

World Happiness Report Visualization

Process book

Authors:
Matthieu Burguburu
Marta Knezevic
Vinko Sabolcec

Introduction

Beginning of the idea

We began the project by choosing the data we want to visualize. After brainstorming, exploring and proposing various ideas, we decided to work with data from the “World Happiness Report”, publication that measures the subjective well-being of individuals across countries and regions.

When we chose the dataset, we asked ourselves: “What is the story we want to tell?”. We started exploring existing works and we noticed that most of the works focus on visualizing the data for only one year. To come up with a novel idea, we decided to visualize the data over a longer period of time to show the changes and evolvement of happiness all over the world. Moreover, we came up with another novel idea; to explore the differences between happiness scores of neighbouring countries and to see which neighbours differ the most.

The dataset we found on Kaggle website contained data for World Happiness Reports from 2015 to 2022. Due to variety of variables measured across the years, we decided to select only a part of the data i.e. only variable measuring happiness score across the years. For visualizing happiness of neighbouring countries, we found a countries borders dataset which contains information of all neighbouring countries across the world. To enhance the story by showing which factors include the happiness and if there is a secret to being happy, we decided to use the full data from World happiness report 2023 which contains happiness score and the following additional measures: GDP per capita, healthy life expectancy, social support, freedom to make life choices, generosity as well as corruption.

One of the most important decisions our team made during brainstorming was to create interactive visualizations which can be used for exploring the data and finding out the key elements to happiness. Therefore, the story which follows the visualizations encourages users to explore the visualizations and provides interesting facts for guidance of exploration.

Data analysis

The datasets we found required cleaning and preprocessing. Since the dataset from World Happiness Report doesn't include data for all countries, we decided visualizing only the ones that exist in World Happiness Report data. Some of the countries changed their name during the period from 2015 to 2022 which required renaming the countries so that data is consistent over the years. Moreover, since some of the countries weren't officially recognized, we had to manually insert the country and its neighbours to the country borders dataset



Choosing the visualizations and technologies

After choosing the datasets and doing the preprocessing of the data, we brainstormed about what kind of visualizations we want to make. Taking into account the story we wanted to tell, we decided to create following five visualizations with their purposes:

Leaderboard

- Interactive visualization that shows top 10 happiest countries, their scores and positions on leaderboard that change over the time
- Purpose of the visualizations is to show the happiest elite of the countries, to see how they compete on the leaderboard and which countries are always in the top 10

World Map

- Interactive visualization of the world map that shows how happiness changes over the time across whole the world
- Purpose of the visualization is to show how different parts of world differ in happiness score, to show regions that are happiest and least happy and how happiness of each country changed during the selected period of time

Correlations

- Interactive visualization of 2D graph with happiness of Y axis and another measure from 2023 data on the X axis
- Purpose of the visualization is to see if there is a correlation between some measure and happiness and to see how it affects the happiness score

Network graph

- Interactive visualization of a graph with countries as nodes and links between neighbouring countries
- Purpose of the visualization is to see if some neighbouring countries form a cluster with similar scores of happiness or another measure (e.g. corruption)

Parallel coordinates

- Interactive visualization of all the measures from 2023 dataset put on different axes
- Purpose of this visualization is to select certain range of variable and to see if there are trends and patterns across the measures from World Happiness Report

To make the website whole, we decided on the extra idea of putting interesting facts about Finland, the world's happiest country which serves as an ending to our story and provides further information worth of exploring to find out the secret to being happy.

Before starting implementing the visualizations, we chose technologies we are going to use in our project. The technologies are listed below:

Python

For exploratory data analysis, as well as transforming the initial dataset into a more practical form for the website.

HTML/CSS/Javascript

For the structure, layout and interactivity of our website.

Vite

For development, as it provides practical build tools (like Hot Module Replacement) and a dev server. Vite is also used as a build system for building the production-ready website in the end.

React

For the foundation of our website. It provides ways to structure the code into smaller components that can be combined. Each member worked on their components and using React made it possible to combine all of them in the end with ease. Moreover, some of our members were already familiar with React.

D3.js

For transforming our data into beautiful and interactive visualizations. D3.js was new to all of us and introduced a learning curve but proved to be a powerful tool to bring our ideas to life. We used readily available extensions such as d3-geo and d3-color to increase our productivity.

Implementation

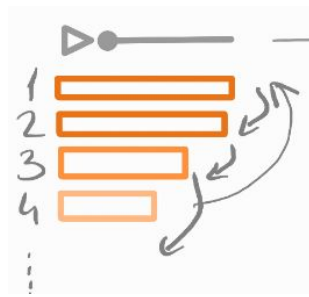
After everything had been prepared for starting the technical part of the project, we began working on the website. The process was divided into three main parts. First, we implemented each visualization in separate React components. This way we ended up with modular parts that can easily be maintained and moved across the website.

While working on visualizations, we discussed ideas for website layout and colors. The initial design for the whole website was made using Figma. Our color choice was the inferno color scale from D3.js library. Due to a lot of countries with similar measure scores, it was important to choose a color scheme that can depict small changes of measures when projected on a color scale.

Here we describe journey of creating each visualization.

Leaderboard

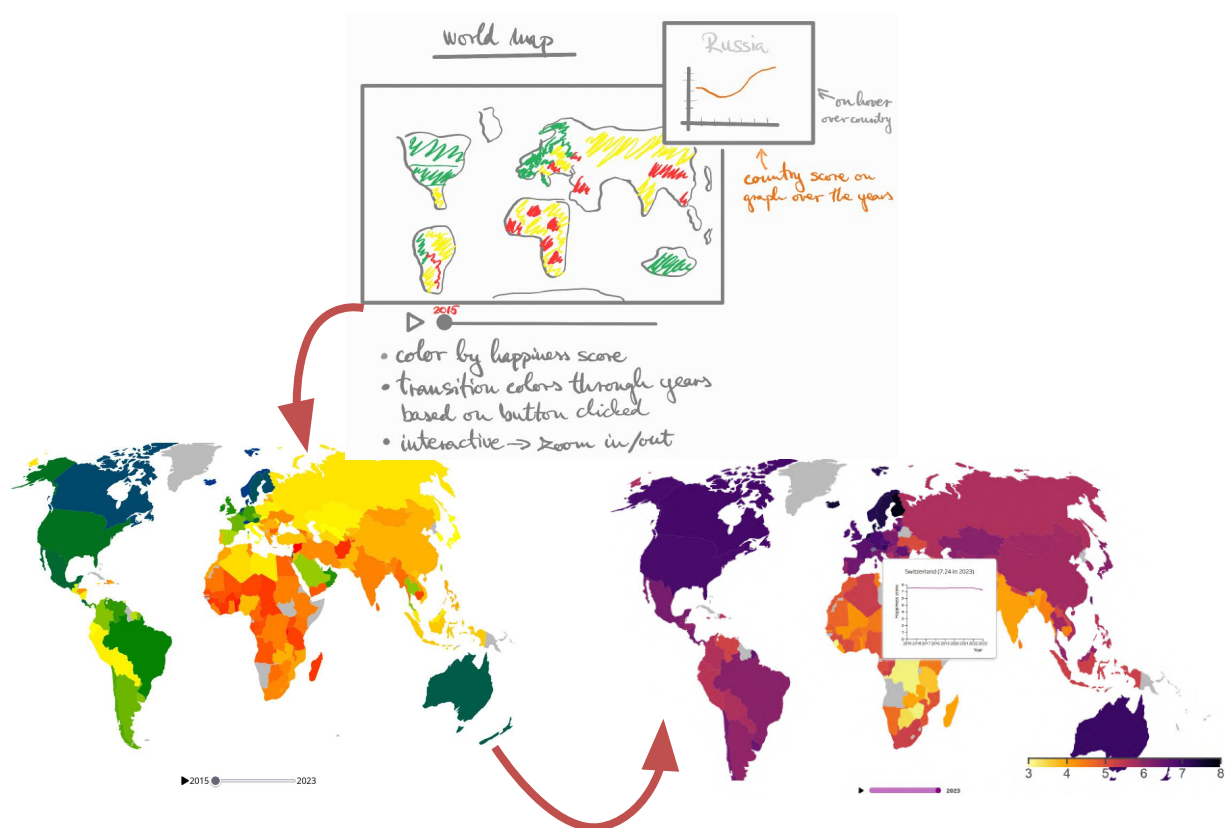
Our first visualization is the leaderboard of top 10 countries. We decided to do a leaderboard that shows not only position of the country, but also its happiness score. The changes over the years are captured by position changes of rectangles with country name, as well as color of the rectangle, that shows the relative difference of the happiness score. Therefore, the wider range of color scheme in last year denotes a bigger difference between the first and last country on the top 10 leaderboard. Interactiveness of the visualization comes with the slider that can be played or manually changed to specific year. Our final leaderboard visualization contains all the requirements specified in the beginning.



Initial sketch (left) and final implementation with the color scheme (right)

World map

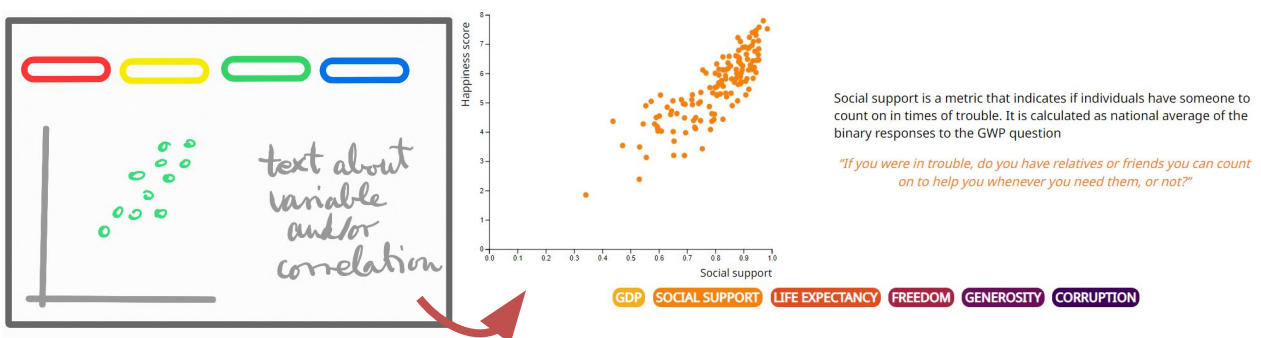
World map was one of the most complex visualizations. We visualize the happiness score on a choropleth map and add a time dimension to it which is changed with an interactive slider. Similar to the leaderboard visualization, the visualization can be played automatically or manually for each year. By hovering over a country, the graph of happiness score over the years for the country is shown. The choice of inferno color scheme was highly influenced by the map visualization since gradient of one or two colors could not depict the small changes of happiness scores across the years. Inferno scale on the other hand captures the differences and shows how certain regions change over time. We experimented with ranges of Y axis on the graph that shows on hover. Although range from minimum to maximum score per country better shows small differences of countries' happiness, we decided to fix the range so that comparison between different countries is possible and more intuitive since all countries have the same range. Countries for which happiness score is not measure were colored gray on the map.



Initial sketch (up), early implementation (left) and final implementation with added hover information, legend, year scroller design and the color scheme (right)

Correlations

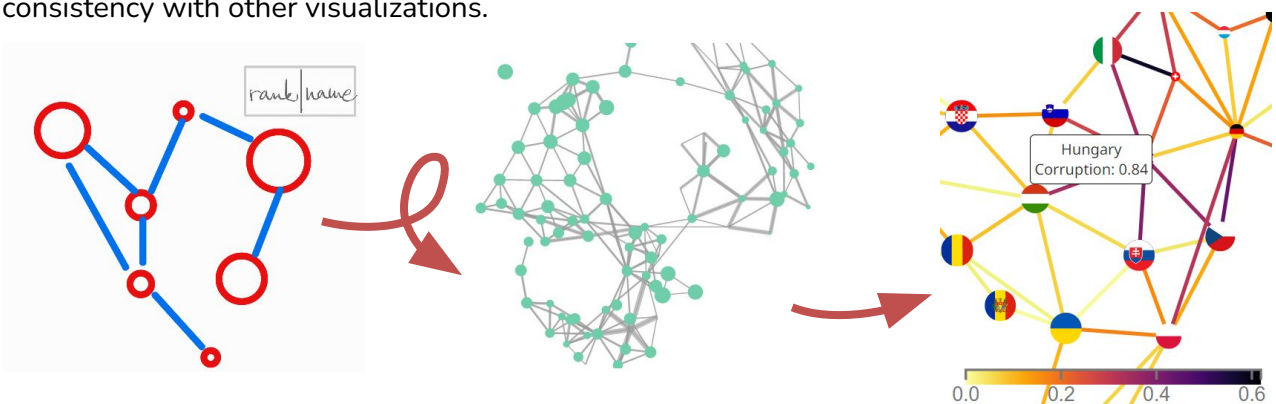
Correlation graphs are a simple visualization of happiness score in relation to other variables. They provide interesting information about correlation between the variables as well as a story about how the measures were computed. Although we proposed adding more dimensions to the graph as an extra idea, we decided not to include it since the idea gives information similar to the information in Parallel coordinates visualization. To enhance our initial idea, we added animation of circles on the graph when the variable is changed and graph is rerendered. The circles are drawn starting from the least happy to the most happy country denoting the pattern of the correlation. Moreover, we added a hover effect on circles which shows name of the country the circle represents. We decided to set the range of X axis from zero to maximum score of the variable. The exception is the graph with generosity scores, since the variable takes both negative and positive values. Lastly, we polished the visualization by adding colors.



Sketch of the correlation plot with buttons for selecting variables (left) and final visualization together with the text information about variable (right)

Network graph

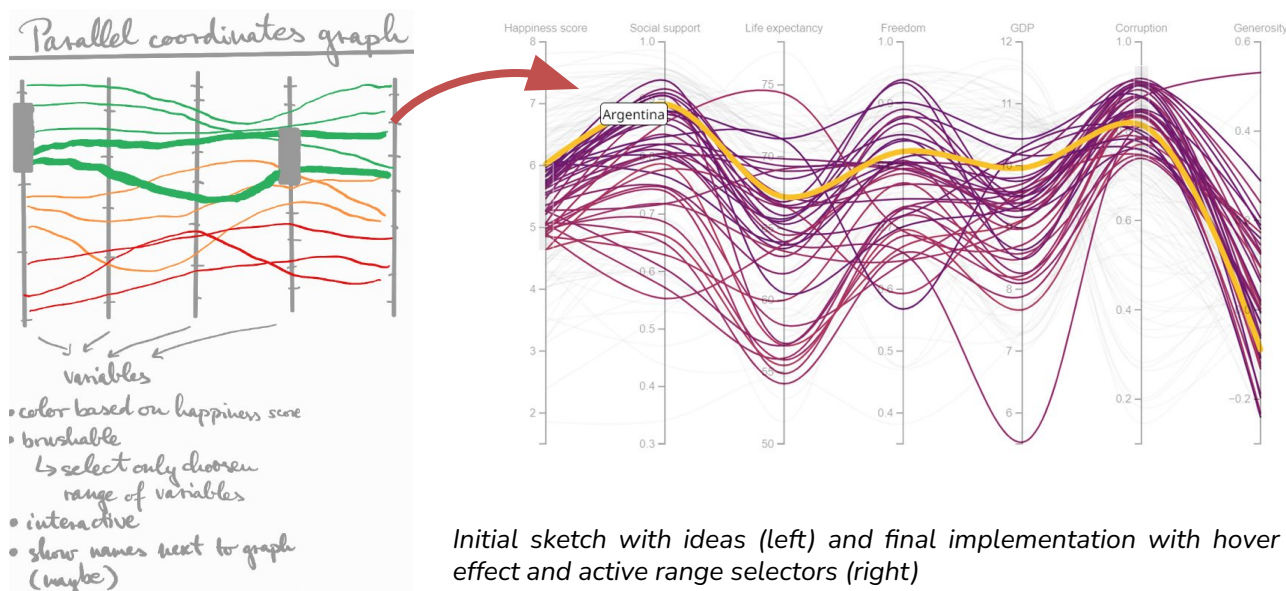
Network graph is by far the most complex visualization we worked on. The initial idea was to create a graph with nodes representing the countries and links between the neighbouring countries. Size of the node on the graph is proportional with the score of the selected variable. As implementation progressed, we decided to add flags of countries on the nodes, so that it's easier to find the target country. Moreover we added additional info into the links of the graph. The color of the links represents the difference between the measures of the neighbours. We discarded the idea of having different weights of the links since it hindered with the visibility and clarity of the graph. The legend for the difference is provided in the corner of the graph. The graph is placed after the visualization of correlations which provides meaning of the scores and their ranges. Therefore, we did not include that information on the graph. The nodes can be moved and dragged across the graph to see neighbouring connections clearly. We discarded the initial idea of having a dropdown menu for selecting the variable and instead put buttons so that we have consistency with other visualizations.



Initial sketch (left), early implementation (middle) and final implementation (right)

Parallel coordinates

Parallel coordinates visualization provides an insight into patterns that emerge across variables by brushing over the graphs axes. Lines of countries that fall into the selected range are colored and hover effect is enabled on them. To remove the selected range of certain variable, user has to click on the axis of that variable. We discarded the extra idea proposed in Milestone 2 of writing all countries that are selected because the list could be potentially long and could shift the layout of the website. Instead, we opted for writing the country name on hover over the line. As in initial idea, the lines on the graph are colored based on happiness score of the country they represent. We discarded the idea of changing the colors based on selected variable because the resulting effects added more noise and it was more difficult to understand the meaning of the visualization.



The second part of project implementation was creating the website with a scrollable layout and all the visualizations included. This part was relatively easy and resulted with all the visualizations incorporated into the website. We put emphasis on consistencies of both colors, elements and layout of the visualizations. Moreover, the consistency of the colors is applied to the process book as well.

Finally, in the last part, we wrote the story that follows the visualizations, gives key insights from each visualization and navigates, encourages and explains to user how to explore the visualizations. As part of storytelling, we provide interesting facts about Finland, the happiest country in the world.

Challenges

Thanks to having experienced members in web development in the team and some good planning, we did not run into major problems that put the project in jeopardy. However, we did run into some technical difficulties during the project. Here is the list of them:

Network graph

The network graph is undoubtedly the visualization that we had the most trouble with. One problem we faced was deciding on the final design even after having the initial idea. As we were new to the D3.js library, it was a lot of work trying to combine force simulations, panning and

zooming and interactivity correctly. It was a process of trial and error which we successfully solved by improving our knowledge of D3.js. Another problem we faced was non-deterministic freezing of the force simulation in the development mode for which we found no solution. However, the freezing didn't occur in the production version.

World map

The world map visualisation exhibited performance issues when zooming and panning using the GeoJSON data. As a result, we have switched to the TopoJSON which improved the performance and also reduced the file size of the data.

Browser support

Another common issue we faced was browser support. Some browsers behaved differently than others on some aspects, for example the size of some visualizations. Some components that we initially used for visualizations were also not available on all browsers, like the flag emojis used in network graph. To solve this issue, we took SVG images of country flags which can be rendered in all browsers. The network graph visualization freezes were also in that category. In the end, we tried to remove as much of these poorly supported elements as possible to make the website experience nearly identical across different browsers.

Peer assessment

This section contains the breakdown of work done by each member. We split the visualisations among ourselves and each visualization has its main author. However, in case some of us encountered a problem, we helped each other and solved it together.

Matthieu Burguburu 🧡

Worked on preprocessing the Country borders dataset. Created first versions of milestone reports. Worked on Network graph visualization. Designed the first mockup of the website layout using Figma. Worked on the process book.

Marta Knezevic 😊

Did research of similar visualizations. Drew the initial sketches for the visualizations by hand. Worked on the Leaderboard and the Correlations visualizations. Created graph that shows on hover over World map. Wrote the data stories for visualizations. Provided facts about Finland. Recorded the screencast. Worked on the process book.

Vinko Sabolcec 😄

Worked on data exploration and setting up the React website skeleton. Worked on the World Map and the Parallel Coordinates visualizations. Helped debugging other visualizations. Iteratively performed code cleanup and organization of code files. Implemented website layout. Hosted the website. Wrote README.md. Created design and layout for the process book.



[Github repository](#)



[Website](#)

We hope you will enjoy exploring the visualizations and find the secret ingredient for happiness.