IN 4086-14 Data Visualization
InfoVis Project:
Group 26
Trends in Alcohol Consumption

Dimitropoulos Georgios: 4727657
Manousogiannis Emmanouil :4727517
Mastoropoulou Emmeleia: 4743539

December 21, 2017

1 Introduction

Data visualization is about making useful pictures. It tries to complement numbers as they are not able able to tell the whole story. Data Visualization has three main goals. The first one has to do with the exploration of the data. The second is associated with the analysis of hypotheses. The last one is about providing a visualization of the results. Information Visualization (Infovis) which is a subset of Visualization, is about non-spatial data. The project which we worked on, belongs to this category. Alcohol consumption is a field of tremendous significance, as this information can be utilized by a variety of organizations. For instance, producers of a specific type of alcohol could use all these data towards making profitable decisions. The purpose of the utilization of such data, is to monitor the magnitudes, patterns and trends of alcohol consumption in the adult population. Hence, in this project we provide this ability through a proper interpretation of our results. Our basic goal is to enable the user with the ability to examine the geospatial relationships, patterns, correlations, similarities and outliers of different types of alcohol consumption per country, for a fifteen year period [2000,2014]. Our visualization design consists of a choropleth chart, streamgraph, line-chart and Pie-chart. Finally, this project belongs to the second variant of the "Data Visualization Assignment 01" through using the D3.js JavaScript library [1].

The rest of the paper is organized as follows: Section 2 refers to the the description of the task by including the design space dimensions proposed in [2]. Section 3 presents the Visualization techniques which we used in our project, namely the choropleth chart, streamgraph, line-chart and Pie-chart. Section 4 provides information about the datasets that we used and the processing operations that we performed on them, whereas Section 5 analyzes the basic design choices of our source code. Section 6 summarizes our work, where interesting conclusion and future work proposals could be depicted. In Section 7 the way of our work is mentioned. Finally, references of suggested bibliography which has been studied in depth are mentioned.

2 Description of the Task

The definition of of the task in our project follows the design space dimensions proposed in [2]

Why?

Why is the user looking at it? This part of the analysis framework has to do with the task abstraction. Through this task, the goals of the user are split into three main parts. The first part has to do with high level actions (the analysis of the data). The second part part associates with middle level actions (search of the data) and the third and last one part is about low level actions (queries). Through our approach we enable the user to compare different countries or regions and indicate geographical spatial differences of alcohol consumption . Our main goal is to give the user the possibility to explore the data and the ability to interconnect different graphs e.g.,alcohol consumption per country/region and per year or detect trends, patterns, outliers and correlations between the data.

What?

What is shown? This part of the analysis framework has to do with the task abstraction. We study the alcohol consumption of four types of alcohols namely the spirits, beer, wine, other alcoholic beverages and the aggregate consumption of these ones. We are focused on the time interval between 2000-2014 for 193 countries.During this information visualization project, we would like to create an exploratory view on consumption of the recorded alcohol per capita. More complex patterns such as correlations and similarities can be found through an exploration of this visualization project by the user.

How?

How is it shown? This part of the analysis framework has to do with the visual encoding and interaction task. We included four different visualization encodings namely a choropleth chart, a line-graph, a pie-chart and a stream-graph. Also, an interaction between the first three of them is provided. More specifically, the user can select different attributes to display from a drop down list through a selection list of wine, beer, spirits, other alcohol beverages and a total of all types in the first HTML page. By, this selection, the user with the mouse hover action, can display a line graph and a pie chart. Each one of them shows specific details of each country. In detail, we chose the pie-chart to display the percentage of each type of the recorded alcohol per capita consumption. At the same time, the user can see a representation in line-graph where we display the consumption during the years. In this way the user can easily change between the attributes and observe similarities through comparison. Moreover, in he second HTML page, through the streamgraph, we give the opportunity for a thorough view at the consumption of all types of alcohols, being clustered per region.

3 Visualization Techniques and Design

Our data consist of the alcohol consumption per type,per year and per country. As mentioned before, we study the alcohol consumption of four types of alcohol namely the spirits, beer, wine, other alcoholic beverages and the aggregate consumption of these ones. We are focused on the time interval between 2000-2014 and for 193 countries. Also, we calculated the mean values of each type of alcohol for each country for this 15-year period. In addition, our data can be categorized using Munzner's categorization technique [3]. More specifically, countries,id of the countries, WHO regions and types of alcohol are categorical attributes whereas the consumption of alcohol, the mean value of the consumption and the year that this consumption had occurred are all sequential quantitative attributes.

The basic information visualization techniques that we use for the visual encoding of our data are the choropleth chart, streamgraph, line-chart and Pie-chart as they are the most suitable for the data which we want to represent. Also, these techniques fit perfectly for the patterns among these data that is desirable to be explored.

3.1 Choropleth Chart

The central task in our work was to understand the spatial relationships, as far as the alcohol consumption is being concerned, among different countries. Hence, the selection of the choropleth chart (Figure 1) is the case in our design choice. Also, our data consist of quantitative attributes (alcohol consumption per type and per year) per region (country). The encoding that we have adopted follows a sequential segmented color-map. The choice of the colors which we used was done after a careful study of [4] and [5].Also, the final choice of the colors had done through exploring color brewer [6]. To be more specific, we chose as basic colors the red, yellow, blue, green and gray. Each of these colors represents the consumption of alcohol per type. Namely, the blue color represents the consumption of spirits, whereas the red one the consumption of wine, the yellow the consumption of the beer, the green one the consumption of other alcoholic beverages and finally the black color reveals the total consumption of all the aforementioned types of alcohol. In order to point out the difference between the values of the alcohol consumption we decided to represent the higher values with a deeper hue in contrast to lower values which are represented by lighter hues. More specifically, we assigned the minimum value of the alcohol consumption of a country to the lighter hue of the specific color that represents this type of alcohol. Similarly, we assigned the maximum value of the alcohol consumption of a country to the deeper hue of that color. The exploration of the relationship between the geographical location and the respective consumption of alcohol is our primary goal. Hence, by looking at the choropleth chart and taking into account the representation of the consumption of alcohol through the above procedure, it is easy to detect the distribution of alcohol among countries.

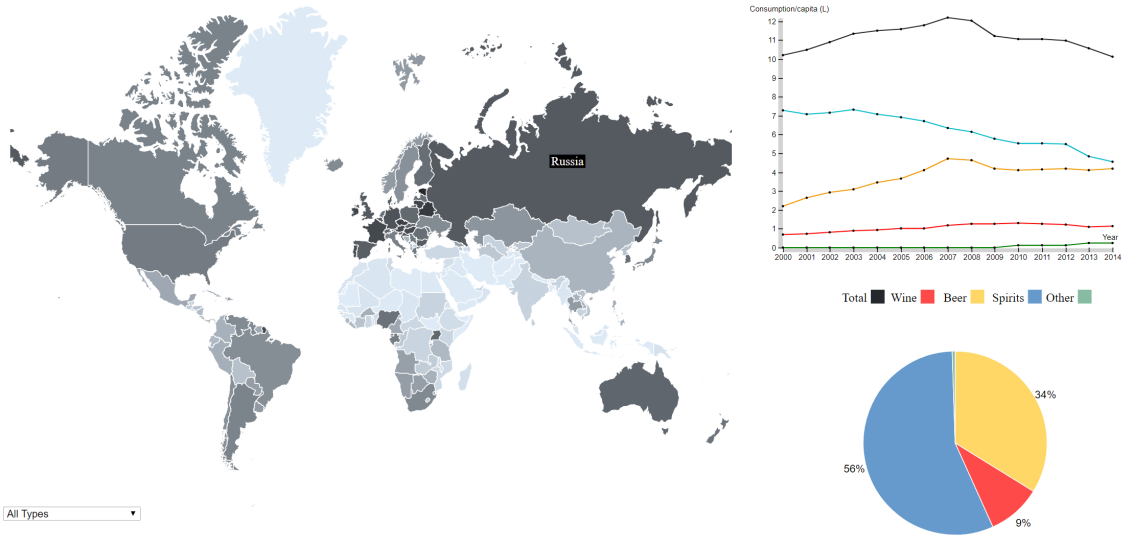

Figure 1: All Types Consumption

3.2 Line-Graph

Our second design choice was a line-graph (Figure 2). In the line-graph, we have a key which is the year that a specific type of alcohol consumption has occurred and one value which represents the exact amount of this consumption. Both these key and consumption values are quantitative attributes which make them a perfect combination for the line-graph representation. We included marks (points) which imply the alcohol consumption for a specific year a and line connection between them as a link. We decided to give a different color in each of these lines (depends on the the type of the alcohol beverage) in order for the results and comparison between them to be more distinguishable. As fas as the channels are being considered,they control the appearance of the marks, align lengths to express quantitative values and are separated and ordered by key attributes into horizontal regions.The shape of these marks are circles and its position implies the corresponding value (higher position implies higher value of consumption). Our main goal, that lead us to the usage of a line-chart, is that through this visualization technique, the detection of the trend of the alcohol consumption through years is an easily manageable task. The connection marks emphasize the ordering of items along the axis that encodes the years, by explicitly showing the relationship between one item and the next one. Hence, this visualization, allows us to study interesting patterns.To conclude, the reason that we chose a line-graph instead of a scatter plot, is that the last one is a better choice for discrete data points, which is not the case in our problem as time is a continuous quantity.

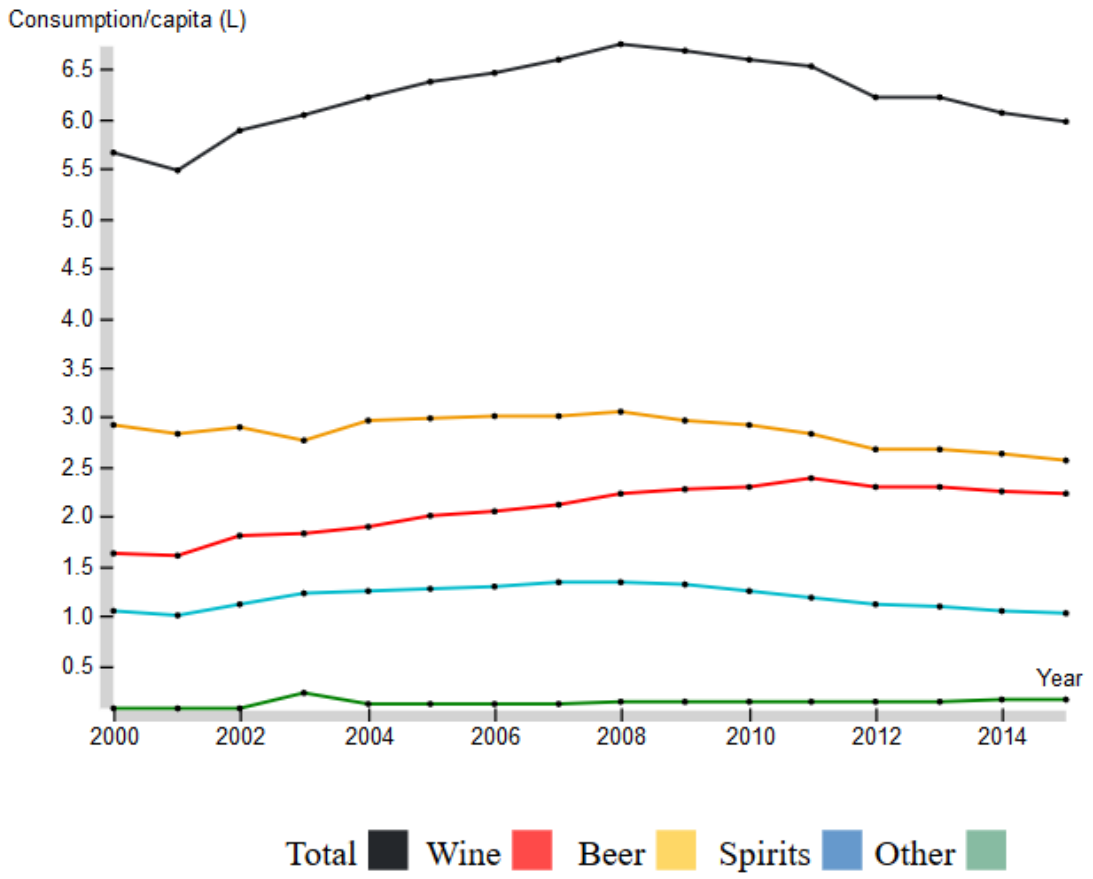

Figure 2: LineGraph, All Beverage Types

3.3 Pie-Chart

The third visualization technique that we use is a pie-chart (Figure 3). This design choice encodes the average consumption of each type of alcohol for each country in the time interval [2000,2014]. It has one key attribute which is the type of alcohol for one specific country and one quantitative value attribute which is the consumption of this type of alcohol. Our main task through this technique is to provide an indication for the part-to-whole judgment for the distribution of all types of alcohol in the aforementioned time interval. Pie-chart utilizes as a channel the angle of each of its components. Finally, the reason behind this design choice has a 2-sided purpose. Firstly, we wanted to provide a thorough view of the average distribution of the alcohol consumption over a large time-period (fifteen years). Secondly, we decided to include both a pie-chart and a line-graph as the first of these two techniques, provides much less accuracy as it indicates a percentage and not the exact values of the alcohol consumption.

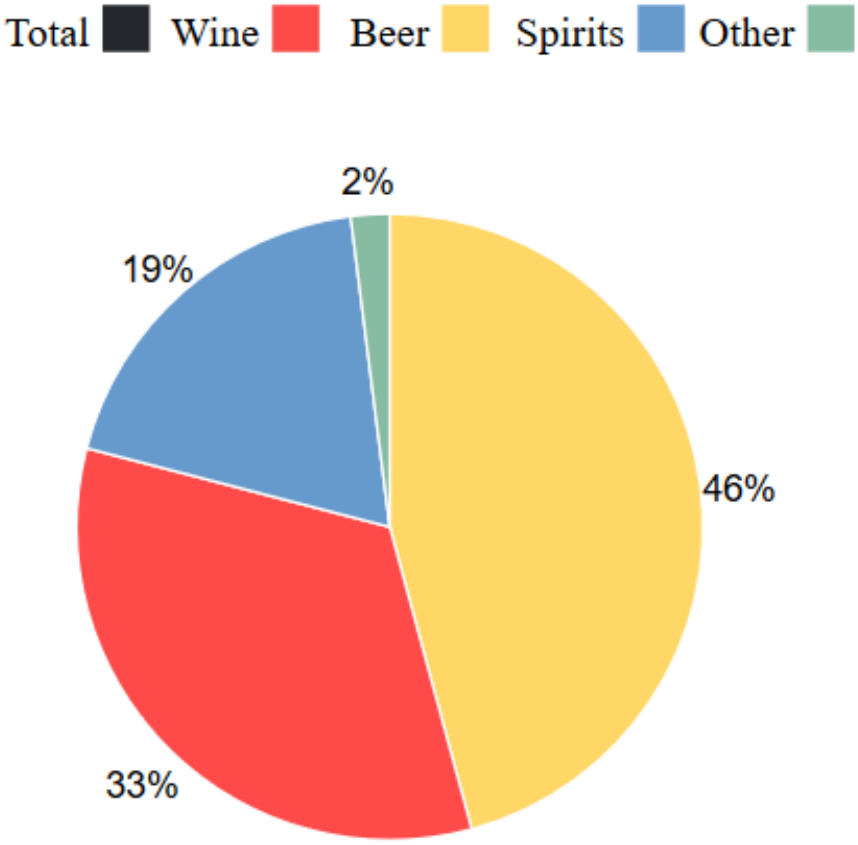

Figure 3: Percentage of alcoholic beverage consumptions

3.4 Stream Graph

Our forth design choice was a streamgraph (generalized stack-graph) (Figure 4) which emphasizes the horizontal continuity vs vertical items. In order to visualize the above streamgraph a clustering per region was performed in the initial dataset by using basic excel functions such as ordering, sum and average. Our goal it to provide the user the ability

to have a thorough view at the consumption of all types of alcohols per region. More specifically, our visualization represents the regions of Africa, America, Eastern Mediterranean, Europe, South-East Asia and Western Pacific. This visualization technique fits perfectly to our dataset as we have one categorical attribute(the regions), one ordered key attribute (years) and finally one quantitative value attribute namely the average consumption of all types of alcohol for the time interval [2001,2012]. This specific interval was chosen as does not contain any outliers (such as empty entries) which facilitates the procedure of representation by a streamgraph. It can be depicted from Figure 4 that the geometry of the streamgraph, namely the height of the layers encodes the the average consumption of all types of alcohol. If a layer is higher than another layer, then this means that the average consumption of all types of alcohol is higher for the this specific region, and for the year that the user has selected to study (by hovering the mouse above this year). Hence the derived data that is available for the user include this layer ordering. Finally, the main reason that we chose the visualization technique of the streamgraph in combination with the line-graph, is that we want to derive interesting results not only for a specific country per year(line-graph) but also for a whole region(streamgraph).

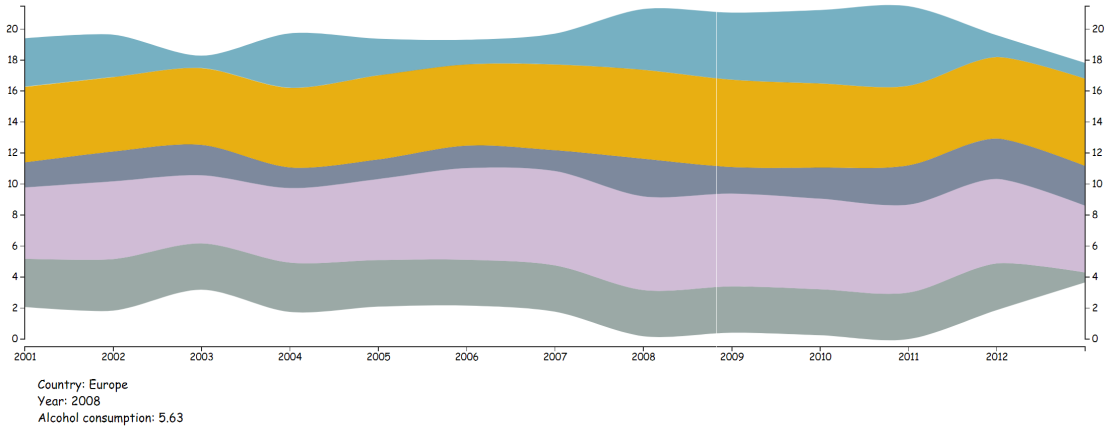

Figure 4: Stream graph representation

3.5 Visual Encoding and Interaction

Our last design choice was to include the capability to the user for a navigation through the above techniques.The design choices that we included are the click and hover operations. Our purpose is the highlight of some specific data, by changing the visual encoding for selection targets like their color, hue and explicitly links between them (Figure 5 and Figure 6). Hence, the navigation of the user into these two HTML pages which we constructed, facilitates the change of their view point and includes operations like slice (show only items matching a specific value for a given attribute, like the alcohol beverage type), cut (show only items within some regions, like a specific country in the choropleth chart or a specific region in the streamgraph). More specifically, Figures 5 and 6 illustrate the choropleth chart along with the scatter plot and pie-chart. There exists a drop down-list in the bottom of the page, where provides the capability to the user to select and study a specific alcohol beverage type among five options. In the sequel, by hovering the mouse over a specific country is able to see the name of that country in the choropleth chart, its alcohol consumption (and its trend) for every year in the interval [2000,2014] in the line-graph and finally its average consumption for this specific beverage type for the given fifteen year interval. In addition, we included a mouse hover operation in the streamgraph as well. The user has the ability to mouse hover a specific year and study the consumption of all types

of alcohol in any of these six regions. Also, by hovering the mouse over a specific region, the layer that represents this region becomes more deeper in its hue in order to facilitate the view point of the user. Taking everything into consideration, our visual encoding and interaction allows the user to interact with these plots through dynamic queries (country, year, region, alcohol beverage type) which promotes the exploration of the data through their processing/filtering, achieved by the above operations.

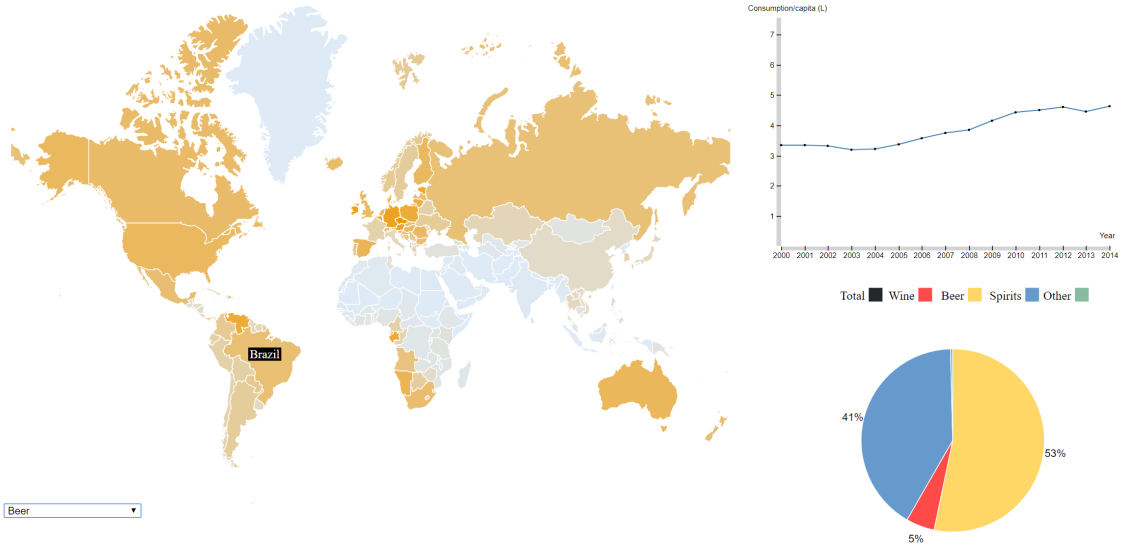

Figure 5: Beer Consumption

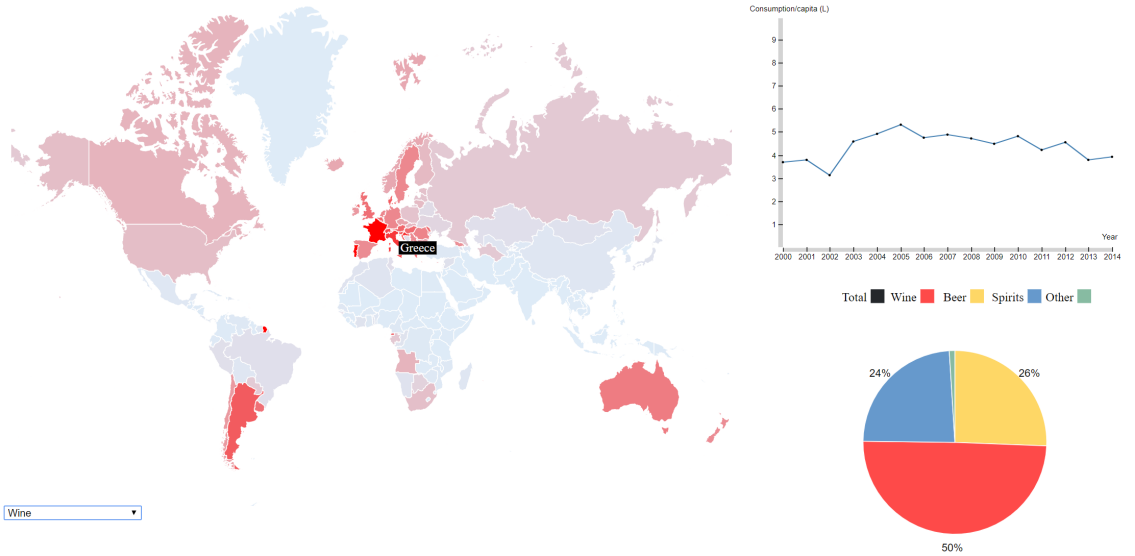

Figure 6: Wine Consumption

4 Data selecting and processing

The World Health Organization was created in 1948 in response to the direct and coordinate international health within the United Nations' system. The data set used for this project was provided by World Health Organization. It contains information about the recorded alcohol per capita consumption, from 2000 until 2014 for almost all countries. The recorded alcohol per capita is part of a core set of indicators, whose purpose is to monitor the magnitudes, patterns and trends of alcohol consumption in the adult population. Specifically, recorded alcohol per capita is defined as the recorded amount of alcohol consumed per capita (15+ years) over a calendar year in a country, in liters of pure alcohol. The indicator only takes into account the consumption which is recorded from production, import, export, and sales data often via taxation. The dataset that was used for the project focused on their consumption in beer, wine ,spirits, other other alcoholic beverages and an aggregation of them [7]. The data is split over multiple CSV files each containing information about one attribute and properly used by the two HTML files.The preprocessing of our data does not include any data reduction or aggregation as we wanted all the available information in order to draw interesting and important conclusions. However, a processing in the initial dataset by using basic excel functions such as ordering, sum and average has been performed in order to optimize the operations of our code.

5 Code Structure

Our visualization techniques were implemented using d3 library of JavaScript. We created two HTML pages, where we display the streamgraph in the first and the choropleth chart with the linked line-graph and pie-chart in the second one.
The HTML page where the world map is displayed, is mainly split into 3 SVGs. The map-graph is appended in one of them, using the **d3.geopath()** function and geographic data obtained from a 'json' file dataset. In the second and third SVG , on the right of the page ,the line-graph and the pie-chart are appended, by linking the selected country from the map with its more detailed profile in terms of alcohol consumption. The line-graph and pie-chart representations are performed in **ShowScatter()** and **ShowPieChart()** functions respectively.Those functions are called, and hence dynamically update their representations, once the user hovers the mouse over a specific country or when a specific beverage type is selected to be displayed from the drop down-list in the bottom of the page. In this way, we ensure the dynamic update of the displayed data in our visualizations, based on the country selected or the beverage type that the user wants to be displayed.
Data processing in our visualization techniques is performed by using one '.json' file for our geographic data, and three different '.csv' files, which have been preprocessed in order to have a convenient format for this kind of visualization.

6 Results and Future Work

After the implementation of our visualization techniques, we were able to draw some very useful conclusions and identify some meaningful patterns about our data. The alcoholic beverage type that tends to be more famous in each country is strongly related, in most cases, with each country's local alcohol industry as well as their culture. For instance, traditionally large wine producers like France, Greece or Italy seem to clearly prefer wine consumption compared to other beverages.The same fact can be also noticed in Russia regarding Spirits, as one of the largest vodka producers worldwide, and Germany regarding Beer.
Furthermore, another interesting result of our visualizations, is that European countries

that were affected most from the economic crisis, for instance Greece, Spain Italy or Ireland seem to have a downward trend in their total alcohol consumption since 2008-2009, when the crisis started.

In general, we believe that this project could extend its functionality and become a useful tool for any potential party in the alcohol production industry, if it is combined with data related to all the alcohol beverage types production throughout the world. A visual correlation between those values, could provide some useful outcomes about any possible investment opportunities in a specific field or at least provide us with a clear view about the industry market worldwide.

7 Individual reports/Project planning

The way we worked as a team in this visualization project can be summarized as follows. In the initial stages, we had a few meetings in order to agree and search for the data sets on which we would work.Based on those datasets,we then considered what we would like to visualize and which visualization techniques could better represent this kind of data. After the whole procedure and the techniques we would use were agreed, we split the technical part of the work and each team member worked on building separate visualization. However, since our techniques should most times link to each other, this was the part that we had to cooperate as good as possible in order to avoid any conflicts when all those parts were connected.Finally, we had our last meetings in order to connect and finalize our code and agree on the report format and structure.

References

[1] *Bostock, M. (2016). D3.js - Data-Driven Documents. D3js.org. Retrieved 20 December 2017, from https://d3js.org/.*

[2] *Schulz, H. J., Nocke, T., Heitzler, M.and Schumann, H. (2013). A design space of visualization tasks. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2366-2375.*

[3] *Munzner, T. (2014). Visualization Analysis and Design. CRC Press.*

[4] *Teuling, A. J., Stöckli, R.and Seneviratne, S. I. (2011). Bivariate colour maps for visualizing climate data. International Journal of Climatology, 31(9), 1408-1412.*

[5] *Borland, D. and Taylor, M. R. I. (2007). Rainbow Color Map (Still) Considered Harmful. IEEE Computer Graphics and Applications, 27(2).*

[6] *Generalized Set of Color Schemes. (2016). Personal.psu.edu. Retrieved 20 December 2017, from http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html.*

[7] *Global Health Observatory data repository. Recorded alcohol per capita consumption, from 2000. Retrieved 20 December 2017, from http://apps.who.int/gho/data/node.main.A1026?lang=en.*