**Data Visualization Process Book**

By **Visual Tension**

Syed Saqib Ahmed (103488324)

Joshua Stone (7696663)

COS30045 Data Visualisation   
Monday 12 30 EN305   
Semester 1, 2023.

Mercury Link

word count

Table of Contents

[**1 Introduction** 3](#_heading=h.gjdgxs)

[1.1 Background and Motivation 3](#_heading=h.30j0zll)

[1.2 Visualisation Purpose 3](#_heading=h.1fob9te)

[1.3 Project Schedule 3](#_heading=h.3znysh7)

[**2 Data** 4](#_heading=h.2et92p0)

[2.1 Data Source 4](#_heading=h.tyjcwt)

[2.2 Data Processing 4](#_heading=h.3dy6vkm)

[**3 Requirements** 5](#_heading=h.1t3h5sf)

[3.1 Functional Requirements 5](#_heading=h.4d34og8)

[3.2 Non-Functional Requirements 5](#_heading=h.2s8eyo1)

[**4 Visualisation Design** 6](#_heading=h.17dp8vu)

[∙ Prototyping (Sketches) 6](#_heading=h.3rdcrjn)

[Design 1 6](#_heading=h.26in1rg)

[Design 2 6](#_heading=h.lnxbz9)

[Design 3 6](#_heading=h.35nkun2)

[Design 4 6](#_heading=h.1ksv4uv)

[∙ Iteration 6](#_heading=h.44sinio)

[Survey 6](#_heading=h.2jxsxqh)

[Feedback 6](#_heading=h.z337ya)

[Amendments 6](#_heading=h.3j2qqm3)

[∙ Use of design guidelines 6](#_heading=h.1y810tw)

[**5 Validation** 7](#_heading=h.4i7ojhp)

[**6 Conclusion** 8](#_heading=h.2xcytpi)

[**7 References** 9](#_heading=h.1ci93xb)

# **1 Introduction**

## 1.1 Background and Motivation

Both climate change and migration are key topics of interest and intense debate takes place frequently. This site aims to provide users with an overview of data relating to this issue from sourced from highly credible sources. (insert references here)

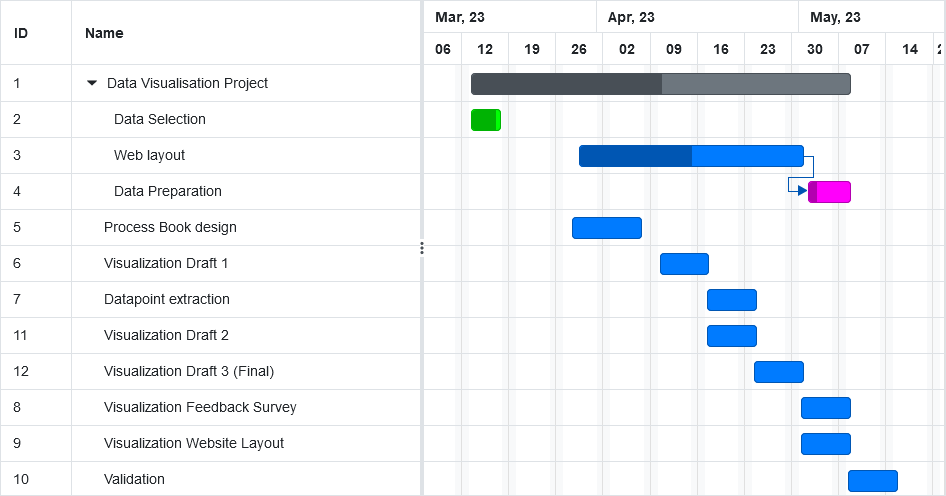
## 1.2 Visualisation Purpose

This page targets the wider population who has little to no idea about the specifics of the data involved when examining migration as it relates to climate change.

Students may use our page as a starting resource to understand or highlight key topics of interest and establish a general understanding of the relationships between variables examined.

Members of the wider population may use this as a second tier information source after having interest sparked by other content, for example news articles. It aims to expand their understanding of the relationships between migration patterns and climate change.

## 1.3 Project Schedule



# **2 Data**

## 2.1 Data Source (Josh)

From where and how are you collecting your data? Provide a link to your data sources. What type of data set is it (e.g., table, network, field)? What are the attributes in your data set and what type of data are the values (i.e., categorial, ordinal, interval, ratio/quantitative)? Is there any data in the set that will not be included in your visualisation? Why?

Data for this assignment has been collected from reputable public sources.

Data for this assignment has been collected from reputable public sources.

Migration Data has been gathered from the OECD publicly available statistics.

<https://stats.oecd.org/Index.aspx?DataSetCode=MIG>

Due to data collection issues around migration and reporting of migration statistics by governments not all countries have reported migration information for every year however this dataset offers a reputable gauge of overall migration information. It not only details migration by source country but also by destination allowing us to examine both inflows and outflows of migration. The period covered by this dataset is from 2000 to 2020 with countries coded using ISO3 labeling making it easy to link datasets together.

| CO2 (ISO3 Coding for country) | Categorical |
| --- | --- |
| Year | Interval |
| Country of birth/nationality | Categorical |
| Value | Ordinal |
| COU | Categorical |
| Country | Categorical |
| Variable | Categorical |
| GEN | Categorical |
| VAR | Categorical |

Population Data OECD

| Location (ISO3 code) | Categorical |
| --- | --- |
| Time | Interval |
| Indicator | Categorical |
| Value | Ordinal |
| Indicator | Categorical |
| Frequency | Categorical |
| Subject | Categorical |

Climate data has been sourced from the IMF from <https://climatedata.imf.org/pages/climatechange-data>. Both the recorded annual surface temperature change dataset and climate related disasters dataset have been combined with the OECD migration information to allow us to contrast migration data with clear metrics related to climate change.

| Country | Categorical |
| --- | --- |
| ISO3 | Categorical |
| Indicator | Categorical |
| Unit | Categorical |
| CTS\_Name | Categorical |
| F2000-F2021 (each of these columns has temperature change from previous year) | Ordinal |

Climate Related Disasters Frequency

| Country | Categorical |
| --- | --- |
| ISO3 | Categorical |
| ISO2 | Categorical |
| Indicator | Categorical |
| Unit | Categorical |
| CTS\_Name | Categorical |
| F2000-F2021 (each of these columns has disaster count for each year) | Ordinal |

## 2.2 Data Processing (Josh)

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented? Will you be deriving variables? Describe clean-up process that was implemented.

The data in this dataset is of high quality however there is some clean up still required. Null values are present in ordinal rows in all datasets, for disasters null values are assumed to be representative of 0 so these null values can be filled however in other cases like migration data a 0 value cannot be assumed as there is not reliable migration data collected for all countries in all years so null values cannot be assumed to be 0. These are instead handled on a case by case basis in javascript.

Some preprocessing has taken place using python and pandas so that the data more easily aligns and is easier to work with. The tables in question have been merged into a single consolidated dataset with the F2000-F2021 columns pivoted to align with the relative year labels for migration information.

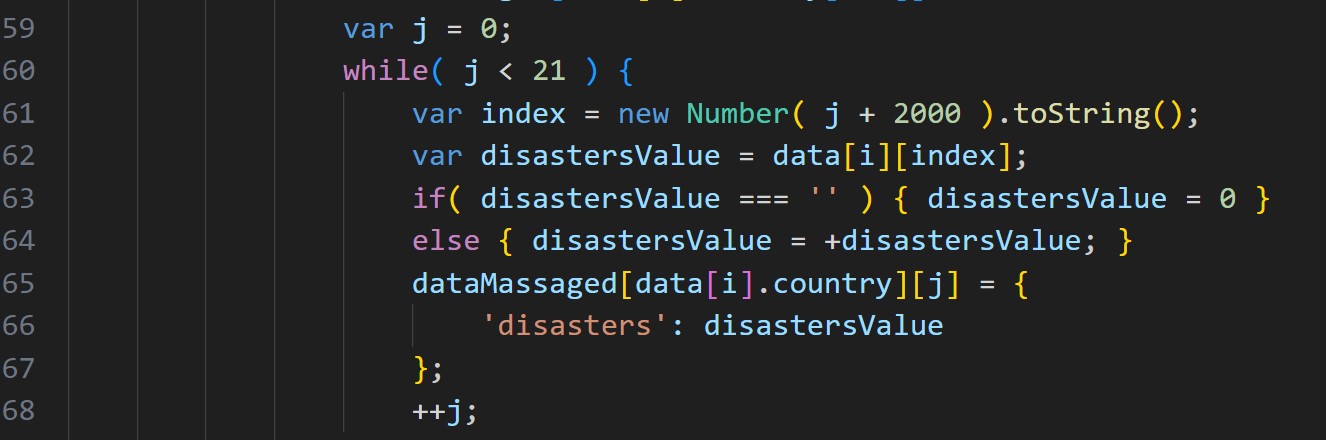
Unneeded columns and unused values have also been trimmed from the dataset with the merged countries aligning based on ISO3 coding used in all datasets.

#insert a table to show all trimmed columns and new data structure.

Total migration for relevant years has been summed and in javascript processing has been performed to show both delta change year on year for temperature and cumulative temperature change from base year 2000.

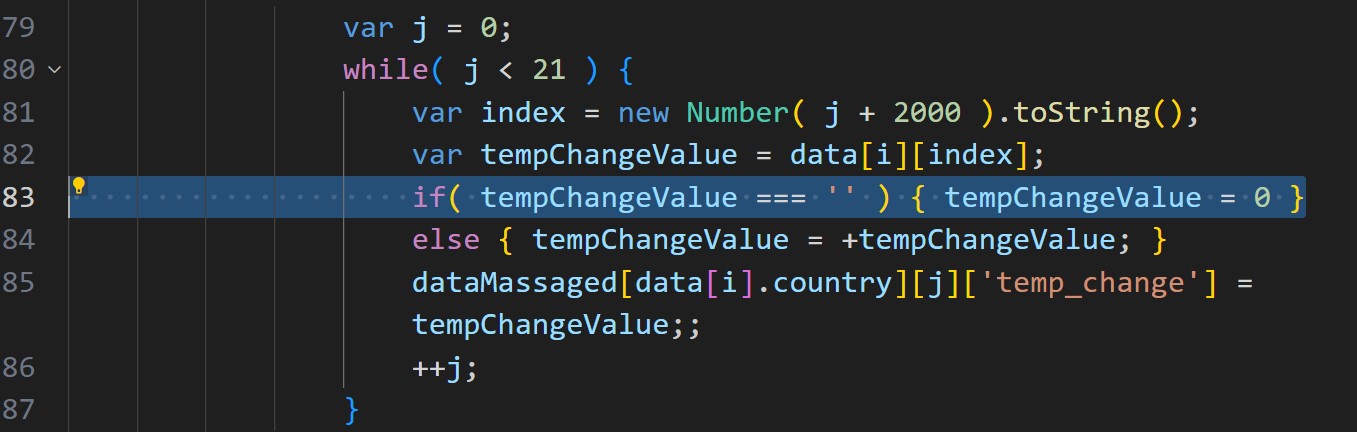
## 2.2 Data Processing (Syed)

1. Dealing with Null values in the CSV files :-



The code uses a while loop to iterate through each row of the CSV data. If the country in the current row is included in a predefined array called "countries", the code creates a new empty array called "dataMassaged" with the country name as the key.

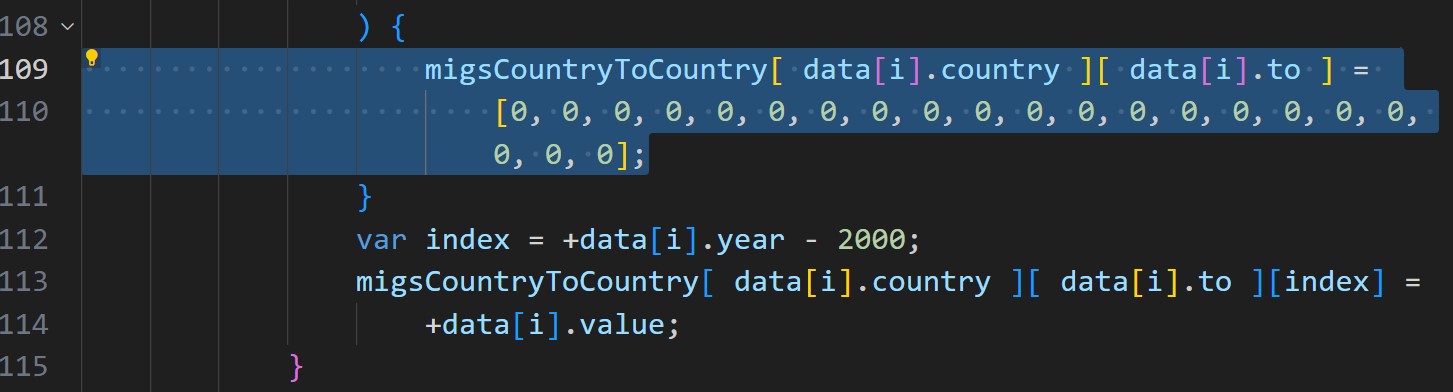
Then, Inside this loop, another while loop is used to iterate through each year from 2000 to 2020. For each year, the code gets the value of the natural disasters data for the current row and year, which is stored in a variable called "disastersValue". If the value is empty, the code sets it to 0.



The script is still iterating through the same row of the CSV data that matches one of the countries in the "countries" array and Inside this loop, another while loop is used to iterate through each year from 2000 to 2020 again. For each year, the code gets the value of the temperature change data for the current row and year, which is stored in a variable called "tempChangeValue".

If the value is empty, the code sets it to 0. Otherwise, the code converts the value to a number using the "+" operator and assigns it to the "tempChangeValue" variable.

The code then assigns the value of "tempChangeValue" to the "temp\_change" property of an object inside the "dataMassaged" array. The object is located at the index of the current year minus 2000, and its "temp\_change" property is being updated with the current temperature change value.



For the total outflows of migration a while loop is used to iterate through each row of the CSV data. If the country in the current row is included in the pre-defined "countries" array, the code proceeds to manipulate the data.The code checks if the current country in the row is not already in the "migsCountryToCountry" object. If it is not, the code creates a new object with the current country as the key in the "migsCountryToCountry" object.

Next, the code then checks if the "to" country in the current row is not already in the "migsCountryToCountry" object for the current country. If it is not, the code creates a new array of length 21 (corresponding to the years 2000-2020) as the value of the "to" country key in the "migsCountryToCountry" object for the current country.

The code then gets the index of the current year (i.e., the difference between the current year and 2000) and assigns the migration value for the current row to the corresponding index in the "migsCountryToCountry" object.

1. Dealing with irregular country name in each CSV file

There is inconsistency in naming of the country in each csv file, for example, in the "mig.csv" file, China is split into several entries, such as "China P.R.", "Hong Kong", and possibly "Chinese Taipei", which may cause issues when trying to combine or compare migration data for China as a whole. Similarly, Congo is split into "Congo, Dem Rep of the" and "Congo, Rep of", which are two different countries with distinct borders, so it's important to distinguish between them.

Korea is split into North and South Korea, which is a well-known political division. Serbia and Montenegro went through a period of political instability and territorial changes in the early 2000s, which could explain the "weird political tension" comment. Sudan and South Sudan are indeed two separate countries that split from each other in 2011.

It's important to be aware of these differences in country representation when working with multiple data files, as they may affect the accuracy and reliability of any analysis or visualization that uses the data.

Hence, to overcome this problem I had to manually change the name in the csv file to make it more coherent and easy to understand with the most recent legal country name. For an Example, “Afghanistan” to **"Afghanistan, Islamic Rep. of"** and “Turkey” to **“Türkiye”**.

# **3 Requirements**

## 3.1 Functional Requirements

Functional requirements are critical for our data visualization project website since they give a clear path for the webpage's development and deployment. These criteria provide the foundation for the website's design and functioning, ensuring that the webpage serves the demands of its users while also serving as a valuable resource for data visualization and analysis. These are some of the functional requirements we think are in line with our project:

**Data integration** is a crucial functional requirement on a data visualization project website because it enables the website to bring together data from multiple sources and provide users with a comprehensive view of the data. By integrating data from different sources such as CSV, Excel, SQL databases, APIs, and other data sources, the website can provide users with a more complete and accurate picture of the data, enabling them to make better decisions. Data integration also improves data quality by ensuring that the data is up-to-date, consistent, and reliable. Additionally, data integration enables the website to scale and grow over time, as it can accommodate new data sources and formats as they become available. This enhances the website's value proposition and makes it more attractive to users.

**Data visualization** is an important functional need for a data visualization project website because it allows users to quickly interpret complicated data and make educated decisions. Users may immediately spot patterns and trends that might be difficult to detect in raw data by presenting the data in a visual style utilizing graphs, charts, and maps. This increased comprehension can lead to better decision-making and, eventually, inspire innovation. Data visualization also improves communication by allowing users to communicate their findings in a more accessible and intelligible manner with others. Furthermore, data visualization makes data more fascinating and engaging, improving user engagement and allowing users to dig deeper into the data.

**Navigation** for the data visualization project is another important sector since it allows for quick and simple access to the website's information. A well-designed navigation system assists users in swiftly finding the information they want, therefore improving their online experience. Navigation improves the usability of the website by making it easier for users to execute activities like discovering data sets, filtering data, and accessing visualizations. Furthermore, it promotes website expansion by allowing for the addition of additional material and features without becoming cluttered or difficult to use. Good navigation may help improve accessibility for disabled people, making the website more inclusive. The website may assist visitors in exploring its material more quickly and engaging with the data and visualizations in a more meaningful way by offering simple and intuitive navigation, eventually increasing the user experience and accomplishing the website's aims.

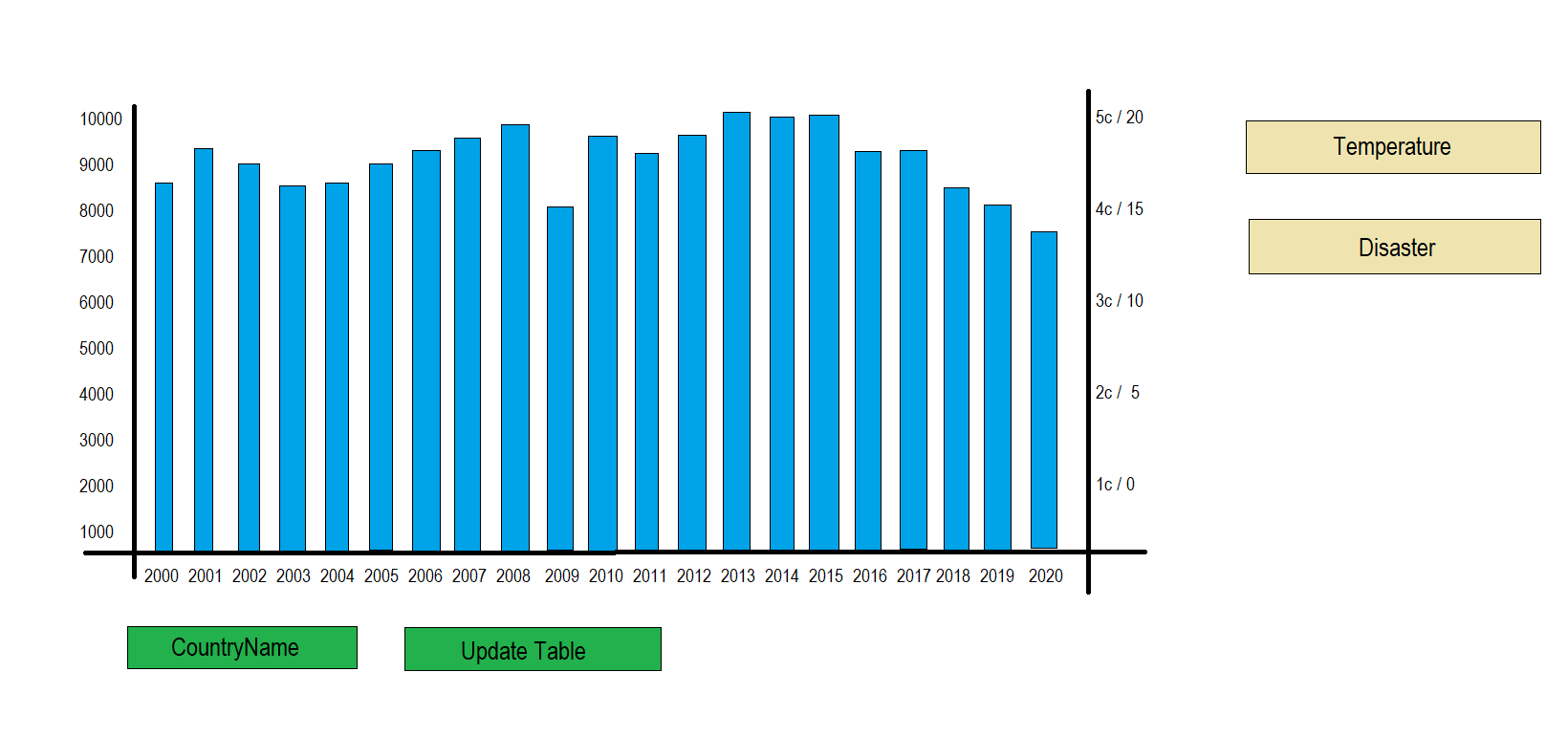
## 3.2 Non-Functional Requirements

Non-functional requirements for a data visualization project website are significant since they describe how the site should run and behave. Non-functional criteria include, among other things, user accessibility, interactive features, performance, and user experience. Non-functional requirements establish how effectively the website should accomplish what functional requirements indicate and may have a big influence on the user experience and the success of the website. Non-functional criteria also guarantee that the website complies with industry standards and best practices, such as those governing accessibility, data protection, and security. Here are some of them that we are focusing on for our project:

**UI design** is a critical non-functional need, as color, typography, layout, and visual hierarchy are all examples of good UI design aspects. The purpose of user interface design is to produce a visually beautiful and user-friendly website that allows users to quickly browse and interact with information. A well-designed user interface may improve the user experience, boost user engagement, and raise user happiness. It can also have an impact on the website's credibility and brand image, since a badly designed website may cause consumers to doubt the website's authenticity or professionalism.

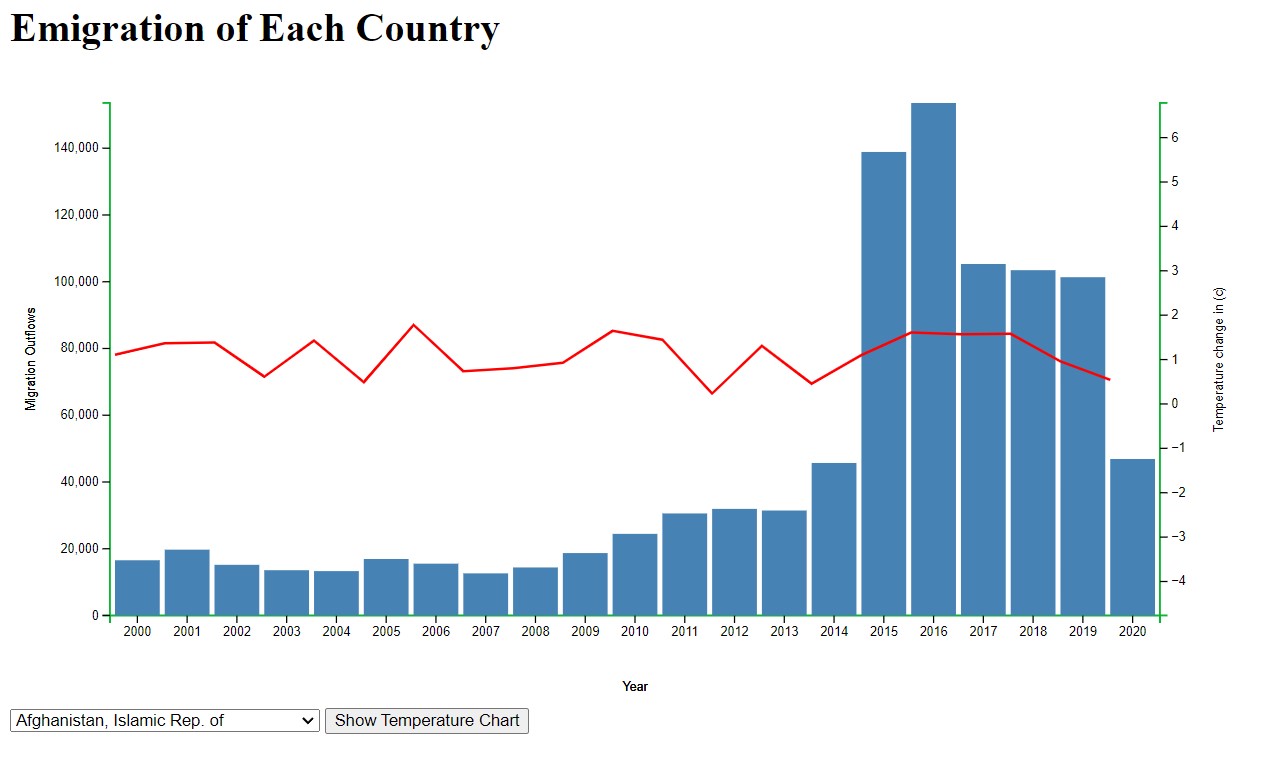
**4 Visualisation Design**

### Prototype 1



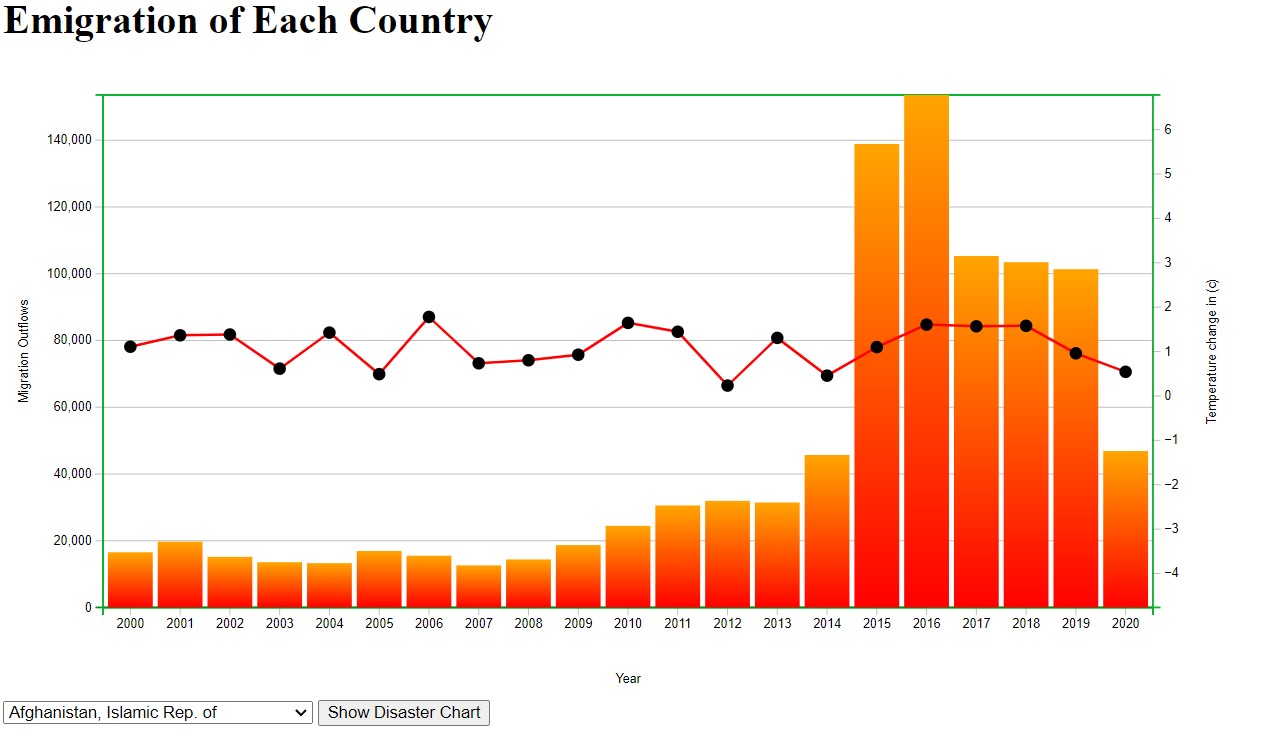
Initial Sketch,

### Iteration 1



Final working Model

### Iteration 2



### 

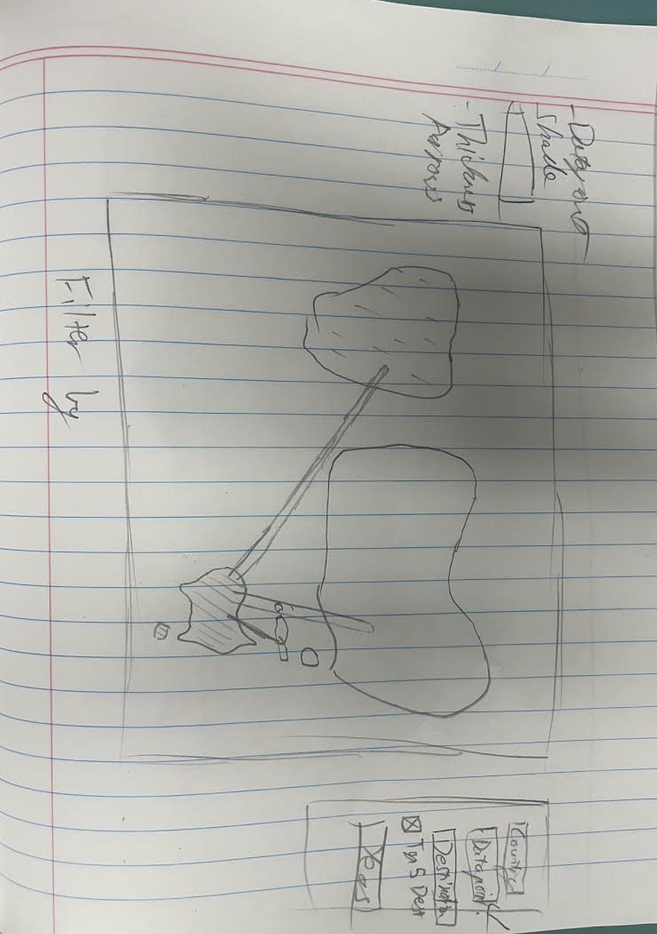
### Iteration 4 (Final)

### 

### 

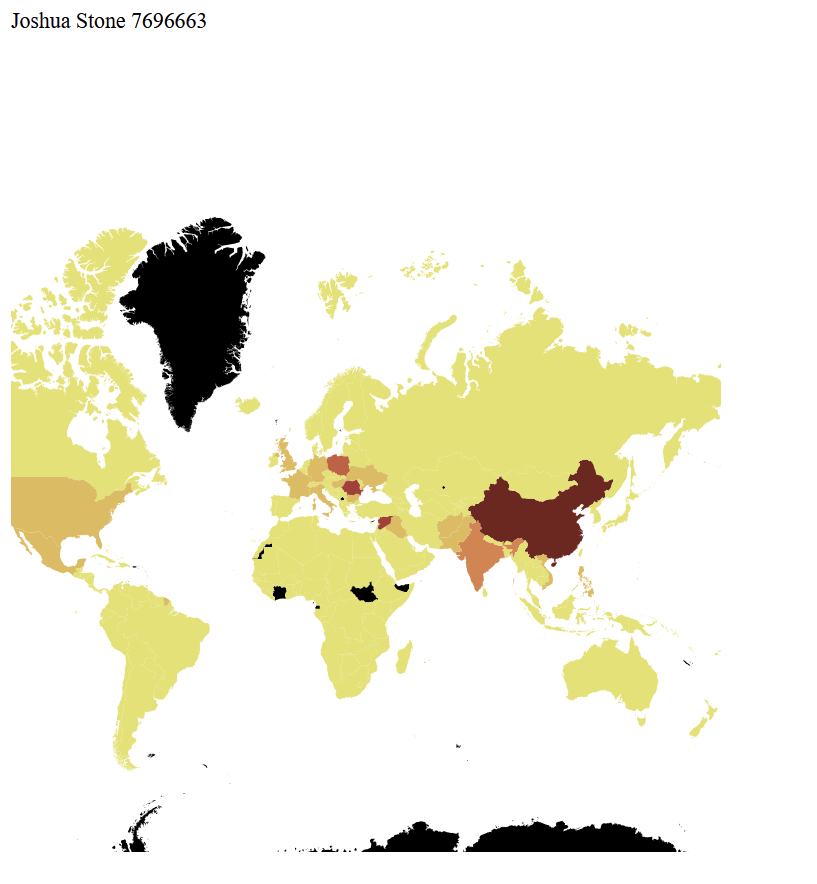
### Prototype 1- Heatmap

Initial draft



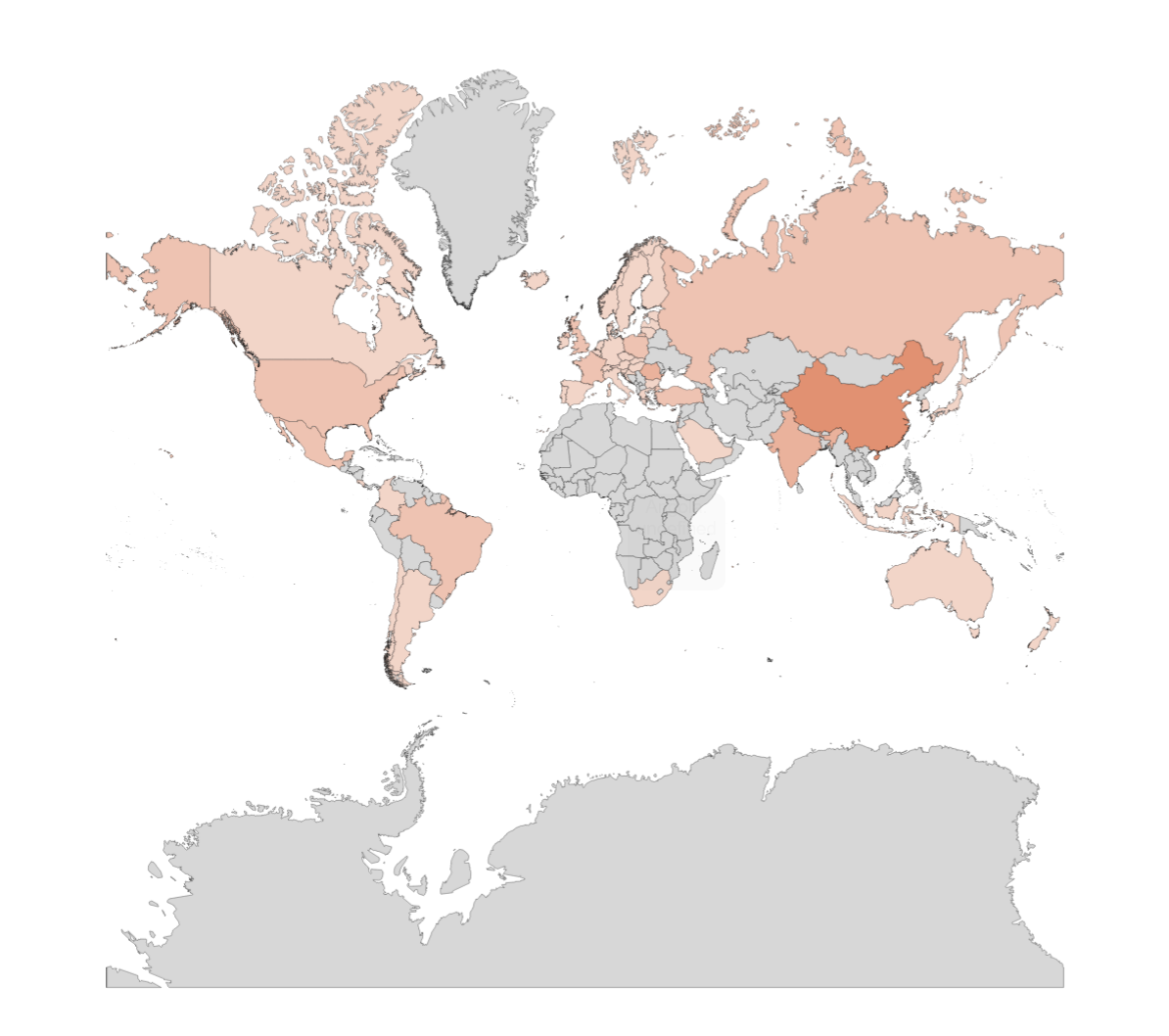
### Iteration 1

Initial D3 Implementation lacking features/color scheme but drawing from data/geojson

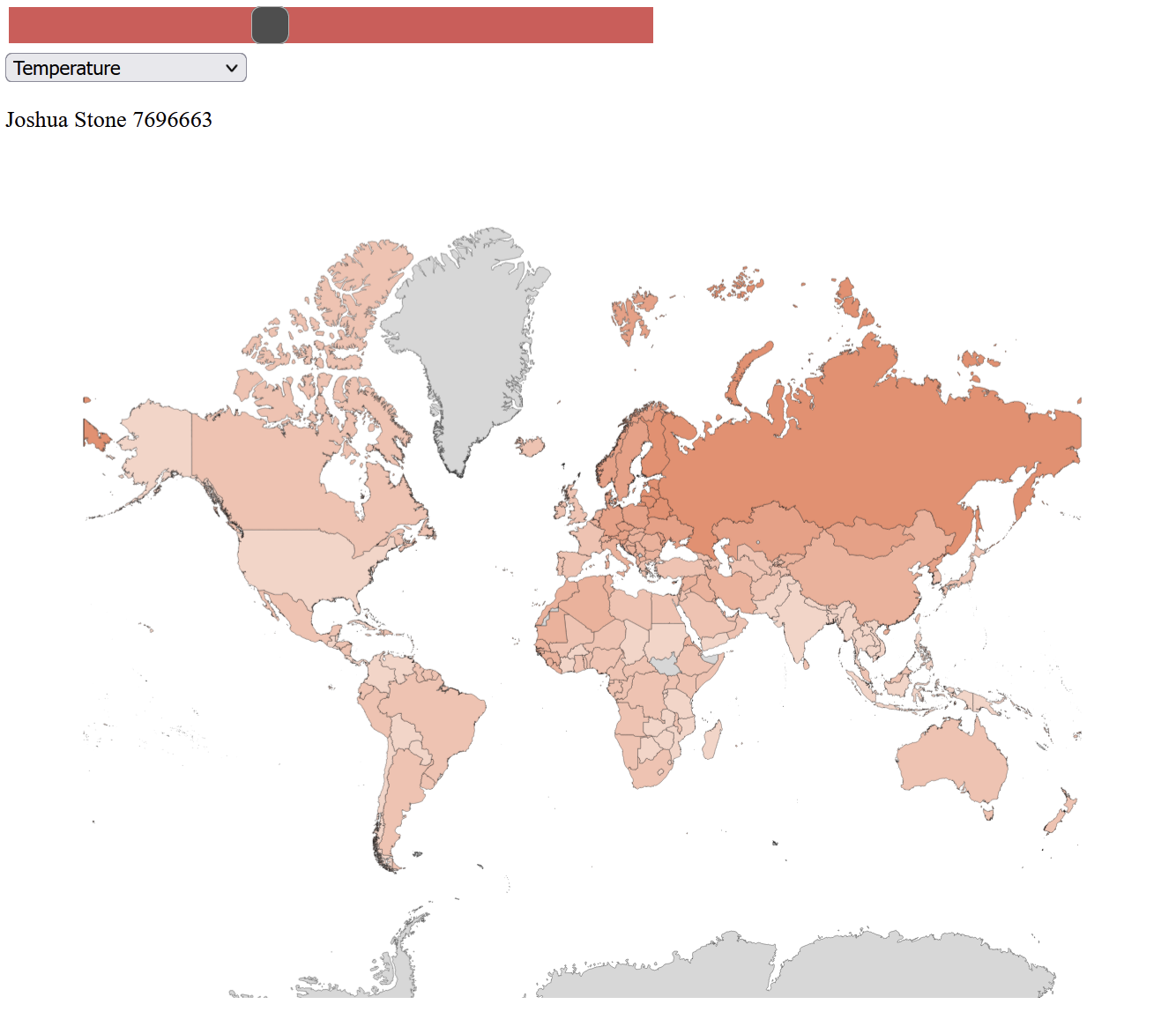


### 

### Iteration 2



Draft 4



### 

### Amendments

## Use of design guidelines

Check/Justify visualisation choice

Define data and data types in use in visualisation

Specify mark, channels, how it works with cognition/perception

.

# **5 Validation**

Test your visualization with users and report the results.

### Survey

<https://forms.office.com/Pages/ResponsePage.aspx?id=eXV_35w-fkq4REICgPU4WdZKoAelzMtFmR1klTNdB_5URFQ2NDBMSkpZM0pFSUpKT0FIVEpORlBTRS4u>

# **6 Conclusion**

The data visualization project involved the creation of multiple charts to display and analyze data. The process started with the selection of appropriate chart types for each data set. This was followed by importing and cleaning the data using various data processing techniques.

Next, the data was plotted using D3.js library in JavaScript. Different types of charts, including bar charts, line charts, and scatterplots, were created to display different types of data. Each chart was styled with custom colors, fonts, and axis labels to make the data more readable and understandable.The charts were then fine-tuned with additional features, including hover-over effects, tooltips, and legends, to provide more interactive and user-friendly experiences. Finally, the charts were tested and evaluated to ensure their accuracy and functionality.

Throughout the project, a high level of attention was given to the accuracy and readability of the data. The charts were designed to provide the audience with a clear and concise understanding of the data being presented. The process involved several iterations, from the initial chart design to the final version, to ensure that the charts effectively communicated the insights from the data. Overall, the project demonstrated the power of data visualization in making complex data sets more accessible and understandable to a wider audience.

# **7 References**

References consulted (blogs, books, academic papers, discussion/help forums - for   
both design and programming)