1. Title, Group and Motivation

Disaster Risk

Steven Lin
Mason Wang

The chosen topic is on world disaster risk. The targeted audience of the dashboard is for policy makers, activists, and people interested in the subject. By using a dashboard, we can highlight details in an interactive and customizable manner, allowing users to focus on what they need or are otherwise interested in.

2. Datasets

https://data.humdata.org/dataset/worldriskindex
The world risk dataset covers the majority of the data used for the project. This dataset covers the summary statistics risk, exposure, and vulnerability, disasters that make up the exposure index, and vulnerability subcategories such as socio-economic development, susceptibilities due to past events, healthcare capabilities, investment capabilities as well as further subcategories such as population exposed to differing disaster strengths, life expectancy, gross income per capita, refugees and asylum seekers, rule of law and so on. In total there are over 150 features not accounting for the normalized columns.

Before any other observation, the dataset first has its normalized columns stripped, the top 5 entries are explored and the missing entries counted.
To explore the dataset, a single country is selected at a time from a list of unique country codes, then the relationship between the variables are explored through correlation analysis as well as line plots of variables over the years. Multi feature line plots are used a lot to explore subcategories as well as how that translate to the aggregate indices. Some oddities noted throughout exploration is the presence of columns with the same name, further exploration shows that these columns are the same but in different units.

As with cleaning, besides pairs of normalized and percentage columns, there is not much error in the dataset. This makes sense as this dataset is used extensively for world risk index' own report which greatly surpasses this project in scope.

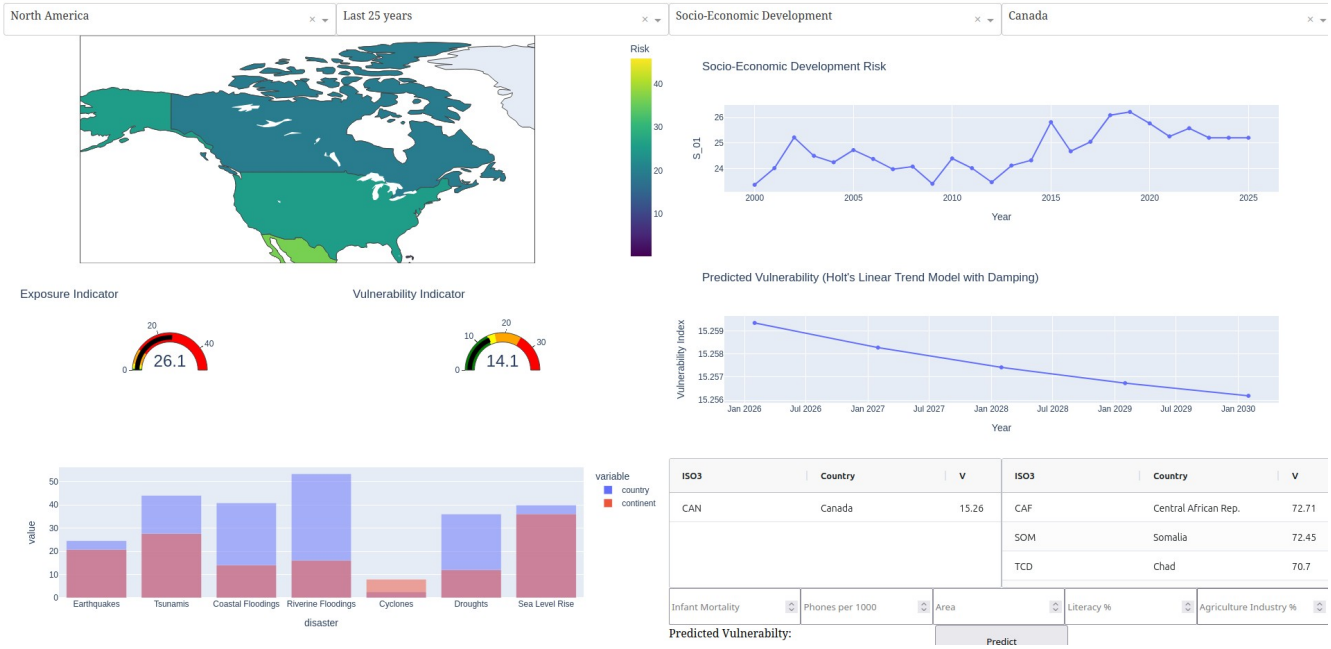
https://www.kaggle.com/datasets/fernandol/countries-of-the-world
Countries of the world is an aggregation of country statistics over several decades. It contains additional factors of a country outside of immediate disaster and disaster related statistics such as continent, area, coastline, climate, arable land, and industry split. These statistics provide an additional comparison points that may be useful. Some of the features overlap with the vulnerabilities subcategories.

A quick initial overview of the data shows many missing entries and that many features contain artifacts which forces them into the wrong data type. A uniqueness cheque was done on the country names, a consistency check was done on percentage land type (crops, arable, other) to make sure values added up to 100 and for percentage industry (agriculture, industry, and services) to add up to 1, range checks were done on other variables such as population, literacy percent, and phones per 1000 as well.

Country and regions names had leading and trailing whitespace stripped, commas were removed from numerical entries and type casted into floats. Regional mean imputation was done to fill missing entries in net migration, infant mortality, gdp per capita, literacy %, phones per 1000, land arable %, land crop %, land other %, climate, birthrate, deathrate, agriculture, industry, and service. The agriculture, industry, and service split of Monaco specifically was over 1.01 in total, and was scaled down to match 1. ISO3 codes were added by fuzzying searching the name using pycountry. Missing codes were manually added added to data. Duplicate codes from fuzzy searching was manually corrected.

Following the cleaning the dataset was visualized using a correlation map, highly correlated variables were plotted using scatter plots, and box-plots of select variables were plotted by aggregating on region.

3. Dashboard layout and overview



The design of the dashboard did not shift much from the initial proposal. The resulting design is split into 4 main quadrants each intended to answer a question: is there a risk, what is the risk, why is there risk, and how to fix this risk? The idea is to present a top-down, left-to-right chain of answers with the key motivating points presented first.

For interactivity, the choropleth map allows user to move the map, zoom in, and change continents with drop-down selector, the line graph is able to zoom in, and the tables are able to be sorted by pressing on a statistic.
For data overview, we have a choropleth map for displaying aggregated risk and country statistics over countries of the world, indicators for comparing aggregated exposure and vulnerability against continent, the vulnerability subcategories line graph displays trends over the years.
For user guidance, we used titles, colour choices, and legends throughout the design.
For decision making features, we have the choropleth map which shows the risk of the country and the indicators which compares the score of the country against the inter-quartile of countries in the continent, all of which is colour coded to highlight higher levels.

4. Panel Explanations

5. Resources

https://dash.plotly.com/
https://dash.plotly.com/tutorial
https://dash.plotly.com/layout
https://dash.plotly.com/basic-callbacks
https://dash.plotly.com/interactive-graphing

https://weltrisikobericht.de/worldriskreport/
https://www.cia.gov/the-world-factbook/about/archives/2021/references/guide-to-country-comparisons/

https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards-design-tips
https://help.tableau.com/current/pro/desktop/en-us/dashboards_best_practices.htm

6. Conclusions and Future Work

Throughout this project, we have explored how to combine information from multiple datasets, how to design dashboards, how to use and apply python panel technologies to build a working dashboard for the project. Outside of the technical matters, we were pushed into developing project management, team cooperation, and presentation skills over the course of the project. In the conclusion of the all this effort, is the complete disaster risk dashboard that included as much useful information that is needed to formulate a decision.

The world risk index dataset is a truly huge dataset to work with. From the overview of the dataset, we can identify 3 layers of aggregation. In this project, we have explored the top 2 layers: the risk-exposure-vulnerability trio of indices, and their disasters (e.g. hurricanes, flooding, earthquakes) , susceptibility (e.g. socio-economics, vulnerable population from wars/diseases), coping ability (e.g. government and heath-care), and adaptability (e.g. reaserch and investment) sub-columns. These sub-columns further split down into individual features such as the hurricane aggregate in disasters would have hurricanes of various strengths as their individual feature. These features could allow for an additional layer to the analysis if time permits.

Some suggestions from colleagues include adding additional countries to the line graph for comparison, adding an indicator for the currently selected country for indicator and bar graph, adding a line to tell the user to click on a country to update. Some features that were appreciated was the time-series projection of vulnerability, the organization of the dashboard, or the use of gauge indicators to highlight decision making variables.

7. Activity table

Activity/Task	Description	Hours planned	Done by	Actual Hours
Dataset Selection	Exploring different datasets to identify interesting and usable datasets for project.	4	Together	6
Project Proposal	Writing a proposal for the project using the selected datasets	5	Together	4
Exploring Python	Reading articles	3	Mason	4

Panel	regarding panel and associated technologies. Exploring and learning panel via experimentation on notebook.			
Designing Dashboard	Reading articles regarding dashboards. Designing a dashboard layout. Incorporating article tips into layout design.	3	Steven	3
Exploratory Data Analysis	Exploring data relations using techniques taught in Lab 1.	4	Together	6
Data Cleaning and Imputation	Cleaning and verifying data using techniques taught in Lab 2. Then imputing missing values using techniques taught in Lab 3.	8	Steven	6
Displaying Data	Working the data into a presentable form through techniques taught in Lab 1 and python panel technologies. This section covers the indicators, bar graph, and tables that was otherwise deemed minor commitments.		Together	5
Regression/ Classification	Creating models used to predict selected values using techniques taught in Lab 3 and 4.	3	Together	5
Geospatial Analysis	Lab 8, creating a choropleth map to display risk and various country	5	Steven	6

	statistics.			
Time Series Analysis	Lab 6, creating time series line graphs projecting vulnerability and/or associated variables.	5	Mason	6
Preparing Presentation	Preparing a presentation on the project.	8	Together	8
Documentation	Write documentation for the code. Make sure notebook is up to standard.	4	Together	3
Report Writing	Writing a report regarding the project. Exact focus of the report is to be determined.	8	Together	
Total		60		