INFO3300- Project 2 Written Description

04/09/19

Aileen Cai (ac952), Nate Schickler (njs59)

Data Description

Data Source: https://www.kaggle.com/unsdsn/world-happiness/version/2?login=true#2017.csv

"The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 158 countries by their happiness levels, was released at the United Nations at an event celebrating the International Day of Happiness on March 20th. The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields - economics, psychology, survey analysis, national statistics, health, public policy and more - describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and shows how the new science of happiness explains personal and national variations in happiness. The Happiness ScoreA metric was measured in 2015-2017 by asking the sampled people the question: 'How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest.'"

Data Processing

Our datasets for the project include three world happiness csv files for 2015, 2016 and 2017. The 2017 csv file did not contain region as a key or value, so the 2015 and 2016 files had to be processed before adding that element to the data object. Additionally, the world-110m-country-names.tsv files contained the id for each country that had to be mapped with the countries in the world-110m.json file. Separate dictionaries were created to store country happiness score and happiness rank based on the country id specified in the topojson. During the data filtering process, the factors (GDP, freedom, etc.) had to be casted as a Number and ensure it did not contain null values. The data includes several variables of interest to our project. First is happiness score and rank, where each country is given a score and ranked against other countries. These will be used to visualize countries on the map. We used d3 and topojson to form the global map svg. The columns such as GDP, family, life expectancy, freedom, and trust are used in the scatterplot to show and explore regression and correlation with happiness rating.

Visual Design Rationale

We chose to display the data in two primary visualizations. One is the global map with a color key for happiness score and the other is a scatterplot that groups countries by continent using

color. In the scatterplot we can interactively display different variables in the X axis to show correlations with happiness. In this way, it is possible to explore and visualize different metrics for measuring and approximating happiness and see correlations in the data. The map and continent grouping in the scatterplot allows us to see geographically how happy people are around the world. A continent on the map that is generally darker in the graph will be happier than a lighter one, and a color group of dots in the scatterplot that denotes a continent that is higher than the other dots is visually happier than one that is lower. The scatter plot regions were given distinct colors so that users can locate them in the graph. The circle sizes were also relatively small so that the data points don't overlap as much. Both x and y axis scales are linear so that the points are spaced out and it would be easy to transition between different factors. Our design (color scheme) and layout were in part inspired by
https://mimno.github.io/showcase/project2/workforce/

Interactive Design Elements

The choropleth map contains buttons that allow users to choose a specific year ranging from 2015-2017. Some of the affordances that went into this design include making the buttons a different color (black) when hovering so that mouse movement is more salient, and changing the color of the year button when it is selected (blue). This lets the user know that the year they selected will show the data for the choropleth map and scatterplot below it.
The map also contains more information as the user hovers over each country. The text box is translucent so that the countries behind it are still visible and creates a contrast on the words. Since the hover features are usually not intuitive at first, instructions were added after the buttons so the user can move their mouse to see the happiness score and rank.
The scatter plot data is also dependent on the year selected so the button interactivity are slightly different for the factors (eg: gdp, freedom, etc..). In this scenario, the selected factor buttons will not stay blue after the user selects it, but the color will switch to black as the mouse hovers over. This design was implemented so that it is more obvious that the choropleth map and scatterplot data are from the same year.
Additionally, the data points in the graph contain the name of countries if the user moves their mouse on the graph. Instructions were not included this time because this seemed more intuitive after the experience with the choropleth map above it. During the demo session, we received a lot of feedback about how the region hover feature was not intuitive because it did not contain button styling. Most people thought it was just a static label instead of an interactive feature. In order to make this element more discoverable, the list of regions where displayed on the side of the graph as buttons that users can hover over to see more specific data points.

Dynamic regression lines were also included as the user moves their mouse to a region text. It displays the average regression line (all countries) in the background as a lighter shade, and the

current sub regression line for the region selected. This will help the user compare slope discrepancies with the average correlation.

Visualization Insight

The choropleth map shows the happiness level in each country by their rank and score from 0-10 (happiest). The user can select between three years to see how each country has changed by color. Since the color scheme is sequentially blue, it makes the changes more obvious if the user wants to see the happiness score/ rank over time. The darkest blue indicates the happiest country based on self-reported score. The map shows that countries in northern Europe (Norway, Sweden, Denmark), North America (Canada, US), parts of Latin America (Brazil, Chile) and Australia rank consistently higher. The results are not that surprising since these countries also have higher GDP, sense of freedom and life expectancy. However, it was interesting the visualize how countries decrease or increase in rank over time.

The scatter plot gives some more insight on how happiness correlates with five main factors (GDP, freedom, family, life expectancy, and trust in government). In order to illustrate this, a regression line was added to get the relation between the factor and happiness score. Based on the visualization, there seems to be a similar trend for all three years (2015-2017). The sub-regression lines in European regions are usually above the regression line for all countries, which is reinforced by the higher happiness scores compared to African countries and the Southeast asian region. Factors like GDP, Family, and Life Expectancy seem to have the most variance in happiness level since the index would fluctuate between a score of 3 to 7. Whereas in factors such as Trust and Freedom, the happiness index is only between (5 to 6) or (4 to 6). This implies that those factors are probably do not affect self-reported happiness as much as the other three factors. However, in factors like Trust and Freedom, the average happiness score in African countries are much lower than the average according to the comparison of the two regression lines. Their regression slope is also showing a uniform happiness index, where an increase in a factor (ex: GDP) on the x-axis does not have a significant increase in happiness. On the other hand, Latin American regions seem to have a regression line above the average slope, and have greater variability in the happiness score as the GDP or life expectancy increases.

Work Contribution

We all worked together to brainstorm visualization ideas and mockups, and decided with the help of our TA to pursue the ‘World Happiness Report’ based on the dataset availability and potential to compare unique data points (eg: GDP vs Happiness) and how they relate to one another. Several hours were spent on getting the color to show up based on the data values and interactive features (eg. happiness score and rank). This part of the project took the most time because the data dictionary had to be properly mapped.

Aileen Cai (10 hours)

- Added choropleth maps for each year
- Added interactive features to the map (ex: hover over each country to get happiness rank and score, select year)
- Added scatter plots for each year and each factor for happiness (gdp, freedom, life expectancy, trust and family size)
- Added styling format to visualization
- Written description of the project (part a,b,c and d)

Nate Schickler (10 hours)

- Fixed any bugs from processing the data in csv file
- Added more interactive features (mouseover) to the scatter plot to show country details (ex: gdp score, freedom score etc.)
- Added regression line and sub-regression line for specific regions
- Add more labels for the scatterplot (color coded regions).
- Written description of the project (part a and b)