**SWINBURNE UNIVERSITY OF TECHNOLOGY**

**COS30045**

**PROCESS BOOK**

|  |  |  |
| --- | --- | --- |
| **NAME & STUDENTID** | **:** | **Nguyen Gia Binh - 104219428**  **Nguyen Quang Dao - 104211864** |
|  |  |  |
| **CLASS** | **:** | **COS30045** |
| **LECTURER**    **TUTOR** | **:**    **:** | **HOANG XUAN TUNG**  **(**txhoang@swin.edu.au**)**  **HOANG XUAN TUNG**  **(**txhoang@swin.edu.au**)** |

**MERCURY LINK :** [**Visualization Link**](https://mercury.swin.edu.au/cos30045/s104219428/Main/index.html)

**Hanoi, April 1, 2024**

**TABLE OF CONTENTS**

TABLE OF CONTENTS

**CHAPTER 1: INTRODUCTION**

1.1: Background and Motivation

1.2: Visualization purpose

1.3: Project Schedule

**CHAPTER 2: DATA**

2.1: Data source

2.2: Data processing

**CHAPTER 3: REQUIREMENT**

3.1: Must-have Features

3.2: Optional Features

**CHAPTER 4: VISUALIZATION DESIGN**

4.1: GEO MAP

4.2: LINE CHART

4.3: STACKED BAR CHART

**CHAPTER 5: VALIDATION**

**CHAPTER 6: CONCLUSION**

**REFFERENCE**

**CHAPTER 1: INTRODUCTION**

**1.1: Background and motivation**

Migration is a natural and inevitable aspect of human history. In the current globalized world, international migration has become more common. In 2020 alone, there were an estimated 281 million international migrants worldwide, comprising 3.6 percent of the global population (UN DESA, 2020). There are both push factors like conflict, persecution and human rights violations that force people to flee, as well as pull factors like better economic opportunities and living conditions that attract migrants.

The situation of forced migration due to the Russian invasion of Ukraine has led to a significant humanitarian crisis, affecting millions of people. As of February 2024, nearly 3.7 million individuals within Ukraine have been displaced internally, highlighting the extensive disruption and displacement caused by ongoing military actions and violence. Additionally, the crisis has forced nearly 6.5 million people to seek refuge in other countries, with Poland hosting the largest number of Ukrainian refugees, approximately 60% of the total. The international community, through various organizations such as UNHCR and UNICEF, has been actively involved in providing humanitarian assistance, including legal aid, mental health support, and basic needs to those affected. More than 14 million people have been displaced since the onset of the conflict, with a substantial number finding refuge in neighboring countries. This displacement represents a significant portion of Ukraine's pre-war population, underscoring the profound impact of the conflict on the country's demographic landscape (UNHCR,2024).

Understanding these complex forced migration dynamics is crucial for policies and international relations efforts to resolve the core Invasion and address the refugee crisis humanely based on international law (UNRWA, 2021).

**1.2: Visualization purpose**

In this assignment, several visualizations are made to explore and analyze geographic patterns and underlying events that contribute to complex issues. By synthesizing location-based data, the visualizations offer potential benefits. The topic we chose is the invasion of Russia in Ukraine. Using interactive map and other data visualizations allows viewers to delve into what's the number of refugees in a certain European country and the internal displacement Demographic of Ukraine. First, the line chart depicts the trend of migration coming from Ukraine each month since just before the war up to now to show how devastating the situation is. Next, the map clearly shows exactly how many Ukrainian refugees a country in Europe has as of the current date. Finally, two stacked bar charts are deployed to give viewers a look into the Ukraine IDP demographic, comparing 2022 and 2024. By combining the geographic map with other visual elements like charts and graphs, you get a comprehensive, multi-faceted picture of the intricate migration and displacement flow before and after the invasion of Russia in Ukraine.

**1.3: Project Schedule**

|  |  |
| --- | --- |
| Week | Tasks |
| 6 | 1. Choose the topic  2. Find data about it |
| 7 | 1. Choose the topic  2. Find data about it |
| 8 | 1. Finalize the topic  2. Clean the data |
| 9 | 1. Start making the demo charts  2. Start sketching the front-end to present the visaulization |
| 10 | 1. Choose what kind of visualization you want to use  2. Visualize the EU map  3. Code the charts out for Russia vs Ukraine  4. Finalize on front-end design |
| 11 | 1. Sorting through the Demo code to finalize the design  2. Bug fixing |
| 12 | 1. Bug fixing  2. Writing the report  3. Conclude the project |

**CHAPTER 2: DATA**

**2.1: Data source**

This analysis on the invasion of Russia in Ukraine utilizes data from a variety of credible sources: IOM (International Organization for Migration), CEDOS (Center for Economic Development Studies), CEEOL (Central and Eastern European Online Library), Humanitarian Data Exchange(HDX) and EuroStat. We'll be meticulously combining and processing this information to create insightful visualizations. Each source brings unique strengths: IOM provides expertise in migration patterns, CEDOS offers insights into Eastern European development, CEEOL grants access to relevant academic research, HDX covers a wide range of categories relevant to humanitarian emergencies and crises, such as population statistics, administrative boundaries, displaced populations,....... and EuroStat supplies official statistics. This multi-faceted approach ensures we leverage the most comprehensive and up-to-date information to illuminate the topic at hand.

Here are the URL to the dataset respective to our visualization:

Line chart:

* <https://data.humdata.org/dataset/ukraine-data-explorer-inputs?>
* <https://cedos.org.ua/en/researches/forced-migration-and-the-war-in-ukraine-february-24-march-24/>
* <https://stat.gov.ua/en/datasets/migration-movement-population>
* <https://cream-migration.org/ukraine-detail.htm?article=3573>

Geomap :

* The geomap itself : <https://github.com/amcharts/amcharts4/blob/master/dist/geodata/es2015/json/region/world/europeUltra.json>
* [https://en.wikipedia.org/wiki/Ukrainian\_refugee\_crisis\_(2022%E2%80%93present)#:~:text=An%20 ongoing%20refugee%20cross%20 began,country%20by%20late%20May%202022.](https://en.wikipedia.org/wiki/Ukrainian_refugee_crisis_(2022%E2%80%93present)#:~:text=An%20%20ongoing%20refugee%20cross%20%20began,country%20by%20late%20May%202022.)
* The data to incorporate with the map: <https://data.unhcr.org/en/situations/ukraine>

Stacked bar chart:

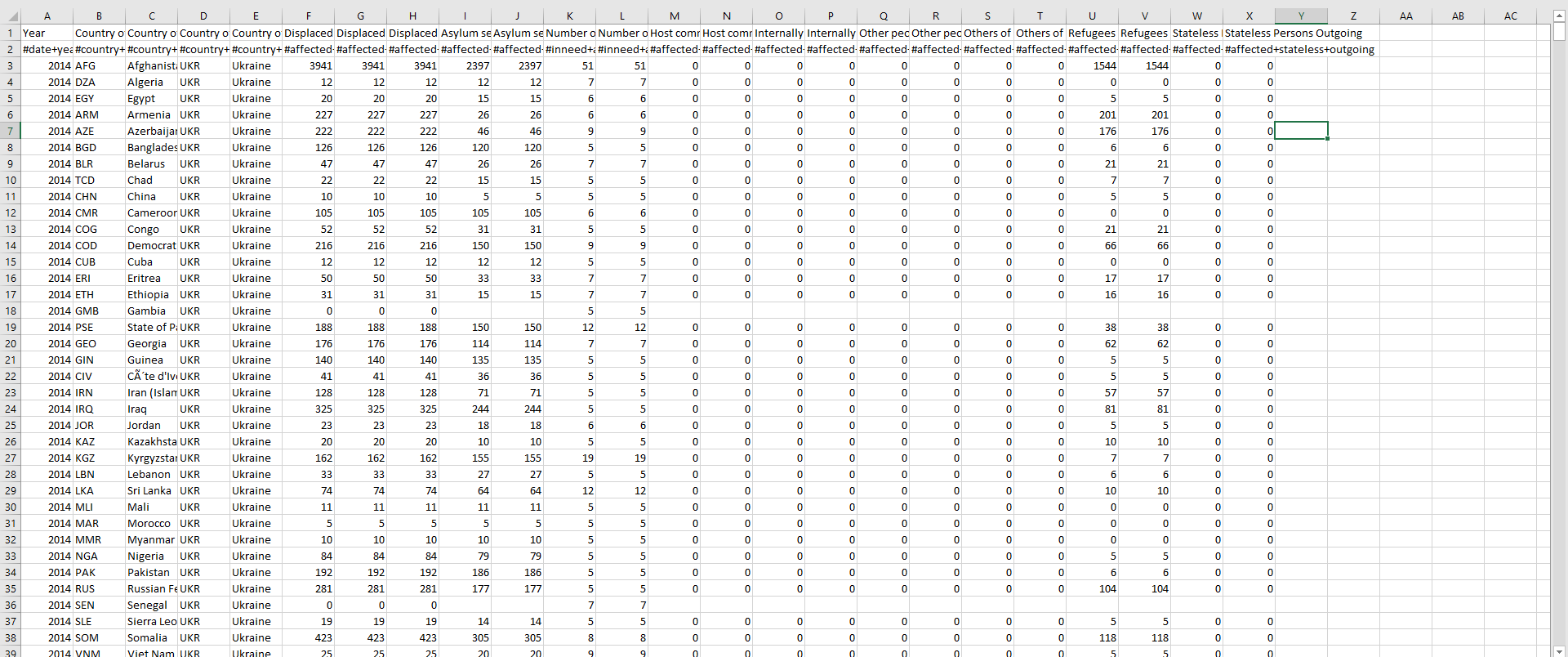
* 2022: <https://dtm.iom.int/reports/ukraine-internal-displacement-report-general-population-survey-round-11-25-november-5>
* 2024:
* <https://dtm.iom.int/reports/ukraine-area-baseline-assessment-raion-level-round-33-february-2024>

**2.2: Data processing**

The datasets obtained from the aforementioned organizations guarantee both high integrity and timeliness. However, it is essential to conduct data cleaning and restructuring on the original datasets to ensure they meet the necessary level of 'conformity' required for visualization in D3.

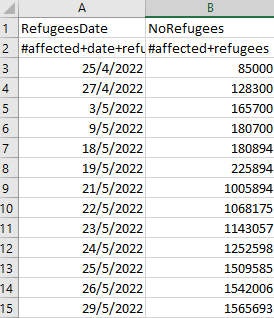
**Line Chart**

For the line chart, the datasets are a little problematic as no one site has the full dataset, each has just a part of it or some set of the dataset is missing so data imputation is a must in order to have the required data for the chart. The original dataset is comprised of daily,monthly figure of ukraine migration number and/or quarterly figure.

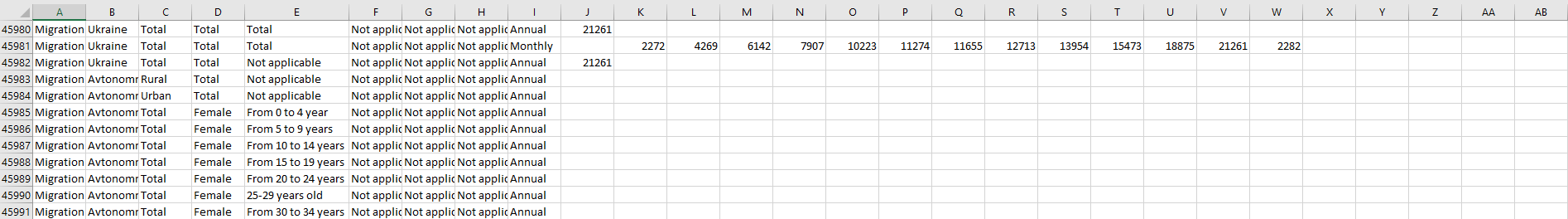


Our main focus is the A,D column as well as the U column (the V column is just a duplicate of the U column) as it tells us which year it is, what country this is from and what is the refugee number. The rest Is pruned to simplify the data structure, improving readability, and also enhance the performance in data processing later on. This table has a significant flaw that is it doesn't have the month date respective to the input of the number. To account for the flaw i use another dataset from [HDX](https://data.humdata.org/dataset/ukraine-data-explorer-inputs?) namely the “**Ukraine Data Explorer - refugees series**” table which has precise date record date to it for cross checking

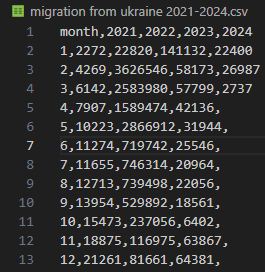
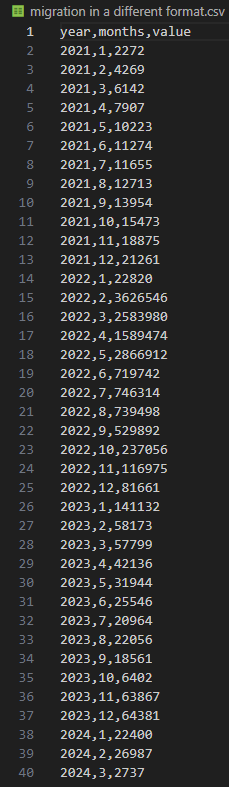
Example:



This table also has its own flaw, it does not have the data from January to April of 2022 so data imputation is required again to fill in. This time it is coming from the [Ukraine government](https://stat.gov.ua/en/datasets/migration-movement-population)



In this dataset, they have the migration number in monthly format and from Jan 2021 to Jan 2022. Since it is coming from the Ukraine government, there is no need for proof checking 2021 data and most importantly January of 2022. The only missing data points left is from 24 Feb 2022 (the day Russia start invading) - end of April 2022. According to [CEEOL](https://www.ceeol.com/search/viewpdf?id=1207439)there are 6,210,526 people migrate from Ukraine in the first two month of the war with 3,626,546 people in the first few day which make March has the number of 2,583,980 and since UNHCR reported roughly 7.7 millions at the end of April, April now has the value of 1,489,474.

That is not the end as we still need to calculate for the rest of the monthly number in “**Ukraine Data Explorer - refugees series**”. It is quite simple as i only have to take the (latest number of a month - the earliest number of a month) to get the final number of refugees. The final calculated value is then written into 2 separate csv files for more convenience during the line chart coding session which will be talked about more in chapter 4.

**Geomap**

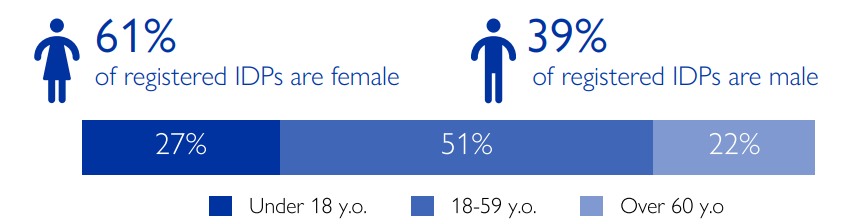


In this geomap, our dataset is quite simple and not complicated for coders. Firstly, we found a lot of sources on google and all of those have a similar number so I decided to choose data from Wikipedia that I think is the most famous page to do research and have a newest update about my project( end of 2023). This table only has two columns that is Name and Number. Name is presented for the name of each country, while number is shown the number of Ukrainian refugees to the EU.

**Stacked Bar Chart**

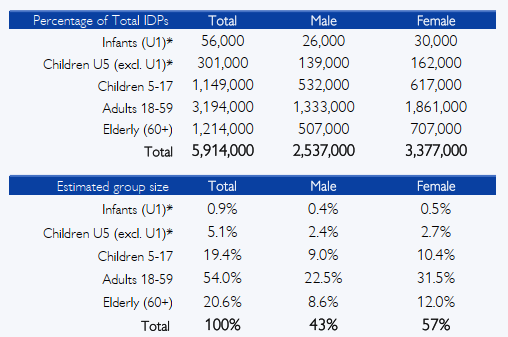
This chart is about Ukraine internally displaced persons (IDP), and the most cite report we can find are “REGISTERED IDP AREA BASELINE ASSESSMENT” and “UKRAINE INTERNAL DISPLACEMENT REPORT GENERAL POPULATION”, both are from IOM and are done by round. For our purpose, we choose the use the round that is conduct at the end of 2022 and February 2024(latest). There are multiple section in the reports but what we need is in the Demographic section which is constant in both type of report.

**REGISTERED IDP AREA BASELINE ASSESSMENT Round 33**:



In the report, IOM recorded 3,413,472 IDP so that make the algorithm used to calculate the exact percentage is (3,413,472 / 100) \* “target”.

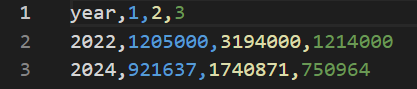
**UKRAINE INTERNAL DISPLACEMENT REPORT GENERAL POPULATION Round 11:**

****

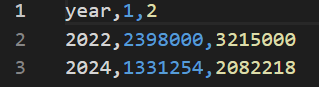
In this report, the only number we have to calculate is the total percentage and recorded number of “Children” category as the report split it into 3 groups.

Finally, the number is put into 2 csv file with the same structure. The type like Child, Adult, Elder is map into 1,2,3 category as Javascript just seem to read it better than text format. The same goes for gender with Male and Female are map to 1 and 2.

age\_group.csv:



gender.csv:



**CHAPTER 3: REQUIREMENT**

**3.1 Must-Have Features**

There are a few features that we consider to be very important to give the reader an easier time understanding of what the chart is for. These features include:

· Interactive visualization:

a) Hover effect: With hover effect, user can see the detail in which a point on the map or a section of a chart represent.

b) Zoom in: This effect allow the user to see the timeline closer especially in line chart or area chart

c) Clickable: Point on the map or sections of the chart may be clickable and would lead to a sub html page with smaller chart or found report representing that point or section

· Additional information:

a) Label: This is a must be to help the reader quickly indentify What the reader is seeing.

· Color: data can be categorize in dozens of type and column, the contrast between color can make some important element to shine through and get to the reader.

**3.2 Optional Features**

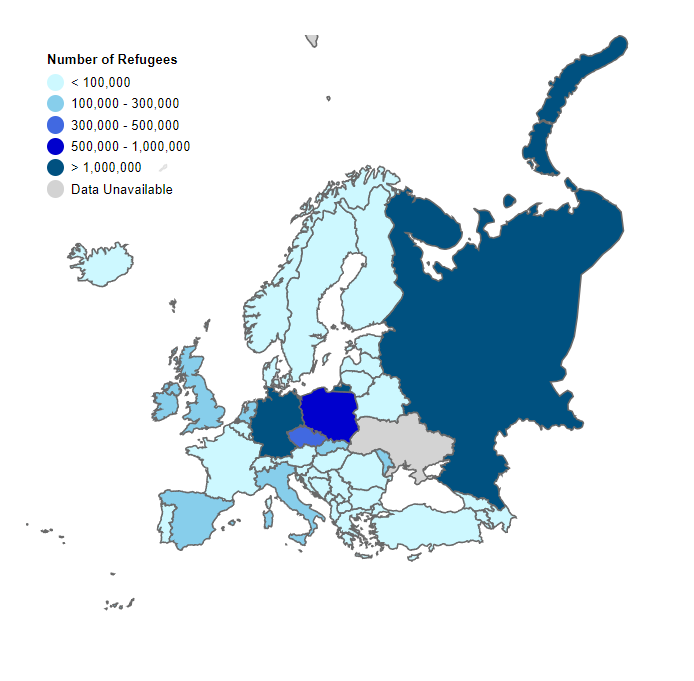
d) Tooltip: in the first design, sub pages would be use to give a more detail view on the points or section of the data but if we use tool tip, there would need to be a sub page and therefore maintaining the page integrity would be easier but at the same time tooltip is relatively hard to implement effectively and easily in my experience.

e) Guide: There should be a guide either by a button or render out next to the chart to give the reader an easier time indentify each section.

**CHAPTER 4: VISUALIZATION DESIGN**

The final version of our visualization can be access through Mercury :[**Mercury**](https://mercury.swin.edu.au/cos30045/s104219428/Main/index.html)or through our Github repository: [**GitHub**](https://github.com/nguyengiabinh/COS30045-Group)

Since the early stage of the visualisation design process, we aim to provide a simple, yet interactive and/or intuitive layout that would suit our targeted audiences which is the general public and not for experts.

**4.1 Geomap**

In this part, we decide to make a geomap that shows how many Ukrainian refugees are in the EU. For the concept, we try to make the EU map for the first look by using the europe.json file that we found on google. Secondly, we try to combine between the database and the colour of the map as a picture below. Finally, we add some functions to help viewers understand more about the geomap. For example, when viewers hover on each country, it will show the number of Ukrainian refugees to this country and it also changes the colour.

Conclusion:  
Step 1: create a EU map

Step 2: link dataset with the map

Step 4: link color value to number of refugee in the dataset

Step 5: link them both to the map

Step 6: check and debug the final result

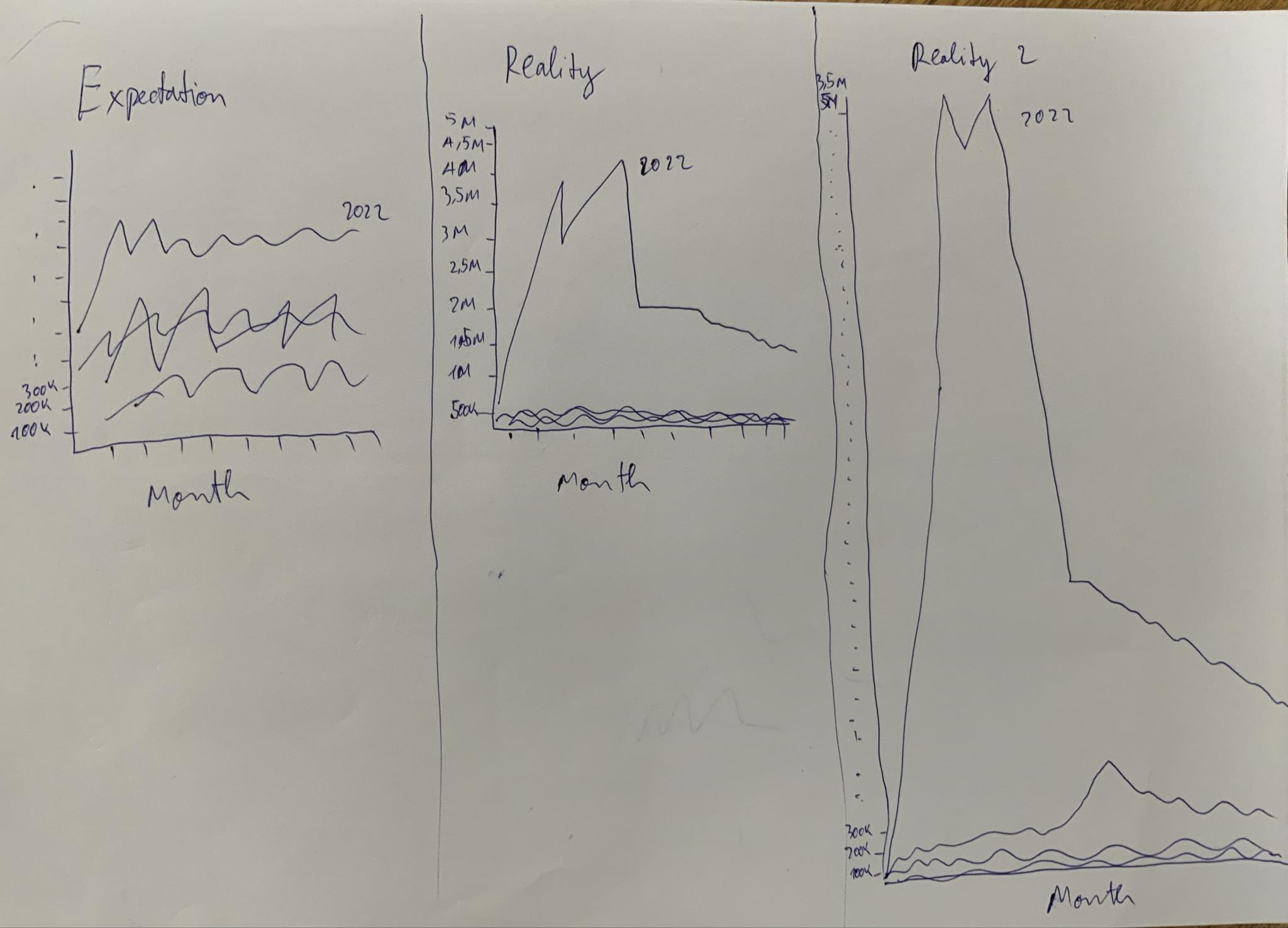
**4.2 Line Chart**

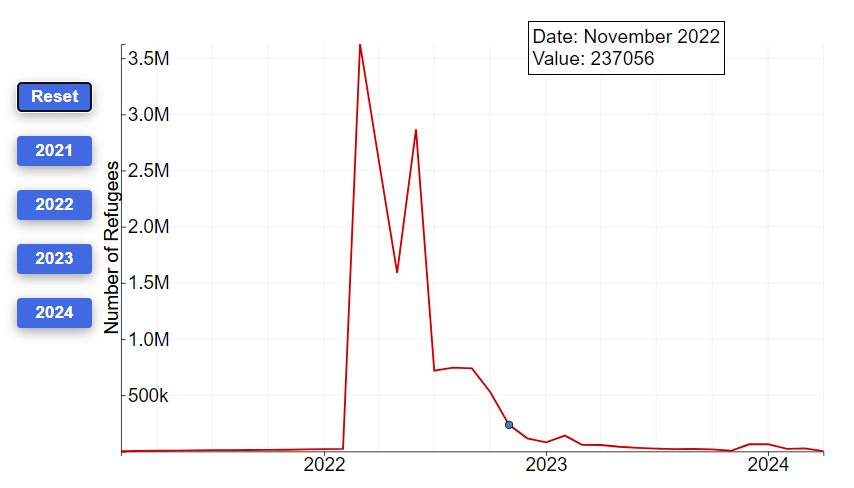
The general idea of this line chart visualization is to reveal the trend of migration from Ukraine to countries in Europe territory, viewer should be able to view different line representing different year, particularly 2021,2022,2023 and 2024. The general design is consist of:

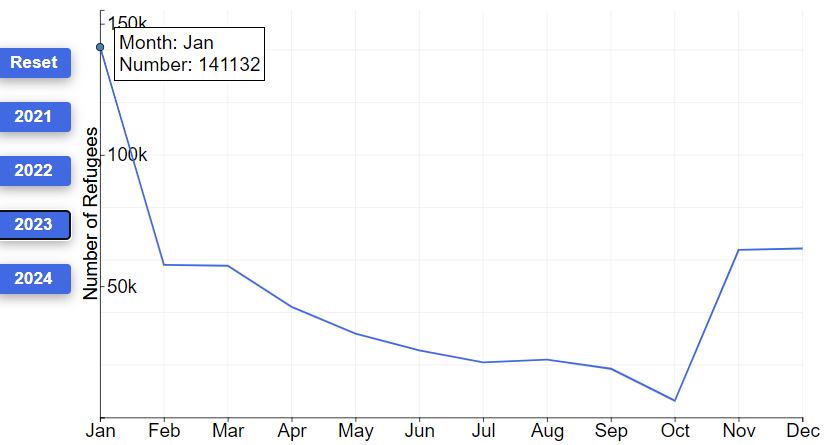
* X-axis for month or year
* Y-axis for the value of each datapoint that mek up the line
* Line: Same color when group in one chart and the line of 2022 is a different color since that is the year Russia start the invasion
* Tooltip: the tooltip will follow viewer cursor and display the information of the closest datapoint. What display is consist of what time of the year it is or what month this is as well as the data value correspond with that datapoint

In the early stage of designing the line chart, i expected the line to be beautifully intertwined with each other, different color for different year and hover effect which will fade out everything else except for the line in which the viewer are hovering over. The idea was there but the expectation was too high as in reality, the data for 2022 has massive number that straight up put the datapoint of other years clustering together at the bottom of the chart. Another variation was tried that would stretch the y axis out to accommodate the massive size of 2022 value but it end up longer than the screen size or if we scale it down, it would be too small to see the details as well as having little to show for nearly 90% of the chart .

This is what it would look like if the line chart was to visualized with separate line for separate years on the same canvas:



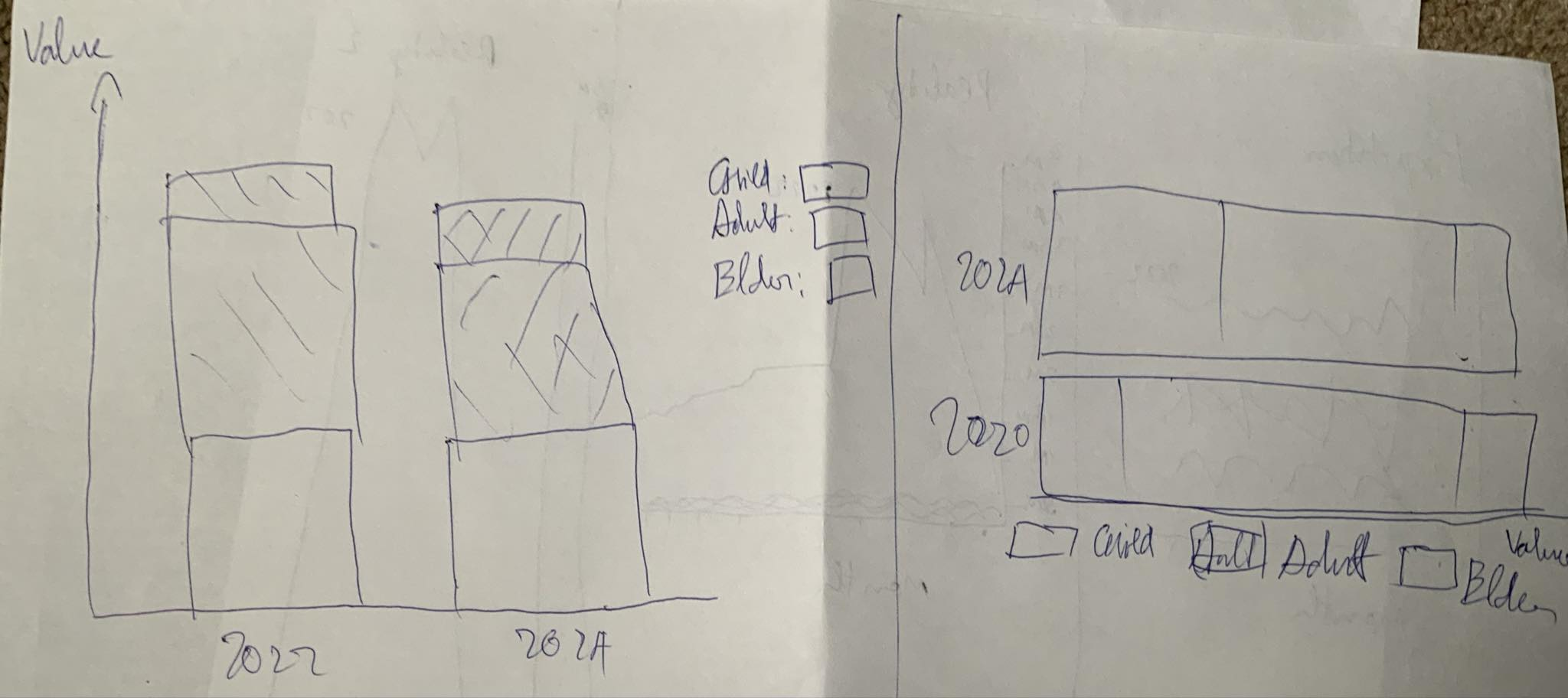
In the end, i decide to split the chart into 5 different line chart that would be view through pressing buttons , the button will delete everything of the previous chat except for the container drawn since it is constant and then draw the new chart in it. We still have one chart to show the whole period since 2021 but this time it is drawn on one chart and the axis is scale down tremendously. Here are some examples:



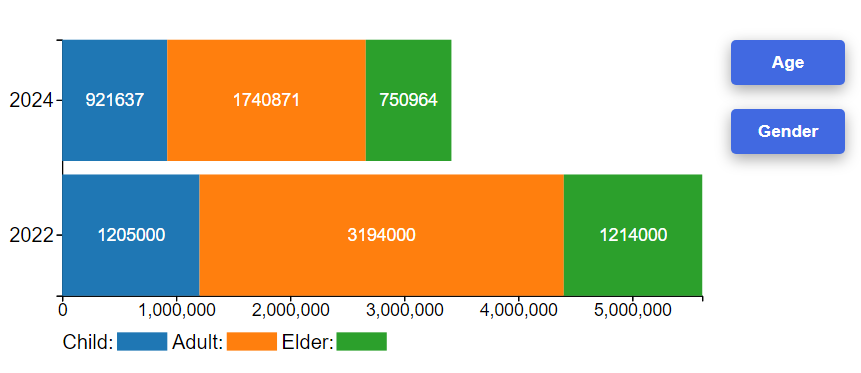
**4.3 Stacked Bar Chart**

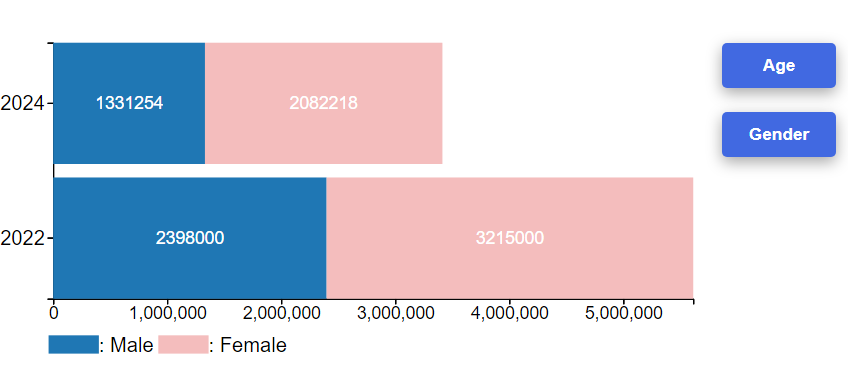
The idea for this chart is vey straight forward, we want to compare the IDP Demographic in 2022 with 2024 to see if it has decreased and what is the ratio between age group and gender. We have 2 variation to consider, one being vertical bar and the other horizontal. The vertical was ok to be honest but it has a lot of space that need to be improve. One of which that is very significant that is there are only 2 bar total so making them stand vertical would risk having too much space wasted and if we make them get closer together then it would just look very cramp and unpleasing. The solution to most of the problem is making a horizontal bar chart instead, this way less space would be wasted since they got close together vertically but horizontally they still take a lot of space. Furthermore, the y-axis is virtually invisible yet viewer can tell which bar is 2022 and 2024 **.**

Here is the sketch:



For the official design, i chose a color scheme that is bright and distinguishable, blue and green have a vulnerability for potential colorblind people so i chose a lighter shade of green and since there are 3 categories i can put orange in between blue and green to further drive the contrast up as well as the chart distinctability. The other stacked bar chart is vastly different as it is depicting Male and Female. Luckily, i dont have to think too much about the color scheme since the color blue for male and light pink for female have already been recognize to be the official colors for the two gender.





**CHAPTER 5: VALIDATION**

**1. How easy was it to understand the information presented in the line chart**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **1** | **2** | **3** | **4** | **5** |
| **Huy** |  |  |  |  | **x** |
| **Tuấn Đức** |  |  |  |  | **x** |
| **Mai Hạnh** |  |  |  |  | **x** |
| **Khôi** |  |  |  |  | **x** |
| **Sơn** |  |  |  |  | **x** |

**2. How would you rate the visual design of the map in terms of its color choices and layout?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **1** | **2** | **3** | **4** | **5** |
| **Huy** |  |  |  |  | **x** |
| **Tuấn Đức** |  |  |  |  | **x** |
| **Mai Hạnh** |  |  |  | **x** |  |
| **Khôi** |  |  |  |  | **x** |
| **Sơn** |  |  |  | **x** |  |

**3. How would you rate the interactivity level overall**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **1** | **2** | **3** | **4** | **5** |
| **Huy** |  |  |  | **x** |  |
| **Tuấn Đức** |  |  |  |  | **x** |
| **Mai Hạnh** |  |  |  | **x** |  |
| **Khôi** |  |  |  |  | **x** |
| **Sơn** |  |  |  |  | **x** |

**4. How did the color choices across the charts affect your ability to interpret the data quickly?**

|  |  |
| --- | --- |
| **Name** | **Answer** |
| **Huy** | **Easily distinguishable** |
| **Tuấn Đức** | **good** |
| **Mai Hạnh** | **Pretty good** |
| **Khôi** | **Easy enough** |
| **Sơn** | **ok** |

**5. Were there any elements in the charts that you found distracting or unnecessary?**

|  |  |
| --- | --- |
| **Name** | **Answer** |
| **Huy** | **X-axis value of stacked bar chart look unintuitive** |
| **Tuấn Đức** | **The year in paragraph need to be more clear (bold font)** |
| **Mai Hạnh** | **Preferably something less bright for map hovering effect** |
| **Khôi** | **Spacing issue in button and y axis of line chart** |
| **Sơn** | **Spacing issue in button and y axis of line chart** |

**CHAPTER 6: CONCLUSION**

Since the invasion of Russia in Ukraine on the 24 of February 2022, migration from Ukraine to European countries and Ukraine internal displacement numbers have been higher than ever, especially in the first two months of the invasion. The migrant number experienced a downward trend after about a few after the invasion started

To present the phenomena above, we collected, extracted and analyzed corresponding datasets and visualized them into 3 distinct kinds of chart. We use line chart to illustrate the trend of migration from Ukraine to European countries in 5 different time periods which are 2021-2024 and each year of that period separately. Geomap is deployed to give viewer a better insight on how many refugee there are in each European countries. Additionally, we made a horizontal stacked bar chart to showcase the percentage of Ukraine IDP Demographic, particularly age group and gender. Finally, a good visualization rarely stand apart and for that we also include text explanation with story telling sprinkled on top of it right beside each chart,map for better understanding of what we are trying to present.

We also make a basic survey to find out how the general public view of our visualization. Some of the people that took the survey are business analysis students and some are IT students. Results were pretty good but clearly there are space to improve though they specified this is more of just their preference.

Overall, we have gain valuable experience in designing and doing a visualization project. It has enhanced our analytical and technical skills, teaching us the importance of transforming data into compelling visuals.

**REFFERENCE**

*Europe - mixed migration flows to Europe, yearly overview (2023)*. Europe - Mixed Migration Flows to Europe, Yearly Overview (2023) | Displacement Tracking Matrix. (n.d.). https://dtm.iom.int/datasets/europe-mixed-migration-flows-europe-yearly-overview-2023

Filipchuk, L., Lomonosova, N., Syrbu, O., & Kabanets, Y. (2022, May 4). *Forced migration and the war in Ukraine (February 24 - March 24, 2022)*. Cedos. https://cedos.org.ua/en/researches/forced-migration-and-the-war-in-ukraine-february-24-march-24/

Koroutchev, R. (n.d.). *UKRAINIAN MIGRATION DURING THE FIRST YEAR AFTER THE BEGINNING OF THE RUSSIAN ARMED CONFLICT IN 2022UKRAINIAN MIGRATION DURING THE FIRST YEAR AFTER THE BEGINNING OF THE RUSSIAN ARMED CONFLICT IN 2022*. PDF Viewer. https://www.ceeol.com/search/viewpdf?id=1207439

*Statistics explained*. Statistics Explained. (n.d.). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Temporary\_protection\_for\_persons\_fleeing\_Ukraine\_-\_monthly\_statistics

*Ukraine - internal displacement report - General Population Survey Round 11 (25 November - 5 December 2022)*. Ukraine - Internal Displacement Report - General Population Survey Round 11 (25 November - 5 December 2022) | Displacement Tracking Matrix. (n.d.). https://dtm.iom.int/reports/ukraine-internal-displacement-report-general-population-survey-round-11-25-november-5

*Ukraine - area baseline assessment (raion level) - round 33 (February 2024)*. Ukraine - Area Baseline Assessment (Raion Level) - Round 33 (February 2024) | Displacement Tracking Matrix. (n.d.). https://dtm.iom.int/reports/ukraine-area-baseline-assessment-raion-level-round-33-february-2024

*Situation Ukraine refugee situation*. UNHCR Data Portal. (n.d.). https://data.unhcr.org/en/situations/ukraine

*Home*. Humanitarian Data Exchange. (n.d.). https://data.humdata.org/dataset/ukraine-data-explorer-inputs

International Migrant Stock 2020. (n.d.). https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2021/Jan/undesa\_pd\_2020\_international\_migrant\_stock\_documentation.pdf

Canton, H. (2021). United Nations Relief and Works Agency for Palestine Refugees in the Near East—UNRWA. In *The Europa Directory of International Organizations 2021* (pp. 286-289). Routledge.