Delft University of Technology

Delft University of Technology

IN4086-14 Data Visualization

World Happiness Report Exploration

Group 28

Authors:	*Student Number:*
Chia Lun Yeh	4718836
Fenglu Xu	4579976
Po Shin Chen	4703308

February 9, 2018

Contents

1 Introduction

What is happiness? Happiness can be interpolation in at least two ways: it is an emotion "are you happy?", or an evaluation "are you happy with your life overall?" Both of two provide important information that can be useful for decision makers. After all it is often how people feel in the moment that determines how they behave but its the second, evaluative, use of the word, that is more important when thinking about human development and progress.
In this decades, happiness is increasingly considered an important and useful way to guide public policy and measure its effectiveness. Therefore, in this project, the World Happiness Index was analyzed and visualized which aims to measure the happiness of countries and it defines the following themes as preconditions for a happy country: Economy, Health, Family, Freedom, Generosity and Trust.

The emphasis of this project is laid on the second variant of the assignment, creating a interactive visualization design with the help of $D3.js$. This report is further organized as follows. In section 2 the problem is formulated based on data selection and the task description. In section 3, our visualization design is discussed together with the justification of our choice. And the setup of the code is briefly demonstrated as well. In section 4, we make observations about the data with the help of our visualization technique and some interesting conclusions are drawn. At the end, the individual reflections on this project are attached.

2 Problem Formulation

The design space of visualization tasks is huge, and many designs are ineffective. For this reason, we perform a solid analysis of the data and task before implementation. In this section, we answer two main questions: "What is shown?" and "Why is the user looking at it?". The analyses can be termed as data abstraction and task abstraction respectively.

2.1 Data Selection

The source from which the data is acquired can be found in [1]. The original survey, in which 155 countries were ranked by their happiness levels, was conducted by the Sustainable Development Solutions Network (SDSN). There are six factors that contribute to the final score of happiness:

- Economy: GDP per capita in terms of Purchasing Power Parity.
- Health: healthy life expectancy at birth.
- Social Support (shown as Family): measured by the response to the question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"
- Freedom: measured by several questions about whether an individual has the freedom to make life choices.
- Generosity: measured by the response to the question "Have you donated money to a charity in the past month?"

- Trust: perceptions of corruption throughout the government and business.

The choice of each explanatory factors and the detailed formulation of the final scores are explained in the World Happiness Report that is released each year [2]. In addition to these six perspectives that were surveyed, two more attributes are also included in the data.

- Dystopia Residual (shown as Dystopia): it reflects the extent to which the six variables under- or over- explain the final happiness evaluation, and differs for each country.
- Region: countries are categorized into smaller regions such as Western Europe, Central and Eastern Europe, Africa, etc.

The final happiness score is the sum of all the factors including dystopia residual. In this project, data of 81 countries for the most recent 3 years are used.

2.2 Data Abstraction

We analyze the dataset and data following Munzner's categorization [3].

2.2.1 Datasets

The dataset can be seen as a three-dimensional table. Each country with all its calculated scores is one item in the table, and the eight attributes mentioned in 2.1 are the attributes in the table. Three years of data form the third dimension. The dataset is a static file.

2.2.2 Attributes

All the attributes that contribute to the final happiness score are sequentially-ordered and quantitative. The attributes country and region are categorical.

2.3 Task Abstraction

In order to understand why users look at the data, we need to identify the users and their goals. We follow selected content in [3] and [4] to analyze the task.

2.3.1 User: Who is executing a task?

Tasks are executed by people who are interested in characteristics of well-being in each country, which are politicians, sociologists, and scientists. General public can also be the user.

2.3.2 Goal: Why is a task pursued?

The goal of our visualization is for exploratory analysis. This means that users navigate through the data in order to form hypotheses based on their observations. Users can also search in the data for known or unknown target at known or unknown location. The visualization should thus show details of the data accurately, while allowing users to easily summarize the data, make comparisons among factors, countries, and regions, as well as identify outliers and correlations.
Based on this, we list some questions that might be interesting for the users.

1. How does the happiness level of a specific country rank globally and region-wise?

2. Which region is the most happy or unhappy about their life quality? What factor contribute to the result?
3. What is the change in happiness level for the recent three years of a specific country or region? Are some factors particularly more influential than others?
4. Which countries are the outliers in a region? Which factors contribute to that?
5. Which factor correlates with the final happiness score most? What are the correlation between two factors?

One or two of the questions can be readily answered from the data itself, but most of them require visualizations in order to show the insights.

3 Visualization Design

After having a concrete understanding of the data and task, we can then decide how to show the data in order to meet the task requirements. In this section, we motivate the idioms chosen to visualize the data, as well as the interactions designed. We also briefly explain the implementation.
Following the effectiveness principle, we want to encode the attributes with highly-ranked channels. For ordered attributes, they are position (aligned or unaligned) and length; while the most effective ones for categorical attributes are spatial region and color hue. We first analyze the possibilities of different idioms. Table 1 shows the design matrix of the task. Requirements are the information we tend to obtain in our task and according to the table, we figured out that stacked bar chart can achieve four of them out of seven. Therefore, the main design of our presentation is the stacked bar chart and the other information will be visualized by radar chart and scatter plot.

Table 1: Decision Matrix

Requirements	Map	Stacked bar chart	Radar chart	Line chart	Scatter plot	Heatmap
Sort the country by rank		yes				
Show the distribution of score by region	yes	yes				
Show the score by year	yes	yes				
Show the recalculated score and rank with customised weight		yes				
Show the distribution of score among each attribute for per country			yes			
Show the trend in score within three years for per country			yes	yes		
Show the relation of score of each attribute among countries					yes	yes
View the outliers of the countries					yes	

3.1 Stacked Bar Chart

In the stacked bar chart, vertical line marks are applied and the visualization channels are color hue and the length of the bars. Data are separated by attribute "Country" horizontally and are aligned vertically, viewers can sort the data by alphabetical order or by rank of the total happiness score according to their preferences. As mentioned before, our data is combined with different kinds of attributes(e.g. economy, health, trust of government,etc.), and we want to observe some informations from the data(e.g. Requirements in Table 1. In order to present as much information in one view with clean and clear, we decide to implement stacked bar chart as our main display method according the the decision matrix in Table 1.
By adding different hue colors to discriminate between each attributes, stacked bar chart allows us to clearly identify both the total happiness score and the value of all attributes in one view.

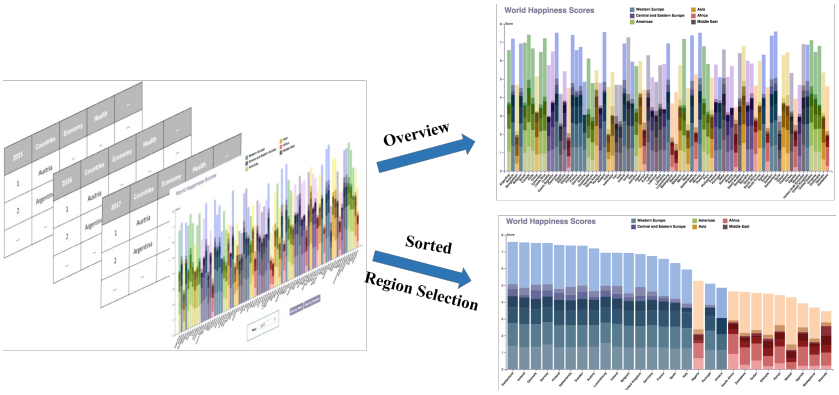

Figure 1: Stacked bar chart

First, comparing the stacked bar chart with normal single bar chart, stacked bar chart can display multiple information instead of only one information in the normal bar chart. Although with the increasing number of the attributes, there will be some visual clustering of the stacked chart, here we only have seven attributes, which is moderate for us to present. Moreover, the viewers are enabled to tune the weights of the attributes and change sorting methods. The display will be the same as normal bar chart if we leave one attribute alone and tune others to zero. Thus, using stacked bar chart is indeed better than normal bar chart.
Second, comparing the stacked bar chart with world map, we can easily notice that world map is unable to display the contribution of each attributes, which is similar to the normal bar chart. The advantage of the world map visualization is to easily observe the data in spacial aspect, however, we can still use different color to separate the relationship in different regions. Therefore, eventually we decide to implement stacked bar chart combining with interactions and several selections for visualization.
Owing to the fact that we have several information to display, we implement some interactions for viewers to tune by what regions they want to look at and by which attributes they care. The simple result for tuning and removing is shown in Figure **??**. In order to let viewers know not only ranks of the happiness and the contribution from each attributes but also the general distribution of each region, the color was encoded for viewers to distinguish both of them. Furthermore, we also applied tooltips for users to have a quick view of the stack when the mouse hovering to it.

3.2 Radar Chart

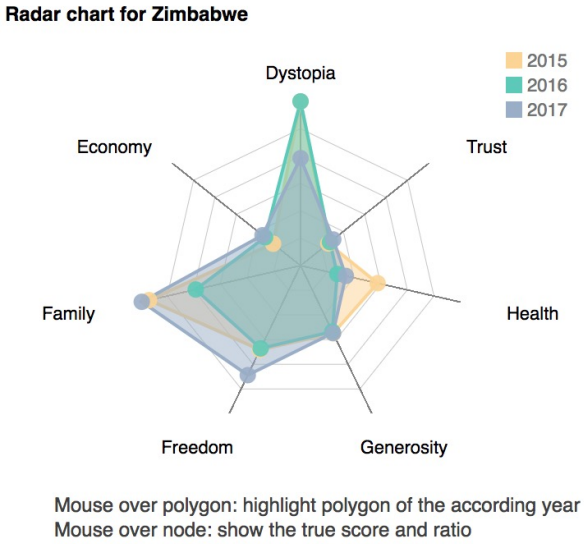

Figure 2: Radar chart for Zimbabwe

The radar chart as shown in figure 2 is a chart that consist of a sequence of seven equi-angular spokes, called radii, with each spoke representing one of the component for happiness score.The data length of a spoke is proportional to the magnitude of the variable for the data point relative to the maximum magnitude of the variable across all data points.A line is drawn connecting the data values for each spoke. This gives the plot a star-like appearance. In fact, the radar chart is chosen to illustrate the changes for one specific country in each component of happiness score during three years from 2015 to 2017, therefore there are three star-like pattern in the chart.We use radar chart other than line chart which seems to be more frequently use to show the trend in timeline. Because we only have three-year data here, it makes no sense to make a curve fitting for each component.Specially,we didn't use the the true scores for the components of happiness to plot the data points instead we use data after the transformation as below:

$$ratio = \frac{the\,score\,of\,component\,x\,in\,year\,y\,for country\,z}{the\,maximum\,score\,of\,component\,x\,in\,three\,year}$$

There is mainly three reasons for transformation:

- The range for score of each component varies and is unknown, therefore we have to normalize them for better comparisons.
- If true scores are used to plot the point, the distribution of them are skewed and it adds difficulty to observe the data.
- Using the maximum score of each component for three year as denominator bring extra merits.It not only allows track the change of one certain country among years but also permit the comparison between one country with the country of max score. In this case, the observer knows its level in world and the space to improve.

Three polygons constructed by points and lines with different colors are present three consecutive years, yellow, cyan, gray for 2015, 2016 and 2017 respectively. Once the cursor of mouse hover over one of polygons, the one selected will be highlighted.The fill of the other two will be clear but the line and points remain. In this way, observer is able to see the data of interested year and compare them with other two year as well. When the cursor hover on the point, the true score for it will show on the screen. In fact,there are two weakness of the radar chart. When the data needed to be illustrated is more than 3 three years, it is hard to show a trend clearly in a radar chart. And if the data of each attribute for different are considerable close, It may be difficult for the observer to distinguish the data point in the chart.

3.3 Scatter Plot

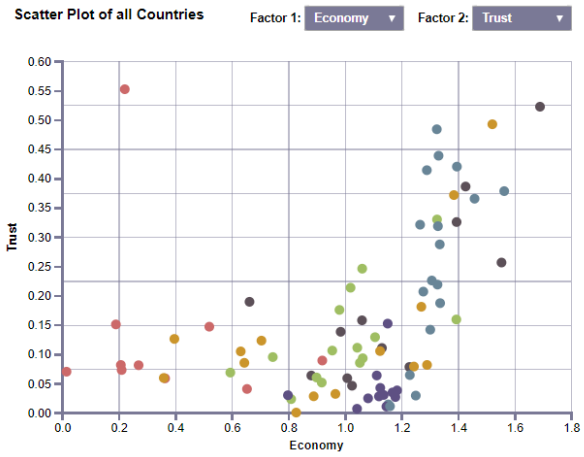

Figure 3: Scatter plot showing the Economy as the x-axis and Trust as the y-axis

The use of the scatter plot is for user to explore correlation among different factors and the final happiness score. A scatter plot uses points as marks and vertical and horizontal positions as channels to encode the data. The user can utilize the drop-down selection to choose the factors he wishes to analyze. We also use color hue to encode the categorical attribute region. In the case, the visualization is especially easy for the user to identify outliers, compare different countries and regions, and find the relationship and trend among factors. Since the amount of the data is not manageable, the scatter plot does not present visual clutter. Heatmap and correlation matrix can also be used to visualize relationship between variables, however, scatter plot can show the outliers more clearly, and thus meet our target.

3.4 Interaction

The default view of the visualization is the stacked bar chart with all the factors of the same weight. This directly shows the happiness score of each country, and gives the user an overview of the data. We follow the principle mentioned in [5], which states clearly that "overview first, zoom in and filter, details on demand".
From the default view, the user can interact with the data by adjusting the slide bar for each of the factor. Since happiness is somehow subjective, users can have various priorities on the

factors. The slide bars therefore provide the interactivity and flexibility for users to evaluate countries based on their definition of happiness.
More details are shown if the user hover over the bars of the countries and attributes in which they are interested. When user click on the bar in the chart, corresponding radar chart is shown. Therefore, the details are shown on demand. Some other interactive options for the user include buttons to sort the stacked bar chart by rank or by alphabetical order, and selection of different years of data.

3.5 Implementation

The dataset required little pre-processing, therefore, it's mainly done in Microsoft Excel. The code itself consists of three main parts: drawing the stacked bar chart, drawing the radar chart, and updating the data and plots when the weights of values are changed. The plots are created using the d3 library. The radar chart was coded following an demo using d3.js but we made many changes according to our project. We read many examples and followed part of them in order to complete the project. However, we make sure that each line is understood, and avoid copying the whole code from others.

4 Results

4.1 Task validation

Our final design achieves the goal and gives answer to those questions put forward in section 2.3.2. Our final design allows users navigate through the data in order to form hypotheses based on their observations and to easily summarize the data, make comparisons among factors, countries, and regions, as well as identify outliers and correlations.

1. The stacked bar chart shows the distribution of happiness level around the world by sorting with respect to the rank.

2. From the radar chart, the change in happiness level for the recent three years for a specific country can be observed and the comparison between factors is visualized as well.

3. The outliers in a region can be found by observing the selected region in the stacked bar chart, or by looking at the scatter plot.

4. By observing the scatter plot, the correlation between factors with the final happiness score and correlation between two factors can be illustrated. Accordingly the most influential factor is found.

4.2 Interesting Finding

1. In average, Western European countries rank the highest in the world while African countries rank the lowest. Most Western countries rank high in factors Family, Health and Freedom compared to other regions. African countries have especially poor Economy and Health.

2. The trend for recent three years can be observed from the radar chart. For example, looking at the radar chart of Hungary, we can see that the Freedom, Health, and Generosity aspects worsen throughout the year while the Family and Economy aspects improve.

3. From figure 3, an outlier can be clearly identified. In this case, the country Rwanda has especially high trust compared to countries of the same economy level.

4. The total happiness score is most related to the factor Economy, as can be seen from the nearly linear relationship in the scatter plot. Alternating the selection and playing around with different combinations of factors, we notice that the other factors are not really correlated. Among them, Economy and Family has the strongest correlation.

5. The happiness rank changes quite dramatically when the weight of factors are changed. For example, if we don't consider material aspects and focus on more spiritual aspects such as Freedom, Family, Trust, and Generosity, the ranks of countries like Mexico, Costa Rica, and Nigeria improve much.

5 Indivisual Reflection

Reflection of Chia-Lun Yeh It is the first time for me to actually work on a Javascript project and using the D3 library. We spent some time constructing the first version of stacked bar chart together. Then I work on the color scheme and tooltip of the chart. The color scheme is really tricky since we want to show two categorical attributes at the same time. To better understand the correlation, I also revised the scatter plot that we built during the practical session. Throughout the project, I didn't only worked on improving my coding skills, but also critically judging various visualization designs. As all of us are somewhat new to the language and structure, it's interesting to learn and discover together. Lastly, I wrote most of the second section, the part regarding scatter plot, interaction, and findings of the report. It's difficult to separate the task as all of us contribute quite equally to the task.

Reflection of Fenglu Xu I have enjoyed this project as I learn quite a lot about JavaScript and D3.js. Initially, we spent some time together to select the topic and searched for the available dataset. Then based on our goal and understanding of the data, we decided the approach to visualize the data. I worked mostly on the radar chart to illustrate the score change in three years for per country and make it interact with the bar chart. It was pretty interesting to transform your interpretation of data into visible things. And getting the elements to interact among each other was even more fascinating. Besides, I maintained the layout of report and mainly write the introduction and the part concerning the radar chart. In conclusion, I have to say it was a pleasant experience to work on InfoVis assignment, which was absolutely based on the good cooperation with my group members.

Reflection of Po Shin Chen I used to learn simple HTML/CSS before but have never implemented them for visualization. We spent plenty of time understanding the structure of d3 and JavaScript together to come up with the first stacked bar chart. Then I majored on interactions for viewers (switching years, sorting, tuning attributes weights and adding/removing data by regions) and improvement of the stacked bar chart. Owing to the data structure that we have, it's quite hard for me to achieve update and remove exhibited data. Moreover, there are lots of matters to be considered during each interaction(for example, the options the user has chosen should be updated when switching the data). After this assignment, I obtain a concrete design structure and knowledge about D3 and JavaScript by working and discussing with my members. Eventually, I wrote part of section 2 and first half of section 3.

References

[1] Kaggle, explore happiness data.

[2] World happiness report 2017.

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

[4] Nocke T. Heitzler M. Schumann H. Schulz, H.J. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19, 2013.

[5] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. 1996.