INFO 3300 - Project 1 Writeup

Daniel Rusk
Cornell University
Ithaca, NY
dgr73@cornell.edu

Kevin Sun
Cornell University
Ithaca, NY
ks933@cornell.edu

Sarah Shin
Cornell University
Ithaca, NY
sjs457@cornell.edu

Valeria Marqués
Cornell University
Ithaca, NY
vmm49@cornell.edu

THE STORY

The overall story that we wanted to portray through this graph was the story of how overall happiness changed in Western Europe from 2019 to 2021, which also happens to be pre and post the COVID-19 pandemic. Our goal was to provide a comprehensive story of the shifts in happiness using our map visualization and our scatterplot. The bar chart that we also included was to give specific insight into how a specific variable generosity has shifted within this period of time to provide a more comprehensive view of the story. We used the variable "generosity" which measured whether or not the country donated to a charity within the past month. The reason we chose the generosity variable was to delve deeper into whether or not countries, despite the overall impact of the pandemic, continued to donate to charity organizations. What we have noticed from the data is that all countries have shown a decrease in the amount of donations. This was an interesting trend to observe because we also have found that most of the countries overall have gotten happier after the covid pandemic, with the exception of Belgium and the Netherlands. However, there still needs to be further analysis done on the two variables to make a conclusion on the relationship between the two variables.

DATA

The data we used came from the 2019 and 2021 World Happiness Report, found in CSV form on Kaggle (2019, 2021), derived from a report published annually by the United Nations Sustainable Development Solutions Network which ranks a country's happiness according to its citizens and seeks to understand this happiness through variables such as a nation's health, economic stability, and freedom. We specifically analyzed two datasets: that from 2019 and 2021 to observe how happiness in western Europe has changed over the course of the Covid-19 pandemic.

The Sustainable Development Solutions Network updated their questionnaire and dataset in 2020, which caused a few major differences in the data from 2019 compared to 2021. Some of these changes were easy to reconcile. For example, the "Score" data point was changed to be called "Ladder Score" for clarity. We fixed this by simply renaming the column labels to match each other. However, other changes were more difficult to deal with.

In 2019, the World Happiness Report measured trust (government corruption) by asking "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" and looking at the percentage of respondents who said that most people can indeed be trusted. However, in 2021, this variable was altered by asking "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?" And averaging the two responses. We decided as a group that these two data points, while both interesting, were not equivalent and could not be used in our analysis.

Finally, the size of the data set was also a major issue. With so many countries, and some only appearing in one year and not the other, we found it difficult to extract meaningful data on such a large scale. To remedy this problem, we decided to narrow in on a small region of the world with complete data and perform a more in depth analysis that would be impossible on a global scale. Due to our familiarity with the region and their interesting performance in the data set, we chose to perform our analysis on the following 10 countries commonly considered to be "Western European": Austria, Belgium, Czechia, France, Germany, Ireland, Luxembourg, the Netherlands, Switzerland, and the United Kingdom.

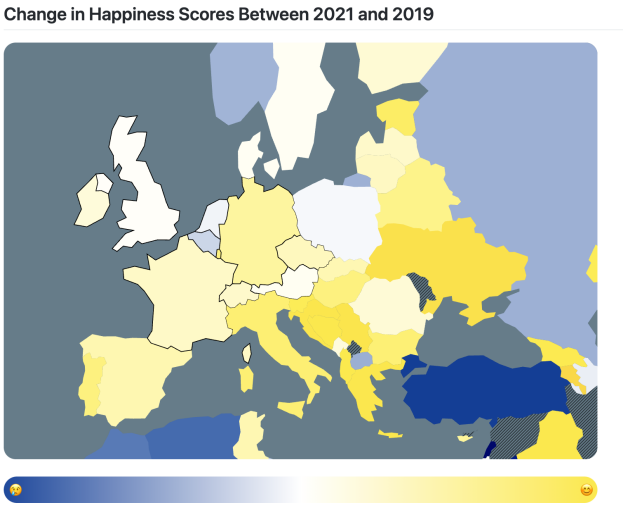
Within each record there was also an abundance of data from which we decided to take the most interesting or informative variables. The most important variable used was the "Score" or "Ladder Score." This variable asked respondents to imagine a ladder, with the top rung (10) being the best life they could possibly have and the bottom rung (0) being the worst. Then, where they are presently on the ladder. The reported score then is the average of a country's responses. We also wanted to look at the different variables associated with happiness and see how they had changed during Covid. Of the six associated variables, we decided that generosity would be an interesting metric to analyze. This variable asked "Have you donated money to a charity in the past month?" Scores were 1 if yes and -1 if not. The reported score is then averaged over a country's responses.

The flags used in the second graph were originally from countryflags.com, used with their open license. The map comes from the d3 library.

DESIGN RATIONALE

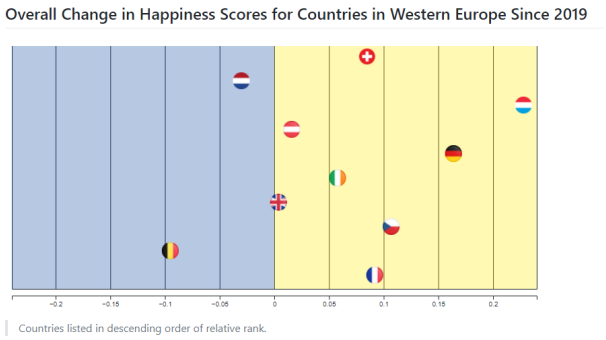
Based on our data, we designed our graphs around readability at a moment's notice. With a narrowed dataset, we could focus our audience's attention on the several key aspects we wished to highlight. To serve this purpose, we designed three different graphs: a map, a scatterplot, and a bar graph.

Map



For this graph, we decided to crop from a larger world map to only display greater Europe in an effort to focus the attention of the audience. Furthermore, because we noticed that it would be disorienting to display only ten select nations out of the whole, we included the complete map and made the design decision to include a thin dark border around western European countries in order to differentiate which countries acted as the primary focus in our dataset. As our primary color scheme, we used blue to depict sadness and yellow, happiness. We'd also considered red rather than blue for anger, but we decided to go with the latter as the lack of happiness was not necessarily a sign of discontent. We also considered non-emotion related color spaces, however it was much more readable when the colors had an emotional context. As there weren't quite enough nations to make granular categories, we opted for a continuous divergent color scale. To differentiate the lighter colors in the center of the gradient from "a lack of country", we gave the background a slightly blue grey, but one whose saturation could not be confused with that of a country. For the several nations whose data was not in the dataset, we implemented a striped texture to signify a lack of information. Finally, we found that emoji were more quickly understood as part of our legend than text, and lent itself well to the emotional connection we're trying to forge.

Flag Graph

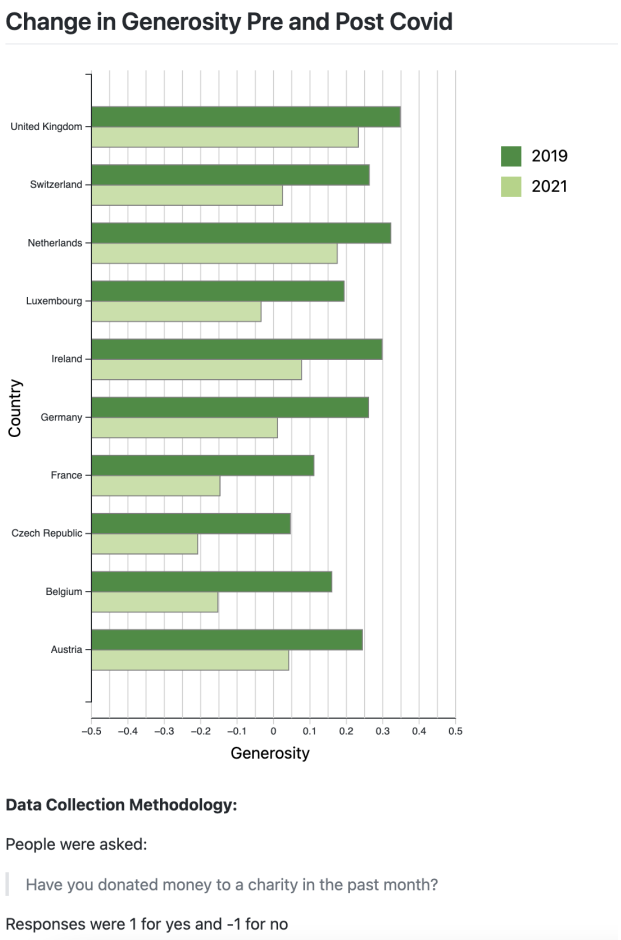


One design choice that we made for this graph was to use the flags of the respective countries for our marks. We felt that this was a better choice than using a legend and colors to make the graph look cleaner. Instead of having floating labels which could look messy, or a legend which would cause the user to go back and forth between checking which color is associated with which country, we decided that using flags was the best choice because we were drawing from an already existing association between countries and their flags.

Furthermore, another intentional design choice was making the background of the graph blue and yellow. This continued the expectations set up in the previous graph, and allowed the viewer to more immediately see which countries had improved (positive values) and which had worsened (negative value) over the pandemic.

Another data visualization choice that we made was to order the y position on the graph. We decided to distribute the points across the graph and have the countries be displayed in descending relative order of their overall 2021 happiness rank. There is no numerical value associated with this axis because only relative rank was used, and evenly spaced across the graph to prevent cluttering and overlap. We believe this better shows the viewer the overall story of not only how happiness had changed for each country, but where they stand relative to each other.

Bar Chart



One design choice we made for the bar chart was to use side-by-side bars for each year. In this way the viewer can easily compare the values from year to year. Another choice we made was to use horizontal bars. We did this because we wanted the viewer to easily read the values corresponding to each country. We decided the horizontal axis was easier for the eyes to track and detect differences between year to year. We used the colors of dark and light green intentionally. For one, the colors are in good contrast to one another so they are easy to distinguish. Secondly, the green represents money and we see over time as generosity generally decreases between 2019 and 2021. The color green also fades to lighter green showing that people were less generous. Lastly, we used green because it adhered to our above color theme of blue and yellow, and we wanted to make the overall three graphs cohesive.

Beneath the x axis we included a small blurb explaining how the data was measured. We did this to make sure the viewer understands the values each bar is representing. With no context, values such as -0.2 and 0.3 have no meaning and do not contribute to what we are trying to depict through the visualization.

CONTRIBUTIONS

Overall breakdown of work was decided by an initial meeting where we all shared our strengths and weaknesses as individuals. We found that Kevin and Sarah were more comfortable with front end/design, whereas Danny and Valeria had more of a technical/backend background. We took this into account by splitting up the work initially by having Valeria and Danny work on getting the csv files and importing the data, and having Kevin and Sarah work on finalizing the sketches and overall color scheme of the graphs. From then on, we came to a consensus on which subsets of data we would focus on, and which designs/sketches we would go forward with. Then, we split work depending on which graphs needed to be done. Throughout the entire time, we kept in communication with one another to update our progress. At the end, we all held a meeting to discuss and finalize design choices and come to a consensus on the presentation of all of our graphs.

Danny

I first started my work by locating the data sets and adding them to the github that Valeria set up. Once the data was in, I began working on my visualization for the variable "generosity". I sketched up a few ideas and then began coding. I went through several versions of bar charts: two single bar charts, two vertical bar charts, one vertical double bar chart, and finally the one horizontal double bar chart that we ended up using. I then played around with color schemes and polished off the graph before adding it to the group file. Finally, I wrote my design rationale for the graph. All in all, I spent approximately 8 hours on my whole part of the project.

Kevin

I worked with Sarah on the initial selection and design of our visualizations, sketching out various options for graphs to use, and how we could best tell the story we were attempting to convey. Afterwards, I prototyped a method for visualizing on a map and designed a data ingestion and filtering system that we'd revise into our final product. Finally, I refactored and integrated the three graphs into one final project, reworking colliding variables and tags, along with a final cohesive design between our individual elements. Overall, I spent approximately 8 hours on this project.

Sarah

Firstly, I contributed to making the design sketches of the graphs with Kevin. Together, we drew sketches and thought out the best ways to portray the story of the happiness dataset. Furthermore, I, along with Valeria, worked with creating the scatter plot graph with the maps. As this graph required more attention due to importing the images of flags as markers and overall more thought into the scaling of the points, we decided to work on the graph together. Initially, getting the data filtered took more time because initially I used the Promise package within d3. However, later I switched back to using d3.csv().then() for a more smooth filtering. Furthermore, the process of scaling the x axis required a lot of going back and forth and trying new ways of scaling (such as simply multiplying the difference by 1000, multiplying by the chartHeight, etc) before arriving at the current xScale we have currently. Lastly, getting the images of the flags to import and figuring out how to put images on the graph using .attr('xlink:href') required lots of looking through documentation. I also helped to organize the best times to meet as a group and setting timelines to complete our work and lastly, contributed to writing specifically the story and design rationale for two of the graph parts of the write up. I approximately spent 8 hours on this project.

Valeria

I set up the github repository and made sure all members had access and could edit the project. With Kevin, I imported the data and set up the basic "skeleton" structure of the project. I took on a more flexible role on the team, helping out, debugging, and workshopping the visualization with other team members. I made sure to document where our sources came from and analyzed what each data point represented and how it was calculated. I also fixed up discrepancies in the data set, found appropriate and legal images for the flags and cleaned up the overall code and file structure to make sure our project looked polished and professional. As a group, we went over final touches and color scheme changes. About 8 hours in total were spent on the project.