Visualizing Swedish election data

Emma Forsling Parborg

Abstract—Election results concerns every citizen and is an important subject. This project aimed to create an interactive web application for visualizing Swedish election results and their correlations with average income and unemployment rate. Data mining has been implemented to find similarities between different municipalities. Several visualization tools have been used to present the data in different ways.

Index Terms—Information visualization, Data mining, Multivariate data, Parallel coordinates, Scatter plot, Donut chart, DBSCAN.

1 Introduction

The initial idea for this project was to visualize data that most people might be interested in. Election results is something that concern every citizen and therefore it was chosen as the basis for this project.

Visualization concerning election results has been done before. Such applications have focused only on the election results, while this project aimed to expand upon this. Data that might have a correlation to the results have been introduced in this project. To investigate similarities between municipalities, data mining has been implemented as well. The final product was an interactive web application.

2 Background and Related Work

Before this project was implemented, different web applications that visualized election results were studied. Two examples that were found were tools from *Valmyndigheten* and *Aftonbladet* [6] [1]. Both illustrated the results in a bar chart with a difference in how much information was available and how it was accessed. *Aftonbladet* used an interactive map in which the user could explore the data by selecting municipalities, or regions within them. *Valmyndigheten* focused on providing more types of data and did not offer the same kind of interactions as *Aftonbladet*.

3 Data

Two data sets have been used in this project. The first data set that is used contains 290 municipalities for three years: 2002, 2006 and 2010. Each data item contains the nine different parties, the average income, total population, number of unemployed residents and unemployment rate. The data is stored in a .csv file, and all datums are strings. Most of the data was extracted from *SCB*, Statistics Sweden [5]. The number of registered unemployed residents in each municipality was extracted from *Arbetsförmedlingen* [2]. The unemployment rate was then calculated using this data and the population data from SCB.

The second data set that is used is the map data for Sweden which is stored in topojson format and was extracted from *Global Administrative Areas* [4]. Each datum in the map data forms a JavaScript object consisting of both the geometry and the name of each municipality in Sweden.

4 Method

The application consists of two major parts: data mining and information visualization tools.

- *Emma Forsling Parborg is studying at Linköping University, Sweden, e-mail: emmfo731@student.liu.se.*
2015-03-18

4.1 Data mining

The data mining algorithm that is implemented in this project is DBSCAN. This is a density based clustering technique that groups together nearby data that is closely packed [7]. Data points in low-density regions are not clustered, and are considered as noise.

Advantages with this algorithm is that it calculates the number of clusters by itself. This is achieved by providing the algorithm with parameters that decide the cluster properties: a euclidean distance threshold and a minimum number of points value.

The algorithm loops through all unvisited data points and looks for all points that are within the distance threshold. Those that are, are called neighbor points. If the number of neighbors is greater than the minimum number of points, the current point is considered a core point. If a neighbor to a core point, doesn't have enough neighbors itself, it is considered a reachable point and will be a part of the cluster too. Points that are not reachable or core points are considered as noise.

In this project the minimum number of points was set to 3 and the distance threshold was set to 0.5 using normalized data. The data items that were accounted for in the clustering where the election results and nothing else.

4.2 Information visualization tools

To display the information, a scatter plot, a parallel coordinates plot, an interactive map and a donut chart were used. These tools display the data in four different ways. Each tool collaborates with the others, so when the user selects a municipality in one view the choice is displayed in the other views.

In the related work mentioned previously, bar charts were used to display the election results. In this project, a donut chart was chosen to replace this technique as it more distinctly portrays to the user that the sum of all parts constitutes 100 percent [10]. A drop-down list containing all municipalities is located near this chart, in order for a user to select a municipality. If nothing is selected the donut chart will display the overall election results for the entire country, see Figure 1. When a municipality is selected, the name and year will be displayed in the center of the donut chart. By default the year is 2010, and the user can change it from a drop-down list located to the left of the municipality list. A tooltip for each part of the donut is provided on mouse hover, which displays the party name and the its election results. A legend with the colors for the parties is also provided and is displayed to the left of the chart. To the right of the chart the average income and unemployment rate for the municipality are displayed.

The interactive map is used for displaying geographical data and provide another means of navigation [11]. A satellite projection has been used on the map to make the country look like it has been tilted, with the purpose to decrease the amount of space it occupies on the screen. When a user chooses a municipality the area is

highlighted with a distinct color and the map is centered on it. When a municipality is deselected, the map will return to its starting-point. At the top left corner of the map, two radio buttons are displayed which enable the user to switch the color of the map. Each municipality on the map is by default colored after the party that has the majority of votes. The user can choose to have it colored by clusters. This can be seen in Figure 2.

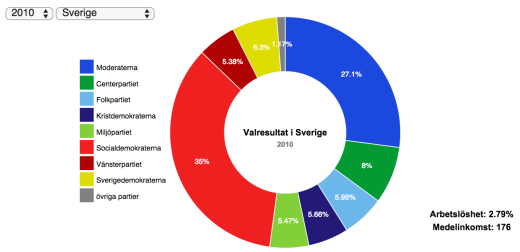

Fig. 1. The view displaying the dount chart and two drop-down lists. In the first one, the year can be changed, and in the second, a municipality can be chosen. In the middle of the image lies the donut chart, displaying the overall election results in 2010 for Sweden. To the left, a legend displays which parties the colors represents, and to the right the unemployment rate and average income are displayed in text.

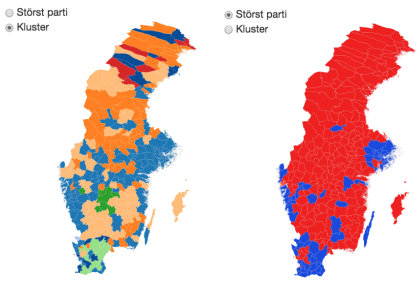

Fig. 2. The view displaying the interactive map and the two radio buttons for changing colors. This image displays the election results in 2010. To the left the map is colored after the clusters and to the right it is colored after which party that has the majority of votes.

The scatter plot was used to display the unclustered data, see Figure 3, since it is a conventional approach to represent bivariate data [11]. Two drop-down lists were used for the scatter plot, one for each axis, in which the user can choose which parameter that should be displayed. Each dot represent a municipality, and the color of it is the same as in the map. A tooltip, which is activated on mouse hover, is provided for each dot, displaying the corresponding municipality and the axis values.

The parallel coordinates plot with reorderable axes was used to display the clustered data, since it is a popular and valued technique for representing hypervariate data [9] [8] [12]. Each line represents a cluster that has been calculated with DBSCAN for each year and is colored individually. From the first data set, all data items except for the number of unemployed residents and the total population are displayed, since these data items where used to calculate the unemployment rate. When a line is selected, the opacity for the remaining lines are decreased. Brushing, a simple but powerful technique [11], is used on the axes, so that the user can select areas of interest and see which lines that are within them.

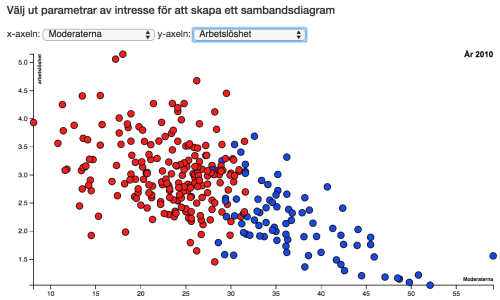

Fig. 3. The view displaying the scatter plot and two drop-down lists in which the parameters for the axes can be changed. This scatter plot displays the election results in 2010, with the parameter *Unemployment rate* on the y-axis, and the percentage of votes for the party *Moderaterna* on the x-axis.

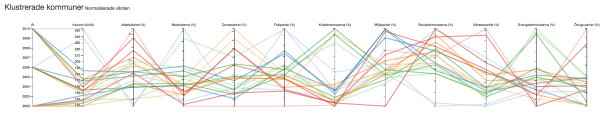

Fig. 4. The view displaying the parallel coordinates plot, containing the clustered data. From left to right, axes displaying: Year, Average income, Unemployment rate, *Moderaterna*, *Centerpartiet*, *Folkpartiet*, *Kristdemokraterna*, *Miljöpartiet*, *Socialdemokraterna*, *Vänsterpartiet*, *Sverigedemokraterna*, *Övriga partier*.

5 Implementation

The project is implemented as a web application, using mainly JavaScript and the d3 library. The advantages with creating a web application is that it works on multiple platforms. The d3 library enables visualization of large data sets in web applications [3]. This library facilitates the creation of different visualization tools. There are however disadvantages as well, concerning the rendering speed, since all the graphics are implemented in software.

6 Results

The final application is interactive and consists of four different views, as previously mentioned, which are interconnected and present the information about a municipality in different ways.

If the user chooses a municipality, either from the map or from the drop-down list, the region is highlighted in all the views. Information about the chosen municipality is shown in the donut chart. The parallel coordinates plot tells the user if it belongs to any cluster. The scatter plot, if axes are chosen, displays how it compares to other municipalities.

If the user chooses a cluster in the parallel coordinates plot, all municipalities belonging to that cluster are highlighted on the map. The values for the normalized cluster can be seen on the axes. To derive the exact values for each line a tooltip is provided on mouse hover, which describes the normalized cluster data and which municipalities that belong to that cluster.

7 Conclusions and Future Work

The clustering of data using the DBSCAN algorithm was successful. Distinct clusters have been found, in which certain results separate them from the other municipalities. For example Figure 2 clearly presents two green clusters. The bright green cluster corresponds to a relatively high representation for the party *Sverigedemokraterna* and the dark green for the party *Kristdemokraterna*.

The data mining is only implemented on the election results for each year. Another approach that would be interesting to implement is to enable temporal clustering by keeping the data for all the years together. The data mining does also not account the average income and the unemployment rate. This is something that could be interesting to further develop and also adding other aspects like education and population density.

The donut chart has the disadvantage that it is harder to compare the pieces that are not adjacent. However since the election results are printed on each part of the donut and the tooltip also provides the same information, this disadvantage is minimized.

The colors for the clustered data have not been colored in a way that describes to the user what makes them similar. They are not colored after parties, but might be interpreted as if they were. A user might think the orange color is a mixture between the parties with the colors red and yellow. This flaw is something that can be improved in future work.

Concerning the map projection, the advantage with this approach was that it preserved screen space. However, the projection altered the shape of the municipalities in different degrees, which gives the country an odd shape that might be perceived as incorrect rather than tilted. The map projection is therefore something that could be improved in future work, to produce a more accurate portrayal of the country.

REFERENCES

[1] Aftonbladet så röstade sverige i valet 2014. http://www.aftonbladet.se/nyheter/valaret2014/riksdag/Valresultat/. Accessed: 2015-03-04.

[2] Arbetsförmedlingen tidigare statistik. http://www.arbetsformedlingen.se/Om-oss/Statistik-prognoser/Tidigare-statistik.html. Accessed: 2015-03-14.

[3] D3 data-driven documents. http://d3js.org/. Accessed: 2015-03-04.

[4] Global Administrative Areas boundaries without limits. http://www.gadm.org/country. Accessed: 2015-03-04.

[5] Statistiska Centralbyrån statistikdatabasen. http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/?rxid=6d8e2509-93b0-4e35-bdea-acc8724ae974. Accessed: 2015-03-14.

[6] Valmyndigheten val till riksdagen - röster. http://www.val.se/val/val2014/slutresultat/R/rike/index.html. Accessed: 2015-03-04.

[7] M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI*, pages 226–231, 1996.

[8] A. Inselberg. Multidimensional detective. In *IEEE, Proceedings of Information Visualization*.

[9] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.

[10] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive P*, 5(1):61–77, 1991.

[11] R. Spence. *Information Visualization Design for Interaction*. Pearson Education Limited., England, 2007.

[12] E. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.