Country Recommender

Visual Analytics 2021/22

Concept & Target users

A tool to simplify **data-driven** comparisons between countries whether it be to gain insights in a casual manner or to make better decisions based on existing data.

The tool is geared towards any individual or organization that wants to choose between countries on single or multiple criteria. This can include choosing:

- A travel destination;
- Where to start a particular business;
- Targets for international aid or promotion of social causes;
- Targets for products or advertisements;

Goals

With the onset of the Web, the amount and access to data concerning countries has increased to a point where comparing this data especially from multiple sources is restricted to company or academic efforts.

With this in mind, we designed our tool guided the following goals:

- Combining multiple datasets must be a simple process even if it comes at a cost of precision loss;
- It must be somewhat agnostic of the user's goals so as to extend its usefulness to multiple scenarios;
- Insights must be gathered without much latency, since the user is incentivised to experiment with multiple filters and parameters, thus constantly changing the underlying data;
- Different visualizations must be used to represent the data in multiple ways and, as such, under different "perspectives".

Related Works and Tools - Our World in Data

Our World in Data offers an extensive repertoire of **data visualizations and articles** regarding research and data into countries. Their goal is to expose data that would otherwise be under obscure datasets or on extensive academic papers, so it can be used to understand our world better.

Some important distinctions between their work and ours are:

- Their visualizations are **less interactive**, usually only allowing the user to set simple filters like date ranges, or countries to consider;
- Visualizations tend to have **low dimensionality**, pertaining only to a few attributes, all usually under the umbrella of a single domain;
- The focus is more on **conveying and explaining gathered insights** through articles with the support of carefully crafted data visualizations to aid understanding;

Related Works and Tools - Eurostat

Eurostat is the statistical office of the European Union that also releases reports containing useful **data visualizations and insights** taken from multiple researches and datasets, through the DataEuropa platform.

Compared to our work, some differences include:

- Although many datasets include data from countries outside the **European zone**, there is a particular **focus on** this area;
- Much like Our World in Data, the focus tends to be on a **few of dimensions within a given field**;

Datasets

Our World In Data provides a wide range of datasets on all countries around the world. Possessing over 3000 datasets on over 300 different topics, it's able to fulfill most of the dataset needs of the project.

```
[
  (...)
    {
        "Entity": "Australia",
        "Code": "AUS",
        "1st (Incomes across the Distribution Database (2016))": "9283.941406"
    }
  (...)
]
```

Datasets

Our World in Data, however, was not able to fulfill all the needs:

- **Language** - another dataset was retrieved in order to obtain data on communication. Namely, a dataset of English speakers in each country
- **LGBTQI+ Rights** - while Our World In Data did provide some data on LGBTQI+ Rights, those only included the USA. Another dataset was found on sexual orientation and gender identity laws of all countries.


```
[
  {
    Country: 'United States',
    'Eligible population': '296603003',
    'Total English speakers': '283160411',
    'Total English speakers (%)': '95.46',
    'As first language': '234171556',
    'As first language (%)': '79',
    'As an additional language[note 1]': '48988855',
    'As an additional language (%)': '16.5',
    Comments: '(...)',
    Q3_iso_alpha: 'USA',
  }
]
```

```
[
    {
        "country": "Zimbabwe",
        "HAPPINESS": "3.692",
        "Per capita GDP": "$2,147.00",
        "% with tertiary educ.": "0.38",
        "GRI": "2.8",
        "Religiosity": "88%",
        "SHI": "2.1",
        "legality of sam-sex sexual-activity": "0.25",
        "marriage/ civil unions": "0",
        "same-sex coupleadoptions": "0",
        "serving in the military": "0",
        "antidiscrimination laws": "0",
        "gender ID markers": "0",
        "Democracy Status": "A",
        "electoral process and pluralism": "0.5",
        "political culture": "2.5",
        "civil liberties": "3.24",
        "Satisfaction with GDP per capita": "1.151",
        "Social support": "1.479",
        "Healthy life expectancy": "0.599",
        "Freedom to make life choices": "0.399",
        "Generosity": "0.065",
        "Perceptions of corruption": "0.025"
    }
]
```

Categories

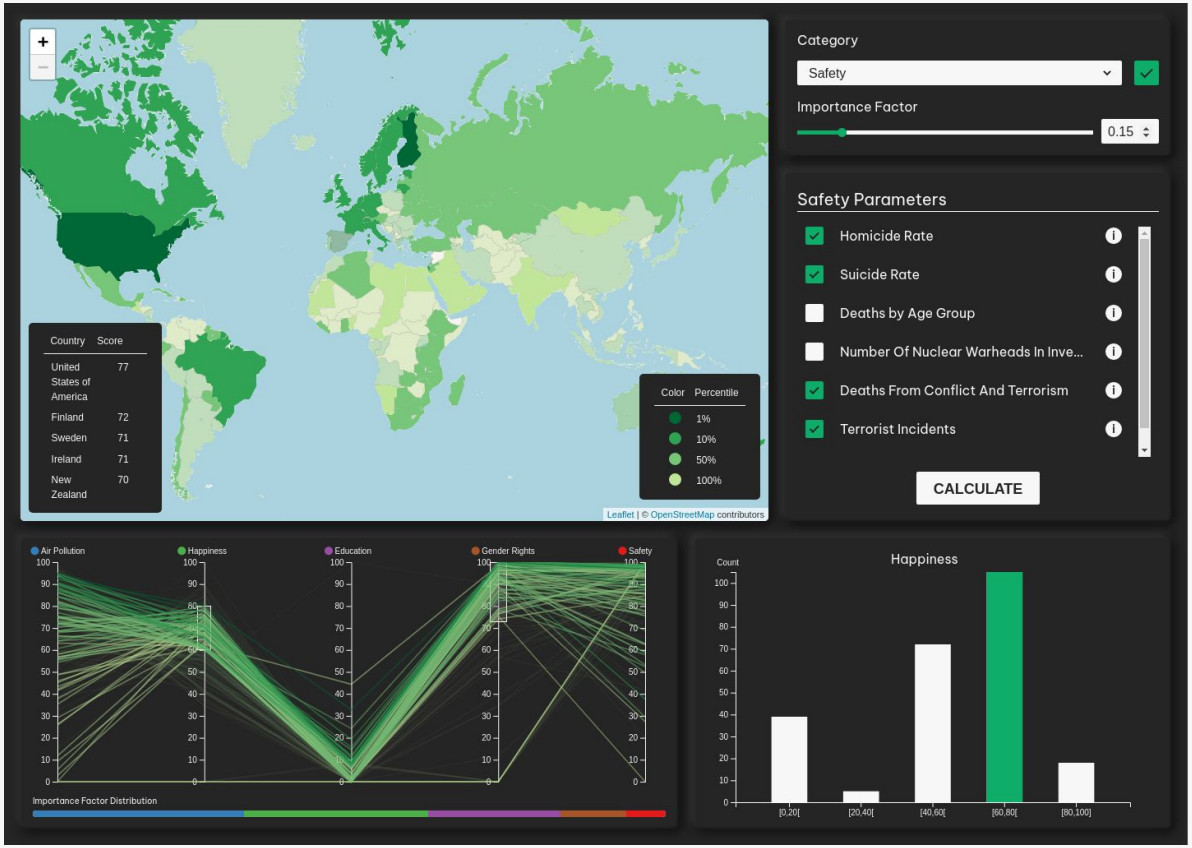
- Air Pollution
- Natural Disaster
- Climate Change
- Food
- Income
- Education
- Political
- Gender Rights
- Language
- Safety
- Happiness
- LGTBQI+ Rights
- Health

Analytics

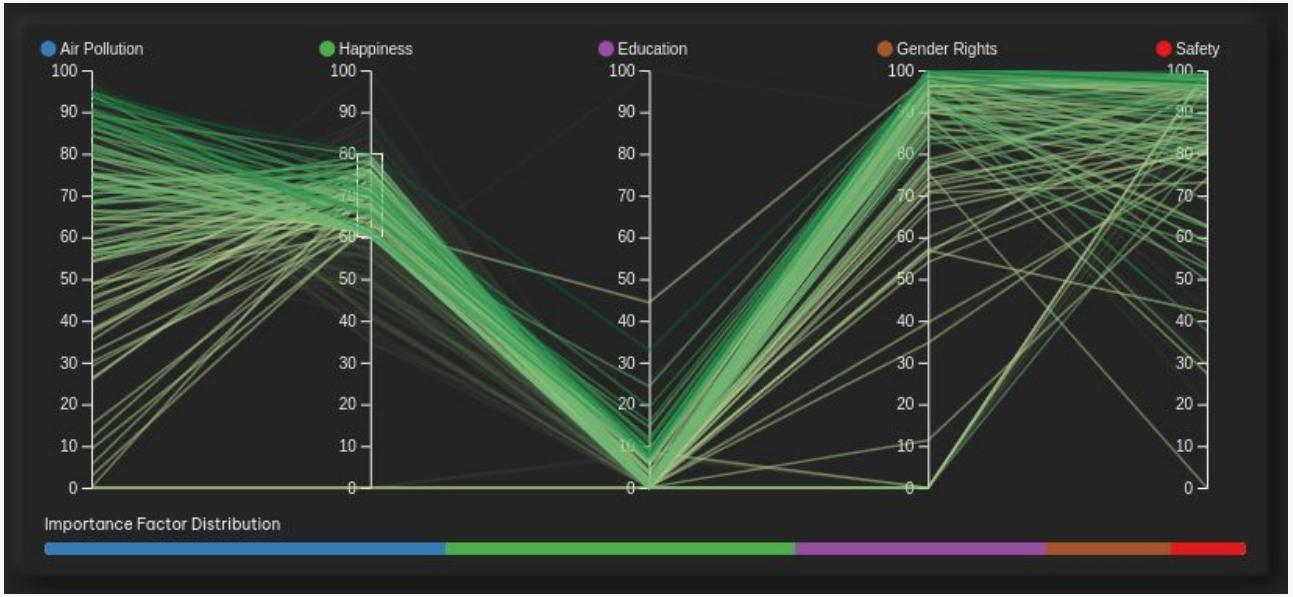
Each category had several datasets. The goal was always to produce a single score out of each category, so that it could easily interpretable by users.

In order to make this happen, the variance matrix generated by the PCA algorithm was used. Since the variance details the “importance” of each field, this could be used to generate a single value, even if this come with a loss.

The Tool

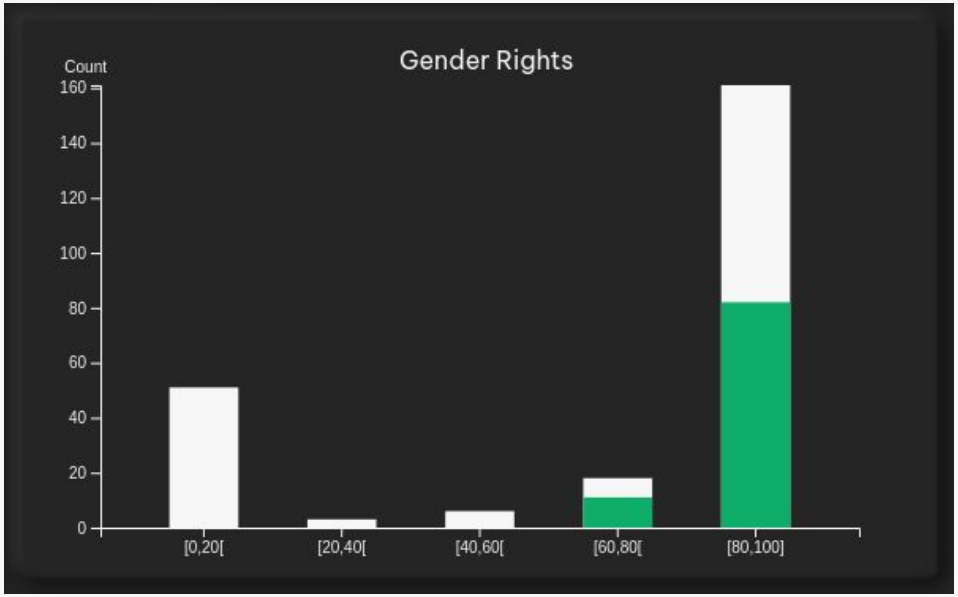

Country Score
United States of America 77
Finland 72
Sweden 71
Ireland 71
New Zealand 70
Color Percentile
1%
10%
50%
100%
Leaflet | © OpenStreetMap contributors
Category
Safety
Importance Factor
0.15
Safety Parameters
Homicide Rate
Suicide Rate
Deaths by Age Group
Number Of Nuclear Warheads In Inve...
Deaths From Conflict And Terrorism
Terrorist Incidents
CALCULATE
Air Pollution
Happiness
Education
Gender Rights
Safety
Importance Factor Distribution
Happiness
Count
[0,20[
[20,40[
[40,60[
[60,80[
[80,100]

Visualizations - Parallel Coordinates



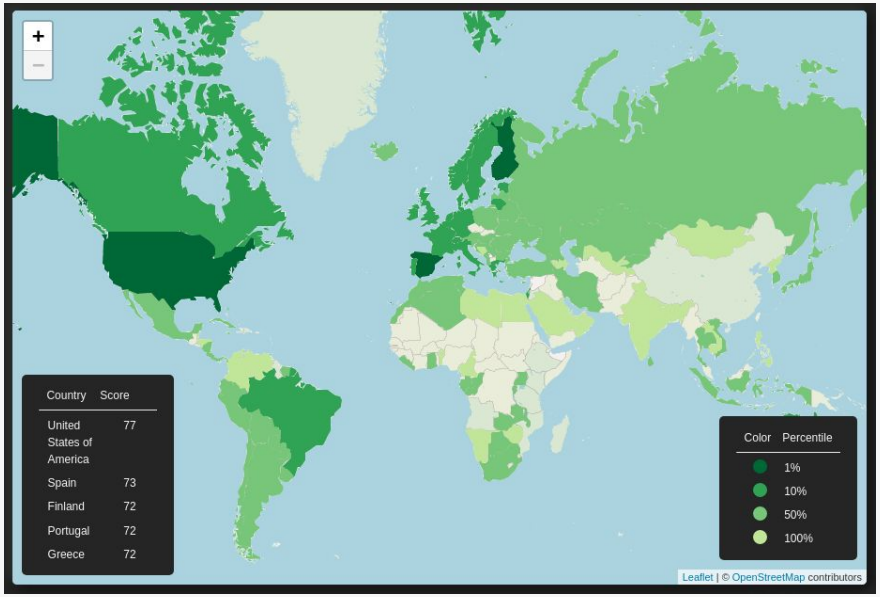
- Each axis is a category with a color matching the IF distribution in the bottom (ordered) taken from a qualitative set of colors from colorbrewer2.org;
- Clicking on a category's name updates the Histogram panel;
- Each line is colored based on the corresponding country's percentile or a bright pink if it is the currently hovered country;
- Supports brushing to select ranges of scores. Lines outside the range have their opacity greatly reduced;

Visualizations - Histogram



- Divides the scores into ranges of 20 and shows the number of countries with a score in a given range for that category;
- Clicking on a given bar filtered the countries to only those that have the category's score within that range;
- If a filter exists due to the Parallel Coordinates, if the range of values doesn't encompass the whole range, the proportion of number of countries that are fall into the filter's range within the given Histogram's range is represented through a green bar with a height matching that ratio. The rest is colored white;

Visualizations - World Map



- Similar to the PC, each country is colored based on the percentile of their overall score (matching the respective bottom right legend) taken from a sequential set of colors from colorbrewer2.org;
- If a country isn't within an existing filter, its opacity is greatly reduced;
- The legend at the bottom left breaks down the Top 5 countries and respective scores of the currently selected ones. If a country is hovered, the legend updates to show its score breakdown instead;
- Supports zooming and panning;

Insights

An example of interesting insights drawn from exploring our application are:

- A high quality democracy does not necessarily result in a high quality education system.
- Safer countries are usually the happiest, but higher income countries are not necessarily so
- Western Europe provides a much higher quality of life, generally speaking

Further work & Conclusion

Although we are pleased with our work, the tool could be further augmented with certain features like:

- Offer other dimensionality reduction algorithms like **t-SNE** or **MDS**;
- A **dropdown** for selecting the Histogram's category, besides clicking on the Parallel Coordinates axis labels;
- Support the selection of two categories and replace the Histogram with a **Scatter Plot** showing the correlation between the two categories while also supporting brushing for setting ranges for those categories;
- Adding a filter for **passport strength** to better suit users looking to use our tool for travelling.

In conclusion, following the established guidelines and our initial goals, we believe we **developed the tool** we envisioned and it can now be used to **fulfill the use cases** we set out to help.

André Filipe Magalhães Rocha - **1991375**

Hanna Eide Solstad - **1991023**

Tito Alexandre Trindade Griné - **1992095**