

# Factoring What Makes People Happy

## Process Book

Team Happiness

CS 4802 - February 28, 2018

Samuel Coache, Samantha Comeau, William Schwartz

# Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Overview</b>	<b>3</b>
<b>Background and Motivation</b>	<b>3</b>
<b>Related Work</b>	<b>4</b>
<b>Questions</b>	<b>4</b>
<b>Data</b>	<b>5</b>
<b>Exploratory Data Analysis</b>	<b>6</b>
<b>Design Evolution</b>	<b>9</b>
Inspirations/ Visualization tools you explored	13
Change in Focus and Design Over Time	13
Cluster Map Progression	13
Bar Chart Progression	15
<b>Implementation</b>	<b>16</b>
<b>Evaluation</b>	<b>17</b>

## Overview

For this project we created a webpage that allows people to interact and see why some countries are happier than others. To complete this task we started collecting readily available country specific metrics that we believed could cause a certain country's inhabitants to be relatively more or less happy. The data that we gathered was the reported average reported happiness by country, total population, Gross Domestic Product (GDP), inflation, unemployment, and alcohol consumption, all of this data was gathered from the year 2008 to 2016. We wanted to present this data in a way that would allow people to engage, and explore on their own. We broke our visualisation into three distinct sections which are explained below.

The first is a choropleth map showing the reported happiness of each country using a blue gradient. Darker blue being happier, and lighter blue meaning a country is less happy. Users are able to click through years, and see how reported happiness changes, as well as compare happiness visually to adjacent countries, by clicking on the country. Hovering over a country brings up a tooltip displaying the reported happiness value.

The second interaction was our most experimental interaction. We created a cluster chart of all the countries with reported happiness data. Using the same blue color scale from the map to relate to the reported happiness. We also arranged the data in this chart to show the countries sequentially organized from unhappy to happy. Hovering over these points displays a tooltip, which like previous displays the country name and the reported happiness score. Sliders for each of our factors are presented on the side of cluster chart and are used to filter countries that are displayed, and give the user a chance to explore and look for correlation between factors on their own.

The third section, is a series of sorted colored bar charts, where the color corresponds to the value, presenting the data this way gives the user somewhat of an indication of the distribution of each of the factors very quickly. The graphic also serves to show how countries rank in specific categories. When hovering over a specific bar, the bar gets highlighted, and the same country also gets highlighted in all the other factors charts to see how this country ranks within each metric. Like in the other sections the mouse over also displays a tooltip, to show the country name as well as well as displaying the value of that factor selected.

## Background and Motivation

The motivation for choosing this project was to determine what factors make people happy. We saw a dataset of how people ranked themselves on their happiness and decided we wanted to try to correlate that data to other factors, like the regions favorite extracurricular activities, historical events, family size, alcohol consumption, etc, and try to determine what factors are most impactful on how someone ranks their own happiness. We have reviewed some of the other happiness reports out there, and believe they report the data, but don't got far enough exploring what social, economic, and environmental factors contribute to those results. We believe that we can present this information in a visually appealing and informative manner to help distinguish some of the ambiguity.

To determine what factors make people happy we compared a dataset where people ranked their own happiness and correlated that data to other factors, like the countries GDP, unemployment rate, alcohol consumption, etc, and try to determine what factors are most impactful on how someone ranks their own happiness. Using our visualization we present this information in a interactive, visually appealing, and informative manner for you to explore the connections between our variables.

## Related Work

We had quite a few sources of inspiration for this project as well as references that we used create this visualization. The following sources that we used for inspiration, as well as serve as example are:

<http://www.oecdbetterlifeindex.org/>  
<http://alignedleft.com/tutorials/d3/scales>  
<https://bl.ocks.org/mbostock/1021953>  
<https://www.kaggle.com/undsn/world-happiness/data>  
[https://data.worldbank.org/indicator/SP.POP.TOTL?order=wbapi\\_data\\_value\\_2011+wbapi\\_data](https://data.worldbank.org/indicator/SP.POP.TOTL?order=wbapi_data_value_2011+wbapi_data)  
<https://bl.ocks.org/sgcc/251f79f68a502c1607c1a1ee3cae0f6f>  
<https://hi.stamen.com/forcing-functions-inside-d3-v4-forces-and-layout-transitions-f3e89ee02d12>  
<http://alignedleft.com/tutorials/d3/scales>  
<http://ieeexplore.ieee.org/document/1532144/figures>  
<http://www.dashboardinsight.com/dashboards/strategic/panopticon-horizon-graphs.aspx>

## Questions

Some guiding questions we are trying to answer through our exploration data visualizations are as follows.

- What makes people happy?
- How does happiness compare across the world?
- What differences between countries cause certain countries to be happier?
- How does self perception (pessimism, view of others) affect accuracy of how happy people think they are?

We then broke these questions up into different categories that we think would affect a person's happiness. These categories and questions related to them were as follows.

### Extracurricular participation

- What do people in these countries do in their free time?
- Do they participate in country-wide events like the Olympics or Car-Free days?
- How much alcohol do they consume?

### Life factors

- What's the overall population of their country?

- Pessimism, this could affect their self reflection on how much they can impact the world etc
- How large is the average family?
- How long are they expected to live?
- How much alcohol per person is consumed

## Monetary factors

- GDP
- What is the unemployment rate?
- Has inflation affected their happiness?

A benefit of comparing these different categories is that it allows us to see more of the whole picture. Rather than solely focusing on something like socio-economic status or unemployment rates we can see how personal life and business life affect humans happiness, because overall, everything in their life is affecting them somehow. We hope to use these categories to compare the different countries to learn what is most benefiting their happiness and what is making them less happy. For example say America has a binge drinking problem and their happiness score is very low, and France doesn't drink as often and their happiness is high. One intermediate conclusion could be that drinking affects happiness. Adding these intermediate conclusions together with the rest of the categories conclusions will show us the big picture on what factors truly contribute to an individual's happiness.

Over the course of this project we had to modify our questions to reflect the datasets that we could find. We had originally planned on answering all of these questions however finding worldwide data for some of these metrics was near impossible. Some new questions we began to consider include:

- Does the individual factors affect the happiness more than countrywide factors?
- Is there a real way to rank a countries happiness?

These questions led to more self reflection on our part rather than research because its hard to research these topics without conduction our own surveys.

## Data

Our original inspiration came from a Kaggle.com dataset that represented happiness scores. This dataset didn't provide enough information for what we wanted to do so we found the World Happiness Report which had more data and more years of data for us to use. We then found a website by the Pew Research Center that had data by country. We chose to take the data for population, inflation, gdp, and unemployment from this site. These are some metrics that we thought would best correlate with the happiness data. We also used an alcohol consumption dataset that we found from the World Health Organization. We had other datasets that we were planning on using in our website but we decided that these metrics would be most relevant to the story that we were trying to tell.

To filter and organize these datasets we decided to write some python scripts. We imported all these datasets into the script using the python package *csv*. We then made a new csv for each year that we wanted to present in our website. Doing this allowed us to reload each of the graphs with a new csv which was the easiest

way for us to allow the user to change the selected year and have the webpage subsequently update every graph to correspond with the correct data.

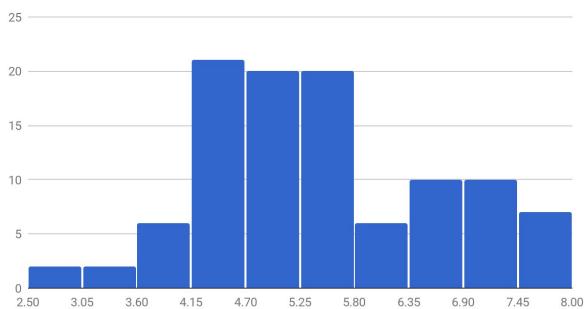
The links to find these datasets are listed here.

- <http://www.pewglobal.org/search/?query=happiness>
- <https://www.kaggle.com/unsdsn/world-happiness/data>
- <http://worldhappiness.report>
- <http://www.imf.org/external/index.htm>
- [http://www.who.int/gho/alcohol/consumption\\_levels/adult\\_recorded\\_per capita/en/](http://www.who.int/gho/alcohol/consumption_levels/adult_recorded_per capita/en/)

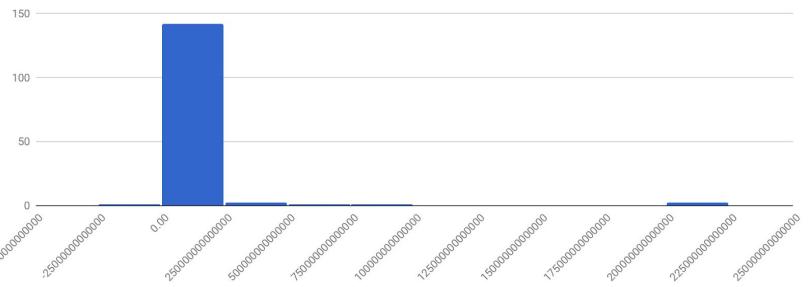
## Exploratory Data Analysis

Initially we imported our filtered data sheets into Google Sheets so we could figure out what would be the best way to display our data. Some of these are shown below.

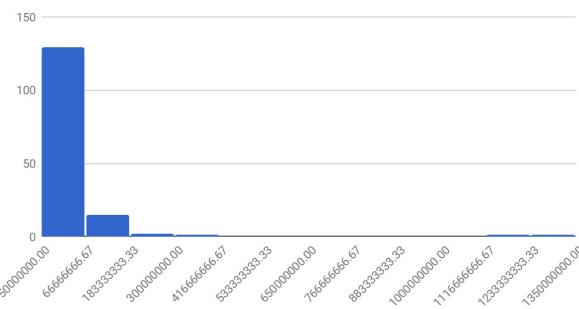
2008 Happiness Histogram



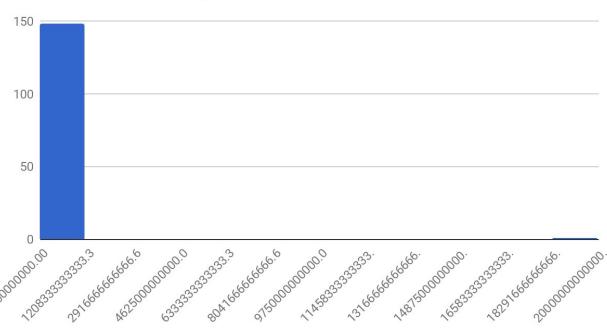
2008 GDP Histogram

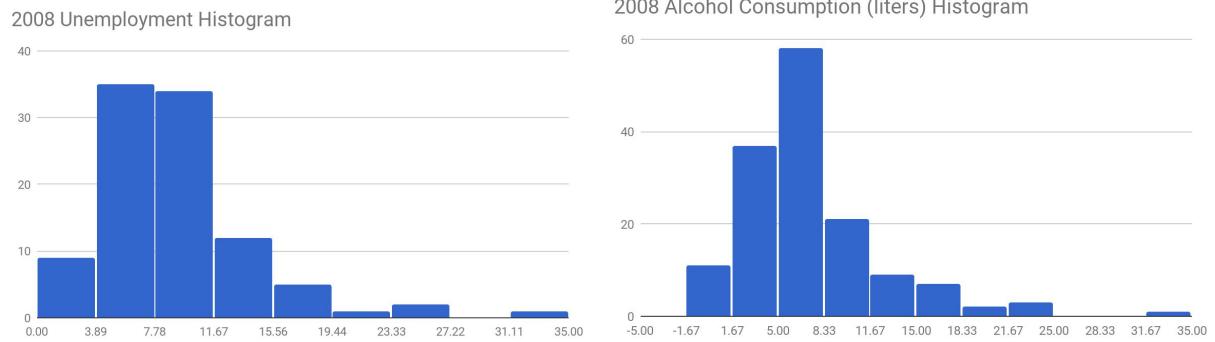


2008 Population Histogram

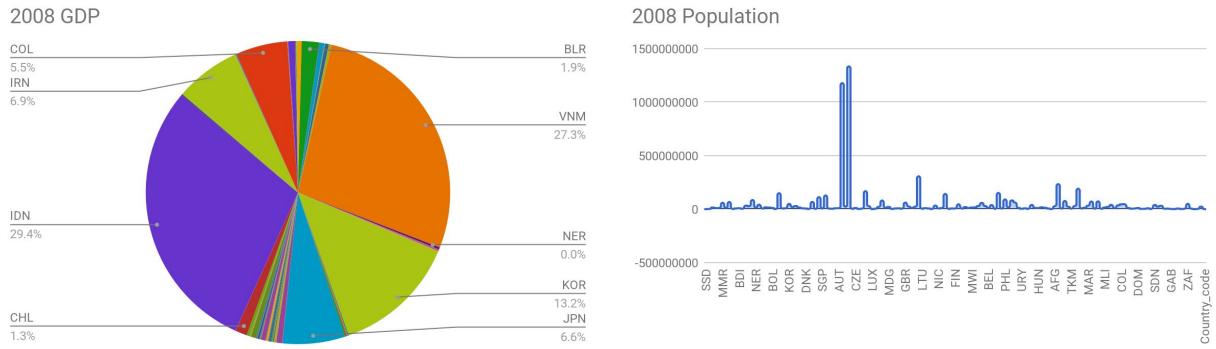


2008 Inflation Histogram





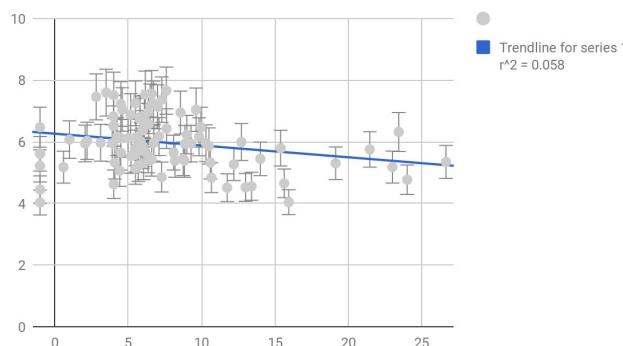
The majority of our data analysis was centered around figuring out how to display the data in a way that made sense from a users perspective. We created these histograms the get a better idea of how the data was distributed so that we could scale the data appropriately. We had a limited number of colors that we wanted to use for our colored bar charts and wanted to make sure that the data fit into the categories defined. We ended up using logarithmic scales for all of our colored bar charts, as well as the sliders used in the cluster chart.



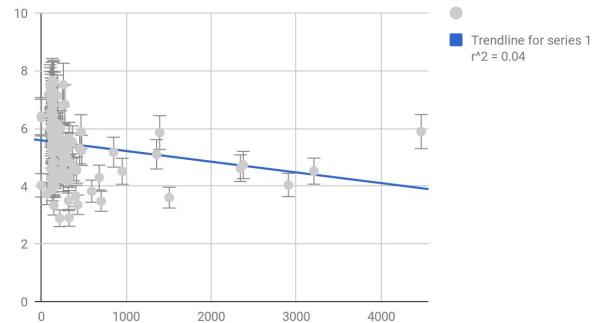
This data exploration confirmed most of our suspicions about the different factors. Creating these initial visualizations helped to guide us in how we scaled the data. As well as get an initial feel for the data.

All of the following scatter plots have happiness on the Y axis and the Title value on the X axis:

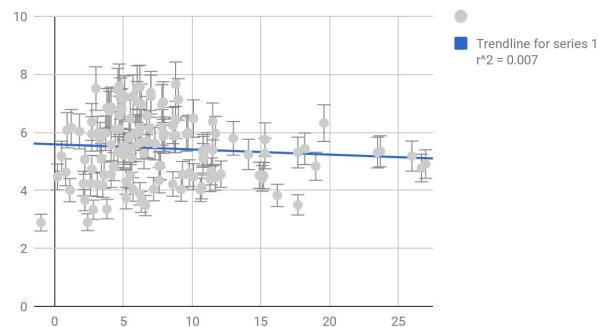
Unemployment 2016



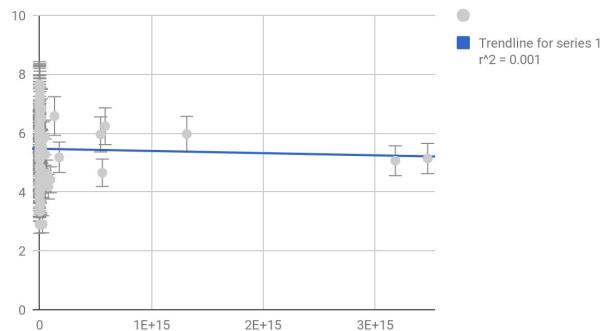
Inflation 2016



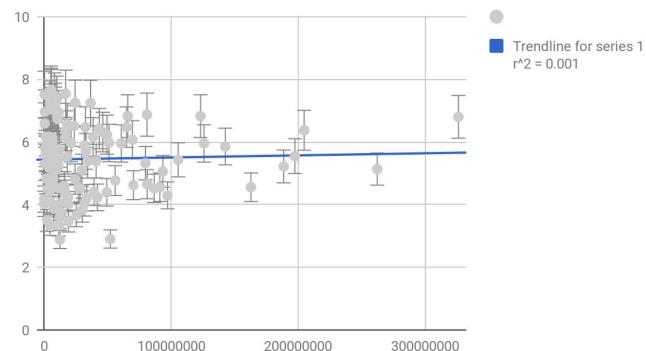
Alcohol 2016



GDP 2016



Population 2016



One of the big hurdles that we ran into initially was, after collecting all of our data creating these scatter plots comparing our factors to happiness appear to have no to very little correlation. Based off the charts seen above.

**Unemployment:** From these charts we can see that there is a slight correlation between increased unemployment and decreased happiness.

**Inflation:** There is another slight trend relating increased inflation to decrease in happiness.

**Alcohol:** A very slight trend relating increased alcohol consumption and a decrease in happiness.

**GDP & Population:** GDP and Population have almost no correlation

Due to the fact that we could not find any very apparent correlation of our factors relating to happiness, we decided that we wanted to make our data visualization extremely exploratory in hopes that the users are able to search and test for their own refine the scope of the search by compounding the factors, and creating filters so that they could see if compounding factors would produce any interesting results.

## Design Evolution

To start designing our web page we decided to use the 5 Design Sheet Methodology to brainstorm our ideas and condense them into a design we liked.

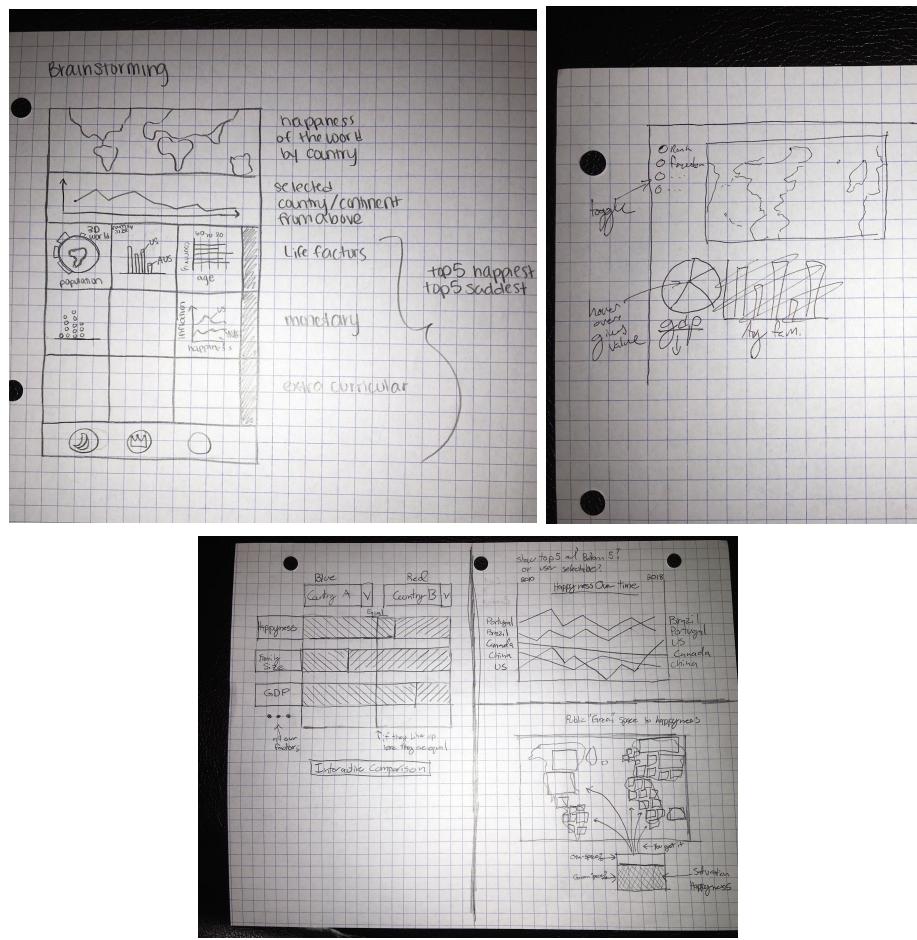
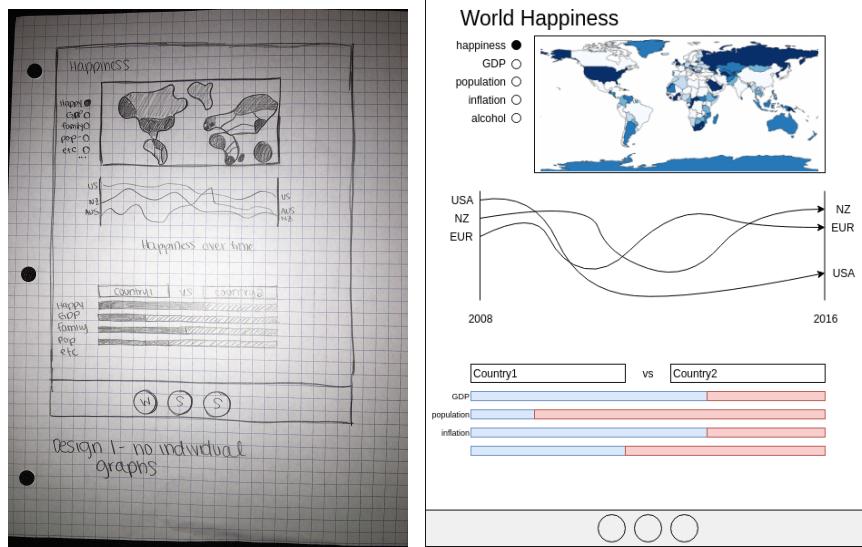


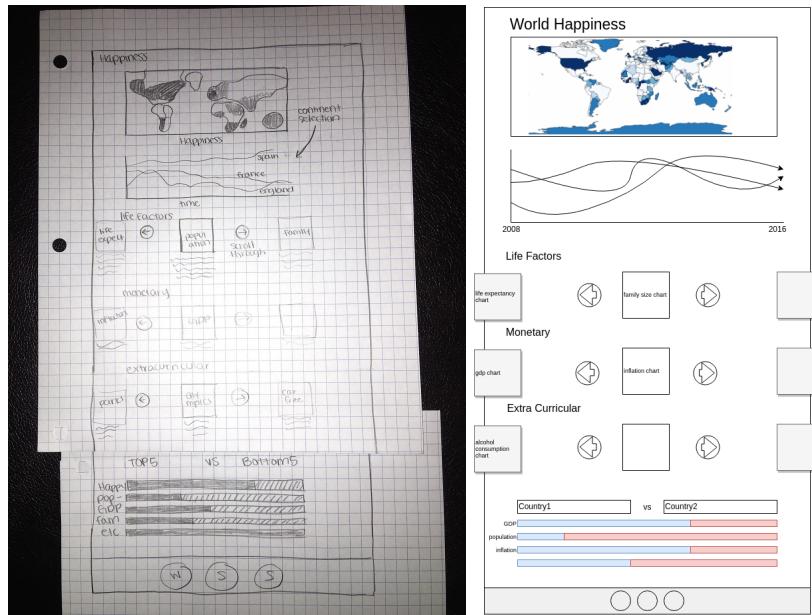
Figure 1. Brainstorming using the 5 design sheet methodology.

Our brainstorming session gave us these ideas on how we wanted to visualize happiness and the different factors that we thought might affect happiness.



*Figure 2. Potential Design 1.*

In this first design we wanted to allow for the user to easily compare everything. We thought that allowing the map to change focus based on the selection on the left would allow the user to easily compare all the countries by just seeing how the shading changes based on the selection. The other comparison method here was the direct country comparison at the bottom. This would allow the user to select two countries and see how each of the individual factors compare between the two countries.



*Figure 3. Potential Design 2.*

For our second design we focused more on the individual factors and their effect rather than comparing them between countries as much. The charts in the middle allow for the user to see each of the factors for the

countries in individual graphs and loop through the different categorical views like looping through life factors to get the population, family size and other related metrics.

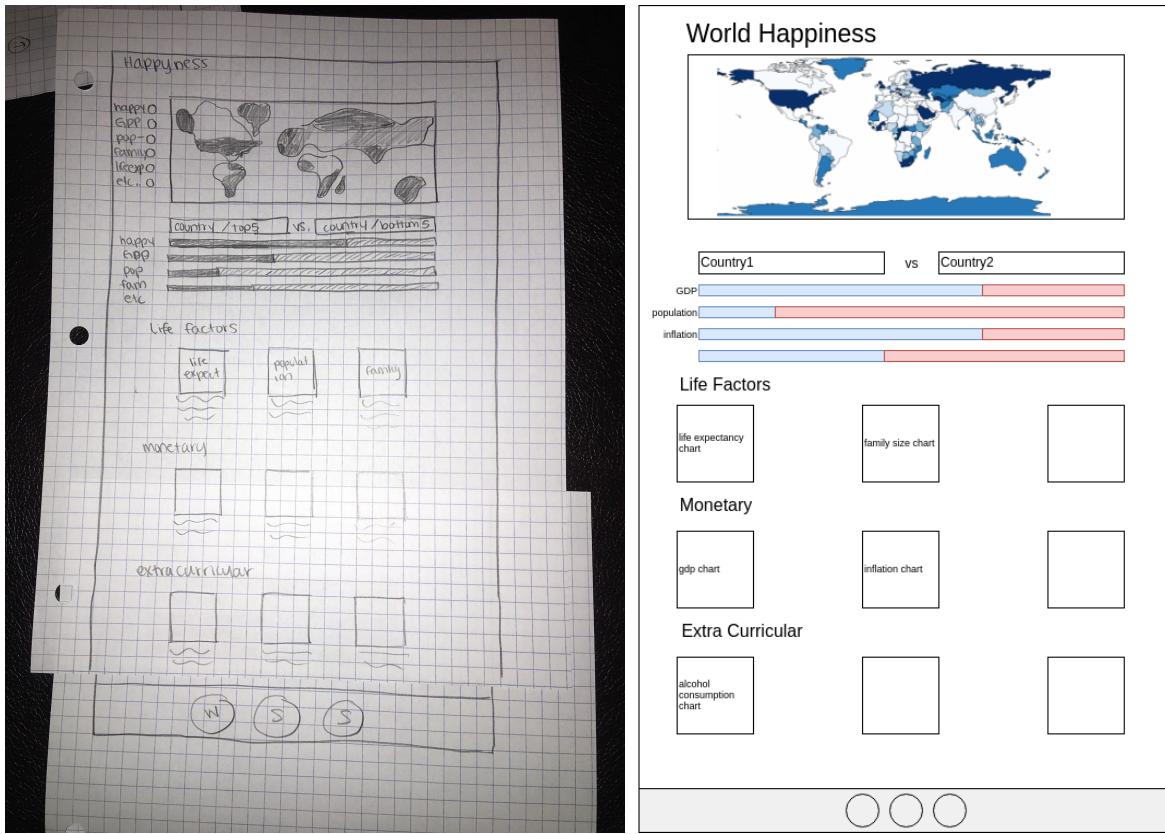
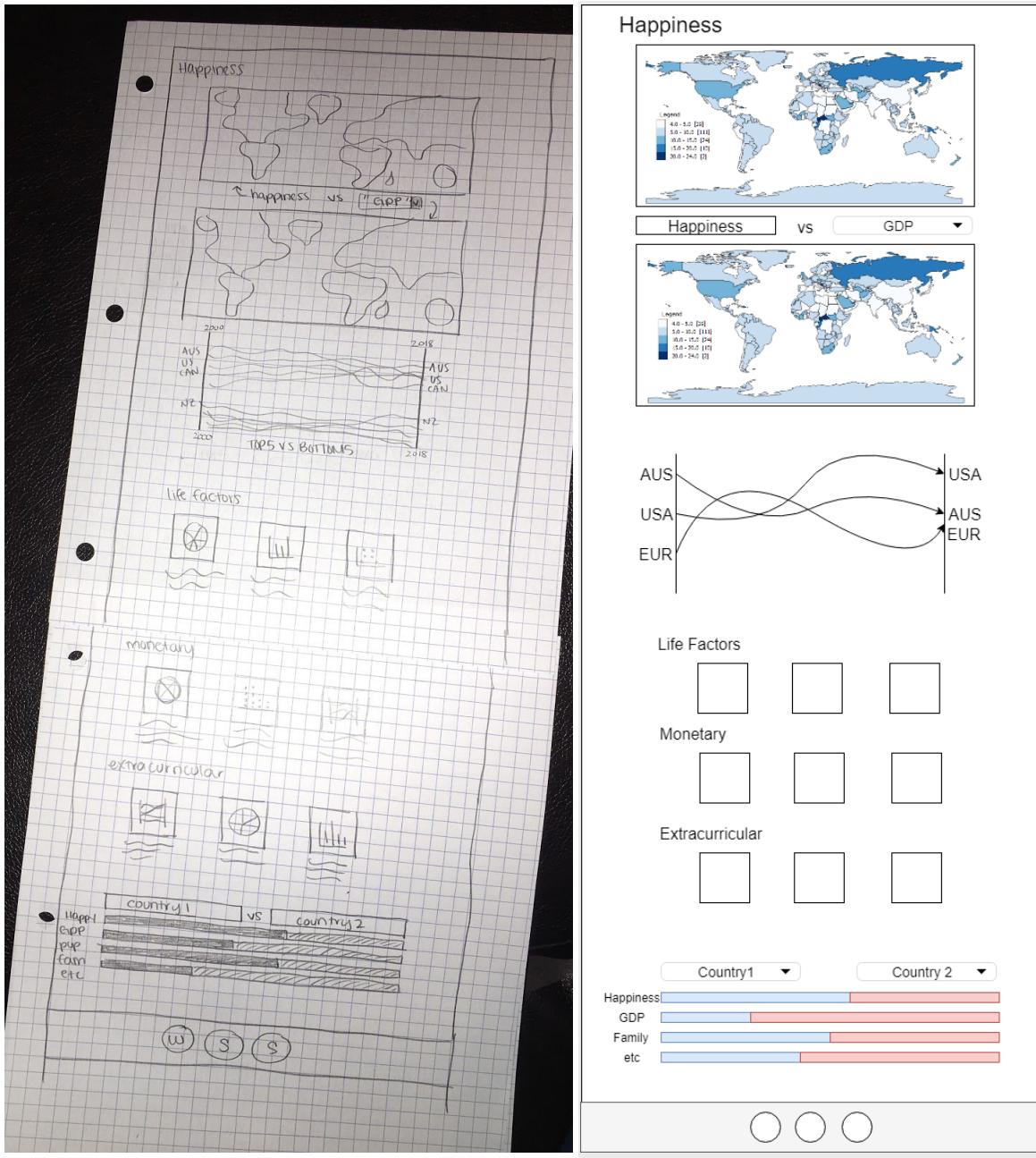


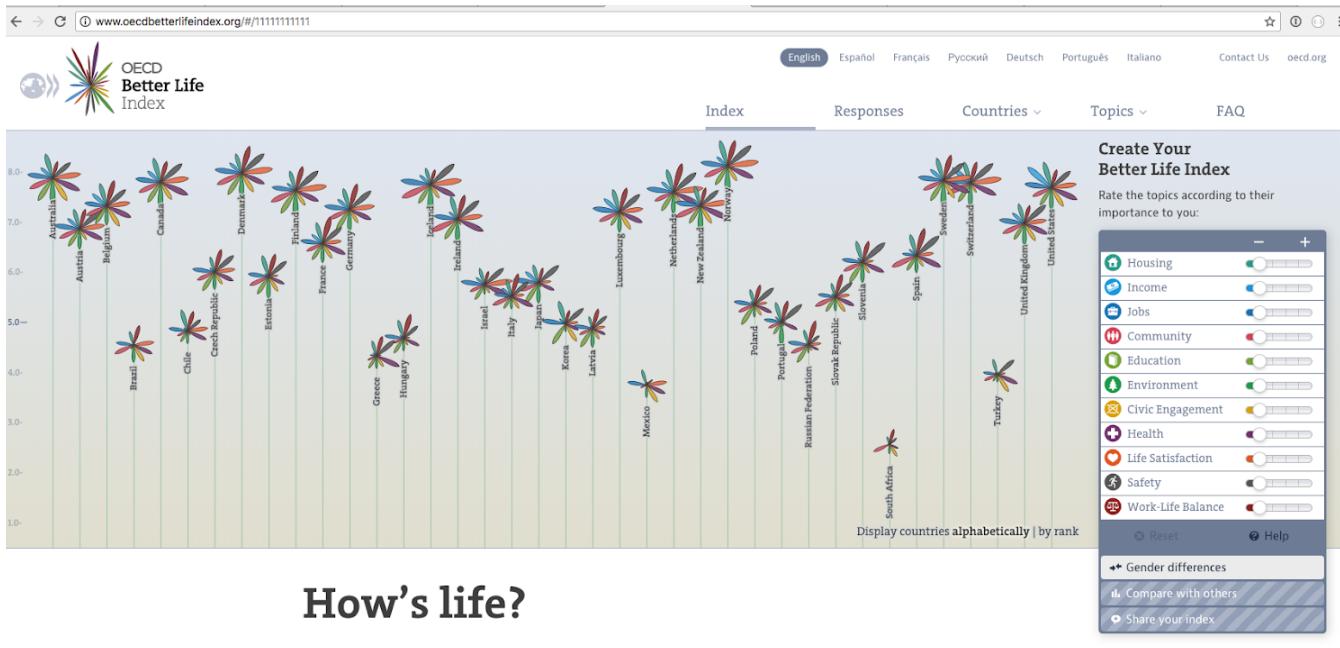
Figure 3. Potential Design 3.

For our third and final potential design during this stage of the 5 sheet design we decided that we wanted to do something unique and present the data in ways had not been realised previous and get people to do some of their own exploration with our data. We wanted to present more graphs



This is the final design we came up with during our original 5 Design Process. This design changed a lot over time as others input ideas for our project to us and as we explored our data more.

## Inspirations/ Visualization tools you explored



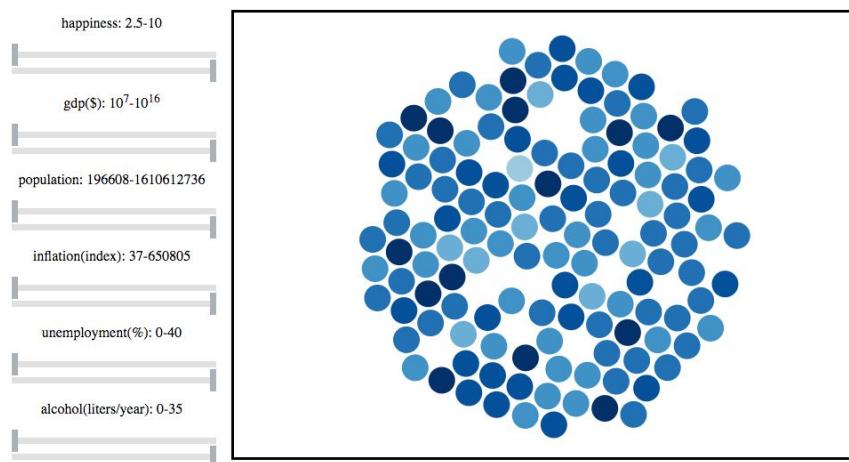
**How's life?**

<http://www.oecdbetterlifeindex.org/>

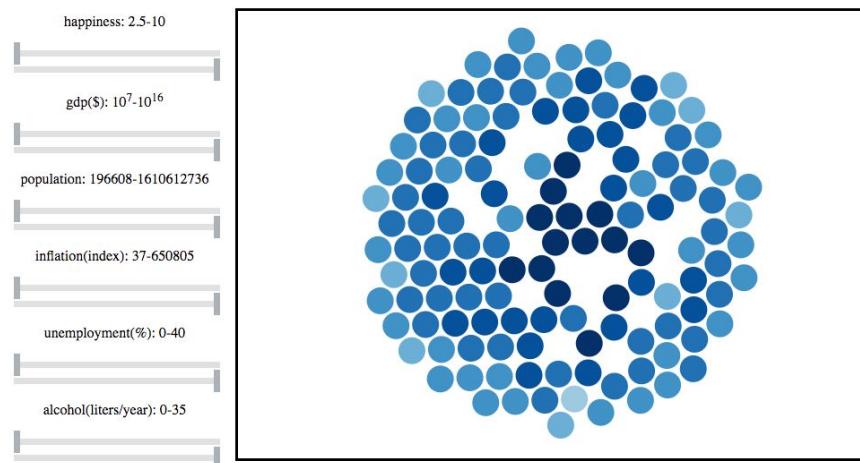
We really liked this chart and how people were able to interact with it, we wanted to portray a similar effect though our cluster map, and while our cluster map is unique, it is not executed as well as this visualization.

## Change in Focus and Design Over Time

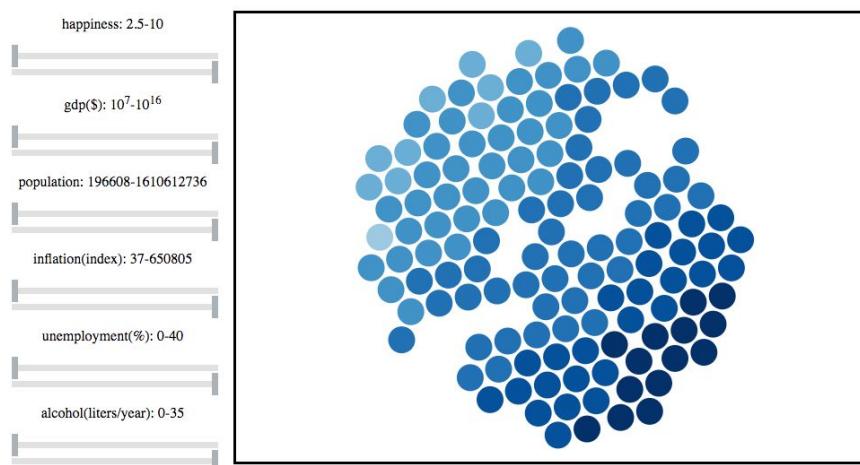
### Cluster Map Progression



*Version 1 - Cluster chart*



*Version 2 - Cluster chart*

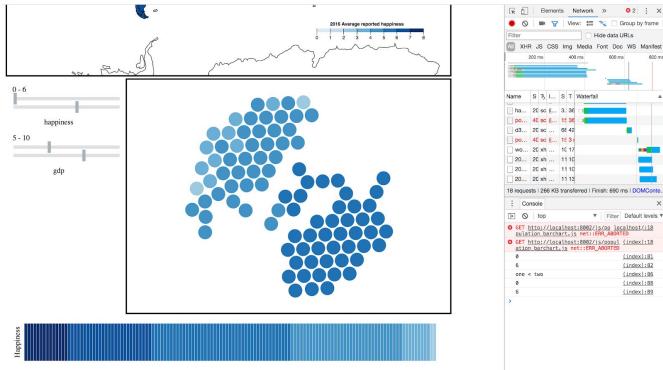


*Version 3 - Final Cluster chart*

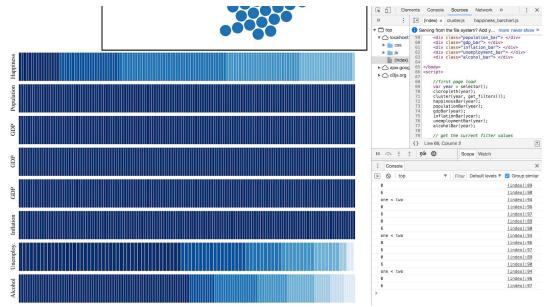
We went through multiple design iterations of the cluster char to try to figure out how to make the chart easiest to read, first all the points were just loaded randomly, then we spend quite a bit of time to try to get the happiest countries to be in the center, and finally we decided that it was easiest to interpret the cluster if they were arranged from least happy top left to most happy bottom right. D3 does not support this kind of ordering of custer charts, so it took quite a bit of time and some funky math to get these clustering effects.

## Bar Chart Progression

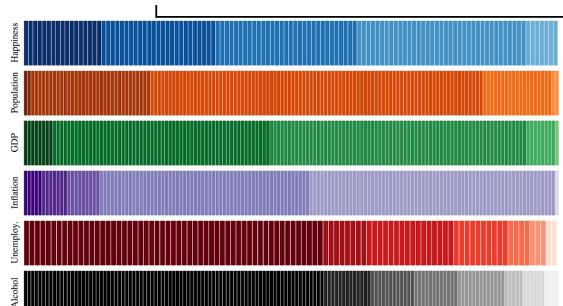
### Version 1 - First bar chart



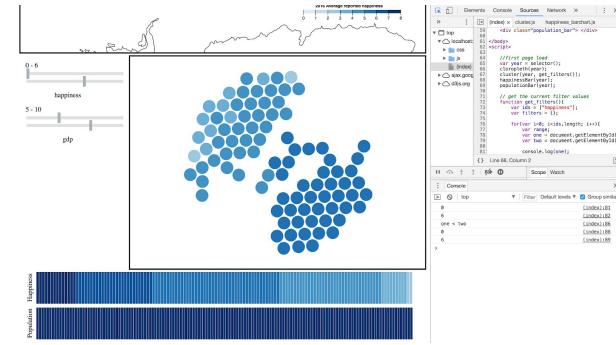
### Version 3 - Generating all charts



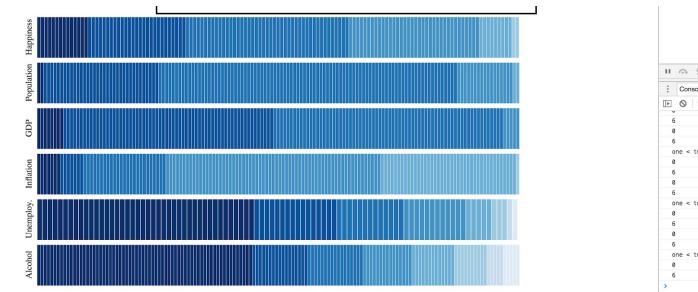
### Version 5 - Different Color Scales



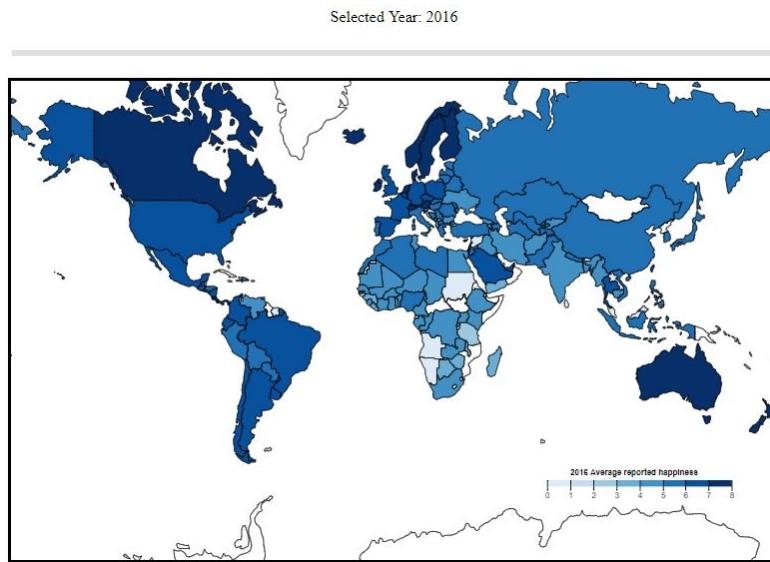
### Version 2 - Loading different years works



### Version 4 - Working scales

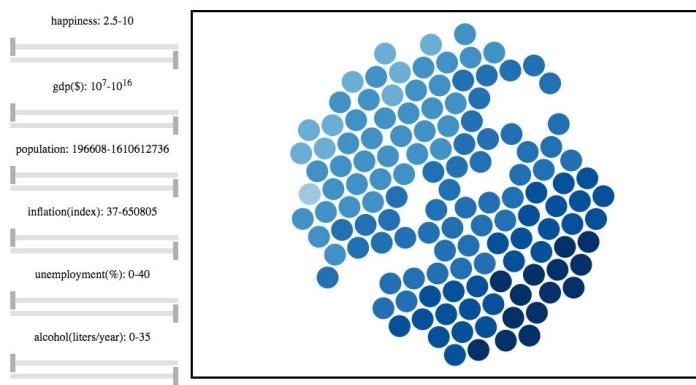


## Implementation

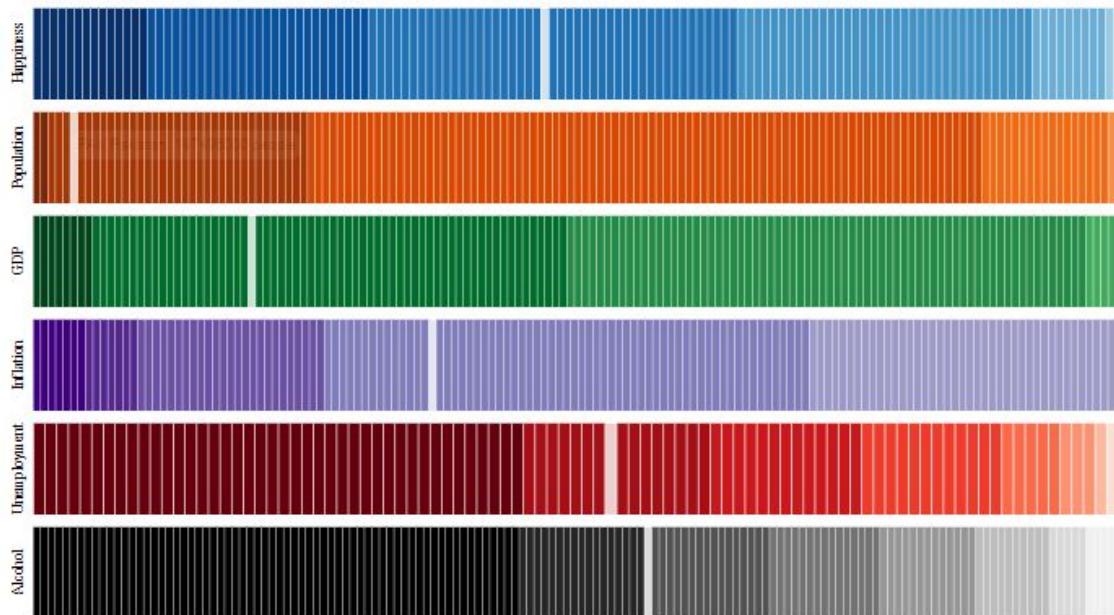


*Figure 4. Choropleth showing happiness.*

The primary visualization is the choropleth map of happiness by year. You can select individual countries and toggle the year between 2008 and 2016 at the top. Below this is a force directed graph of all the countries where the coded color matches that of the choropleth map. Beside it is a series of filters we created that allows you to define the parameters and will exclude, from the directed graph, the nations that do not meet the criteria. We chose to do this because this was the best idea we had for data exploration as it allows you to see the nation's happiness compared to others based on the filter values.



*Figure 5. Cluster chart and filters.*



*Figure 6. Bar charts and hover feature.*

A series of bar charts were chosen because they visualize the variables that we deemed most relevant for comparison to the slightly ambiguous “happiness score”; ambiguous only because it was based on a self-scoring system. The five comparison variables are Population, GDP, Inflation and Unemployment rates, and Alcohol Consumption in liters per person. As you can see above, in the section “Bar Chart Progression”, the bar chart we are creating is based on a gradient of colors instead of the heights of each bar. Between versions 3 and 4 we developed a more accurate scale using logarithms to more accurately scale each distinct section in the charts to make it consistent with the values that are displayed. Finally, we chose a new set of colors for the fifth version and tried to choose colors that matched the variables where applicable, such as green for GDP. Each chart takes in a year as its parameter and updates as the year is selected at the top of the page, accurately displaying the information to match the data displayed in the choropleth map and cluster chart.

## Evaluation

From this data we found that happiness is a hard metric to quantify and explain using other factors. Although we thought we were going to have a full picture through using metrics from multiple categories (life, monetary and extracurricular) it was still an insufficient and incomplete picture. One factor that we think held us back was the fact that our happiness data was a self reflection of each human who took the survey. If we had collected our own happiness data we could make more conclusions on the type of people taking the survey like personal things that they thought affected their own personal rating of themselves. If we had the chance to take the data ourselves it would be interesting to first ask the person how happy they would rate themselves then ask them things about themselves that they personally think affected it. Some ideas that come to mind would be asking them questions about their family life, work life and general culture in the area that they live.

When we were developing this project we were constantly thinking about the original questions we posed. The most important question we had was *What factors influence happiness?* When we were exploring our own data we found that unemployment seemed to be correlated to the happiness of countries. As unemployment grew, the happiness scoring went down. This could be caused by the subset of people who took the survey but we think that it makes sense that countries with large unemployment rates would be less happy overall.

Our website as a whole gives an interactive and interesting way to interact with the data we had and allows any user to make their own conclusions about happiness in different regions in the world. Our goal was to allow the user to explore the data and become inspired by their own life to see if they agree with national averages. If we were to improve our website one thing we would want to improve is adding more data. This would give the user a fuller picture of the countries status. Another thing we would want to add is making this more narrative. Giving the user some direction while interacting with the data would also add to the story we were trying to tell. Originally we didn't want to make any conclusions from the data because we thought it would taint the users ability to naively interact and make their own conclusions. However, after receiving some feedback we realize that we might have left the question too open ended. Some technical things we would add would be allowing for the user to directly compare two countries. We think that this would make things simpler. Overall we are happy with how the project came out given the time we had and the limited data resources.