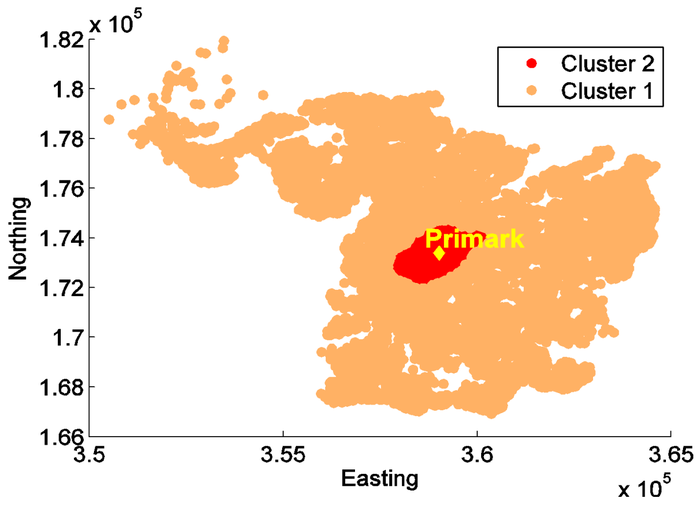
Does geography define who we are?

Clustering techniques applied to the gathered data on social features of Bristol established that Bristol is naturally split into the City Centre and the Suburbs (2). The division was reconstructed with 88% accuracy with a single geographical feature: distance to *Primark* (1). Principal component maps (3) detected geographical borders separating areas with extremely different social characteristics.

(1) Bristol postcode clusters vs. distance to *Primark*.



(2) Bristol postcode clusters mapped.



(3) Second principal component map of Bristol.