Happiness Around the World

Info 247 final project by Natalia Timakova and Avi Dixit

Table of Contents:

Introduction 2

Project Goals 2

Data 3

Tools used 4

Related Work 5

Links 12

Process 12

Final version 16

Usability testing 22

Contribution of each team member 24

Introduction

GDP has always been an imperfect measure of people's well-being, and with advance of the digital age it became only worse. While internet is eating the world, many services that were for a fee become free, which takes them out of the GDP metrics but not out of people's lives. With Wikipedia, Google maps, Gmail we become better off, but GDP doesn't count it in. This is just one example of the reasoning behind the discussion about shifting from classical GDP to Gross National Well-Being or Gross National Happiness as a holistic measure of nations' development.

In this light, it would be instrumental to understand what constitutes happiness and if there are patterns and correlations between nations' happiness indices and these constituents. The Sustainable Development Solutions Network, a global initiative for the United Nations, attempts to do just this. Starting 2012, SDSN publishes the World Happiness Report containing not only latest happiness index, but also metrics on GDP per capita, social support, freedom to make life choices, healthy life expectancy, generosity, and perception of corruption. With the OLS regression model applied to these data for several years, SDSN found certain correlations between happiness scores and the abovementioned metrics. These metrics are now called predictors and are believed to be responsible for approximately 3/4 of the happiness score worldwide.

In our project, we attempted to visualize the World Happiness Report 2017, with the focus on revealing the components of the happiness score and their contribution to the rating.

Project Goals

As our project goal evolved, we transitioned from a simple per chapter visualization of the report to more analytical approach. Eventually, we decided to create an exploratory tool for people interested in understanding how a happiness score can be explained and what can be done to improve a happiness rating. The narrative of the visualization guides a reader from a simple stating of the happiness ranking through the concept of happiness predictors and to the actual impact of these predictor, in general and using an example.

Since the audience is the general public, we wanted to make our visualization accessible, easy to comprehend, entertaining, and yet thought-provoking.

At a certain point, we even considered donating our final work to the UN SDSN so they could publish it along the report at http://worldhappiness.report/ed/2017/. Depending on the results of this evaluation, we still might want to do it.

Data

In our visualization, we used Gallup World Poll data on happiness, freedom to make life choices, social support, generosity, and perception of corruption, as well as log GDP per capita, healthy life expectancy at birth, and population size metrics from the World Bank, World Health Organization, and UN Population Division. The data covers all available countries and years 2006-2016. Everything but the population size was found in the appendix to the World Happiness report. Link to the data are here: http://worldhappiness.report/ed/2017/. The population data for the relevant years for all countries was provided by the UN Department of Economic and Social Affairs: https://esa.un.org/unpd/wpp/Download/Standard/Population/.

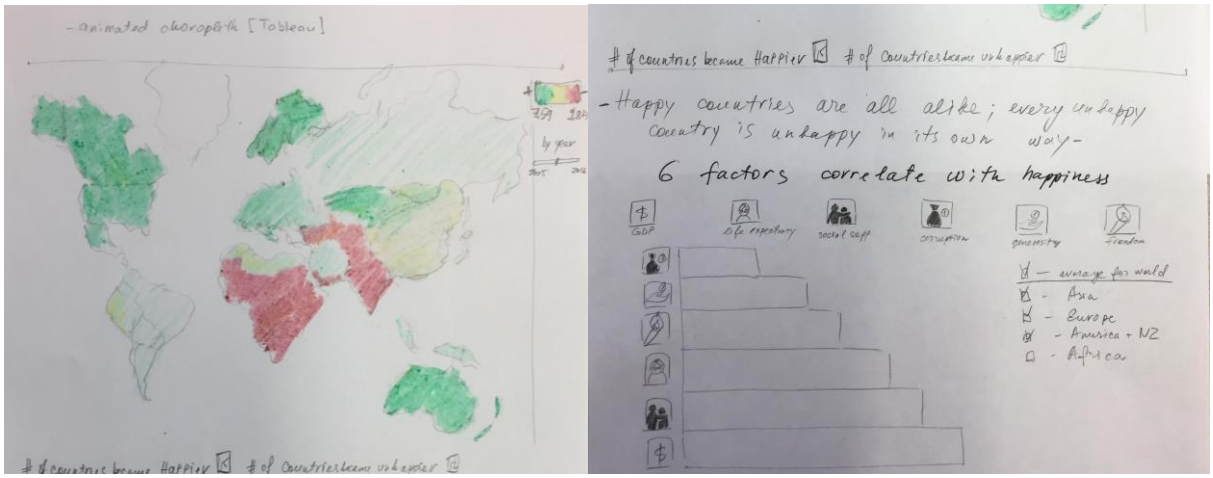
There were no significant gaps in data that could put our work at risk. The list of participated countries was inconsistent throughout the decade, but still covered the majority.

Cleaning and mingling the data took some time, especially for the bubble charts in the chapter *Can We Predict the Happiness*. This part of the visualization required our data to be in a JSON format for which we wrote a Python script that converted the CSVs to the JSON format required by each of the bubble charts.

It was also challenging to understand how the OLS regression coefficients were calculated by Helliwell et.al. in the report and what kind of visualization would explain them the best.

Tools used

Pencils were used for early sketches.



We used Tableau for prototyping the charts and in the process of defining the narrative. We found this tool fast and efficient for this purpose. We also returned to Tableau to visualize the interactive bars in the final chapter *Let's Make America Happy Again*. Tableau allowed us to create a Dashboard of two Sheets one of which featured bars, another showed custom made icons that worked like filters for the bars. Icons were downloaded from TheNounProject.

Illustrator was used for editing icons to represent factors in happiness and for creating the isotype representing factor weights.

D3 was used for majority of the visualizations including the bar chart, the scatter plots for the factors, and the parallel coordinates chart.

Bootstrap was used as a web framework for making responsive applications (HTML, CSS, and JavaScript).

Python - to create a CSV to JSON parser.

MS Excel

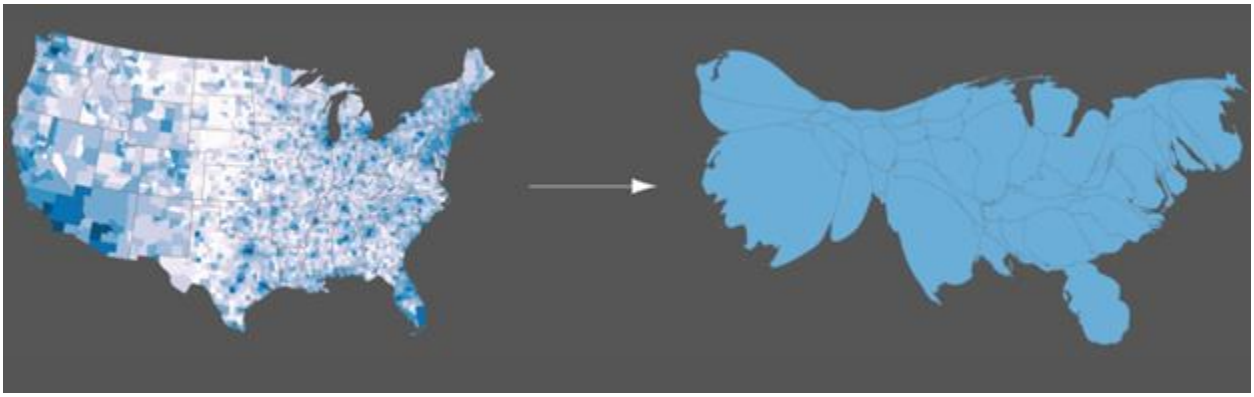
Custom fonts: https://fonts.google.com/. We downloaded Open Sans.

Related Work

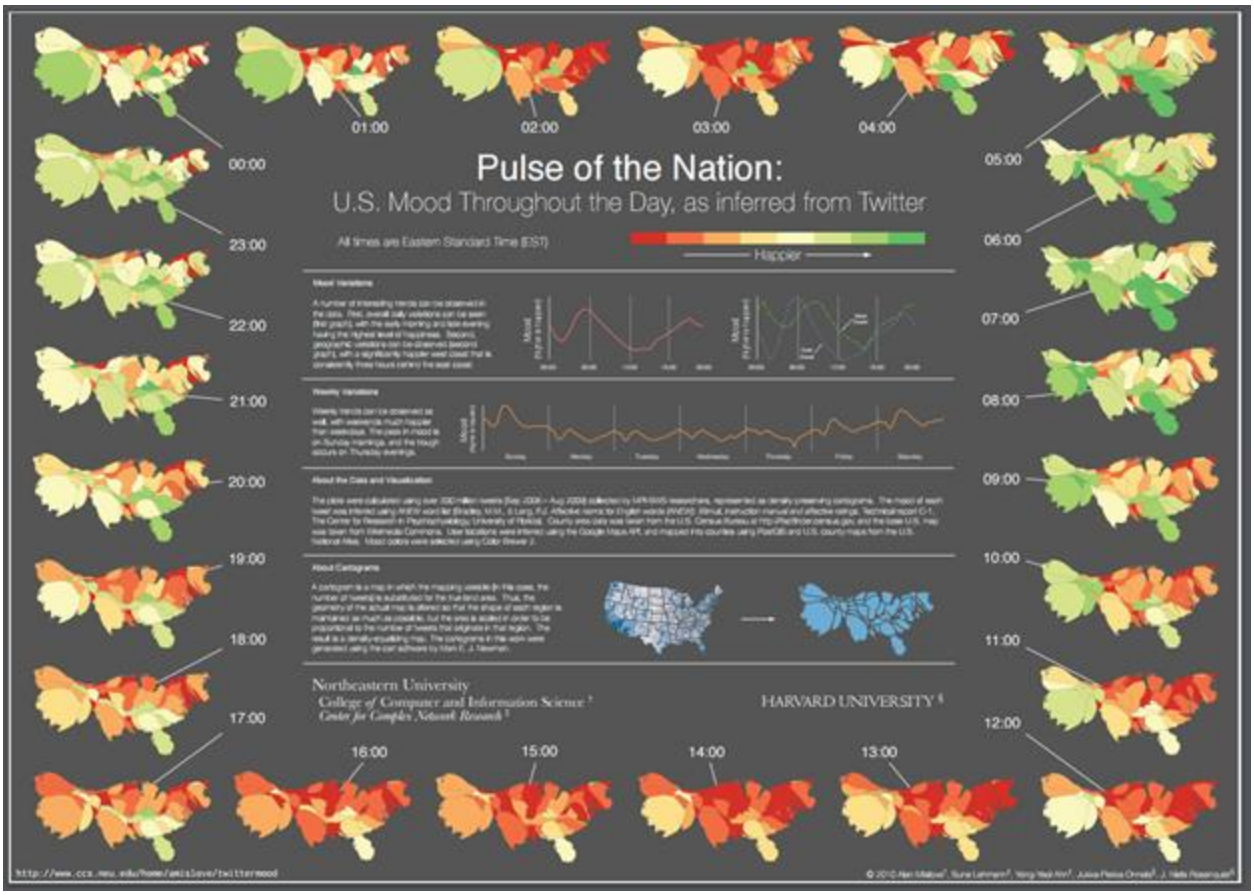
There were several different ideas that we came across when searching for related work, and some of them were very close to the topic we were trying to visualize:

1. Pulse of the Nation (http://www.ccs.neu.edu/home/amislove/twittermood/):

This is a project undertaken by Alan Mislove at Northeastern University. It attempts to visualize the mood of the United States throughout the day using 300 million tweets. One of the interesting things about this visualization is that it uses a cartogram instead of a normal map to account for the different number of tweets occurring from each region. A cartogram is a map in which the mapping variable, tweets in this case, are substituted for the true land area. Thus, the geography of the actual map is altered so that the shape of each region is maintained as much as possible, but the area is scaled in order to be proportional to the number of tweets that originate in that region.



The visualization below shows more happy areas as green and less happy areas as red and plots the distribution throughout the hours, which allows the researchers to observe interesting trends and variations over regions and weeks.

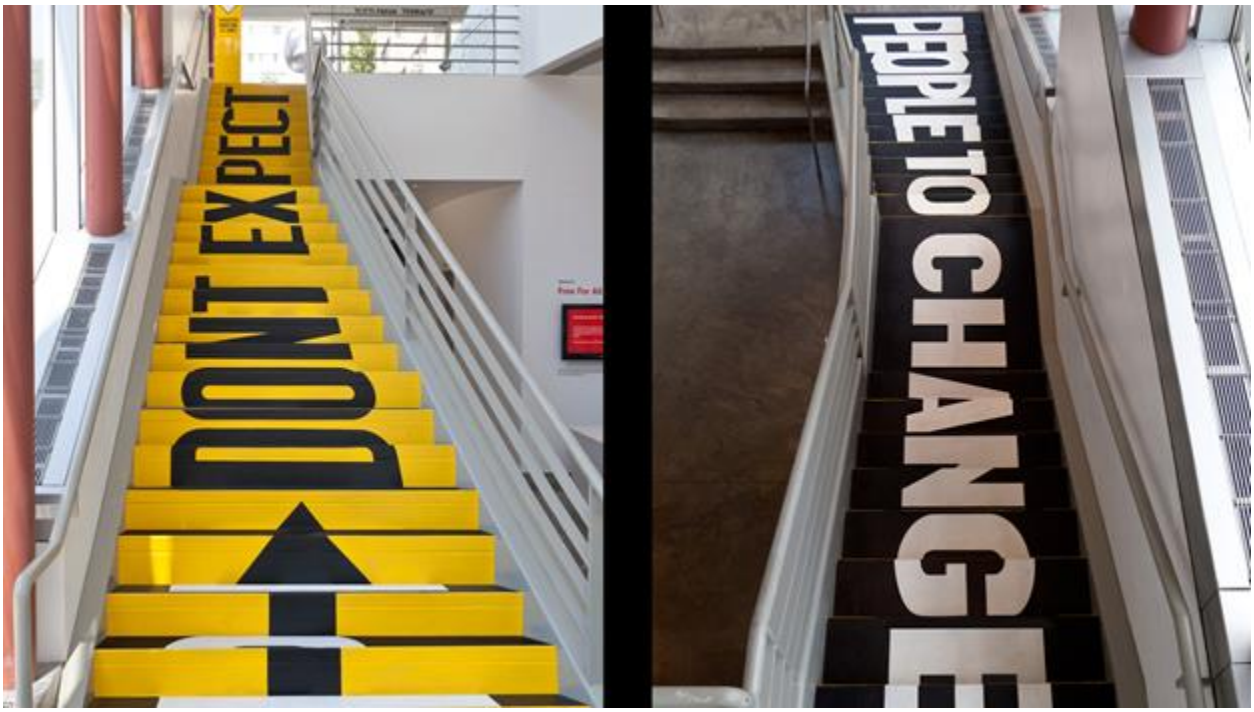


Although this viz fixes a common "bug" of the choropleths (when the geographical size of the region dominates the meaning and small regions "don't matter") we still think it is not good enough for our purpose. If we distort the map of the world according to the happiness score, some small unhappy countries may simply vanish. Also, the countries will not be recognizable by their shapes, so the use if the map loses its meaning.

A cartogram, probably, looks best on the multidimensional infographics, when it is plotted along a choropleth and a proportional symbol map, as Alberto Cairo shows in his *The Functional Art* (p.54).

2. The Happy Show (http://sagmeisterwalsh.com/work/all/the-happy-show/):

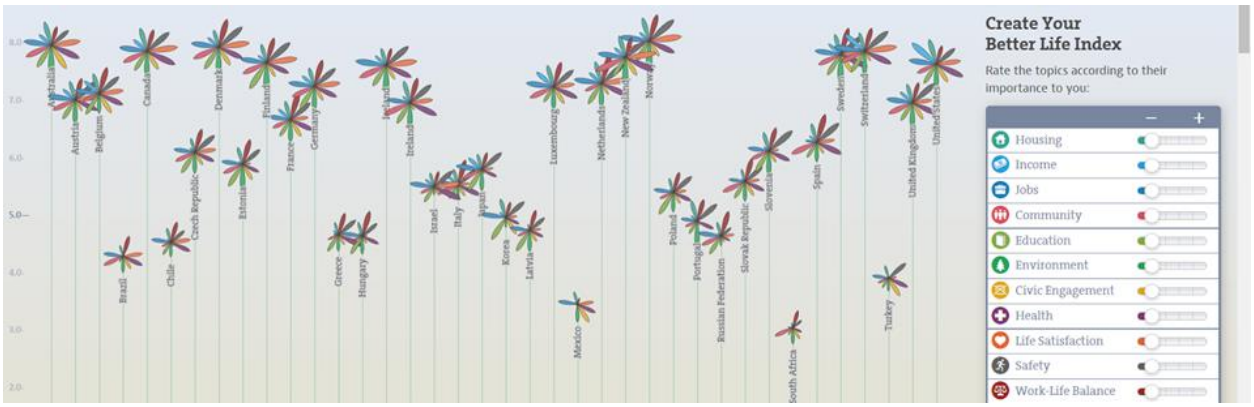
This exhibition by Stefan Sagmeister attempts to delve deeper into what it means to be happy by visualizing his exploration of happiness, and details rules or maxims that the designer thought of in imaginative and interactive forms. Apart from providing a personal narrative on what the designer considers to constitute happiness, the exhibition also had social data detailing the role of gender, age, race, money, and other factors that determine happiness.



While this kind of work was clearly out of scope for this class, it was interesting to see how factors and perceptions of happiness could be "charted out" though physical representations.

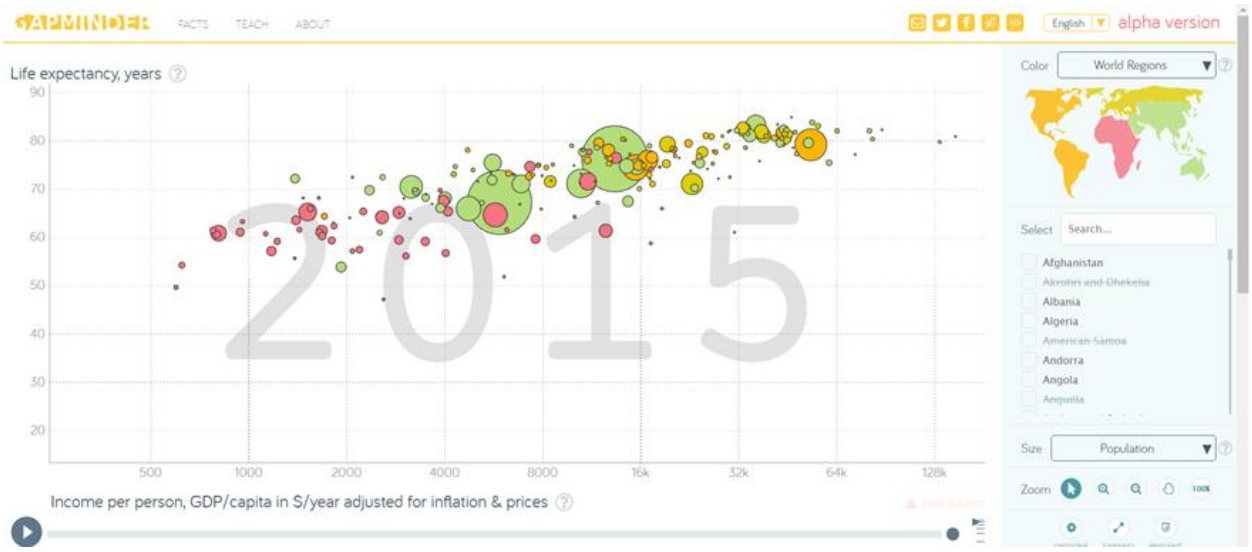
3. OECD Better Life Index (http://www.oecdbetterlifeindex.org/):

It's a well-known interactive visualization created by the Organisation for Economic Co-operation and Development. It is limited to OECD countries and three non-OECD economies - Brazil, Russia, South Africa. In terms of data, Better Life Index has one variable similar to happiness called life satisfaction which they plot along with ratings for housing, jobs, income, environment, safety and some others. They do not try to explain life satisfaction with other subjective or objective metrics, while our Happiness Report has exactly this task.



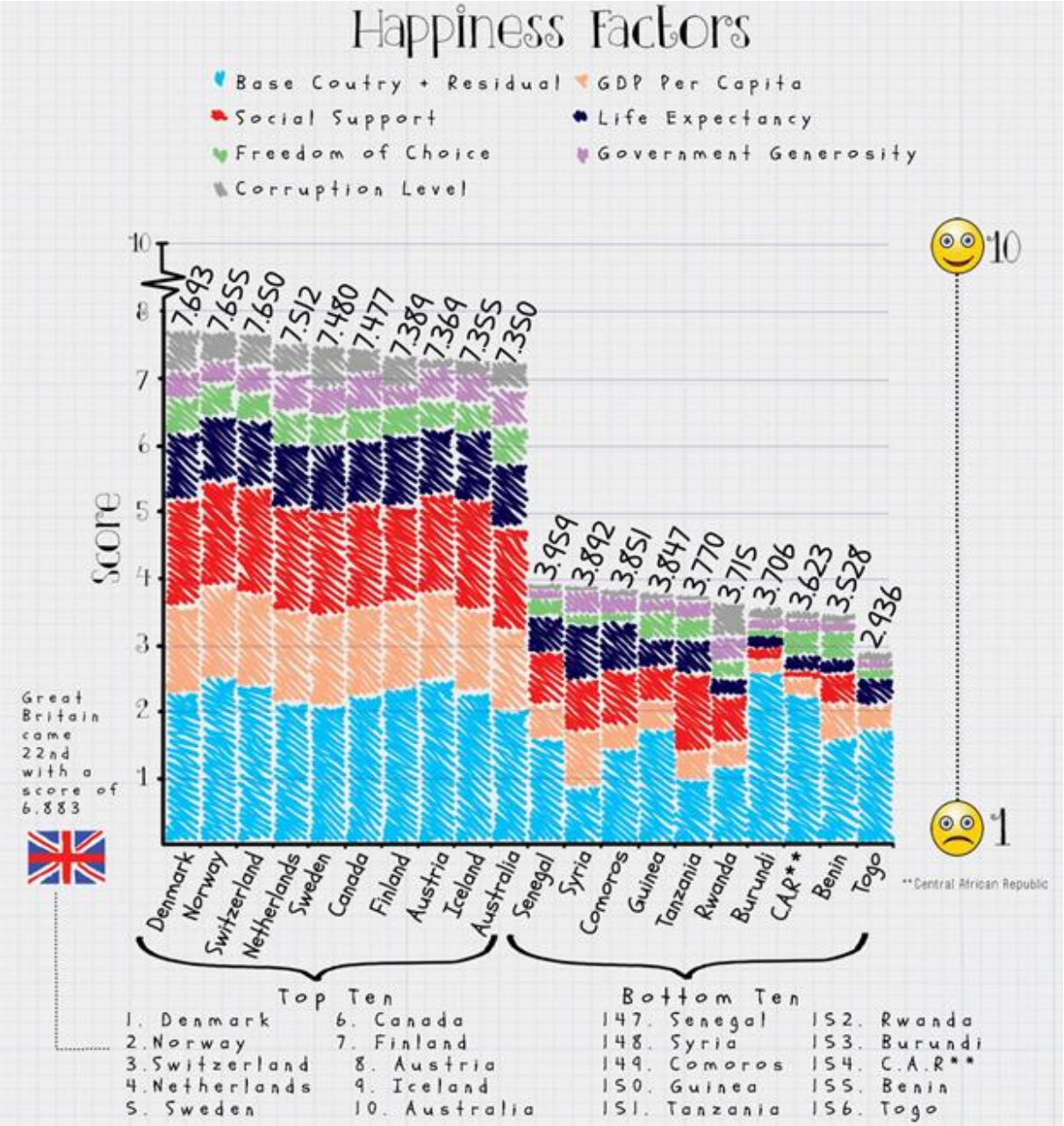
4. Gapminder (http://www.gapminder.org)

Another well-known visualization of the world metrics. Incepted by Hans Rosling and further developed and supported by his son Ola, Gapmider is a tool for educators and a general public. The software in the core of it called Trendalyzer was build by Ola, uses interactive bubble chart as a visualization form, and now belongs to Google.



We found this visualization effective for demonstration correlations between happiness score and each of the six happiness predictors, aesthetically pleasing and also possible to build with D3.

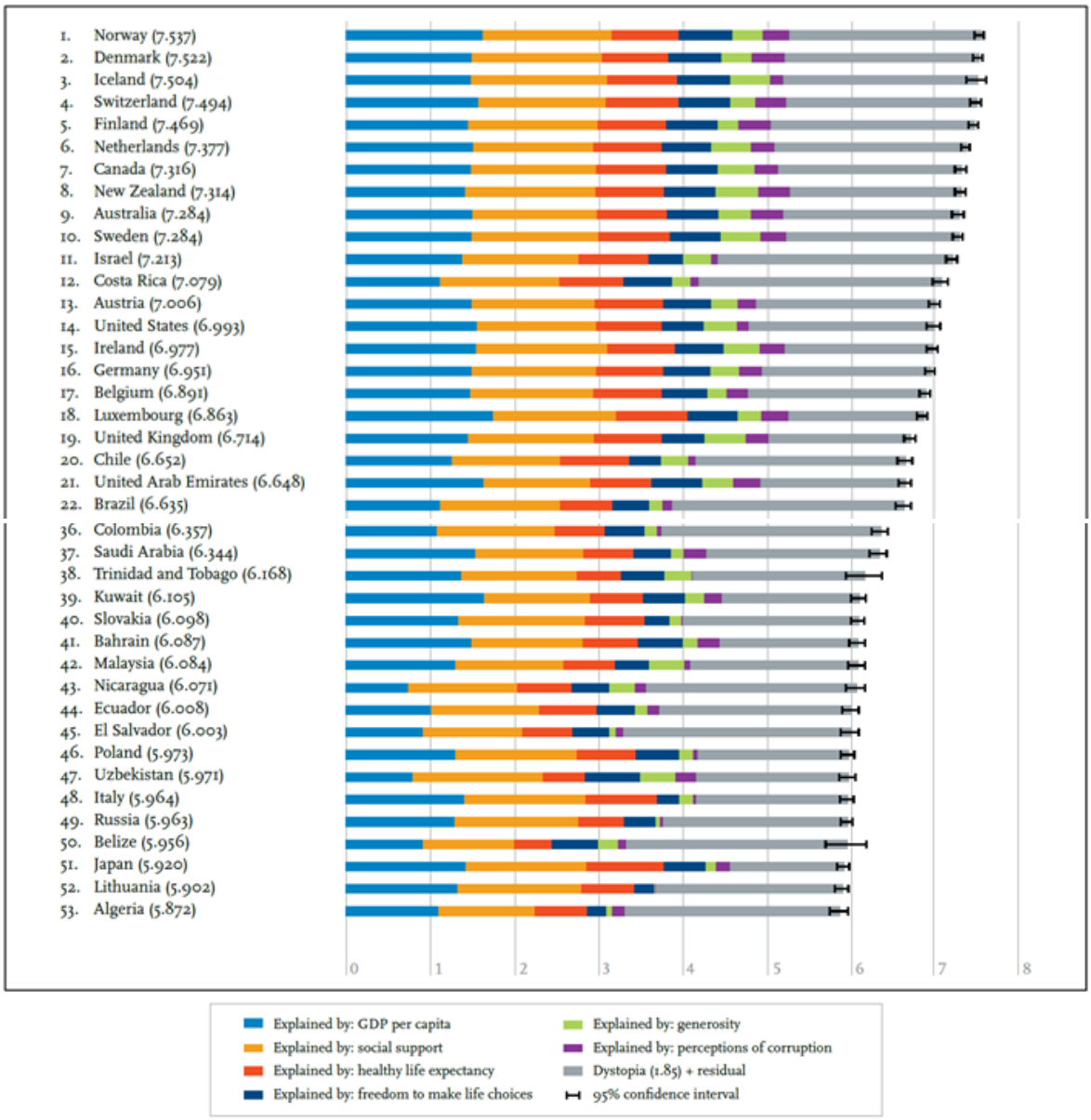
5. Amateaur viz by Jonathan Rolande, refering to himself as Director Of House Buy Fast, http://visual.ly/happiest-countries-2014



Jonathan's work is especially interesting for us because he tried to visualize the World Happiness Report. He also realised the importance of the happiness predictors, although the sketchy style that he had chosen and the format of the stacked bar chart does not help accurately estimate the sizes of the segments and compare them.

6. The original viz of the World Happiness Report 2017 team, http://worldhappiness.report/wp-content/uploads/sites/2/2017/03/HR17-Ch2.pdf

Figure 2.2: Ranking of Happiness 2014-2016 (Part 1)



The core chart is the Ranking of Happiness averaged for three past years. It's a stacked bar chart that divides the happiness score on six segments which can be explained by the happiness predictors, a dystopia segment and also has a confidence interval line. As with the viz above, stacked bars do not allow to compare segments (except for the first one) across the countries. Although this chart contains a lot of information, this information is not easily retrievable.

The Links

Website: http://people.ischool.berkeley.edu/~avi.dixit/
Github link: https://github.com/addt/Infoviz-Final-Project

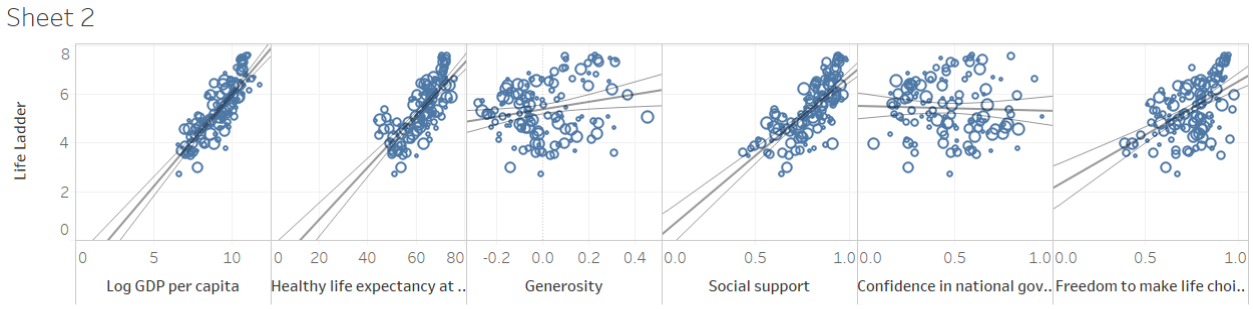
The Process

We started out with the idea to visualize the World Happiness Ranking scores using a choropleth. A map is, probably, the most natural association with the word "world," but rarely a good choice for visualization. We realized our choropleth would make the bigger countries stand out while the smaller countries would be hardly noticeable (people would have hard time finding the happiest Finland!). Color would also bias the perception towards the brightest one, and a subtle difference between the shades would make comparisons between distant countries difficult. Also, as A. Cairo said about similar problem with the information graphics called "The Defense of the Neighbors: An overview of the armed forces of countries surrounding Brazil,"[1] using so much real estate for the map suggests that the main goal of the graphic is to show where the countries are located. That was not our goal.

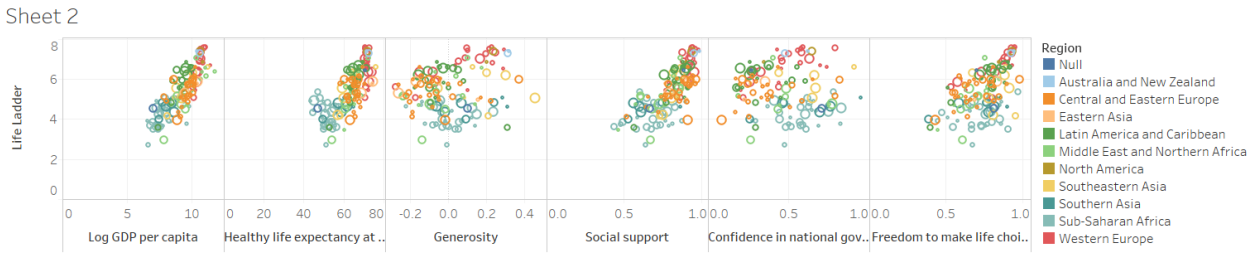
We wanted our opening chart to give a clear impression of what countries were on top of the happiness ladder, and what were in the bottom. A simple bar chart ranged by the score seemed an ideal solution.

To visualize correlations between the happiness factors and a happiness score, we decided to follow Hans Rosling's successful visualization with a bubble chart. This type of charts lets us compare happiness score against all six happiness predictors individually, with account for a population size. Below, you can see the prototypes of the six bubble charts created in Tableau, with the trend lines. The trend lines let us explore our data further and add some thoughtful description in our final visualization.

[1]Cairo Alberto, The Functional Art. An introduction to information graphics and visualization. 2013. pp-26-31



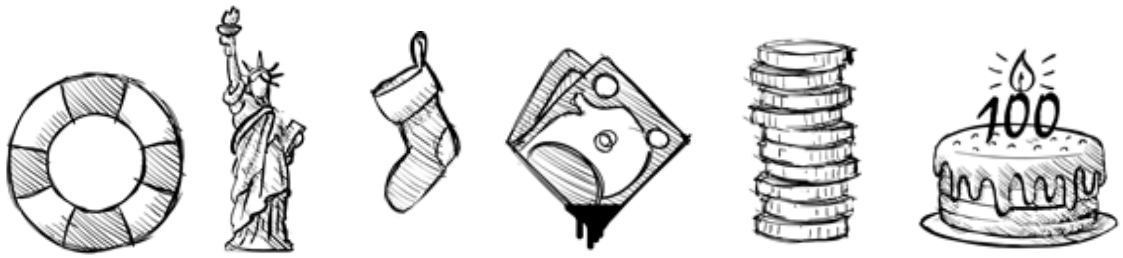
Sum of Log GDP per capita, sum of Healthy life expectancy at birth, sum of Generosity, sum of Social support, sum of Confidence in national government and sum of Freedom to make life choices vs. sum of Life Ladder. Size shows details about 2015 (popsize). Details are shown for Country. The data is filtered on Year, which ranges from 2015 to 2015. The view is filtered on 2015 (popsize), which keeps no members.



Sum of Log GDP per capita, sum of Healthy life expectancy at birth, sum of Generosity, sum of Social support, sum of Confidence in national government and sum of Freedom to make life choices vs. sum of Life Ladder. Color shows details about Region (2016). Size shows details about 2015 (popsize). Details are shown for Country. The data is filtered on Year, which ranges from 2015 to 2015. The view is filtered on 2015 (popsize), which keeps no members.

Bubble chart, although a powerful tool, is limited to three dimensions (two axes and a bubble size). We needed another chart to show all six factors in one plot. Spider chart came to mind, but we quickly realized our data (log GDP, scores for freedom, age for life expectancy, etc.) were of different scale, while spider chart works best when all the axes are of the same scale. A parallel coordinates chart became a better alternative, even allowing for a functionality a spider chart does not grant.

Icons that represent the coefficients and are a unifying element of the visualization were initially in two sets. Set 1 contained these sketchy icons:



(from left to right: social support, freedom to make life choices, generosity, corruption, GDP, healthy life expectancy at birth)

Set 2:

(from left to right: social support, freedom to make life choices, generosity, corruption, GDP, healthy life expectancy at birth)

We tested both sets among our classmates and arrived to a conclusion that set 2 conveyed the meaning better than set 1. Also, people noted that these icons are more "iconic" than the sketchy ones.

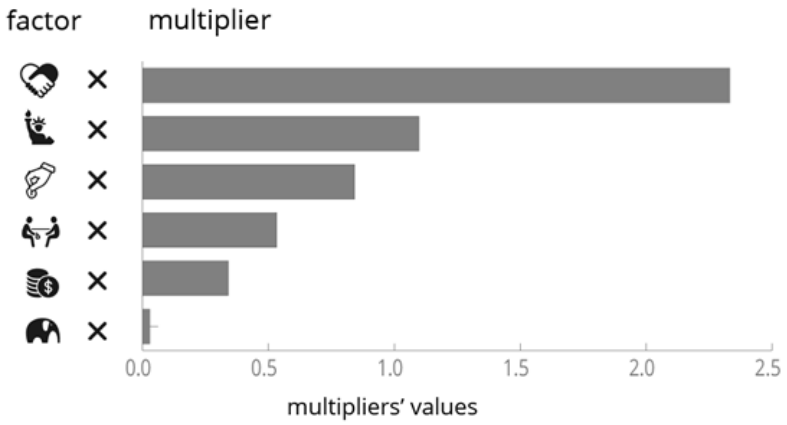
Probably, the most difficult part was visualizing the OLS regression coefficients represented in the table 2.1 of the report.

Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

	Dependent Variable			
Independent Variable	Cantril Ladder	Positive Affect	Negative Affect	Cantril Ladder
Log GDP per capita	0.341 (0.06)***	-.002 (0.009)	0.01 (0.008)	0.343 (0.06)***
Social support	2.332 (0.407)***	0.255 (0.051)***	-0.258 (0.047)***	1.813 (0.407)***
Healthy life expectancy at birth	0.029 (0.008)***	0.0002 (0.001)	0.001 (0.001)	0.028 (0.008)***
Freedom to make life choices	1.098 (0.31)***	0.325 (0.039)***	-.081 (0.043)*	0.403 (0.301)
Generosity	0.842 (0.273)***	0.164 (0.031)***	-.006 (0.029)	0.482 (0.275)*
Perceptions of corruption	-.533 (0.287)*	0.029 (0.028)	0.095 (0.025)***	-.607 (0.276)**
Positive affect				2.199 (0.428)***
Negative affect				0.153 (0.474)
Year fixed effects	Included	Included	Included	Included
Number of countries	155	155	155	155
Number of obs.	1,249	1,246	1,248	1,245
Adjusted R-squared	0.746	0.49	0.233	0.767

Notes: This is a pooled OLS regression for a tattered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 to 2016. See *Technical Box 2* for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels respectively.

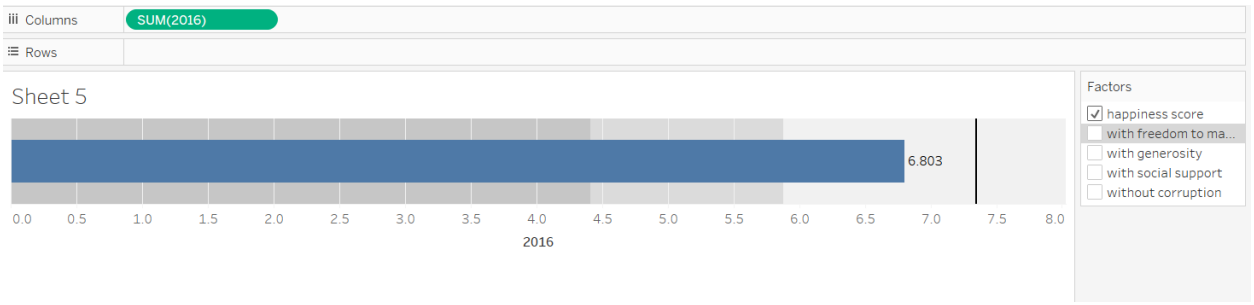
Although one of the main principles of the information visualization is “Clarify, not Simplify”, in this case we decided to sacrifice data about robust standard errors and significance and leave only the coefficients. One of the early versions of this visualization looked like this:



Eventually, we made one step further and calculated the weight of each factor in the happiness score as if the value of the factors were equivalent and they differed only by the coefficient. We compared all the factors against the biggest of them, and visualized the result in an isotype.

In order to make the impact of the factors visually salient, we needed to see them in action. The final chapter of the World Happiness Report 2017 featuring the decline in happiness in the United States came handy.

We decided to plot an interactive bar that would grow with more factors included, all factors being calculated using the coefficients and data from the chapter. One of the versions was a bullet chart (invented by Stephen Few) with 2016 as a main bar and 2006 as a control year (vertical bar). However, with our users in mind, we decided against it. If we designed for statisticians and not general public, we would go with this design though.



Final version

In this chapter, we will review the final version of the visualization and explain the choice of a chart and a tool for each section.

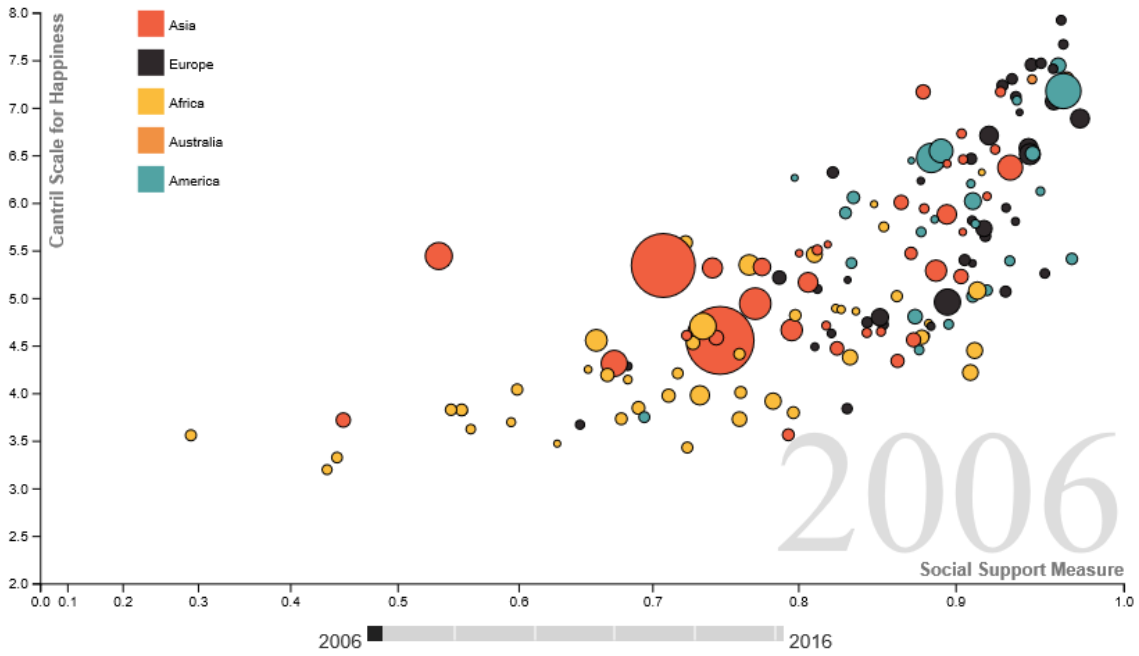
Section: Happiness Rankings
Visualization: Sorted bar chart
Tool: D3

We started by showing users a simple bar chart with rankings of countries according to their happiness index. The aim of this visualization was just to introduce users to the topic and let them explore the rankings by clicking on the year tabs above the chart and scrolling the chart. There were some trends that were apparent in the charts such as the dominance of western

European or North American countries at the top of the rankings and African countries at the bottom. We had initially planned to have this kind of trends highlighted in the visualization so that users don't have to browse through the entire data to infer them, but we just didn't have the time to integrate something like that. Each chart was associated with a CSV file that contained the rankings in a sorted order along with the happiness rating for that country. We used D3 to display this visualization because we wanted to perform some customizations on it that Tableau didn't allow (customized tooltip, row highlight color on mousing over).

Section: Can we Predict Happiness?
Visualization: Bubble charts
Tool: D3

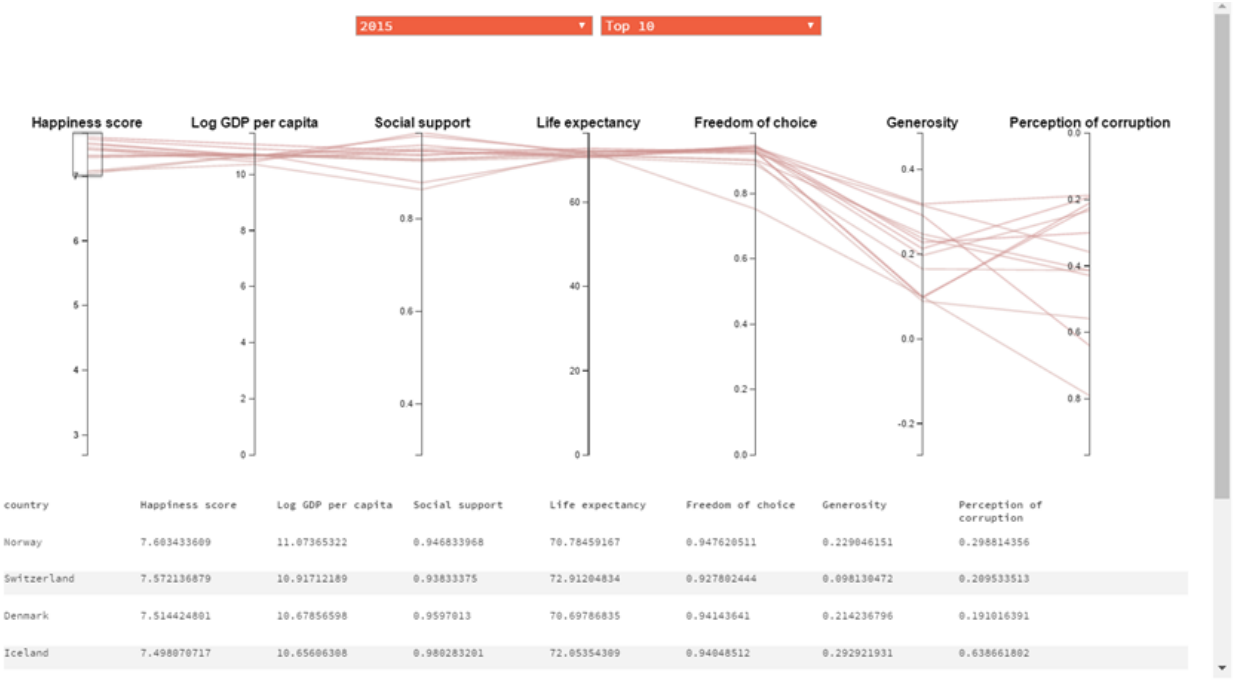
In this next part, we wanted users to explore the factors that are believed to constitute the happiness rating. Icons are supposed to represent the factors. Users could hover over them to see an explanation of how the data for each factor was collected. The layout for this part was created using illustrator and bootstrap frameworks grid system to show them in a specific grid and for displaying the explanation on mouse over.



This bubble chart pops up upon clicking on a factor icon, where the X axis represents a factor, the Y axis - the happiness rating, and the size of the bubble is proportional to the population of the country. We chose this chart because it is space-efficient, and has no problems associated with choropleths where the size of the country is not proportional to the population and thus can provide a distorted view of the actual facts. This graph shows the correlation of the factors with the happiness rating over the years, from 2006-2016. To see the dynamics, you need to use a slider at the bottom of the graph. Further, on hovering the mouse over any of the bubbles in the chart, users are able to see the names of countries, and, upon clicking, see the progress of a country represented by a gray trace line.

This graph was made completely in D3 and used data aggregated from the World Happiness Report and from population metrics of the UN's Department of Economic and Social Affairs. For these charts, we needed to write a Python script to convert the aggregated CSVs into a JSON format.

Section: All Factors in One Plot
Visualization: Parallel Coordinates chart
Tool: D3

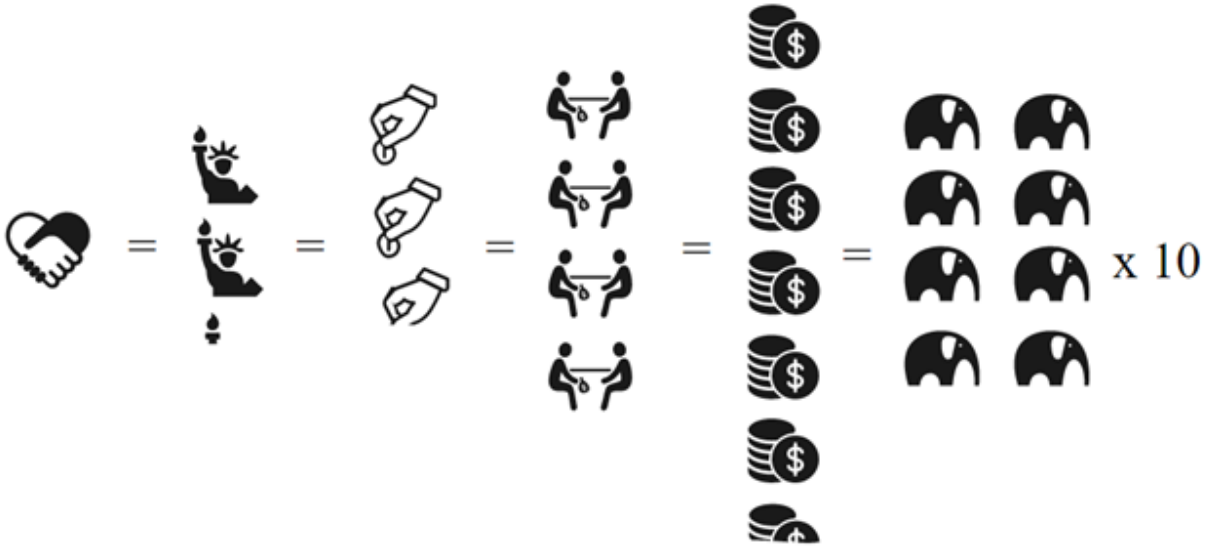


country	Happiness score	Log GDP per capita	Social support	Life expectancy	Freedom of choice	Generosity	Perception of corruption
Norway	7.603433609	11.07365322	0.946833968	70.78459167	0.947620511	0.229046151	0.298814356
Switzerland	7.572136879	10.91712189	0.93833375	72.91204834	0.927802444	0.098130472	0.209533513
Denmark	7.514424801	10.67856598	0.9597013	70.69786835	0.94143641	0.214236796	0.191016391
Iceland	7.498070717	10.65606308	0.980283201	72.05354309	0.94048512	0.292921931	0.638661802

This parallel coordinates chart was chosen because it is friendly to data represented in disparate scales (unlike spider chart) and provides a big picture of the correlations between the happiness score and all six factors. Thanks to this chart, we successfully proved Leo Tolstoy's hypothesis about families extended to countries. Indeed, happy countries are all alike; every unhappy country is unhappy in its own way.

D3 library supports parallel coordinates very well, so we didn't look further than D3 in search for a tool. We think the chart worked pretty well, although there is still room for improvement. For the updated version, we would sort the countries in the drop-down menu alphabetically (effectively removing the "top 10 countries" filter), put the table into a separate div or remove it.

Section: Not All Factors are Created Equal
Visualization: Isotype
Tool: Adobe Illustrator



The goal of this isotype is to visualize the OLS regression coefficients for the happiness predictors described in the previous chapter. So, in a hypothetical country, if all of the factors had the same value in some bridge unit, their impact on happiness would still be different and dependent on the coefficients. For example, the coefficient for social support and consequently its impact on happiness score is equal to 2.332; the coefficient for freedom to make life choices is 1.098. In that case, if we want to compare the factors against each other to see which is more "important", we may say that the value of one social support unit would be equal to the value of 2.332/1.098 = 2.12 units of freedom to make life choices, and so on. These proportions are reflected in the isotype.

We chose isotype for its clarity and ability to directly convey messages. The message we wanted to deliver was about the importance of the social support for happiness. The choice of the tool is pretty straightforward: Illustrator.

Section: Let's Make America Happy Again!
Visualization: Interactive bar chart
Tool: Tableau

Bar chart is commonly used and easily understood by general public. Comparing the length of two bars plotted along each other should be an easy task. In this visualization, we sought to engage a user into a play. In order to make our work truly interactive and also to let the reader change the perspective and "participate" in our research, we created interactive filters disguised as already familiar icons: freedom to make life choices, generosity, social support and perception of corruption. Upon click, these filters add a designated value to the bar above, effectively increasing the US happiness score. The values were calculated by Jeffrey Sachs who authored the chapter Restoring American Happiness in the World Happiness Report 2017.

We used Tableau for this visualization because it allowed us to do it faster than it would probably take us in D3.

Usability testing

To each part of the viz we posed a series of questions we aimed at answering with our work. Nine people we presented with the viz and asked to answer these questions. The results are in the chart below:

Question (answer)	# of correct answers	# of incorrect answers
1. Can you tell what the least happy country was in the world in 2016? (Central African Republic)	9	0
2. Go to the Can We Predict Happiness. Click on the GDP tile, then on the Generosity tile. What pair of variables seem to be directly correlated: Cantrill ladder score & GDP or Cantrill ladder score & Generosity? (CLS&GDP)	9	0
3. Look at India's dynamic on the healthy life expectancy bubble plot. In the past decade, India's ..: a) life expectancy was growing but happiness score declined b) both life expectancy and happiness declined (a)	8	1
4. In All Factors in One Plot section, choose any year and box the happiness axis of the chart around the scores 8-7. Look at the highlighted lines. What four variables are consistently high among the happy countries? (GDP, Social support, Life Expectancy, Freedom of Choice)	8	1
5. Move the box down the happiness axis and look how the lines change. Can you tell that less happy countries are consistently low on all the predictors' axes? (no)	7	2
6. On the Not All Factors are Created Equal, how much units	7	2

of healthy life expectancy one need to offset the effect of one unit of social support? (80)		
7. Let's Make America Happy Again! Todays' US happiness score is below 7.0. If you add to it all the freedom, generosity, and social support the United States lost in the past decade, will the score exceed 7.0? (yes)	7	2
Total:		

Generally, people familiar with statistics and the main types of charts answered all the questions correctly, while people not used to reading charts made some mistakes. As a result, we've learned about several weaknesses of our visualization:

1. A few people thought that the tiles in the section *Can we Predict Happiness* are unclickable until they carefully read the preceding description. This is, probably, because the tiles have a double functionality: they give the information about the predictors upon hovering, so people do not expect them to do something else upon clicking.
2. One person said, she couldn't find India on the plot. Probably, a search findow could fix that.
3. Two people said they would expect the countries in the drop-down menu in the parallel coordinates chart to be sorted alphabetically, not by ranking. We think, it's a fair request.
4. In the section *Not All Factors are Created Equal*, a few people had troubles counting the elephants: in one case the user said the number is 100, apparently counting 8 as 10; in another case the user didn't notice **x10**. This is, probably, because eight is a too big of a number to visually grasp (preattentive properties), even if it's a 2 by 4 squad of elephants. We will try to employ the gestalt principle of proximity and regroup elephants into two groups of four, so that it would be easier to calculate them. Also, we need to make **x10** bigger.
5. For *Let's Make America Happy Again* section, we got the comment that <ctrl> on PC equals <cmnd> on Mac, so we need to account for that in our instructions (we both are PC users).

Apart from usability testing, we collected the feedback from our TAs. Kinshuk suggested we should have a table under the parallel coordinates chart in a separate div element. We agree it would look better and will try to implement this. Keshav recommended to enlarge the font.

Contribution of each team member

Project Tasks	Avi	Natalia
Project Research and Literature Review	30%	70%
EDA and Data formatting	60%	40%
Designing Layout	40%	60%
D3 Charts	70%	30%
Illustrator and Tableau Charts	30%	70%
Front End development	70%	30%
User Research and Usability Testing	50%	50%
Report	50%	50%
Total	50%	50%