Process Book: Data Visualization Project World Refugees Migration

Nguyen Nam Tung (103181157)
Bill Pham (103826322)
Tutorial: Monday 16:30 - 18:30
Semester 1, 2023
Due date: 4 June 2023
Mercury Website link: https://mercury.swin.edu.au/cos30045/s103181157/DataVisualizationProject/src/home.html
GitHub Website link: https://datavisualization-github-io.vercel.app/src/home.html
Word count: 7383

Executive Summary

Our data visualization project aims to shed light on the multifaceted narrative of global refugee migration. By leveraging complex migration data, we have created interactive visualizations that provide a comprehensive, yet accessible overview of the global refugee crisis.

Our first visualization, a dynamic heatmap, presents a global perspective on refugee origins, with color intensities corresponding to the number of refugees from each country. When a user hovers over a particular country, an area chart and a pie chart appear, providing a detailed breakdown of refugee trends over time and the gender composition of the refugees respectively.

Our second visualization is a sankey diagram that delineates the refugee flows from the top three refugee-originating countries to different regions around the world. This visualization facilitates a deeper understanding of the magnitudes and directions of these population movements.

Together, these interactive visualizations enable users to explore the data and gain meaningful insights into the refugee crisis. We hope these insights will inform policymakers, stimulate research, enhance public understanding, and ultimately contribute to solutions that will make a tangible difference in the lives of refugees.

Table of Content

Executive Summary....2
Table of Content....3
1. Introduction....5
1.1. Background and Motivation....5
1.2. Visualization Purpose....5
1.3. Project schedule....6
2. Data....7
2.1. Data source....7
2.1.1. Refugees migrating from a country from 1980 to 2021....7
2.1.2. Demographic data of the refugees from 1980 to 2021....9
2.1.3. Refugees migration flow of top 3 countries with the highest number of refugees in 2021....11
2.2. Data processing....12
2.2.1. Refugees migrating from a country from 1980 to 2021....12
2.2.2. Demographic data of the refugees from 1980 to 2021....14
2.2.3. Refugees' migration flow of top 3 countries with the highest number of refugees in 2021....15
3. Requirements....19
3.1. Must-Have Features....19
3.2. Optional Features....19
4. Visualization Design....20
4.1. Conceptualization....20
4.2. Design Sketching....20
4.2.1. Initial Design....20
4.2.2. Iterative Design....24
4.3. Final Design....28
4.3.1. Design Overview....28
4.3.2. Design Description....29
4.3.3. Design Critique....34
5. Validation....35
5.1. Introduction....35
5.2. Testing....35
5.3. Result....35
5.4. Recommendation....40

6. Conclusion 40
Appendix 42
A. User Validation Form Question 42
B. User Validation Form Result 47
Reference 56

1. Introduction

1.1. Background and Motivation

Migration is a complex and dynamic process. It has played a crucial role in the development and evolution of the world's societies, cultures, and economies since the beginning of time. In recent years, however, its significance has grown due to rising political conflicts, social disruptions, and environmental threats. The most vulnerable group within this larger migration phenomenon are refugees, which are the individuals forced to leave their homes due to conflict, persecution, or disaster.

We currently find ourselves confronting an overwhelming refugee crisis. According to the UN Refugee Agency, by the end of 2021, around 89.3 million people would have been forced to abandon their homes (The UN Refugee Agency, n.d.). As a concerned student, I felt an obligation to leverage my skills to bring this crisis into focus and stimulate informed discussions about the plight of refugees.

This project is motivated by my belief that data visualization serves more than an analytical purpose; it's a powerful storytelling platform. By bringing refugee data to life, my goal is to create a forum where their stories resonate and are listened to.

The project's ambition extends beyond shedding light on the voiceless. It aims to humanize the data, transforming numbers into narratives that can stir a call for change. Moreover, it seeks to foster a deeper understanding of global refugee movements. By doing so, we have the opportunity to better understand their lived experiences, acknowledge the hurdles they face, and advocate for a more welcoming, empathetic world.

The target audience for this project is diverse and expansive, encompassing anyone interested in or impacted by global migration trends, with a particular emphasis on refugees. This can include policy makers who can use the visualization insights to formulate effective strategies and policies, as well as individuals eager to expand their understanding of this critical issue.

1.2. Visualization Purpose

Our data visualizations are primarily designed to unravel the intricate narrative of global refugee migrations. They were designed to answer the many questions such as:

- What are the global patterns of refugee migration and how have they evolved over time?
- How do the demographics, specifically in terms of gender, vary among the refugee populations originating from different countries?
- How are refugee patterns from the top three refugee-producing nations distributed throughout the world's various regions?
- How many males or females migrated from a specific country to another place?

Through our visualizations, we aspire to empower our audience with an in-depth understanding of refugee movements. We believe that this visual approach can help demystify the patterns of forced displacement, contextualizing the gravity of the situation in each country of origin, and highlight the regions shouldering the largest refugee burdens. By focusing on these details, our audience can gain insights into the dynamics of the global refugee crisis. This knowledge could inspire new strategies for managing refugee influx, foster empathy towards affected populations, and ideally contribute to solutions for reducing forced displacement and improving the lives of those impacted.

1.3. Project schedule

Week 1 - Week 3:

- Establish a communication channel and a GitHub
- Brainstorm ideas for the data visualizations
- Begin sketching the design of the visualizations
- Outline the goals, objectives, and scope of the project

Week 4:

- Gather all the data sources that are related to the topic
- Start working on the Process Book

Week 5:

- Choosing the most appropriate datasets
- Start processing the dataset and programming the data visualization
- Continue the work on the Process Book

Week 6 - Week 10:

- Finish data cleaning and processing
- Programming the visualizations
- Continue working on the Process Book

Week 11:

- Finish the visualizations and their interactive feature
- Style the data visualization website with Tailwind and CSS
- Create a user-validation form for the website

Week 12 - Week 13:

- Upload the website to Mercury Server and insert the link to the Process Book
- Finish the Process Book
- Finalise everything: Remove redundant test files, Prepare the submission files
- Uploading the project

2. Data

2.1. Data source

Our project relies on data sourced from the United Nations High Commissioner for Refugees (UNHCR) website, which provides comprehensive data on migration around the world. This dataset is a crucial resource for our project, enabling us to analyze and visualize trends in refugee movements and inform policies and interventions to address the needs of refugees globally. Here is the link to the data source:
https://www.unhcr.org/refugee-statistics/download/?url=cbM1dR

An intriguing aspect of the website is that it enables us to query for the desired dataset, which will save us a significant amount of time during the data cleaning process. For our visualisation website, we have selected the figure for refugees moving from a country of origin from 1980 to 2021, as well as its demographic data (male and female), and the migration flow from three countries with the highest number of refugees (Syria, South Sudan, and Afghanistan) to the different regions of the world (Europe, Asia, etc.).

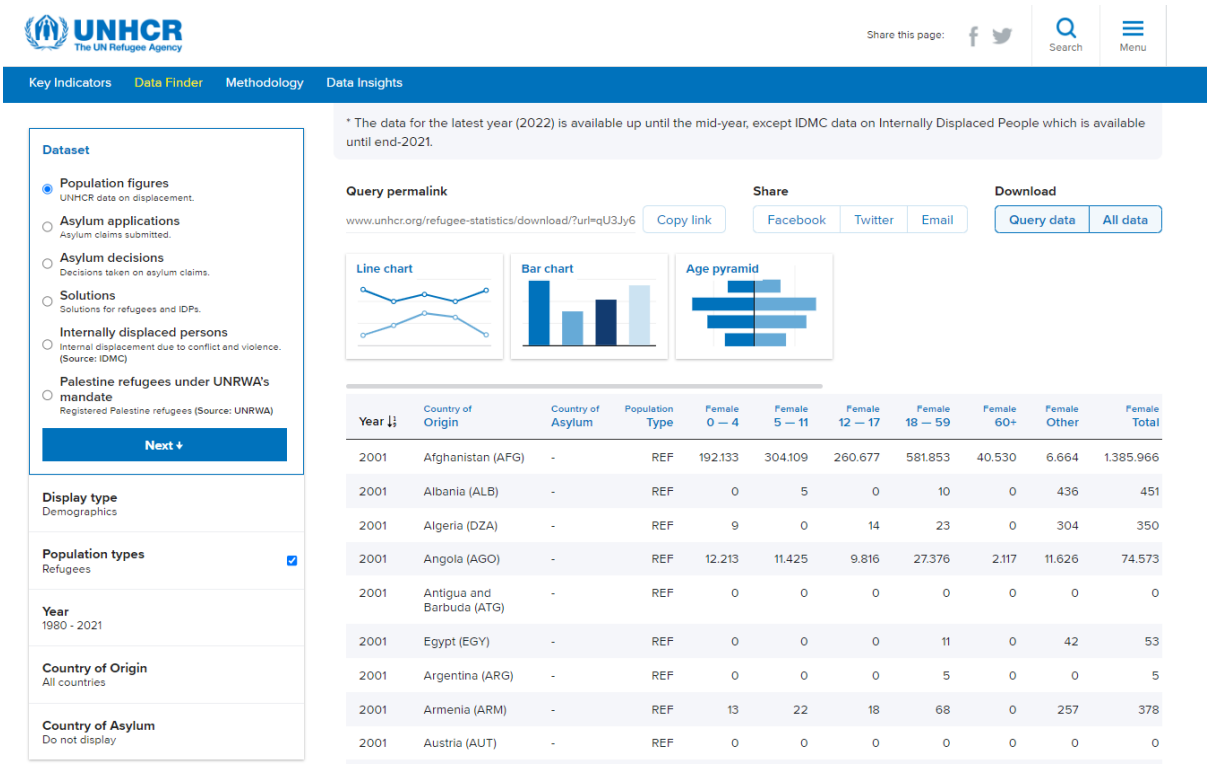


Figure 1: UNHCR website

2.1.1. Refugees migrating from a country from 1980 to 2021

This is how we queried the website's data set for refugees migrating from a country between 1980 and 2021:

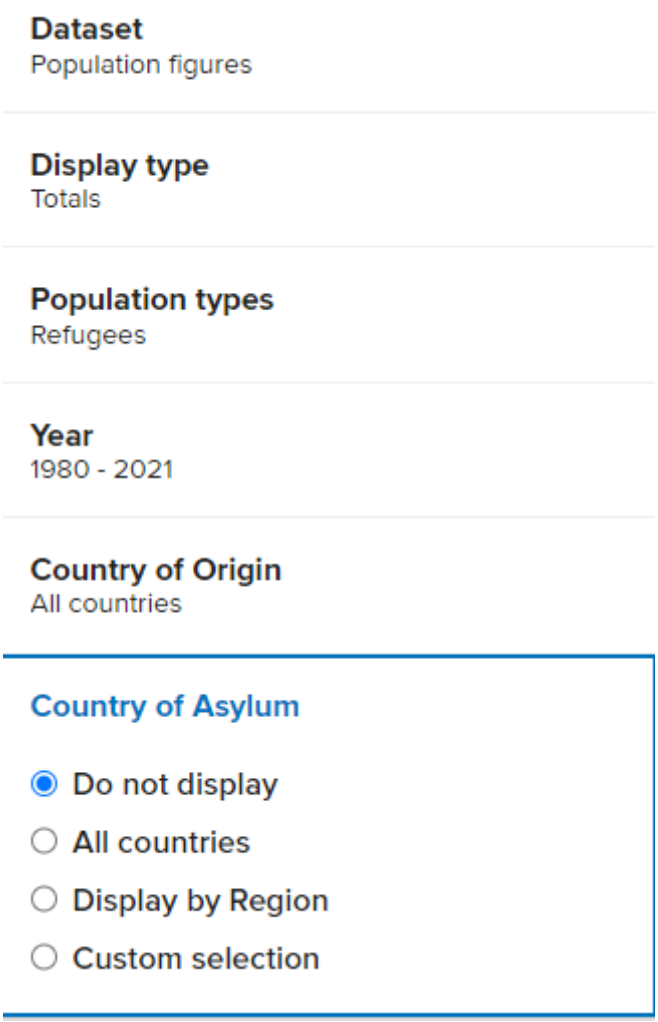


Figure 2: Querying for refugees migrating from a country from 1980 to 2021

The query data was formatted in csv with the following attributes:

Attribute	Description	Type
Year	The year the record occurred	Interval data
Country of origin	Country that the refugees come from	Nominal data
Country of origin (ISO)	The three-letter code that is the unique identifier of the origin country	Nominal data
Country of asylum	The destination country of the refugees	Nominal data
Country of asylum (ISO)	The three-letter code that is	Nominal data

	the unique identifier of the asylum country	
Refugees under UNHCR 's mandate	The count of the refugees	Discrete data

Table 1: Refugees migrating from 1980 to 2021 original dataset fields

2.1.2. Demographic data of the refugees from 1980 to 2021

Here is how we queried the website for the demographic data of refugees leaving a country between 1980 and 2021:

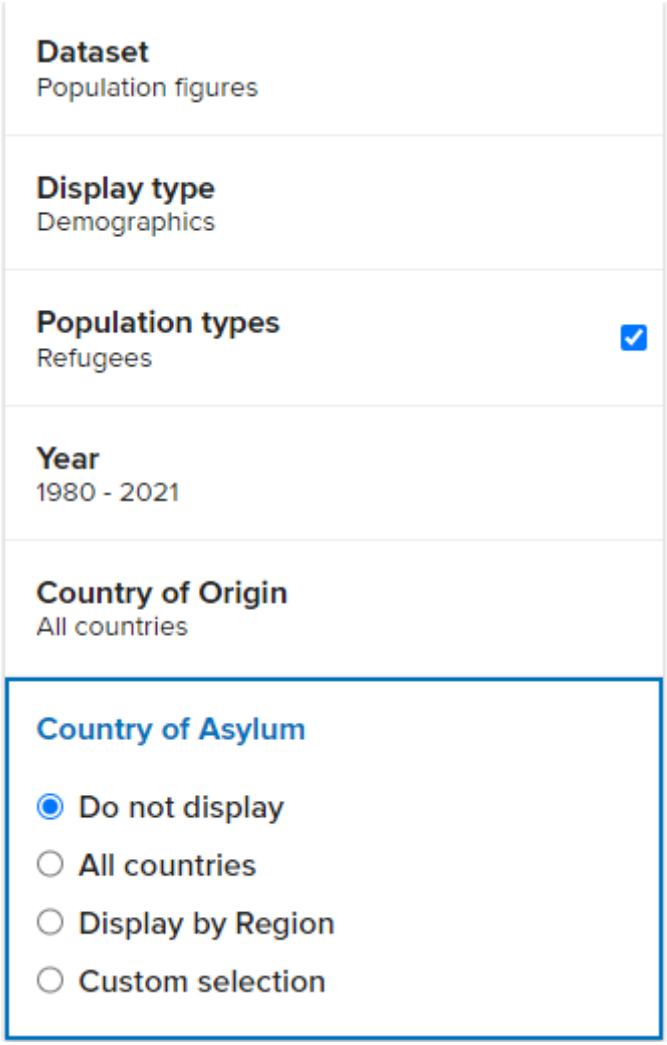


Figure 3: Querying for demographic data of refugees from 1980 to 2021

The query data was in csv file with the following attributes:

Attribute	Description	Type
Year	The year the record occurred	Interval data

Country of origin	Country that the refugees come from	Nominal data
Country of origin (ISO)	The three-letter code that is the unique identifier of the origin country	Nominal data
Country of asylum	The destination country of the refugees	Nominal data
Country of asylum (ISO)	The three-letter code that is the unique identifier of the asylum country	Nominal data
Population type	The type of population	Nominal data
Femalee 0 - 4	The number of females aged 0 to 4	Discrete data
Female 5 - 11	The number of females aged 5 to 11	Discrete data
Female 12 - 17	The number of females aged 12 to 17	Discrete data
Female 18 - 59	The number of females aged 18 to 59	Discrete data
Female 60	The number of females aged 60 and above	Discrete data
Female total	The total number of female refugees	Discrete data
Male 0 - 4	The number of males aged 0 to 4	Discrete data
Male 5 - 11	The number of males aged 5 to 11	Discrete data
Male 12 - 17	The number of males aged 12 to 17	Discrete data
Male 18 - 59	The number of males aged 18 to 59	Discrete data
Male 60	The number of females aged 60 and above	Discrete data
Male total	The total number of male	Discrete data

	refugees	

Table 2: Demographic data of refugees migrating from 1980 to 2021 original dataset fields

2.1.3. Refugees migration flow of top 3 countries with the highest number of refugees in 2021

The website does not provide the dataset for the total migration flow of refugees to a particular region; consequently, we collected the dataset for refugees moving from three countries (Afghanistan, Syria, and South Sudan) to all of the countries in a region and then calculated the total number of refugees. There were eight regions utilized in my dataset: Asia and the Pacific, East and Horn of Africa, Europe, Middle East and North Africa, Southern Africa, The Americas, and West and Central Africa.

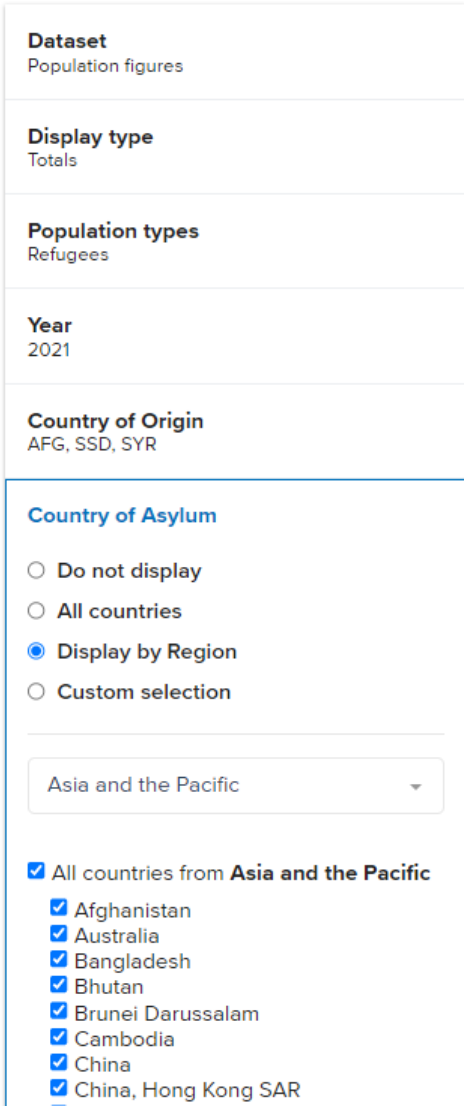


Figure 4: Querying for the sankey data

In Figure, I executed a query to retrieve data from three countries of origin to all nations in the Asia-Pacific region. The same method is used to obtain the remaining areas. Consequently, I have eight CSV datasets, one for each region, with the following attributes:

Attribute	Description	Type
Year	The year the record occurred	Interval data
Country of origin	Country that the refugees come from	Nominal data
Country of origin (ISO)	The three-letter code that is the unique identifier of the origin country	Nominal data
Country of asylum	The destination country of the refugees	Nominal data
Country of asylum (ISO)	The three-letter code that is the unique identifier of the asylum country	Nominal data
Refugees under UNHCR 's mandate	The count of the refugees	Discrete data

Table 3: Sankey data fields

2.2. Data processing

2.2.1. Refugees migrating from a country from 1980 to 2021

Extracted:	United Nations High Commissioner for Refugees				
Date extra	26/05/2023 03:11 UTC				
Last updat	16-Jun-22				
Some data presented may differ from statistics published previously due to retroactive changes or the inclusion of previously unavailable data.					
Data table	Population figures				
Content ty	End year stock population totals				
Data quick	https://www.unhcr.org/refugee-statistics/download/?url=Jng4JY				
Data defin	https://www.unhcr.org/refugee-statistics/methodology/				
Usage lice	https://creativecommons.org/licenses/by/4.0/				
Populatior	REF,ROC,ASY,IDP,IOC,OIP,OOC,STA				
Years:					
Origin:	all				
Country of asylum:					
Year	Country of origin	Country of origin (ISO)	Country of asylum	Country of asylum (ISO)	Refugees under UNHCR's mandate
1980	Afghanistan	AFG	-	-	1734921
1980	Albania	ALB	-	-	1681
1980	Angola	AGO	-	-	449030
1980	Argentina	ARG	-	-	38
1980	Armenia	ARM	-	-	333
1980	Burundi	BDI	-	-	169770
1980	Bolivia (Plurinatio	BOL	-	-	170
1980	Brazil	BRA	-	-	5
1980	Bulgaria	BGR	-	-	329
1980	Cambodia	KHM	-	-	192843
1980	Chad	TCD	-	-	221000
1980	China	CHN	-	-	24
1980	Chile	CHL	-	-	1505

Figure 5: Original dataset

The initial dataset was vast and contained numerous non-essential features, such as the ISO Code. To streamline and focus our analysis, I reduced the dataset by removing certain columns and retaining only the most relevant attributes. Additionally, I removed redundant descriptive sections from the head of the dataset. In order to create a more user-friendly experience while working with our data visualization website, I also condensed the attribute name “Refugees under UNHCR’s mandate” to “Refugees” for ease of reference.

The subsequent challenge was to integrate our CSV data with the JSON data - a necessary step to ensure our CSV data could be appropriately represented on our geographic map. This process proved time-consuming due to disparities between country names listed in our CSV data and those in the JSON file.

To overcome this, I utilized a distinct dataset, identical in nature but exclusive to the year 2021, and projected it onto a test choropleth map. This allowed us to identify any mismatches between the datasets, as any undefined values from the CSV file appeared as a grey area on the map. Using this method, we could visually contrast our map with the original world map, identify any discrepancies in country names, and subsequently update those country names in our primary dataset. This meticulous process ensured a seamless and accurate integration of our data sources.

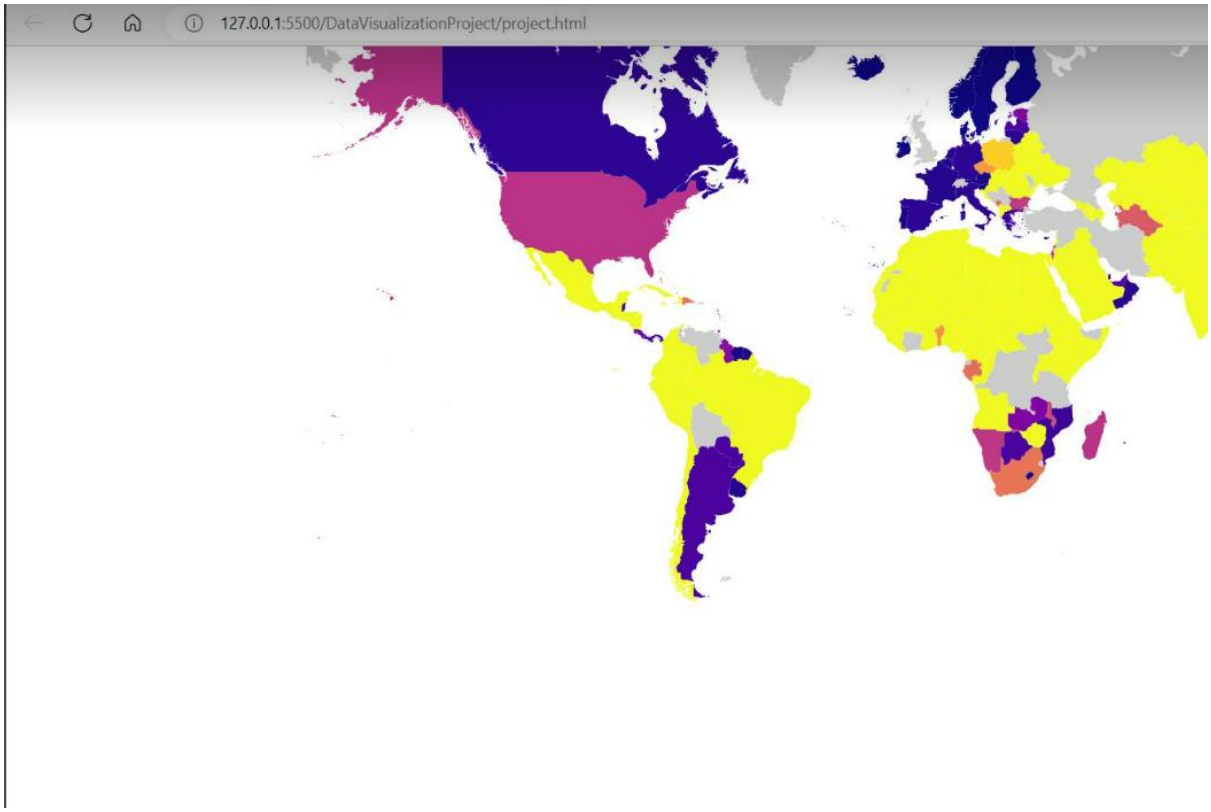


Figure 6: Test choropleth map

Here is our final dataset and its attributes:

Year	Country	Refugees
1980	Afghanista	1734921
1980	Albania	1681
1980	Angola	449030
1980	Argentina	38
1980	Armenia	333
1980	Burundi	169770
1980	Bolivia	170
1980	Brazil	5
1980	Bulgaria	329
1980	Cambodia	192843

Figure 7: Final heatmap dataset

Attribute	Description	Type
Year	The year the record occurred	Interval data
Country	Country that the refugees come from	Nominal data
Refugees	The count of the refugees	Discrete data

Table 4: Final refugees data from 1980 to 2021 field

2.2.2. Demographic data of the refugees from 1980 to 2021

Extracted:	United Nations High Commissioner for Refugees																		
Date extra	26/05/2023 03:16 UTC																		
Last updat	16-Jun-22																		
Some data presented may differ from statistics published previously due to retroactive changes or the inclusion of previously unavailable data.																			
Data table	Population figures																		
Content ty	End year stock population totals																		
Data quick	https://www.unhcr.org/refugee-statistics/download/?url=6eqOTk																		
Data defin	https://www.unhcr.org/refugee-statistics/methodology/																		
Usage lice	https://creativecommons.org/licenses/by/4.0/																		
Populatior	REF,ROC,ASY,IDP,IOC,OIP,OOC,STA																		
Years:																			
Origin:	all																		
Country of asylum:																			
Year	Country of	Country of	Country of	Country of	Populatior	Female 0 -	Female 5 -	Female 12	Female 18	Female 60	Female otl	Female tot	Male 0 - 4	Male 5 - 1	Male 12 -	Male 18 -	Male 60	Male othe	Male total
2001	Afghanista	AFG	-	-	REF	192133	304109	260677	581853	40530	6664	1385966	195429	322895	276762	782758	57368	8718	1643930
2001	Albania	ALB	-	-	REF	0	5	0	10	0	436	451	0	0	0	9	0	653	662
2001	Algeria	DZA	-	-	REF	9	0	14	23	0	304	350	10	11	10	93	0	556	680
2001	Angola	AGO	-	-	REF	12213	11425	9816	27376	2117	11626	74573	11401	12078	10357	22967	1827	14743	73373
2001	Antigua an	ATG	-	-	REF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	Egypt	EGY	-	-	REF	0	0	0	11	0	42	53	0	0	0	21	0	65	86
2001	Argentina	ARG	-	-	REF	0	0	0	5	0	0	5	0	0	0	0	0	0	0
2001	Armenia	ARM	-	-	REF	13	22	18	68	0	257	378	11	23	21	69	5	244	373
2001	Austria	AUT	-	-	REF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	Azerbaijan	AZE	-	-	REF	3725	5436	4672	63529	66040	131	143533	4011	5262	4514	55266	52140	124	121317
2001	Bahrain	BHR	-	-	REF	0	0	0	0	0	0	0	0	0	0	5	0	0	5
2001	Barbados	BRB	-	-	REF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	Burundi	BDI	-	-	REF	36236	33516	28735	73256	3531	95120	270394	36527	34596	29673	77432	3564	77256	259048

Figure 8: Original demographic dataset

The data cleaning process of this dataset was the same as the previous one. We eliminated unnecessary attributes and records, adhered to consistent naming conventions, and aligned the CSV data with its corresponding JSON data. Here is the final dataset and its attributes:

Year	Country	FemaleTot	MaleTotal
2001	Afghanista	1385966	1643930
2001	Albania	451	662
2001	Algeria	350	680
2001	Angola	74573	73373
2001	Antigua an	0	0
2001	Egypt	53	86
2001	Argentina	5	0
2001	Armenia	378	373
2001	Austria	0	0
2001	Azerbaijan	143533	121317
2001	Bahrain	0	5
2001	Barbados	0	0
2001	Burundi	270394	259048
2001	Belgium	0	0

Figure 9: Final demographic dataset

Attribute	Description	Type
Year	The year the record occurred	Interval data
Country	Country that the refugees come from	Nominal data
FemaleTotal	The count of the female refugees	Discrete data
MaleTotal	The count of the male refugees	Discrete data

Table 5: Final demographic data field

2.2.3. Refugees' migration flow of top 3 countries with the highest number of refugees in 2021

Let's take a look at a sample of the dataset for the migration flows of the top three countries with the highest number of refugees in the Asia-Pacific region.

	A	B	C	D	E	F	G
1	Extracted:	United Nations High Commissioner for Refugees					
2	Date extra	27/05/2023 05:10 UTC					
3	Last updat	16-Jun-22					
4	Some data	presented may differ from statistics published previously due to r					
5							
6	Data table	Population figures					
7	Content ty	End year stock population totals					
8	Data quick	https://www.unhcr.org/refugee-statistics/download/?url=ThXw4					
9	Data defin	https://www.unhcr.org/refugee-statistics/methodology/					
10	Usage lice	https://creativecommons.org/licenses/by/4.0/					
11	Populatior	REF,ROC,ASY,IDP,IOC,OIP,OOC,STA					
12	Years:	2021					
13	Origin:	AFG,SYR,SSD					
14	Country of	AFG,AUL,BGD,BHU,BRU,CAM,CHI,HKG,MAC,COK,KRN,FIJ,FPO,GUI					
15	Year	Country of	Country of	Country of	Country of	Refugees under UNH	
16	2021	Afghanista	Australia	AFG	AUS	9648	
17	2021	Afghanista	China	AFG	CHN	30	
18	2021	Afghanista	Fiji	AFG	FJI	8	
19	2021	Afghanista	China, Hor	AFG	HKG	11	
20	2021	Afghanista	India	AFG	IND	9678	
21	2021	Afghanista	Indonesia	AFG	IDN	5828	
22	2021	Afghanista	Iran (Islam	AFG	IRN	778054	
23	2021	Afghanista	Japan	AFG	JPN	84	
24	2021	Afghanista	Kazakhstai	AFG	KAZ	327	
25	2021	Afghanista	Kyrgyzstan	AFG	KGZ	213	
26	2021	Afghanista	Rep. of Ko	AFG	KOR	46	
27	2021	Afghanista	Sri Lanka	AFG	LKA	113	
28	2021	Afghanista	Malaysia	AFG	MYS	1604	

Figure 10: Original sankey dataset

As usual, I first pruned the CSV data by eliminating any irrelevant attributes and rows. Then, Using Excel's SUM function, I determined the total number of refugees for each of the three countries.

```
= SUM(F2,F3,F4,F5,F6,F7,F8,F9,F10,F11,F12,F14,F13,F15,F16,F17,F18,F20,F19,F21,F22,F23,F24)
```

Figure 11: SUM Excel function utilzed

After obtaining the result, I inserted it into the Sankey JSON file. Here is the final JSON file for our sankey diagram.:

```
{
    "nodes": [
        {"node": 0, "name": "Afghanistan"},
        {"node": 1, "name": "South Sudan"},
        {"node": 2, "name": "Syria"},
        {"node": 3, "name": "Asia and the Pacific"},
        {"node": 4, "name": "West and Central Africa"},
        {"node": 5, "name": "America"},
        {"node": 6, "name": "Europe"},
        {"node": 7, "name": "South Africa"},
        {"node": 8, "name": "Middle East and North Africa"},
        {"node": 9, "name": "East Africa"}
    ],
    "links": [
        {"source": 0, "target": 3, "value": 489645},
        {"source": 0, "target": 4, "value": 366257},
        {"source": 0, "target": 5, "value": 343392},
        {"source": 0, "target": 6, "value": 626879},
        {"source": 0, "target": 7, "value": 433530},
        {"source": 0, "target": 9, "value": 453166},
        {"source": 1, "target": 3, "value": 456123},
        {"source": 1, "target": 4, "value": 395712},
        {"source": 1, "target": 5, "value": 524897},
        {"source": 1, "target": 6, "value": 432901},
        {"source": 1, "target": 9, "value": 132621},
        {"source": 1, "target": 8, "value": 420505},
        {"source": 2, "target": 4, "value": 1425953},
        {"source": 2, "target": 5, "value": 634872},
        {"source": 2, "target": 6, "value": 2567210},
        {"source": 2, "target": 7, "value": 1379126},
        {"source": 2, "target": 8, "value": 841704}
    ]
}
```


Figure 12: Final JSON dataset

Our dataset is represented by the JSON file, which is used to generate the Sankey diagram. A JSON file contains two objects: node and link. Node consists of the properties like "node" to signify the unique node identifier and "name" to denote the name associated with each node. In contrast, Link is indispensable for establishing connections between the nodes. It comprises a “value” attribute representing the data volume or value transferred from the “source" node to the “target” node . As an illustration of how the JSON file works, let’s consider Node 0 and Node 3. Node 0 is representative of Afghanistan, while Node 3 corresponds to the Asia and Pacific region. The link established from the source at Node 0 to the target at Node 3 has a

specific value of 489,645. This numerical value depicts the migration of 489,645 refugees from Afghanistan to the Asia-Pacific region.

3. Requirements

3.1. Must-Have Features

We have defined some must-have features without which the project cannot continue and which must take precedence over all other considerations. These included:

Heatmap, line chart, and bar chart:

These visualizations serve as the primary representations in our project. They provide comprehensive insights into the data, enabling a clear understanding of the patterns and trends.

Interactive graph:

The interactive graph enhances user engagement by allowing them to explore and analyze the data more deeply. Within the heatmap, a slider is incorporated, enabling users to slide through different years and observe corresponding data. Additionally, by selecting a particular country, users can view detailed information about that country's data for the current year.

Alongside the heatmap, the line chart depicts the trajectory of refugee movement within each country over time.

Sankey diagram:

The Sankey diagram adds further functionality through the inclusion of tooltips and an outline on hover. This enables users to obtain additional information and insights when interacting with the diagram.

3.2. Optional Features

In addition to the required features we intended to use in the visualization, we are delighted to introduce some optional features that have been included to enhance the overall visual appeal, improve data differentiation, and highlight trends within the data. These optional features are outlined below:

Heatmaps:

To ensure inclusivity, we have incorporated two color ranges within the heatmaps. The first color range is designed for individuals with normal color vision, while the second color range is specifically tailored for individuals with color blindness. Our utmost priority is to provide an equal visual experience for all users, regardless of their visual abilities. Furthermore, we have implemented tooltips using D3, enabling users to access brief descriptions when hovering the mouse over the maps.

Pie Chart:

The pie chart is seamlessly integrated into the same page as the heatmaps, positioned alongside them. This informative pie chart showcases the distribution of males and females among the refugee population, providing a concise representation of gender demographics.

Sankeys:

Our Sankeys feature offers an interactive experience as it allows users to drag and manipulate the visual elements. This dynamic functionality provides users with the flexibility to explore and analyze data relationships with ease.

By incorporating these optional features, we aim to create an exceptional visualization that not only enhances data differentiation but also caters to the diverse needs of users.

4. Visualization Design

4.1. Conceptualization

Our visualization design was driven by well-established principles of data visualization, notably those proposed by experts such as Edward Tufte and Tamara Munzner. These principles guided us in creating a clear, effective, and engaging visual narrative of global refugee movements.

Edward Tufte's principles played a central role in our design process. Tufte advocates for graphical integrity and a high data-ink ratio, meaning the ink used to represent data should take precedence over decorative elements. Therefore, we minimized any redundant or non-data ink, ensuring the focus remained on the crucial information we intended to present. Tufte's notion of maximizing data density, or the amount of data within a given space, also influenced our designs. We aimed to provide comprehensive insights within our visualizations, presenting a dense yet readable data view. (Tufte, 1983)

Tamara Munzner's guidelines were instrumental in choosing our visual encodings and idioms. Munzner emphasizes the importance of matching the type of data with appropriate visual encodings. This conscious selection of encodings and idioms has been instrumental in maintaining the accuracy and effectiveness of our visualizations. (Munzner, 2015)

These guiding principles have ensured that our visualizations are not just aesthetically pleasing, but they effectively communicate the complex story of global refugee movements. By aligning the principles of visualization design with our project goals, we've strived to create a user experience that is informative, intuitive, and impactful. We believe that the data speaks for itself, and our design is simply a tool to amplify its voice.

4.2. Design Sketching

4.2.1. Initial Design

This is our initial website design layout, showcasing the key elements on the first webpage. At the top, we have implemented a user-friendly navigation bar, enabling navigation throughout our visualizations (see Figure 13).

Moving on to the visualization section, we have created a dynamic combination of three graphs. The first graph is a heatmap, followed by a bar chart and a stacked bar chart. Immediately below, we have included a clear and informative legend specifically designed for the heatmaps.

In this initial stage, our focus is to provide data exclusively for the year 2021 in the heatmap, ensuring a comprehensive and detailed representation of the migration patterns.

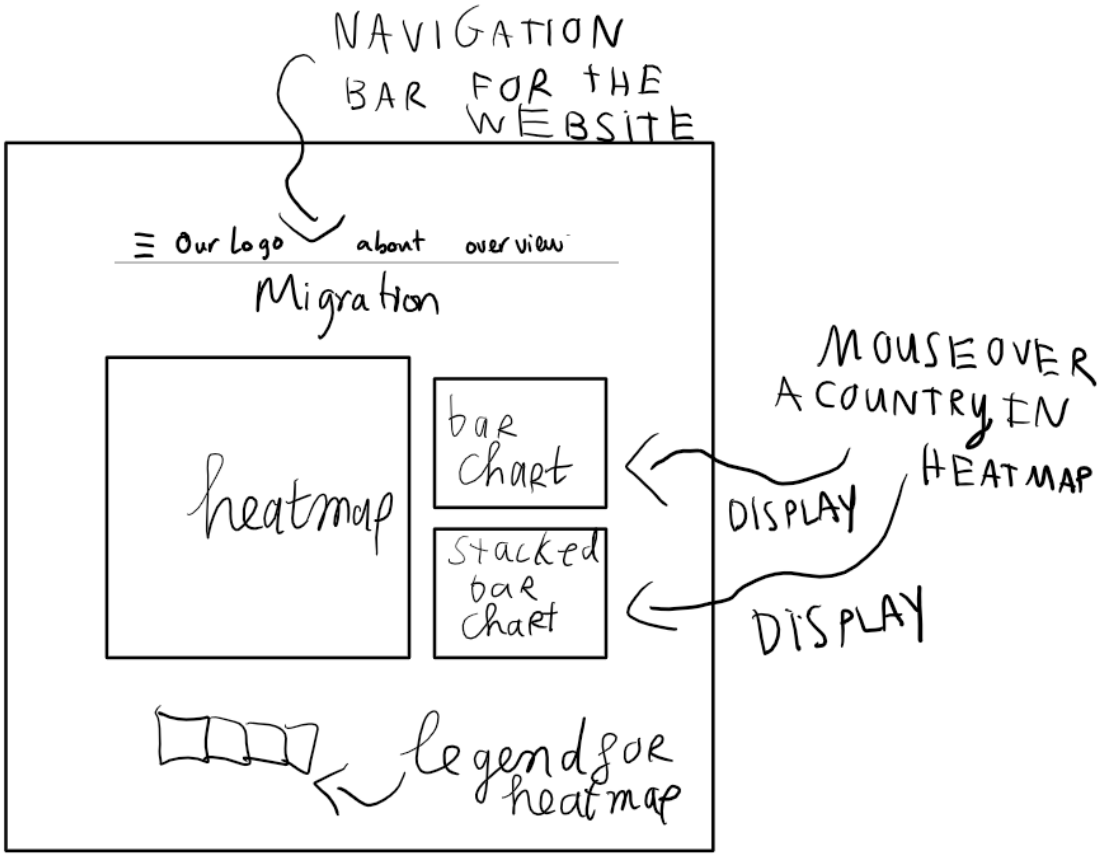


Figure 13: Initial design layout

To effectively display the time-series data, we have chosen heatmaps (also known as choropleths) as our preferred visualization method. Each country's migration flow, measured by the number of individuals, is intuitively encoded into vibrant colors.

Advantage:
Choropleth maps offer a range of benefits for visualizing spatial patterns. By utilizing color or shading to represent distinct values or categories across geographical regions, they enable us to perceive these patterns effortlessly. Viewers can swiftly identify areas of high or low values, gaining a comprehensive understanding of variations across different regions. Furthermore, these maps provide a clear and concise visual depiction of data distribution within a geographic area. Choropleth maps possess the capability to simplify complex data, making it easier to grasp key insights and discern prevailing trends, such as the migration of refugees between countries.

Disadvantage:
However, it is important to acknowledge a limitation of choropleth maps. They heavily rely on the chosen color scheme to effectively convey information. Hence, careful consideration and selection of appropriate colors are crucial to ensure accurate representation and interpretation of the data on the maps.

In conclusion, the advantages of the choropleth maps outweigh the disadvantages, so this will be our main visualization for the web.

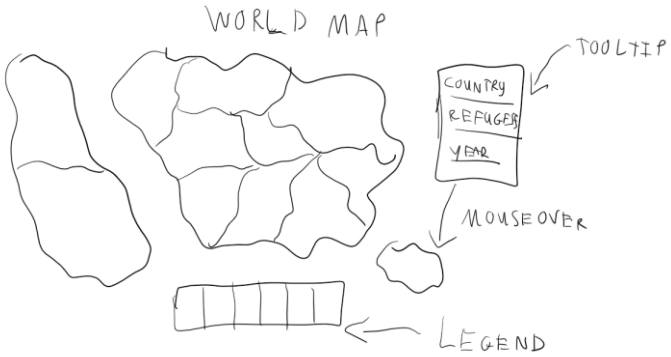


Figure 14: Initial heatmap

The bar chart and stacked bar chart will only appear when a selected country is hovered within the heatmaps.

The bar chart (Figure 15) will be utilized to visualize the changes in the number of refugees over time. The horizontal axis represents the years from 1980 to 2021, while the vertical axis represents the number of refugees.

Advantages:
On one hand, this bar chart effectively illustrates the refugee numbers. It is simple to create and widely used among various groups in the course. By observing the height of each column, one can easily discern the variations in data from year to year. Moreover, bar charts are familiar to end users, even beginners, as they can easily interpret labels and identify patterns and trends.

Disadvantages:
On the other hand, considering the extensive time range covered (1980 to 2021), the bar chart would inevitably contain numerous bars. Furthermore, since the primary area of the website is dedicated to displaying heatmaps, which are our main visualization tool, only a limited amount of space remains for the bar chart. After careful consideration, we concluded that it cannot and should not be used. Attempting to include it would undermine the advantages of the bar chart, resulting in confusion and information overload for viewers, impeding their ability to focus.

In conclusion, after weighing the advantages and disadvantages we discussed, it is clear that the drawbacks of incorporating a bar chart in our visualization outweigh its benefits. Therefore, we need to explore alternative and improved solutions to address this issue.

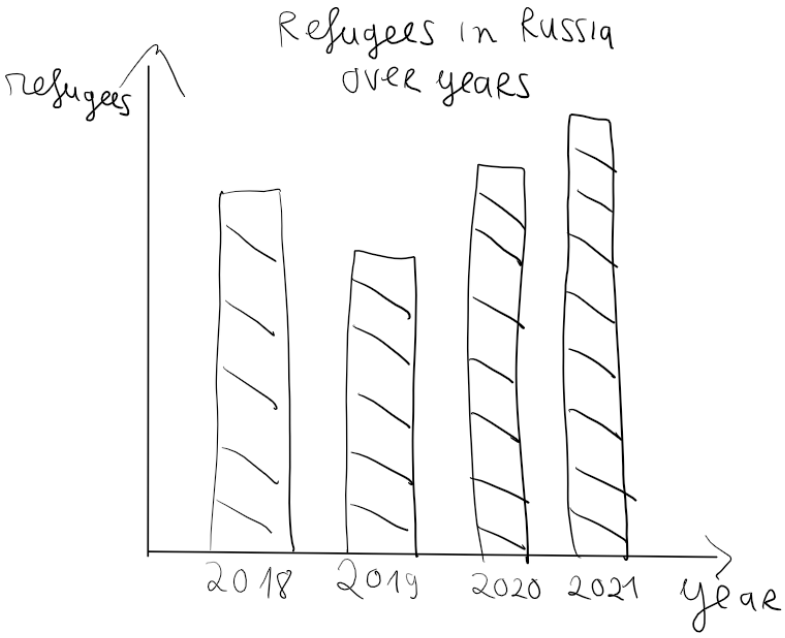


Figure 15: Initial barchart for the design

The stacked bar chart (Figure 16) is utilized to visually represent the distribution of male and female refugees in the selected country. The horizontal axis spans from 1980 to 2021, depicting the timeline, while the vertical axis quantifies the number of refugees.

Advantages:
On one hand, the stacked bar chart effectively illustrates the proportions of male and female refugees within a single bar. This allows viewers to quickly understand the relative distribution of genders within the dataset. Furthermore, the stacking of the bars enables the observation of cumulative totals over time, facilitating the identification of patterns and trends in gender composition.

Disadvantages:
On the other hand, considering the extensive time range from 1980 to 2021, the stacked bar chart may face challenges that outweigh its advantages. With such a large dataset, visual redundancy may arise, potentially making it difficult for viewers to discern meaningful patterns. Additionally, the limited space available on the website could lead to discomfort and confusion for end users if multiple stacked bar charts are displayed simultaneously.

In conclusion, after thorough discussion and careful consideration of the advantages and disadvantages, the stacked bar chart appears to be unsuitable for our current design.

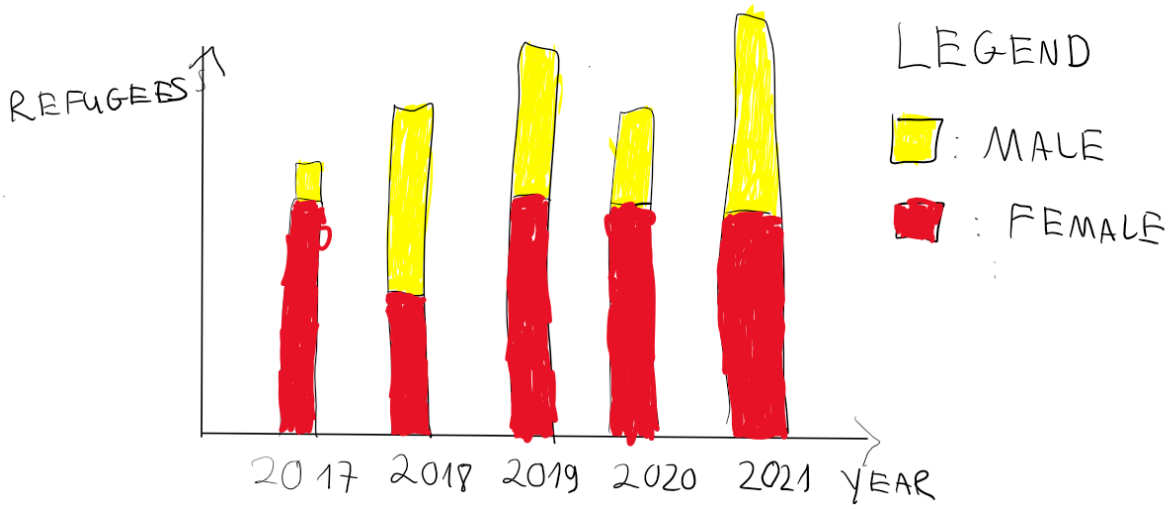


Figure 16: Initial stacked barchart for the design

4.2.2. Iterative Design

In this stage, we have established a comfortable website layout and have decided to maintain it. However, we recognize that there is room for improvement in our graph. We are determined to enhance it further.

Regarding the heatmaps (Figure 17), we have chosen to retain them as our primary visualization. Nevertheless, we have realized that if the heatmaps are fixed to display data for only one year in our main visualization, it would result in the rest of the data being overlooked. Given that the heatmaps are a crucial visualization, we need to find a solution to effectively incorporate all the data from multiple years into our graph.

To address this, we have introduced additional elements to our website. Firstly, we have implemented a years slider, which allows users to dynamically change and view data for different years. Additionally, a dropdown menu has been added to provide users with the option to select between the normal view and a color-blind view, ensuring equal accessibility for all users. Moreover, we have included tooltips that appear when the mouse hovers over a specific area, displaying the name of the country, the selected years, and the number of refugees from that country.

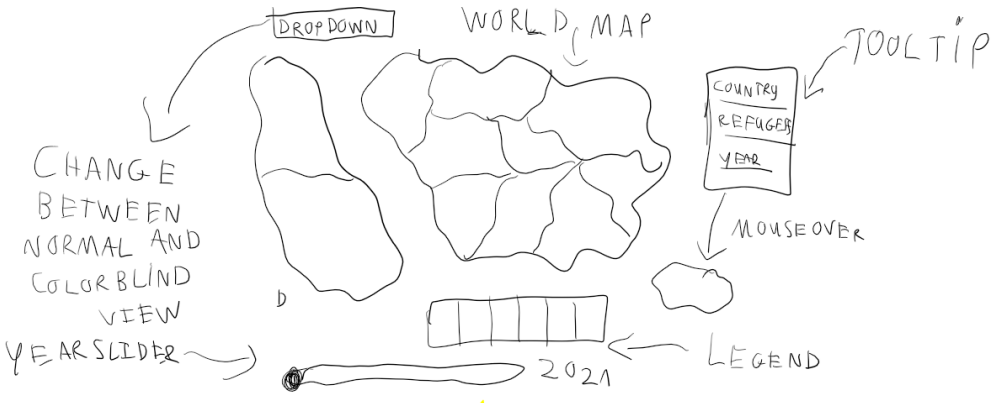


Figure 17: The redesign of the heatmap

Regarding the chart demonstrating the gender of refugees, we have explored two design options: a line chart and a pie chart.
After careful consideration, we have identified two promising designs for our visualization. However, we noticed that the line chart (Figure 18) is already being utilized to address the same question. In other words, we are using different charts to represent the same data, which goes against the purpose of our visualization. Our aim is to employ different chart types to allow viewers to observe, obtain information, and answer specific questions.
To achieve this, we should represent the same data using different formats. For instance, instead of focusing solely on the number of refugees, we can transform the data into percentages and utilize heatmaps to highlight specific years selected by the viewer using a slider. This approach would provide a fresh perspective and enhance the viewer's understanding of the information presented.

After careful consideration, we have determined that the pie chart (Figure 19) is the most suitable choice for our presentation.

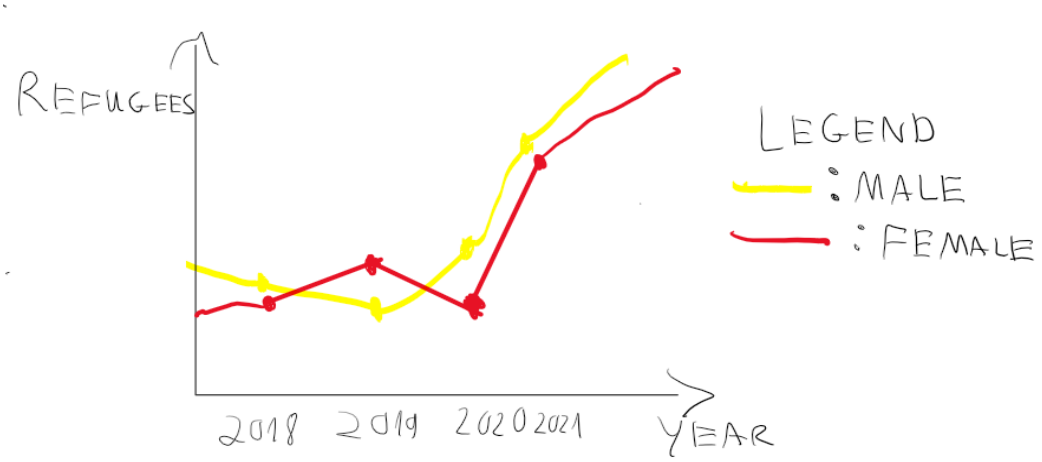


Figure 18: Draft design gender line chart

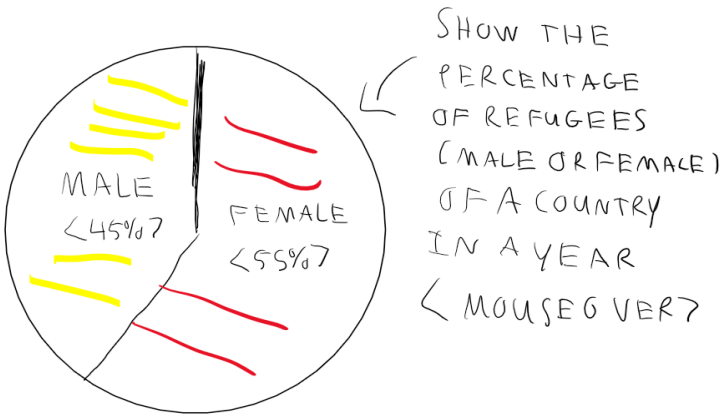


Figure 19: Draft design gender pie chart

In response to the concern regarding the bar chart's limitation of displaying the annual total of refugees, we have devised an effective solution by implementing an area chart. The area chart (figure 20) is a well-suited choice for illustrating trends over time, offering distinct advantages. By utilizing the filled area between the line and the x-axis, the area chart provides a visually comprehensive representation of the magnitude and direction of change. This clear visualization enables viewers to readily identify patterns and gain a comprehensive understanding of the overall trend.

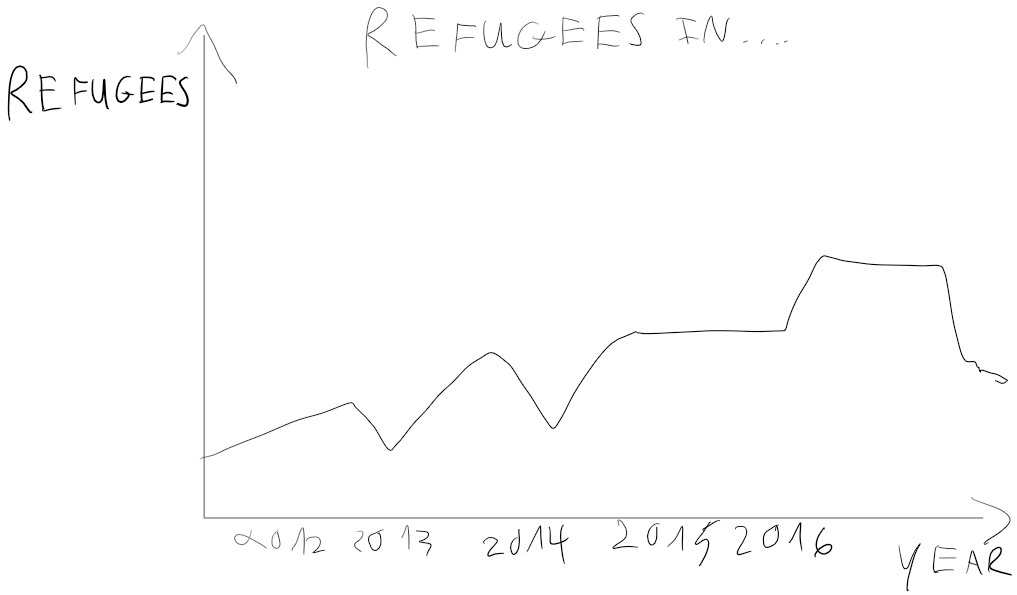


Figure 20: Draft design number of refugees over year in a particular country area chart

Also, we have altered the navigation bar. on that navbar, we can go around the pages.

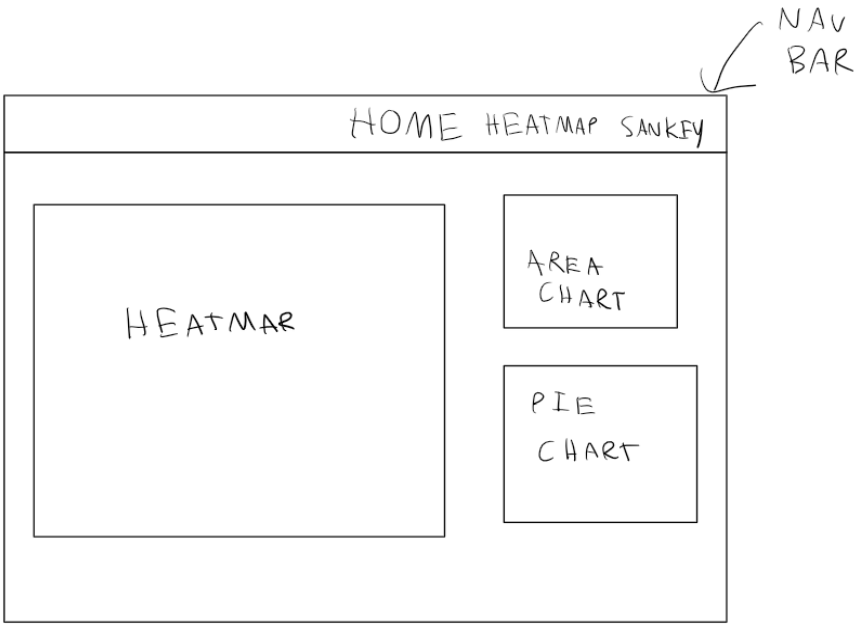


Figure 21: Final navigation bar

After discussion, we came up with a final visualization website (see Figure 22 and Figure 23). We separate the visualization into two pages, the first page contains the heat maps, while the second page contains the Sankey diagrams. The Sankey diagrams we have come up with the top 3 countries that have been migrated to. And the heatmaps in the first pages have a little change, the line graph beside will be used to see the trends of the refugees from the selected country migrate to the other country, while the pie chart shows the percentage of gender in the refugees. This chart looks promising to answer lots of questions such as “how many percent of men migrate from Australia to other countries around the world?”
The overall design will be a little bit different. We are also working on the list that can change the colour from the heat maps that will be able to the colour blind people to see the maps but also for the normal people to clearly see the data that we represent

This is the final layout of the heatmap that we are preparing:

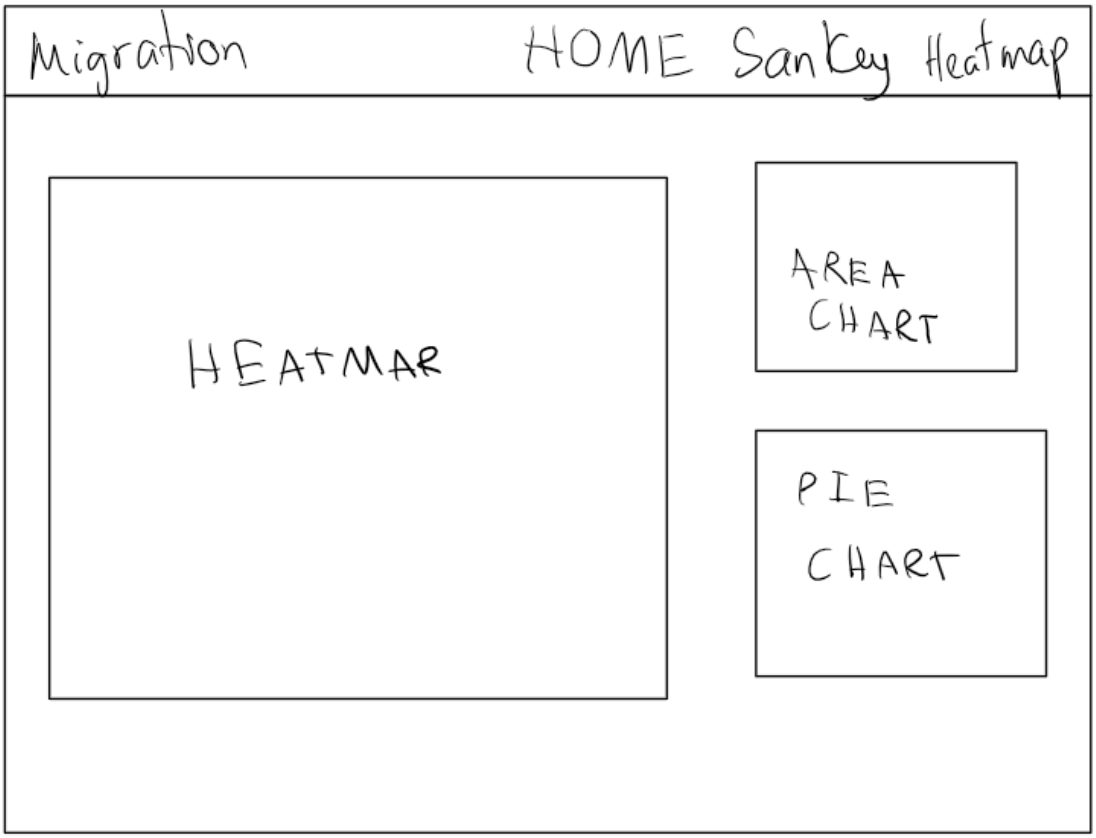


Figure 22 : Final Heatmap page

Here is the final Sankey diagram page:

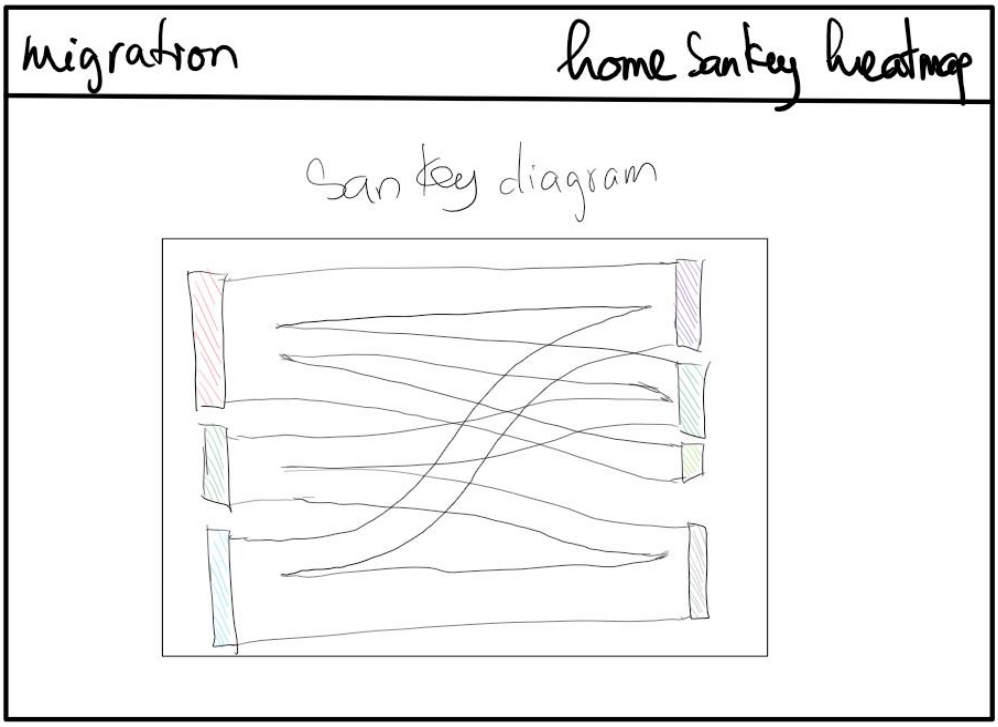


Figure 23: Final sankey website layout

4.3. Final Design

4.3.1. Design Overview

Our final visualization design embodies the principles discussed in the conceptualization part, resulting in two main visualizations: an interactive heatmap and a Sankey diagram.

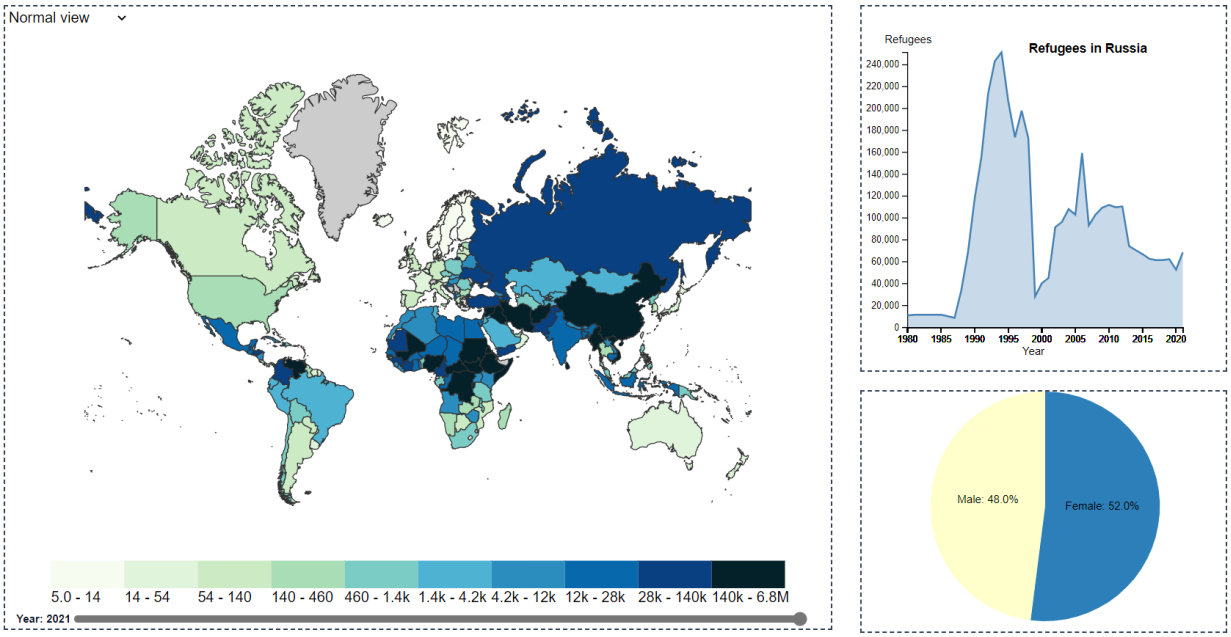


Figure 24: Final visualization design: Interactive heatmap

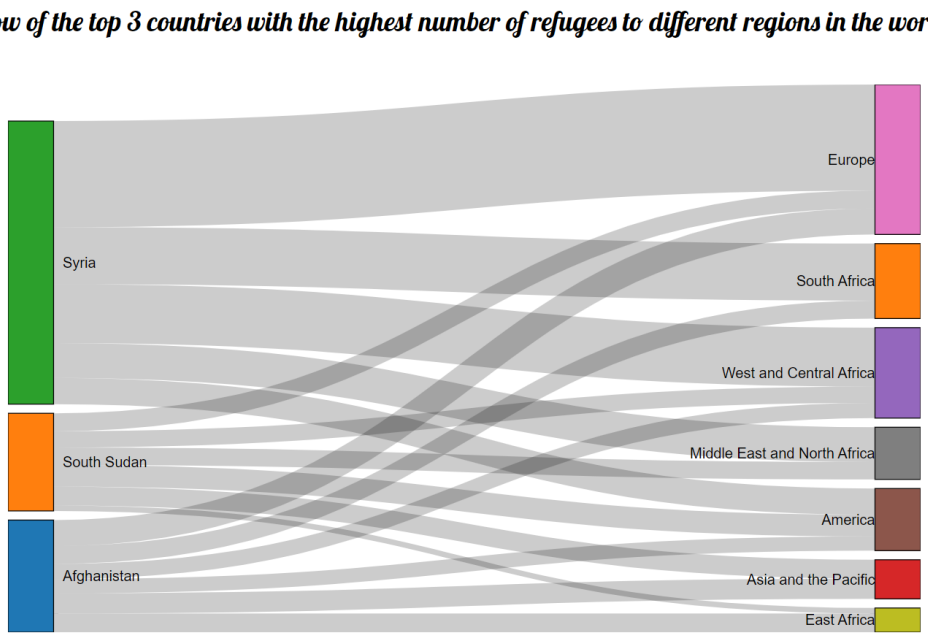


Figure 25: Final visualization design: Sankey diagram

4.3.2. Design Description

The interactive heatmap serves as the central element of our data visualization website. The number of refugees for a specific country is treated as quantitative data and encoded utilizing color saturation. The level of color intensity reflects the data capacity, with the greater the intensity of the colour, the greater the refugee population. To ensure inclusivity and accessibility, we have implemented two ranges of color saturations to accommodate both individuals with normal color vision and those with color blindness.

The first color saturation is designed for people with normal color vision, using a carefully selected color palette that effectively portrays the varying magnitudes of refugee numbers across different countries. The color gradient captures the nuances of the data, allowing users to discern the differences in refugee populations at a glance.

Figure 26: Normal color saturation

In consideration of individuals with color blindness, we have incorporated a second color scale that employs a color scheme specifically tailored to be distinguishable for those with common types of color vision deficiency. This ensures that all users, regardless of their color perception abilities, can accurately interpret and derive insights from the heatmap.

Figure 27: Color-blind color saturation

By incorporating two color saturations, we prioritize inclusivity and ensure that our visualization is accessible to a wider audience. This design choice reflects our commitment to effective communication and data representation, allowing users to explore and understand the refugee data in a comprehensive and meaningful way. The color view of the heatmap can

be easily customized by selecting different options from a dropdown menu located at the top of the visualization (see Figure 28 and 29)

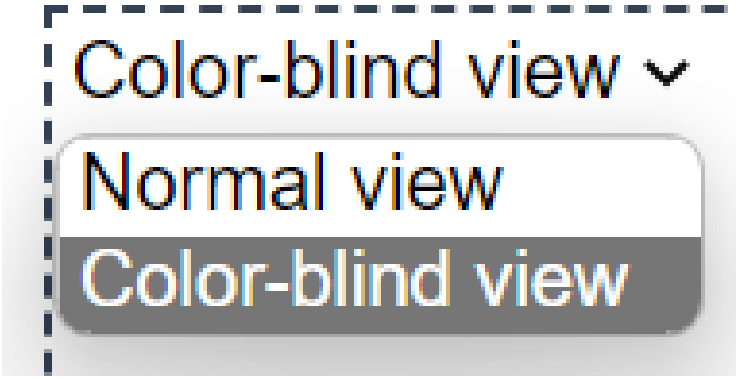


Figure 28: Dropdown to change view

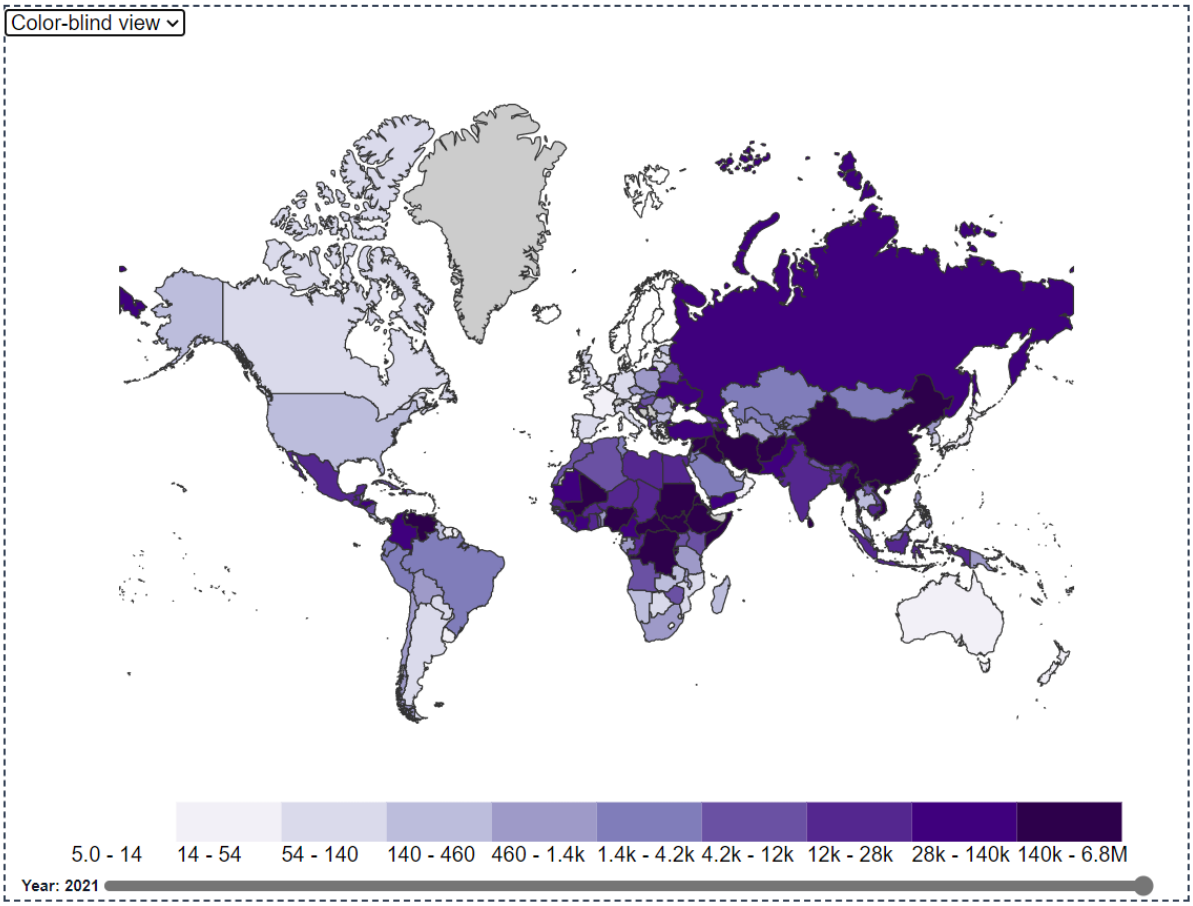


Figure 29: Color blind map

In instances where data is unavailable or missing, the design employs a thoughtful approach to maintain visual coherence and communicate the absence of data. Specifically, when there is no available data for a particular country, it is encoded with a subtle grey color. This deliberate choice ensures that users can differentiate between countries with available data and those without, while maintaining a consistent visual representation across the heatmap.

Figure 30: Null data

The heatmap incorporates a wide range of interactive features to enhance the user experience and facilitate data exploration. When users hover over a country, a dynamic tooltip emerges, providing essential information such as the country's name, the number of refugees migrating from that country, and the corresponding year. This immediate feedback allows users to quickly gather insights about specific countries and their refugee populations

Figure 31: Tooltips when showing on Russia

To provide a comprehensive view of the migration patterns, an area chart is positioned to the right of the heatmap. When users move their mouse cursor over a country, the area chart dynamically updates to display the migration trend of refugees from that particular nation over the years. This visual representation allows users to discern patterns, trends, and significant changes in refugee movements over time, facilitating a deeper understanding of the data.

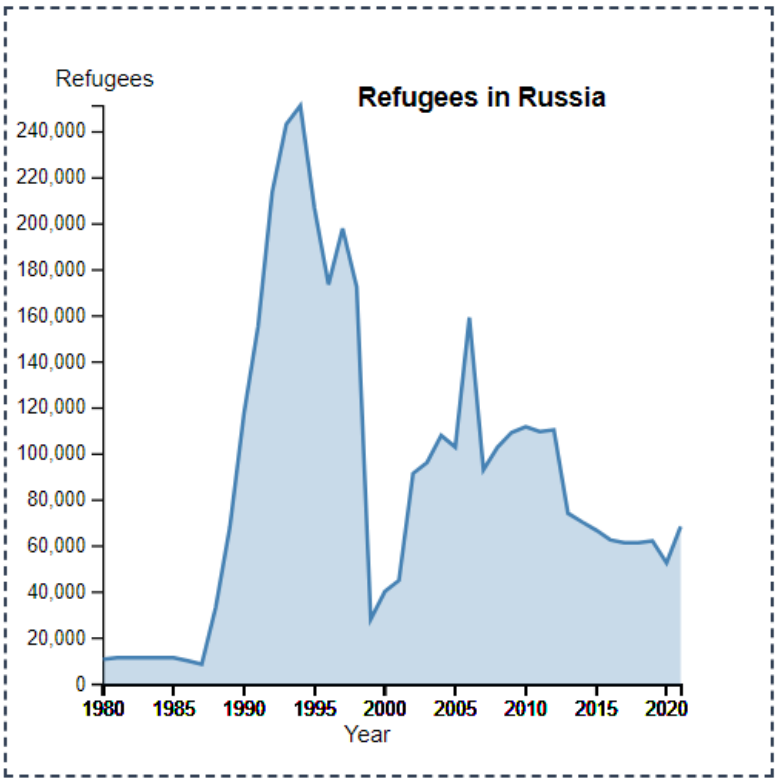


Figure 32: Refugees in Russia (area chart)

Additionally, when users hover over a country, a pie chart appears, illustrating the proportion of male and female refugees within that country. This interactive feature enables users to gain insights into the gender composition of refugee populations, enabling comparisons and analysis across different countries. By visually depicting this information, users can identify gender disparities or trends that may exist within the data.

The pie chart visualization in our design focuses on depicting the proportion of male and female refugees, utilizing the available data attributes. However, it is important to note that the dataset does not include specific attributes related to other sexual orientations, such as LGBTQ.

While we acknowledge the importance of representing diverse identities and experiences within the refugee population, our visualization is limited by the dataset. The absence of explicit attributes related to LGBTQ individuals in the dataset restricts our ability to provide a comprehensive representation of sexual orientation diversity within the refugee population. In future iterations or with access to more comprehensive data, we could explore the inclusion of additional attributes that capture the diversity of sexual orientations within the refugee population. This would enable us to provide a more inclusive and representative visualization that acknowledges the experiences of various marginalized groups.

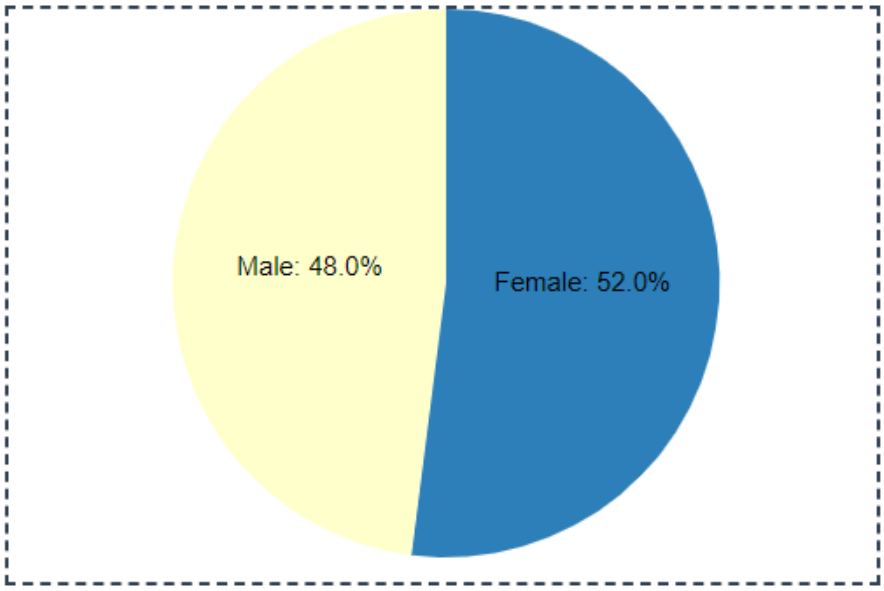


Figure 33: The proportion of gender among refugees come from Russia

Last but not least, the heatmap includes a year slider located at the bottom of the visualization. This interactive feature allows users to change the displayed data on the heatmap based on the selected year. By dragging the slider along the timeline, users can seamlessly explore how refugee migration patterns evolve over time.

Figure 34: Year Slider feature

On the other hand, the Sankey diagram is employed to showcase the migration flow from the top three countries with the highest number of refugees: Afghanistan, South Sudan, and Syria, to various regions across the world. This visualization effectively communicates the paths taken by refugees from their origin countries to their respective destinations. By hovering over a node or a link in the diagram, users are presented with informative tooltips that provide specific details about the number of refugees associated with that particular node or link. Additionally, an interactive feature has been implemented for the nodes, allowing users to drag them up and down within the Sankey diagram.

Figure 35: Tooltip for the link in the Sankey Diagram

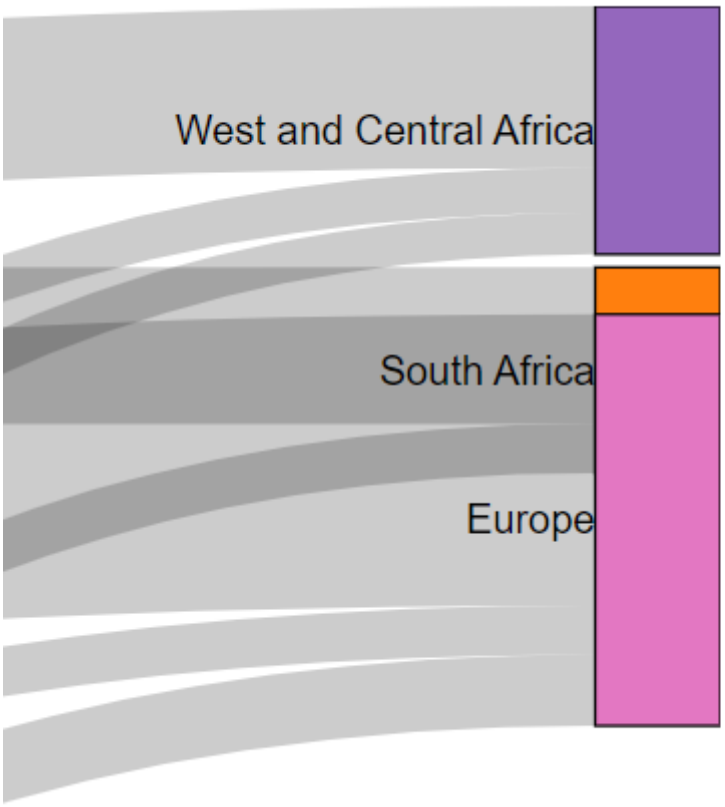


Figure 36 : Dragging feature of Sankey Diagram

4.3.3. Design Critique

The final design adheres to the design principles outlined by Munzner and Tufte, ensuring the efficiency and integrity of the graph. For the heatmap, Munzner's principles guided the selection of suitable visual encodings to represent the number of refugees migrating from different countries. Color hue and saturation were employed effectively to encode the quantitative data, with darker shades indicating higher numbers of refugees. In the sankey diagram, Munzner's principles were also applied by carefully selecting visual encodings that best represent the migration flows, such as using the thickness of the Sankey links or nodes to indicate the number of refugees. The color in both visualizations has been carefully selected to accommodate color-blind individuals, enhancing their usability.

Furthermore, the incorporation of interactive features, such as tooltips and hover effects, enhances user engagement and provides additional information. The inclusion of a dropdown menu for color selection in the heatmap allows users to customize their viewing experience, promoting personalization. Interactive features, such as tooltips and node dragging, are also available in the sankey diagram, further contributing to its interactivity and user-friendliness. Additionally, the design demonstrates a commitment to maintaining graphical integrity by consciously avoiding unjustified 3D visualizations. By opting for 2D representations, we want to ensure clarity and minimize potential confusion arising from issues such as scale distortion (Tufte, 1983) , poor depth judgment, and perspective distortion (Munzner, 2015).

Last but not least, our team has diligently worked to optimize the data-to-ink ratio and maximize data density. Unnecessary visual components and "chart junk" have been

purposefully eliminated from the visualizations (Tufte, 1983). This deliberate simplification allows users to focus on the essential data elements and facilitates their ability to extract meaningful insights efficiently. By reducing visual distractions, the design enhances the clarity and readability of the visualizations, enabling users to quickly grasp and interpret the information without any unnecessary cognitive load.

5. Validation

5.1. Introduction

To validate our data visualization website, we conducted an user study involving a diverse group of participants. After their interaction with our website, we requested their valuable feedback by asking them to complete a validation form. The feedback form was thoughtfully designed to capture their insights, opinions, and suggestions regarding various aspects of our visualizations. Some topics that are asked inside the validation form include:

- Part 1: User Information
- Part 2: Overview
- Part 3: Visualization 1: Interactive Heatmap
- Part 4: Visualization 2 : Sankey Diagram

The complete user validation form can be found in the Appendix of our process book. Alternatively, you can access the form directly through this link: https://docs.google.com/forms/d/e/1FAIpQLSel0MCkbS77JiRL8N4o0zgilG3zB4bjS4xFwv7PXBM2v59cFg/viewform?usp=sf_link

5.2. Testing

Part 1: User Information

In this part, we ask the user some very basic information about themselves including age, English and computer proficiency and whether they are color blind or not. This helped us understand the diverse perspectives and backgrounds of our users.

Part 2: Overview

This section focused on capturing the participants' impressions of the overall design of the visualizations. We asked them to provide feedback on the clarity, aesthetics, and overall appeal of the visualizations.

Part 3: Visualization 1: Interactive Heatmap

In this section, we specifically addressed the interactive heatmap. Participants were asked to evaluate the effectiveness of the heatmap in representing the number of refugees and their migration patterns.

Part 4: Visualization 2 : Sankey Diagram

This section centered around the sankey diagram. Participants were prompted to assess the sankey diagram's ability to illustrate migration flows from different countries to various regions.

5.3. Result

All the results of the questions and the responses from the feedback form will be included in the Appendix section of our process book.

Part 1: User information

The users who accessed our website fall within the age range of 12 to 55 years old, with a notable concentration in the 20 to 22 age group (see Figure).

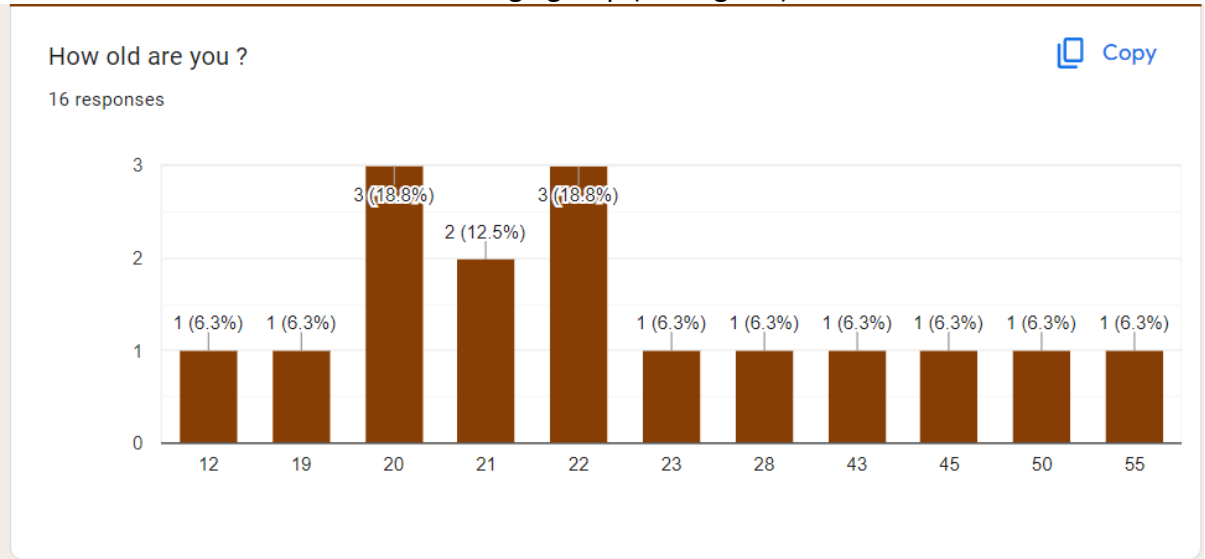


Figure 37: User age

The majority of users who accessed our website demonstrate a high level of English proficiency and possess advanced computer usage skills (see Figure . This proficiency in English enables users to easily navigate through the website and comprehend the information presented. Their high computer usage skills indicate a familiarity with technology and an ability to interact effectively with the website's features and functionalities. By considering the users' language proficiency and computer skills, we have tailored the design and user experience to meet their expectations and ensure a smooth and engaging interaction with the website.

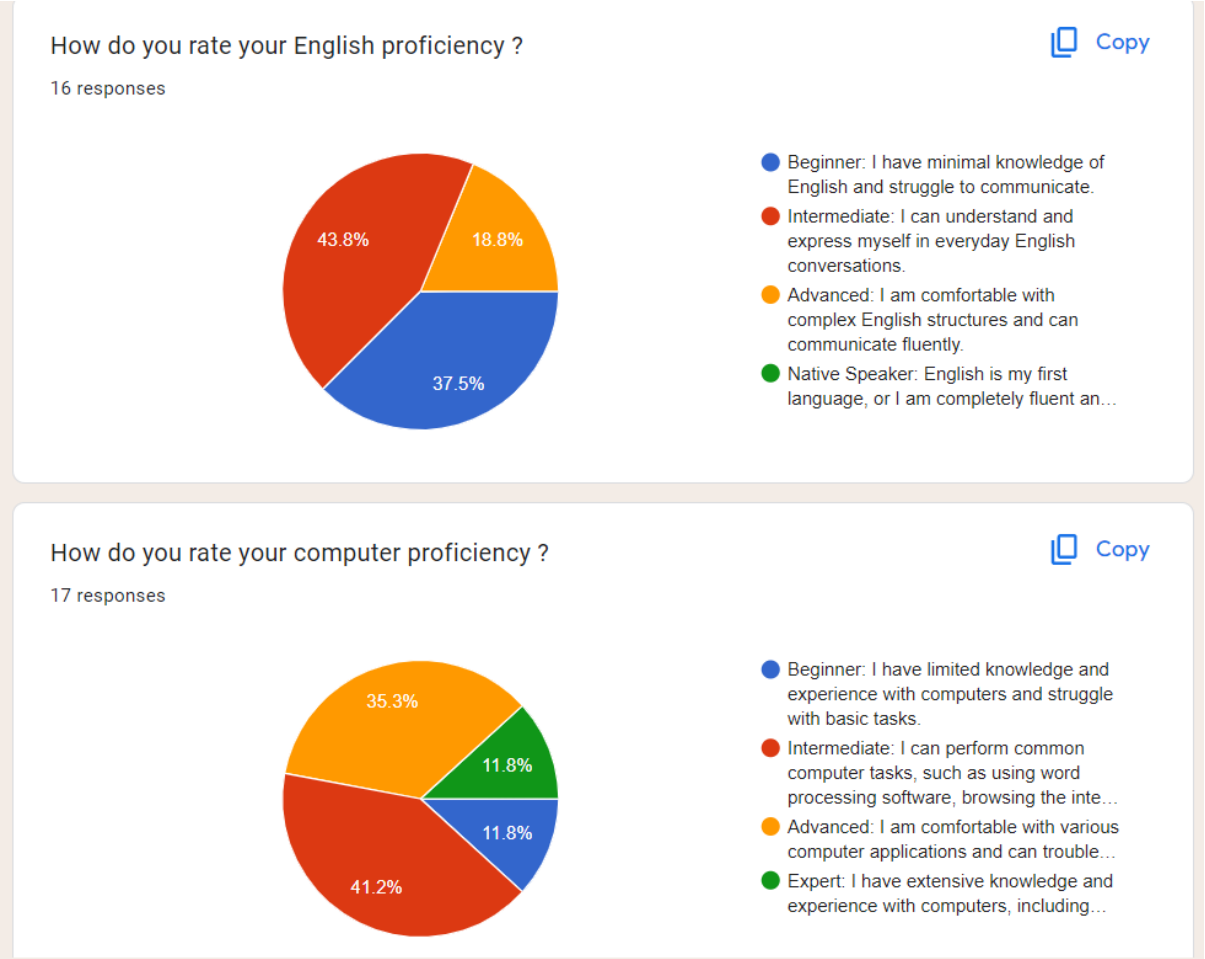


Figure 38: User English lever and Computer Proficiency

Approximately 18.8% of our users are color-blind (see Figure), which indicates a significant portion of our audience. It is crucial for us to prioritize their needs and ensure that our visualizations are accessible and meaningful to them.

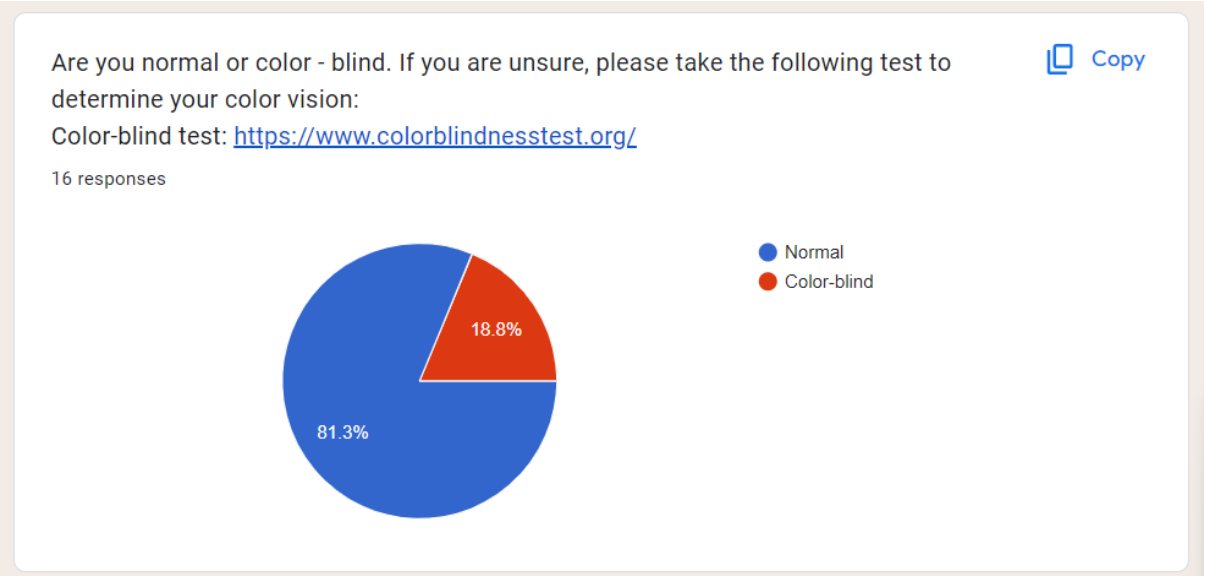


Figure 39: User vision question

Part 2: Overview

We have received a lot of positive feedback and ratings for both the Interactive Heatmap and the Sankey Diagram. The majority of users expressed satisfaction with the overall design of the visualizations. However, it is worth noting that one user provided a below-average rating of 2 out of 5 for the Interactive Heatmap. This feedback should be taken into consideration to help us improve the design process in our future project.

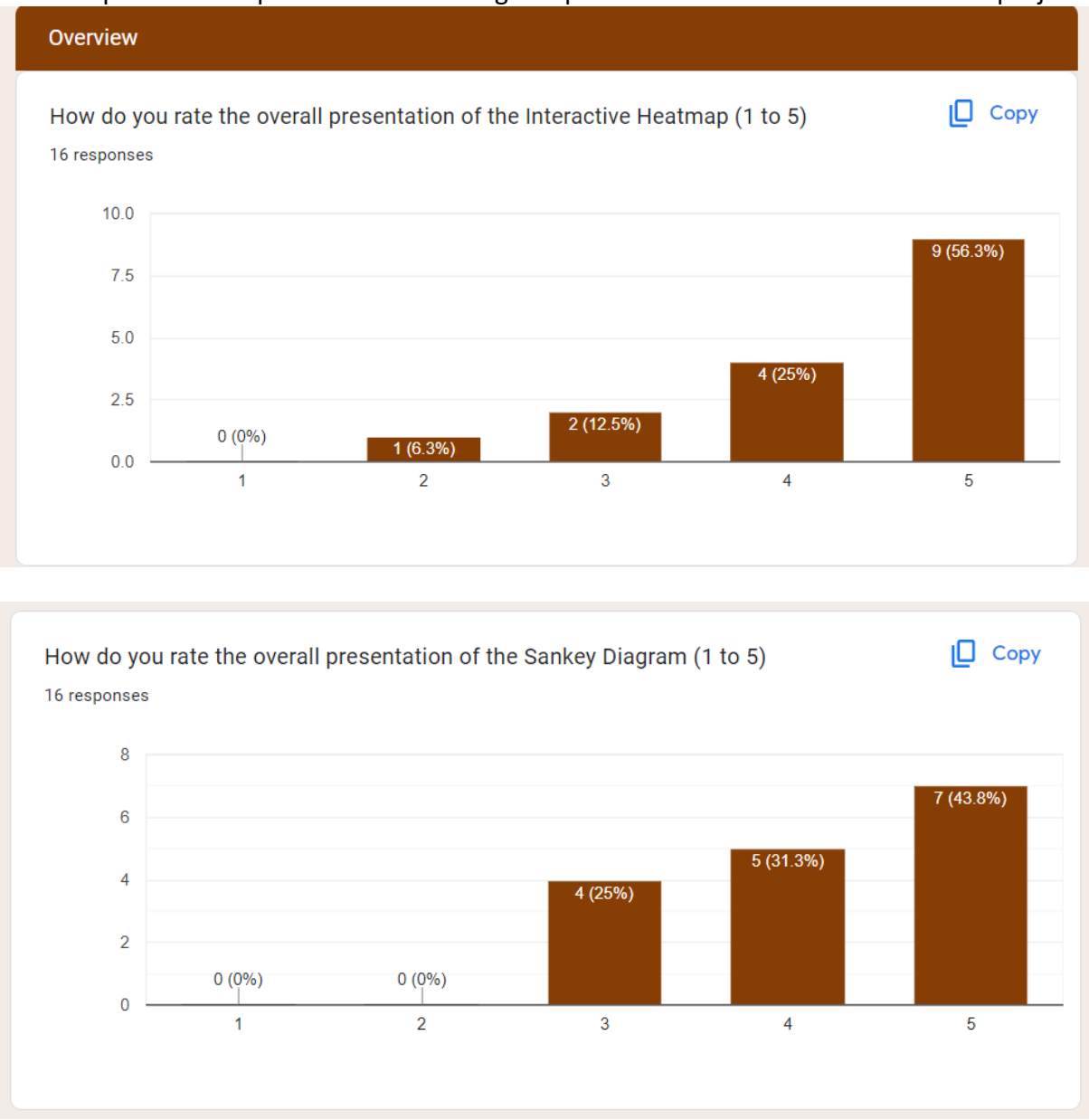


Figure 40: Rating overview

Part 3: Visualization 1: Interactive Heatmap

In the feedback received, the majority of users demonstrated a good understanding and accurate interpretation of the information presented through the heatmap and area chart. Their responses reflected a high level of comprehension and engagement with these visualizations. However, when it came to the demographic question that involved the usage of the pie chart, the correct answer rate was only 62.5% (see Figure . This suggests that there may be room for improvement in the clarity or effectiveness of the pie chart in conveying the demographic information in the pie chart.

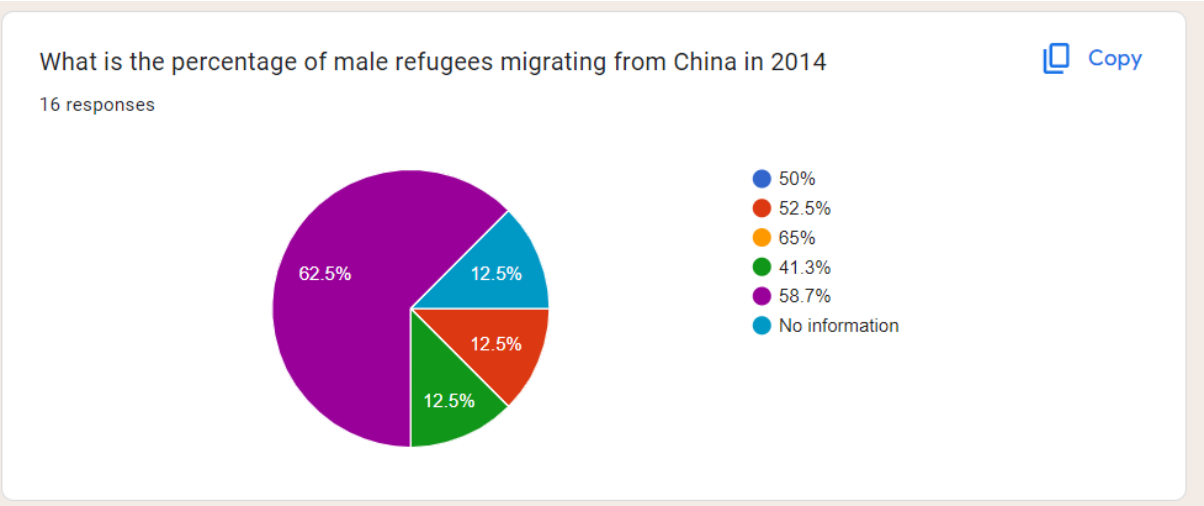


Figure 41: Demographic data answer

Part 4: Visualization 2: Sankey Diagram

For the Sankey Diagram, it is worth noting that all users successfully interacted with the diagram by dragging the nodes and links (see Figure). This indicates a high level of intuitiveness and usability in the design of the interactive features. The ability to effortlessly manipulate the nodes and links demonstrates the effectiveness of the interface in facilitating user interaction and engagement.

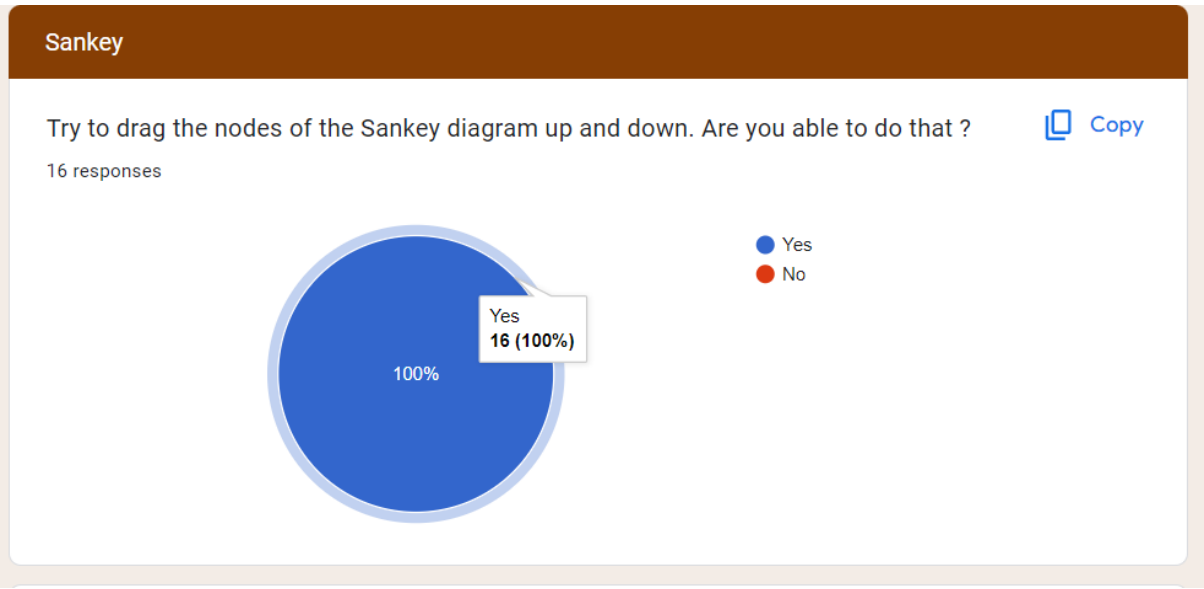


Figure 42: Draggable sankey feature task

However, during user testing, it was observed that the tooltips for the node such as "Syria" in the Sankey Diagram did not function as expected. Approximately 25% of the users reported difficulties in obtaining accurate information about the total number of refugees in Syria due to this tooltip issue.

5.4. Recommendation

To enhance the Interactive Heatmap, we can introduce an interactive feature to the pie chart. When users hover over different segments of the pie chart, the corresponding segment can be highlighted with a tooltip, providing additional information or data specific to that segment. This interactive feature will facilitate a deeper understanding of the data represented in the pie chart and enable users to explore specific details with ease.

In addition, to further improve the usability of the Interactive Heatmap, we have included a user manual on the web page of the visualization. This user manual serves as a comprehensive guide, providing instructions and explanations on how to interact with the visualization, interpret the data, and utilize the available features. By incorporating a user manual, we aim to provide users with clear guidance and ensure they can fully utilize the capabilities of the Interactive Heatmap.

Similarly, for the Sankey diagram, we can enhance the tooltip for the nodes by making it larger to assist users in better understanding the migration patterns and enable them to extract meaningful insights more easily.

6. Conclusion

In conclusion, our process book outlines the journey of our data visualization project, highlighting the key steps, decisions, and considerations involved in creating effective and impactful visualizations. We started by understanding the project's objectives, audience, and data, which guided our conceptualization phase.

Drawing inspiration from renowned visualization experts such as Edward Tufte and Tamara Munzner, we applied their design principles and guidelines to ensure clarity, effectiveness, and meaningful communication of the data. We prioritized graphical integrity, data density, data-ink ratio, appropriate visual encodings, and interactive features to enhance user engagement and understanding.

Throughout the design process, we iteratively refined our visualizations, incorporating feedback and making thoughtful decisions about color choices, layout, chart types, and interactivity. We employed techniques such as layering, separation, and careful selection of visual elements to convey information efficiently and avoid visual clutter.

The final design reflects our efforts to create engaging and informative visualizations. The heatmap effectively represents the number of refugees migrating from different countries, leveraging color hue and saturation to communicate quantitative data. The interactive Sankey diagram visually demonstrates the flow of migrants, providing insights into internal migration patterns.

We also paid attention to the details of interactivity, incorporating tooltips, area charts, and pie charts to provide users with additional information and context when interacting with the visualizations. The use of a year slider in the heatmap allows users to explore the data across different time periods, enhancing their understanding of migration trends over time.

To summarize, our process book showcases a systematic and thoughtful approach to data visualization, incorporating design principles, visualization techniques, and interactive features. It demonstrates our commitment to creating visually compelling and informative visualizations that enable users to gain insights and develop a deeper understanding of the data. Through this process, we have learned the importance of user-centric design, effective data communication, and the power of visualization to tell compelling stories

Appendix

A. User Validation Form Question

Part 1: User information

User Information

We want to know more about you in this section

How old are you ?

Văn bản câu trả lời ngắn

How do you rate your English proficiency ?

- Beginner: I have minimal knowledge of English and struggle to communicate.
- Intermediate: I can understand and express myself in everyday English conversations.
- Advanced: I am comfortable with complex English structures and can communicate fluently.
- Native Speaker: English is my first language, or I am completely fluent and proficient

How do you rate your computer proficiency ?

- Beginner: I have limited knowledge and experience with computers and struggle with basic tasks.
- Intermediate: I can perform common computer tasks, such as using word processing software, browsing...
- Advanced: I am comfortable with various computer applications and can troubleshoot basic technical is...
- Expert: I have extensive knowledge and experience with computers, including advanced software usage, ...

Are you normal or color - blind. If you are unsure, please take the following test to determine your color vision:
Color-blind test: https://www.colorblindnesstest.org/

- Normal
- Color-blind

Part 2: Overview

Overview

Here are the pictures of the data visualization website. Please have a comprehensive look and answer some question:
Here is the link to the website:
https://datavisualization-github-io.vercel.app/src/home.html

How do you rate the overall presentation of the Interactive Heatmap (1 to 5)

	1	2	3	4	5	
Lowest	○	○	○	○	○	Highest

How do you rate the overall presentation of the Sankey Diagram (1 to 5)

	1	2	3	4	5	
Lowest	○	○	○	○	○	Highest

Please provide your comments and thoughts on the Interactive Heatmap and Sankey Diagram

Long-answer text

Part 3: Visualization 1: Interactive Heatmap

Interactive Heatmap

Please utilize the Interactive Heatmap to answer the following questions and please rate the difficulty level of the task.

How many refugees migrating from Russia in 2020 ?

Short-answer text

Is it straightforward for you to find and answer the previous question ?

- Yes
- No

What is the period of year that the number of refugees migrating from Canada reachs the maximum number?

Short-answer text

Is it straightforward for you to find and answer the previous question ?

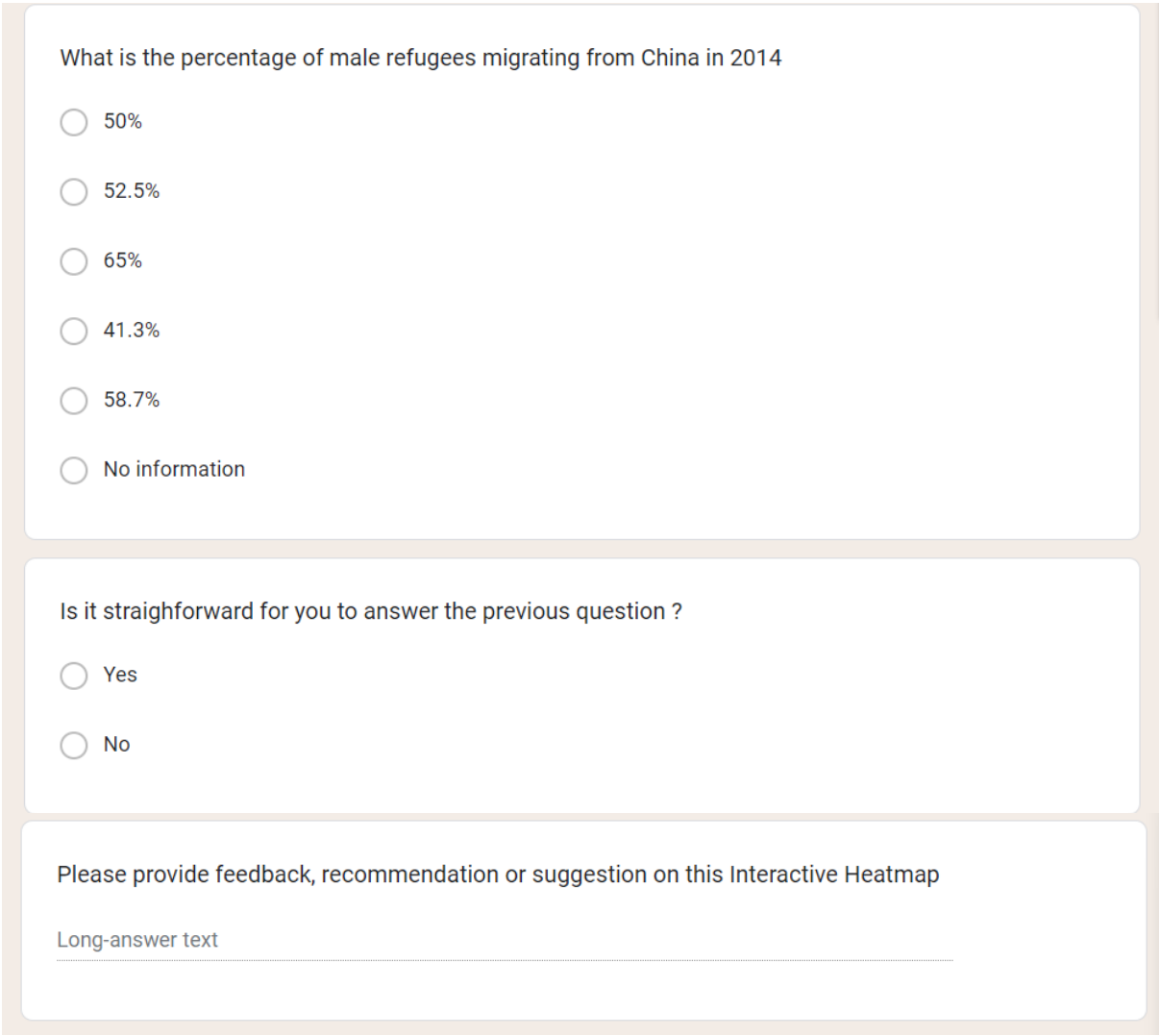
- Yes
- No

Provide the name of a country that has the number of refugees over 140000 in 2021

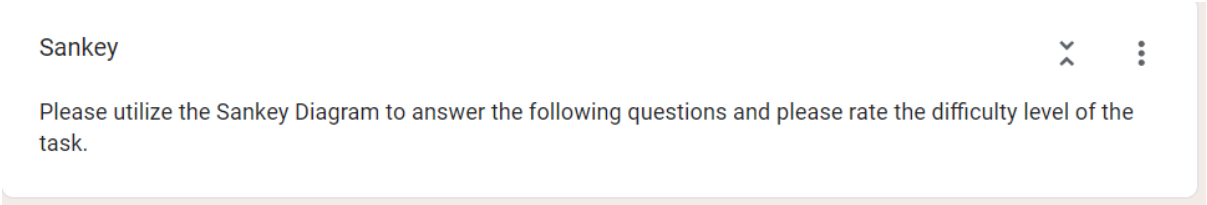
Short-answer text

Is it straightforward for you to find and answer the previous question ?

- Yes
- No



Part 4: Visualization 2: Sankey Diagram



Try to drag the nodes of the Sankey diagram up and down. Are you able to do that ?

- ◯ Yes
- ◯ No

Please rate the ease of doing the task (from 1 to 5)

	1	2	3	4	5	
very difficult	◯	◯	◯	◯	◯	very easy

How many refugees are there in total in Syria

Short-answer text

Is it straight forward for you to answer the previous question

- ◯ Yes
- ◯ No

How many refugees migrating from South Sudan to Europe

Short-answer text

Is it straight forward for you to answer the previous question ?

- ◯ Yes
- ◯ No

Please provide feedback, recommendation or suggestion on this Sankey Diagram

Long-answer text

B. User Validation Form Result

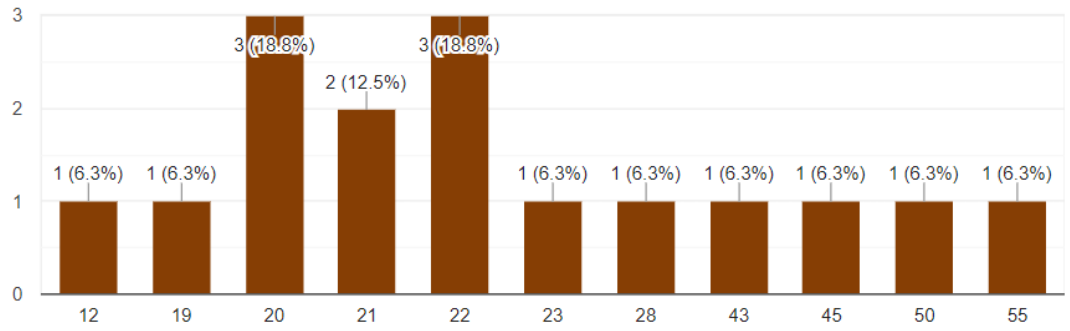
Part 1: User information

User Information

How old are you ?

Copy

16 responses



How do you rate your English proficiency ?

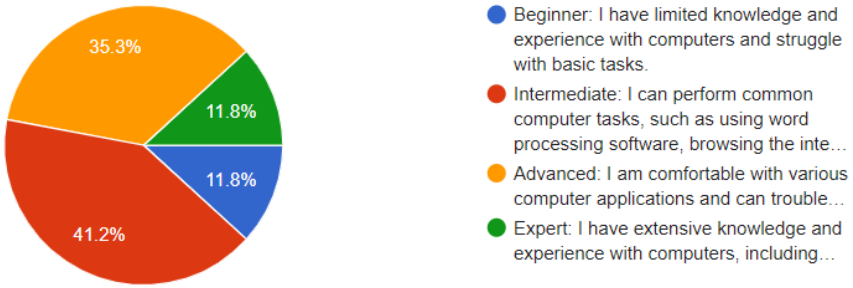
Copy

16 responses

How do you rate your computer proficiency ?

Copy

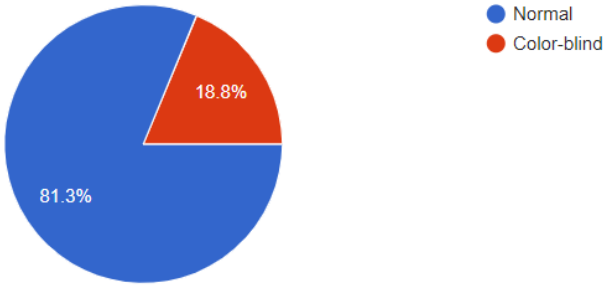
17 responses



Are you normal or color - blind. If you are unsure, please take the following test to determine your color vision:
Color-blind test: https://www.colorblindnesstest.org/

Copy

16 responses

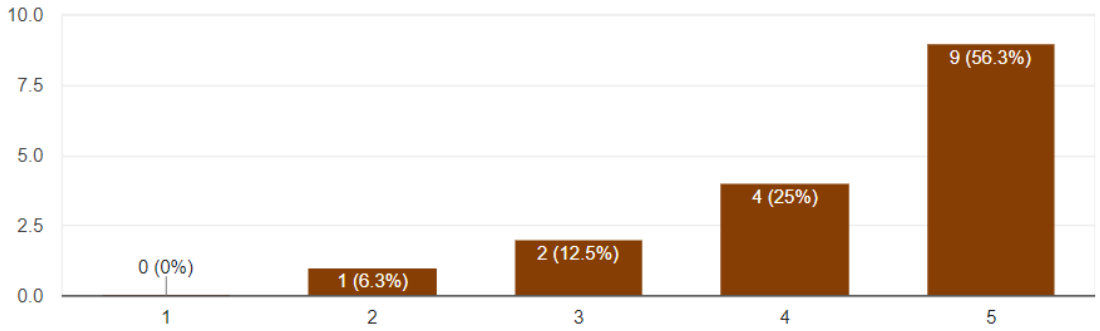


Part 2: Overview

How do you rate the overall presentation of the Interactive Heatmap (1 to 5)

Copy

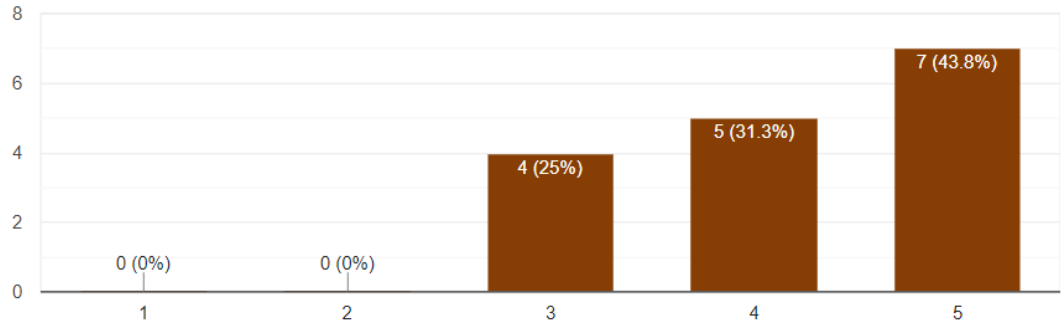
16 responses



How do you rate the overall presentation of the Sankey Diagram (1 to 5)

Copy

16 responses



Please provide your comments and thoughts on the Interactive Heatmap and Sankey Diagram

11 responses

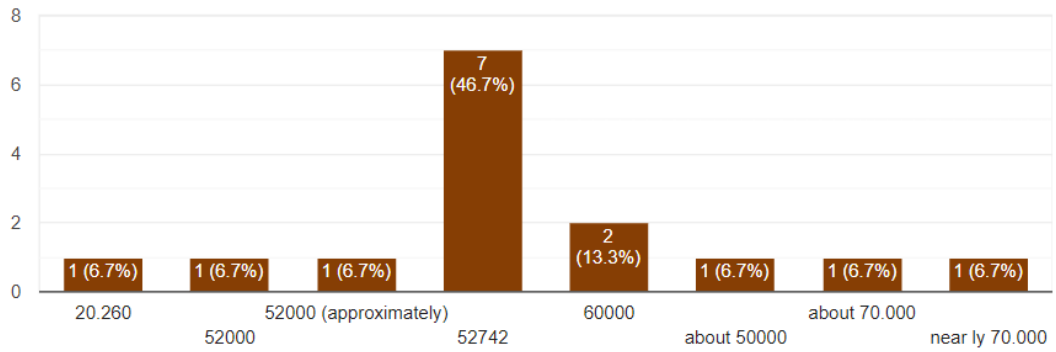
- The overall presentation is good and well- designed
- No comments, overall good layout
- Overall no need to change
- They interesting
- Good illustrations and diagrams. Useful information and data provided.
- Heatmap: Interactive, good
  Sankey: Good
- Overall good, no comments
- I think the heat map is detailed. The sankey is not as detailed
- Exceptional visualization. Easy to understand
- excellent work. I can easily get the information, the trend and compare among the country. the web visual data is friendly and usefully.
- Good

Part 3: Visualization 1: Interactive Heatmap

How many refugees migrating from Russia in 2020 ?

Copy

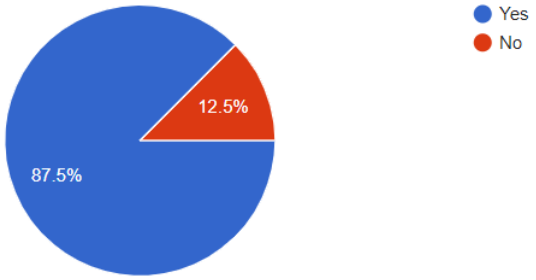
15 responses



Is it straightforward for you to find and answer the previous question ?

Copy

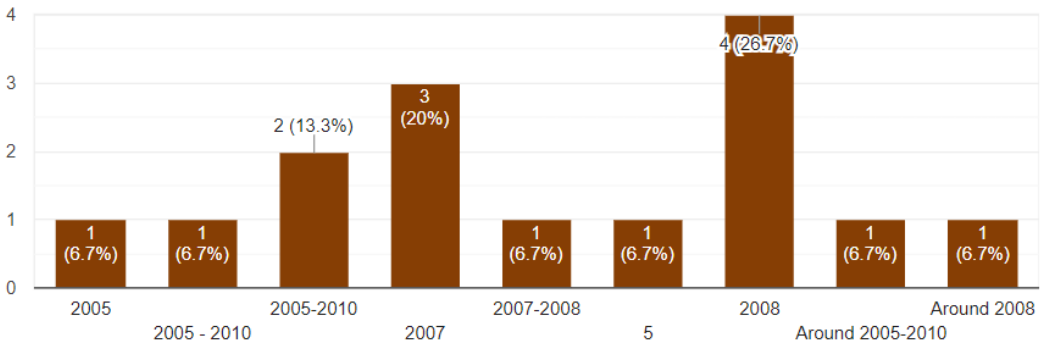
16 responses



What is the period of year that the number of refugees migrating from Canada reachs the maximum number?

Copy

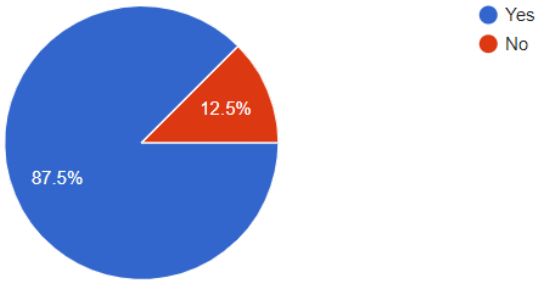
15 responses



Is it straightforward for you to find and answer the previous question ?

Copy

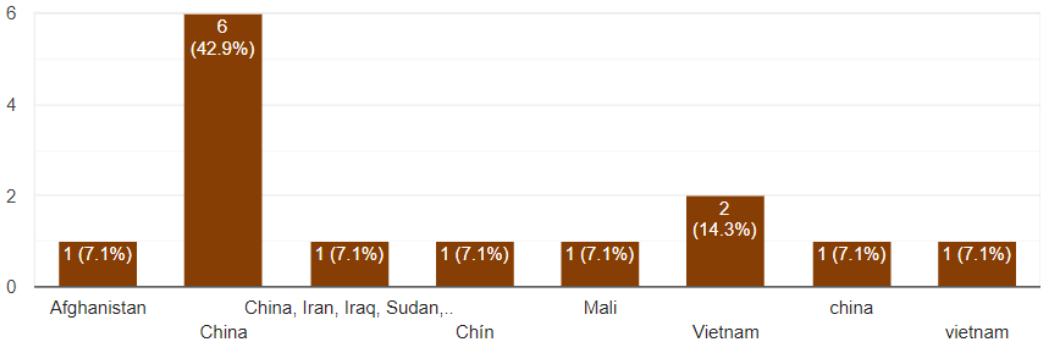
16 responses



Provide the name of a country that has the number of refugees over 140000 in 2021

Copy

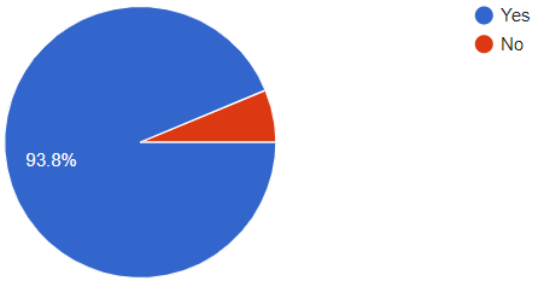
14 responses



Is it straightforward for you to find and answer the previous question ?

Copy

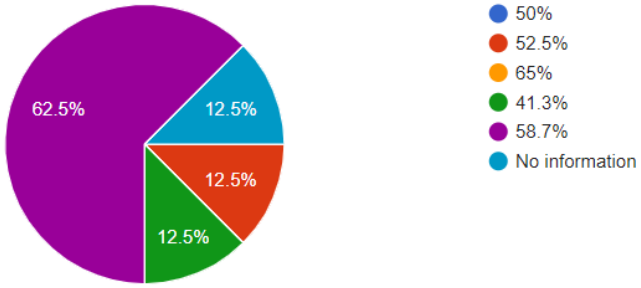
16 responses



What is the percentage of male refugees migrating from China in 2014

Copy

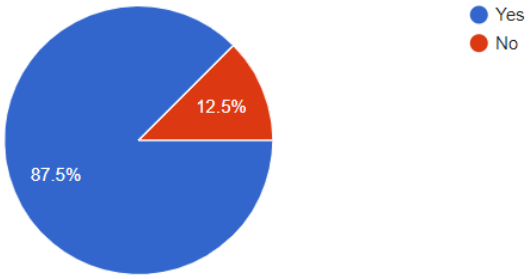
16 responses



Is it straighforward for you to answer the previous question ?

Copy

16 responses



Please provide feedback, recommendation or suggestion on this Interactive Heatmap

11 responses

- good interactive
- good
- They are pretty good, no comment
- Some functions might be added such as: Viewer can see number of regugees in a specific year by cursing the mouse to the chart,... But this is informative enough.
- No comment
- Interactive, color blind can still see
- very easy to find the information and data from the Interactive Heatmap
- The line graph is hard to use because it is hard to see which year the data is from
- Good visualization with great information

The line graph is hard to use because it is hard to see which year the data is from

Good visualization with great information

Clear and easily to get the information

Good

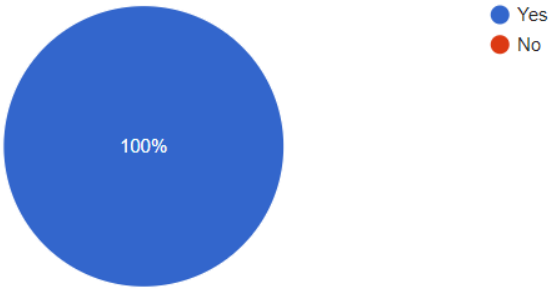
Part 4: Visualization 2: Sankey Diagram

Sankey

Try to drag the nodes of the Sankey diagram up and down. Are you able to do that ?

Copy

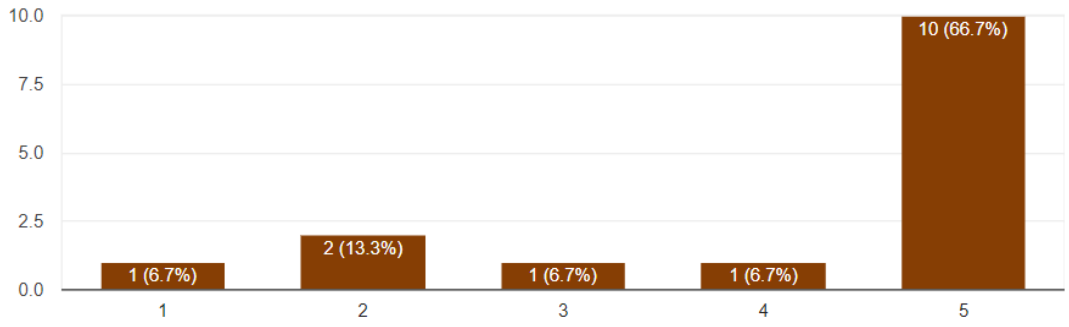
16 responses



Please rate the ease of doing the task (from 1 to 5)

Copy

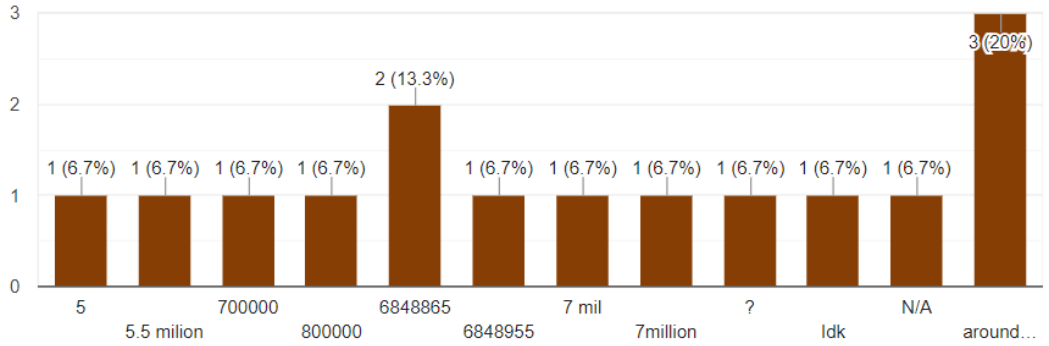
15 responses



How many refugees are there in total in Syria

Copy

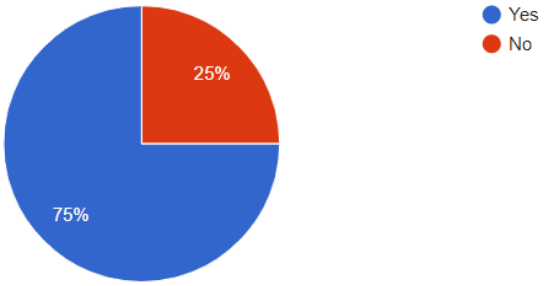
15 responses



Is it straight forward for you to answer the previous question

Copy

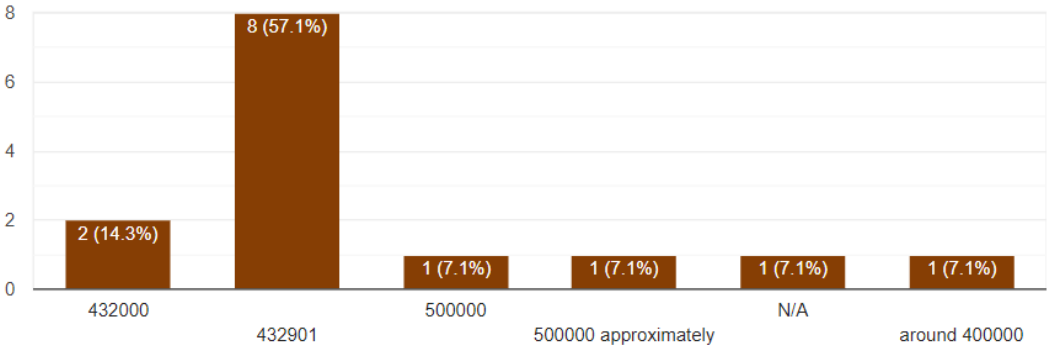
16 responses



How many refugees migrating from South Sudan to Europe

Copy

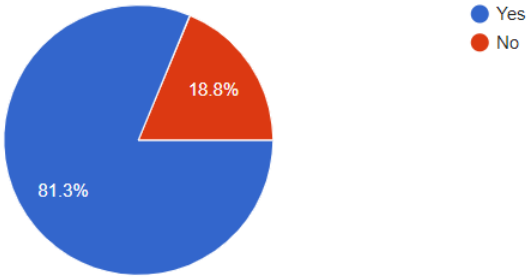
14 responses



Is it straight forward for you to answer the previous question ?

Copy

16 responses



Please provide feedback, recommendation or suggestion on this Sankey Diagram

11 responses

- no rec
- All good
- This is good and it's easy to find information.
- no comment
- No comment
- I cannot see the number of refugees per node, just the edges
- great visualization
- well done
- Overall good

Reference

The UN Refugee Agency. (n.d.). Available at: https://www.unhcr.org/global-trends [Accessed 27 May 2023].

Munzner, T. (2015). *Visualization analysis & design*. Boca Raton: Crc Press, Taylor & Francis Group.

Tufte, E.R. (1983). *The visual display of quantitative information*. Cheshire, Conn.: Graphics Press.