Information Visualization

Julius Kördel

Abstract—Spritvis is a tool that allows a user to explore sales data and alcoholic habits in a interactive way. The tools purpose is to spread awareness of alcohol sales and habits of Sweden and can serve as an complement of the yearly reports provided by Systembolaget.

Index Terms—Multivariate data,Categorical data, Alcohol, Parallel coordinates, Choropleth map, Line graph.

━━━━━━━━━━━━ ✦ ━━━━━━━━━━━━

1 INTRODUCTION

In Sweden the government has a monopoly on the distribution of alcohol in stores. The state owned company is named Systembolaget Aktiebolag. Except selling alcohol Systembolaget also works with informing about the risks associated with alcohol. This project investigated a way to explore and visualize data related with Swedes and alcohol. To do this several data sets and visualization methods are combined.

2 BACKGROUND AND RELATED WORK

Systembolaget does on a regular basis publicize papers and rapports on alcohol consummation and sales statistics. These often comes with basic but precise visualizations for the given context and focus on interesting findings. There exists visualization of Systembolagets products but they tend to be for the purpose of finding similar drinks in taste/style or comparing taste against price. This application visualize the data about sales and habits but also enables the user discover the data themselves something that field of alcohol related visualization related to Sweden lacked.

3 DATA

Systembolaget publicize their sales statistics in volume for different alcohol groups on a quarterly basis [5]. The groups are: cider/cocktails,non-alcoholic drinks ,liquor, beer and wine. All data sets used in this project covers the time span from 2006 to 2017. Which means that the sales statistics contains 6336 entries.

The data set about sales is complemented by 2 others. The first being the average population for counties over a year resulting in 252 entries. This data is distributed by SCB and it is SCB that does statistical work on behalf of the government. The second data set is alcohol habits for counties over same time span as the other data. The data is obtained through the "Nationella folkhälsoenkäten" and it is published by Folkhälsomyndigheten [4]. The data from this data set that was chosen to visualize contains a total of 1260 entries.

4 METHOD

To visualize the data there are multiple tools working together.The application has 4 sections, each 1 with its own purpose/ way of looking at the data and allowing for interaction in each section as recommended by [7]. The first section is for numeric values for each quarter, the second section is a map, the third section is for how sales changed over time but also the alcohol habits and the last section is for parallel coordinates for correlation of different alcohol groups.

4.1 Geographical data

Sales data for counties is well fitted to be displayed with a choropleth Map. A choropleth map is representing data values with a coloring to the area depending on the data value for the area it represent. A big factor for this visualization is how the colors are picked together

---

- *Julius Kördel student at Linkoping University, Sweden, e-mail: julko800@student.liu.se .*

with the corresponding numeric spans these color represent in the data. Through ColorBrewer [2], the color for each span can be picked with a equidistant perceptive distance meaning that the change between all colours are interpreted as equally big.The spans were set through trial and error and the final method is based on finding the largest value for respective category/type for the selected year. Then the spans was chosen depending on the max value. All spans were equally big except the final span that had doubled the size of the other spans,the spans range can be seen in 1.

Span1	0 <x <maxvalue * (1/6)
Span2	maxvalue * (1/6) <x <maxvalue * (2/6)
Span3	maxvalue * (2/6) <x <maxvalue * (3/6)
Span4	maxvalue * (3/6) <x <maxvalue * (4/6)
Span5	maxvalue * (4/6) <x

Table 1. Spans for colors in map

This methods was chosen over just dividing the spans in fifths of maxvalue because it resulted there being more spread in the coloring of counties,over all years and alcohol groups in the data.

4.2 Changes over time

To display quantitative values over time a technique is to use a line graph, plotting data against time as mentioned in [6] and can be seen as state of the art. For this application the line graph display the how the sales for a selected type has changed over all the years in the dataset(2006-2017) for the selected county. The y-axis goes from 0 to a numeric value picked so that the data fits into the graph. The y-axis changes from county to county but this is because readability of how it has changed over time with regard to itself was prioritized in this view, over how it changed compare to others.

4.3 Habits

To visualize the alcohol habits for each county the first approach was to use symbols placed on each county representing its habits. But because of clutter to fit the 3 types of habits into every county on the map a other approach was chosen. This method was to display habits with a bar filled with color. Each colored part was represented a habits percentage of the data for a selected county and to make the percentage more clear the parts contains a number and percentage sign.

4.4 Numeric values and menus

To give the user a numeric overview of a selected county's 4 quarters numeric values for the picked alcohol group is shown in a table. It is also in the section for the table the color coding for the map is explained aswell as were the user chose alcohol group and year to be display through dropdown menus.

4.5 Correlation

Parallel coordinates is a technique used for multivariate data. Even doe all 6 types of alcohol sold is in volume they are distinctly different

from one another in type of product. The parallel coordinates is used to visualize the data for all quarters of the year to show the correlations between different alcohol types for all counties and or between quarters. Each quarter is drawn with a unique color for easier separation of data. If the user wants to find a all 4 related lines to a county in the parallel coordinates, it can click on the county in the map and its 4 lines will be highlighted in the parallel coordinates. If a user just want to see a line clearer or wants to know which county it belongs to it can hover the line and its name will be displayed and its line highlighted and moved infront of all other lines.

5 IMPLEMENTATION

The data has been through a pre-process of converting google spreedsheet data into json format though a pythonscript using the gspread api [1]. This was done for full control of how the data structure the data is stored in and converting data into the format json. The sales data could be divided into year that contains counties that contains sales data for that county and then each type and its value for that county. For same reason process was made for the other data as well.

The product is produced for usage on desktop environment. This is because when visualizing data the need for readability of figures and graphs is made easier for larger screens than for small such as mobil devices. The visual implementations was done with the javascript api d3.js. [3].

6 RESULTS

The project resulted in interface that displays the data in different ways in different sections on the screen. The user will itself choose which data that it wants to look at through interaction.The user decides if it want to look at the sales data with regards to the only volume 1 or with regards to volume per persons in that county 2. The map view is often different with regards to population compared to total volume as seen in the figures for all years and groups as seen in the figures.
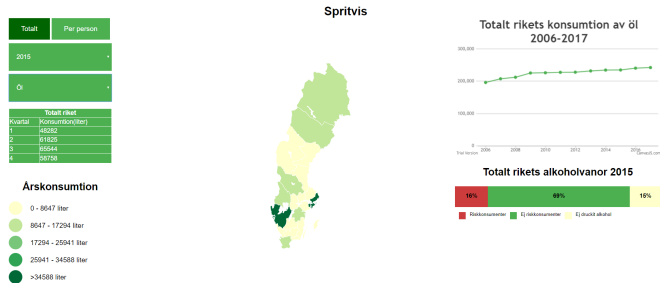

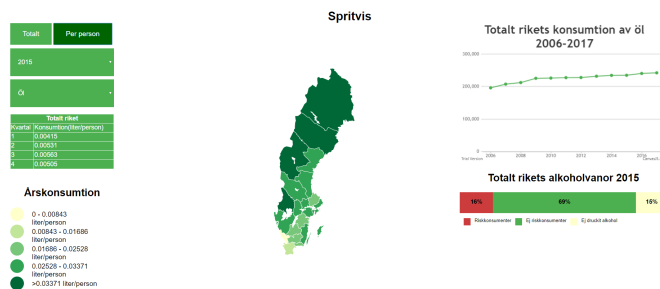Fig. 1. Overview of spritvis with regards to total sales

Fig. 2. Overview of spritvis with regards to sales per person

Which county the different sections such as the bar, line graph and tables are displaying is selected through clicking on the county in the map. Thus demanding that the user has a geographical knowledge or

that the user clicks on one of the counties lines in the parallel coordinates. Which can be hard to find the correct one and time consuming.

The implementation for the parallel coordinates does only display sales per county and not per person in that county. Preferably the possibility of displaying per person should affect this data aswell. It also lacks the possibility to to brush data or zoom which makes it hard to see the clutter areas in the lower areas of the graph 3. It does allow moving the axis which is the way the tool should be used to see the correlation between two different sales of alcoholic groups.
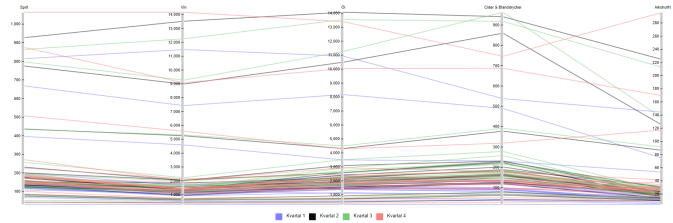

Fig. 3. Parallel coordinates with sales for different alcohol groups for each quarter for the selected year.

7 CONCLUSIONS AND FUTURE WORK

The resulting product allows for a exploitative interactive experience with the data about sales and alcohol habits and can be used to spread awareness.To further increase the accessibility and potential of spreading this awareness, the application solution should be reconstructed for mobile use. This will come with the challenge of showing alot of data on small screen. Therefor alternative and or complementary visualization techniques should be used or reduce the total information showed on the screen at one point, but the reward is a more accessible product.

There is also needed to find a proper method for choosing the numeric spans which decides the color for the map aswell as setting the axis for the line graph for a more consistent experience. This is needed because the current implementation is each year has different color spans and axis length. The current implementation can lead to misleading visualizations and a decrease in the ability for the user to draw correct conclusions with regards to the data show by the product.

For understanding if the current techniques for visualization is good with regards to create awareness,which is its purpose. The interface need to undergo testing.

Lastly if accessible a expert in the subject of alcohol sales in Sweden and alcohol habits should be used to tailor the spans and ranges for map and the graphs. This would most likely contribute to a more justified visualization of the data and increase its credibility.

REFERENCES

[1] burnash,Gspread cgoogle sheets python api. `https://github.com/burnash/gspread`. Accessed: 2019-03-12.

[2] Colorbrewer cynthia brewer, mark harrower and the pennsylvania state university. `http://colorbrewer2.org/`. Accessed: 2019-03-12.

[3] D3js javascript library. `https://d3js.org/`. Accessed: 2019-03-12.

[4] Folkhälsomyndigheten alkoholvanor (självrapporterat) efter ålder,kön och år.andel (procent). `http://fohm-app.folkhalsomyndigheten.se/Folkhalsodata/pxweb/sv/B_HLV/B_HLV___aLevvanor__aabLevvanoralk/HLV_Alkoholvanor_alder.px/?rxid=eefcc3a5-a7fe-4116-867a-8cfaeefa0bec`. Accessed: 2019-03-12.

[5] Systembolaget försäljningsstatistik. `https://www.omsystembolaget.se/om-systembolaget/foretagsfakta/forsaljningsstatistik/`. Accessed: 2019-03-12.

[6] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE TRANS. ON VISUALIZATION AND COMPUTER GRAPHICS*, page 3, 2008.

[7] R. Kosara, H. Hauser, and D. L. Gresh. An interaction view on information visualization, 2003.