COS30045 - Project Process Book

INTERNATIONAL STUDENT MIGRATION AROUND THE WORLD

Ta Quang Tung
104222196

Nguyen Quang Huy
104169507

Due date: Week 12

https://mercury.swin.edu.au/cos30045/s104222196/project

COS30045 Data Visualization Project - International Student Migration Around the World - Origins and Destinations

Ta Quang Tung - 104222196 **Nguyen Quang Huy - 104169507**

Contents

1 Introduction

1.1 Background and Motivation

Thanks to greater economic prosperity and the convenience of international travel, people are becoming more mobile than before. No longer are people tied down to their country of origin - they can now travel the world to look for a better life in other countries. Among these migrants, many leave their homes in search of better educational opportunities. This visualization project aims to provide an in-depth overview of international student migration patterns.

Users: This visualization will be of interest to a wide range of users, but mainly to students considering studying abroad. That said, educators, policymakers, researchers studying migration patterns, and the general public interested in global trends will also find this project useful.

Tasks: Users will be able to explore the data interactively, identify trends, compare different countries, and gain insights into the factors influencing student migration.

Understanding these migration patterns is crucial as it can help in policy formulation, educational planning, and fostering global understanding.

1.2 Visualisation Purpose

The purpose of this visualization is to answer the following questions:

- Which are the top origin and destination countries of international students?
- What are the biggest destinations/origins of each origin/destination country found in the previous question?
- How have the numbers changed for each pair of origin-destination over the years?

For that reason, the completed visualization will provide users with a comprehensive and accessible view of international student migration, fostering greater understanding and facilitating informed discussions on this topic.

1.3 Project Schedule

To ensure the timely and successful completion of this project, we have planned our work as follows:

- Week 1-2: Team formation and project setup
- Week 3-5: Data collection and initial data processing.
- Week 6-8: Data cleaning and process, scratch some mock design.
- Week 8-10: Iterating on the design based on feedback and adding features.
- Week 10-12: Preparation of the Process Book, source code and final submission.

By planning our work in this manner, we aim to avoid a rush before the final project deadline and ensure that we have enough time for each stage of the project.

2 Dataset

2.1 Data sources

The data for this visualization project is sourced from two primary datasets provided by the UNESCO Institute for Statistics (UIS). The original data source is available at http://data.uis.unesco.org. The datasets relevant to our project are:

1. **Outbound internationally mobile students by host region**: This dataset contains the number of students that left each country to study overseas between 2017 and 2023. As the data for 2022 and 2023 are missing for most of the entries, we only use the data from 2017 and 2021. Each row of the table contains the name of the country and the number of students for each year. This data set is used by the three charts of the “Origin” section of our visualization project.

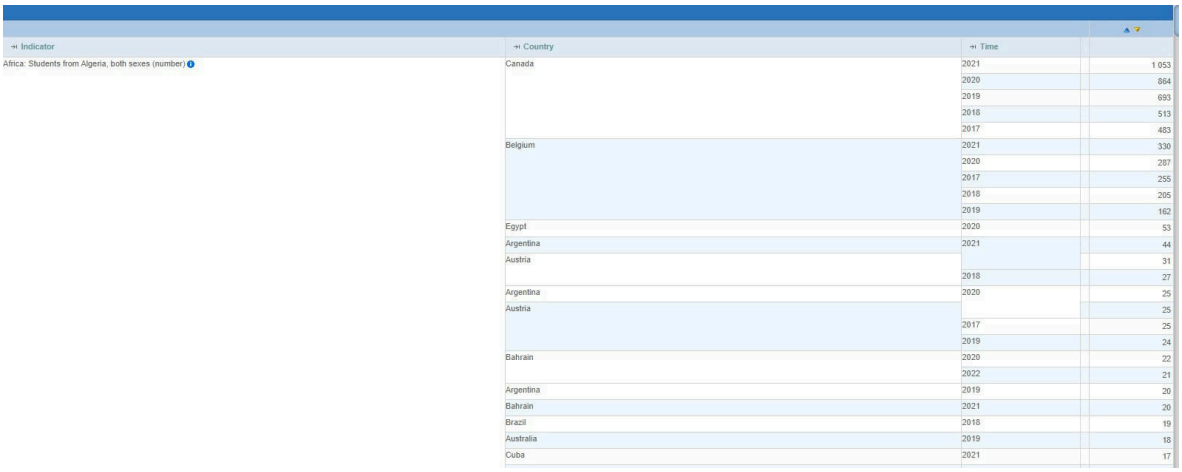
Other policy relevant indicators : *Outbound internationally mobile students by host region*

Customise Export Draw chart

Indicator	Total outbound internationally mobile tertiary students studying abroad, all countries, both sexes (number)													
Time	2017		2018		2019		2020		2021		2022		2023	
Country														
Croatia	9 692		9 754	(*)	10 169	(*)	10 011	(*)	10 440		..		..	
Cuba	2 581		2 476	(*)	2 711	(*)	3 188	(*)	3 326		..		..	
Curaçao	218		233	(*)	239	(*)	235	(*)	150		..		..	
Cyprus	24 456		26 501	(*)	26 161	(*)	25 976	(*)	26 718		..		..	
Czechia	13 200		12 333	(*)	11 890	(*)	12 200	(*)	12 316		..		..	
Democratic People's Republic of Korea	1 490		1 368	(*)	1 016	(*)	1 125	(*)	1 050		..		..	
Democratic Republic of the Congo	11 413		12 284	(*)	12 857	(*)	14 354	(*)	14 581		..		..	
Denmark	5 319		5 946	(*)	5 844	(*)	6 045	(*)	5 827		..		..	
Djibouti	2 311		2 421	(*)	2 473	(*)	2 582	(*)	2 947		..		..	
Dominica	788		964	(*)	1 119	(*)	1 320	(*)	1 458		..		..	
Dominican Republic	4 427		4 643	(*)	4 167	(*)	4 055	(*)	3 820		..		..	
Ecuador	20 998		22 906	(*)	23 414	(*)	23 551	(*)	26 237		..		..	
Egypt	34 770		39 468	(*)	43 828	(*)	47 279	(*)	50 814		..		..	
El Salvador	4 242		4 359	(*)	4 430	(*)	4 547	(*)	4 633		..		..	
Equatorial Guinea	1 539		1 510	(*)	1 217	(*)	1 231	(*)	1 295		..		..	
Eritrea	1 794		1 846	(*)	1 591	(*)	1 528	(*)	1 404		..		..	
Estonia	3 892		3 705	(*)	3 617	(*)	3 523	(*)	3 477		..		..	
Eswatini	2 420		2 404	(*)	2 377	(*)	2 149	(*)	2 145		..		..	

Figure 1: The raw Outbound internationally mobile students by host region dataset.

2. **Inbound internationally mobile students by country of origin**: This dataset contains the number of students from one country to another country between 2017 and 2023. As the data for 2022 and 2023 are missing for most of the entries, we only use the data from 2017 and 2021. Each row of the table contains the origin country, the destination country, the year, and the number of students in that year. This dataset took the most time to clean up as the name of the origin country (column "Indicator" in the figure below) is mixed up with a lot of unneeded text. This data set is used by the three charts of the "Destination" section of our visualization project.



Indicator	Country	Time	
Africa: Students from Algeria, both sexes (number)	Canada	2021	1 053
		2020	864
		2019	693
		2018	513
		2017	483
	Belgium	2021	330
		2020	287
		2017	255
		2018	205
		2019	162
	Egypt	2020	53
	Argentina	2021	44
	Austria		31
		2018	27
	Argentina	2020	25
	Austria		25
		2017	25
		2019	24
	Bahrain	2020	22
		2022	21
	Argentina	2019	20
	Bahrain	2021	20
	Brazil	2018	19
	Australia	2019	18
	Cuba	2021	17

Figure 2: The raw Inbound internationally mobile students by country of origin dataset.

The datasets come in a tabular format, with each row representing a unique combination of origin country, destination country, and year. The attributes in both datasets include:

- Country of Origin: This is a **categorical** attribute that represents the home country of the students. In the bump chart, each country will be represented by a series of bumps (circles) connected by lines. Each series will have a distinct color to tell them apart. In the sankey diagram, each country will be represented as a rectangular node with a unique color. In the line chart, this information is not included.
- Destination Country: This is also a **categorical** attribute that represents the country where the students choose to study. In the bump chart, each country will be represented by a series of bumps (circles) connected by lines. Each series will have a distinct color to tell them apart. In the sankey

diagram, each country will be represented as a rectangular node with a unique color. In the line chart, this information is not included.

- Year: This is an **interval** attribute that represents the year of the data record. In both the bump chart and the sankey diagram, this information is encoded with the x-coordinate. For example, in the bump chart, a bump (circle) further to the right would indicate a later year in the period.
- Number of Students: This is a **quantitative** attribute that represents the number of students who have migrated from the origin country to the destination country in a given year. In the bump chart, the number of students is not represented directly. Instead, the countries are compared and the numbers of students are converted into **ordinal** rankings that are encoded with the y-coordinate. A bump (circle) higher up in the chart indicates a higher ranking. In the sankey diagram, the number of students is encoded with the width of each flow and the height of each rectangular node. In the line chart, this value is encoded with the y-coordinate. Points higher up in the chart indicate higher numbers of students.

We plan to include all the data in our visualization as it is all relevant to our analysis. However, we may choose to aggregate the data in different ways depending on the specific visualization.

2.2 Data processing

The raw data from UNESCO comes in CSV format, with each row representing a unique combination of origin country, destination country, and year. However, to effectively visualize this data, we needed to perform several data processing steps:

- **Data Cleaning:** The raw data contained some missing values and inconsistencies. For example, some rows lack the numbers of students, or these numbers are too small and are marked as "n" (negligible). For these rows, we used Microsoft Excel to filter out all the missing or insignificant values, as they will not be added to our aggregate.
- **Data Transformation:** The raw data was in a long format, with separate rows for each combination of origin and destination country. We transformed this into a wide format, with separate columns for each destination country. This made it easier to analyze the data and create visualizations. We are also mindful of the file format of the resulting data set. For this project, we have extracted a data set for each chart on the website to reduce the need for sorting and filtering data. We used CSVs for tabular data that needed minimal sorting and filtering and JSONs for data that required fast look-ups. An example of JSON data is the data for the line charts. In the snippet below, which is part of the dataset for the line chart of our "Destination" section, the number of students who went from Ukraine to Poland between 2017 and 2021 can be obtained easily with `data["Poland"]["Ukraine"]`.

```
{
    "Poland": {
        "Ukraine": {
            "2017": 34692,
            "2018": 26864,
            "2019": 26938,
            "2020": 27068,
            "2021": 30903
        },
        "Belarus": {
            "2017": 5002,
            "2018": 5258,
```

```
            "2019": 6025,
            "2020": 7472,
            "2021": 8994
        },
        "India": {
            "2017": 2084,
            "2018": 2497,
            "2019": 2153,
            "2020": 2960,
            "2021": 2665
        },
        ...
    }
    ...
}
```

- **Derived Variables:** We calculated some additional variables based on the raw data to aid in our analysis. For example, we calculated the total number of students from each origin country and the total number of students going to each destination country. We also calculated the year-on-year change in these totals to identify trends over time.

3 Requirements

3.1 Must-Have Features

Our visualization project, "International Student Migration Around the World - Origins and Destinations", is divided into two sections, one for origin countries and the other for destination countries. Each section is designed with several must-have features to ensure it effectively communicates the patterns and trends in student migration:

- **Bump Chart**: The first visualization is a bump chart that displays the top 15 origin or destination countries with the most students. This chart should allow users to understand the major players in international student migration. The chart must be able to highlight and blur other items to pull the users' attention when they interact with the chart. The users must be able to inspect each node and discover the number of students in that particular period of that country. Most importantly, clicking on a country in this chart should update the sankey and line charts with the corresponding country data.
- **Sankey Chart**: This chart changes based on the country selected in the bump chart. It shows the flow of students from the selected country to various destinations or vice versa. This chart must be able to allow the users to see the destination or origins of the selected country. When hovering on the link between those countries, there must be a label to indicate the number of students that migrate between those countries. When a link is clicked on, the line chart must be updated with the data for that pair of origin-destination.
- **Line Chart**: This chart updates based on the link selected in the Sankey chart. It shows the trend of student migration to and from the selected country from 2017 to 2021. The chart must be able to have a smooth transition and inherit the color of that country from the Sankey chart so that the user can track the information more easily. When users hover over the chart, there must be a dotted line indicating the exact number of students who migrated in a particular year.

- **Overall website:** The website must be user-friendly and the description of each chart must be written above them to indicate their purpose. Besides that, we also need to ensure that the transitions between each user interaction must be nice and clean.

Besides these basic must-have features, there are also some guidelines for maintaining good visualizations for our website:

- **Axis**: Choosing meaningful axis ranges is crucial for accurately representing our data. The range should be wide enough to include all data points but not so wide that the data appears insignificant or misleading (Kelleher and Wagener, 2011). Specific charts, like bump and line charts, should have axes that are marked. The axes should be appropriately scaled with a sensible number of tick marks to aid in interpreting the chart.
- **Description**: Each chart should include a short description that indicates what is the data currently displayed about and maybe some small guidance to help the user interact with the charts more easily. This will help users gain a more comprehensive understanding of the presented data.
- **Color**: Color has been a powerful tool in marketing for shaping our perception of the world. The effect of color relies on both societal learning and certain assumed common human perceptions (Khan, 2018). On our website, we must carefully pick out the most suitable color palette for each chart to maintain a good visualization.

We are pleased to report that we were able to deliver all these promised features. Each feature presented its own set of challenges, but through careful planning, diligent work, and iterative testing, we were able to overcome these challenges and successfully implement all the must-have features.

3.2 Optional Features

While our visualization project has been designed with several must-have features, we also considered several optional features that could further enhance the user experience and provide additional insights:

- **Data Filtering:** Provide options for users to filter data based on certain criteria, such as the number of students, regions, or specific periods. For example, instead of viewing only the top 15 of the bump chart and the top 10 of the Sankey chart, the user can specify the number of records they would like to see in the graph.
- **Data Download:** This feature would allow users to download the data they are viewing in a common format like CSV or Excel. This could be useful for users who want to do their analysis or use the data in a report or presentation.
- **Mobile Responsiveness**: Ensure that the website is fully responsive and provides a seamless experience on various devices, including tablets and smartphones. This function could be useful for users who want to view our visualization via small-screen devices.
- **Comparison Section:** Allow users to select multiple countries and compare their migration patterns side by side. This seems to be a useful function to have in this project. In this section, the users will pick 2 or more countries and compare them side by side. Other types of charts need to be used in this section.

While these features were not included in the initial version of our visualization due to time and resource constraints, they represent potential areas for future development and improvement.

4 Visualisation Design

The design of our visualization, “International Student Migration Around the World - Origins and Destinations,” is guided by the principles of clarity, interactivity, and user-friendliness. Our goal is to present complex data in a way that is easy to understand and explore. This is especially important considering the primary audience of our visualization is students, some of whom might not be proficient in reading advanced charts.

In designing our visualization, we went through several iterations. We started with rough sketches of our ideas, then created prototypes and asked our lecturer for feedback. Based on the feedback we received, we refined our design until we arrived at the final version.

4.1 Visualization 1 (Bump chart)

At first, we were struggling to find out which chart was the most suitable to answer the question: “What are the top origin and destination countries of international students?”. For that reason, we came up with two main ideas and sketched them in the notebook:

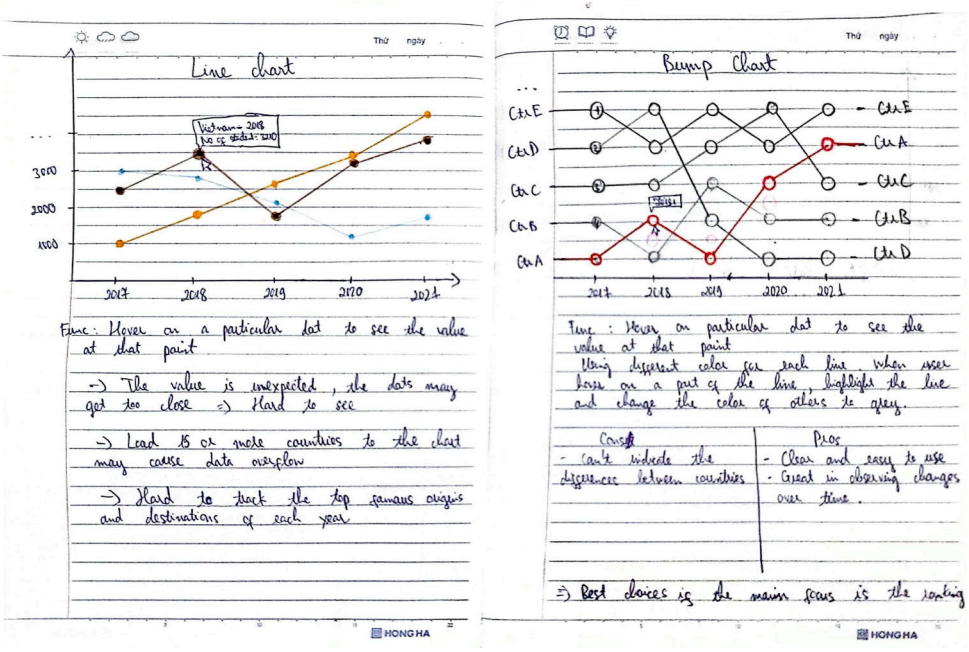


Figure 3: The sketch and quick comparison between line and bump chart

A line graph is a chart that uses lines and points to depict changes over time. It represents quantitative data between two variables with a line connecting successive data points. These two variables are compared on a vertical and horizontal axis (*Line Graph (Line Chart) - Definition, Types, Sketch, Uses and Example*, no date). In this project, The line graph is a good fit for figuring out which countries are major origin and destination points for international students because it’s great at showing trends over time. You can easily see how many students are coming from or going to specific countries over different periods. This helps you spot any patterns, like if more students are coming during certain seasons or if there’s a long-term increase or decrease in student numbers.

A Bump Chart is a special form of a line plot. This chart is well-suited for exploring changes in ranks over time. Using this chart, you can easily compare the position, performance, or rankings of multiple observations concerning each other rather than the actual values themselves (Sridharan and Sridharan, 2018). It lays out countries like nodes connected by lines, letting you compare them all at once.

It's like lining up countries side by side to see which ones have the most student traffic. Besides that, the hovering pop-up gives you a quick idea of how many students are moving between countries. It's a straightforward way to spot the big countries in student migration.

A line graph is great for tracking changes over time of continuous values, while a bump chart is better at showing discrete rankings. Since our question is concerned with ranking over time, we decided that a bump chart was best. Line charts, with potentially intersecting lines that are not evenly spaced, can make it harder to tell which values rank higher than others.

4.2 Visualization 2 (Sankey chart)

After determining what to do with the bump chart, we will have to let the users discover the answer to the question: "What are the biggest destinations and origins of the countries found in the previous chart?". Based on the user's interaction with the bump chart, we have come up with two main ideas for our next visualization: The bar chart and the Sankey chart.

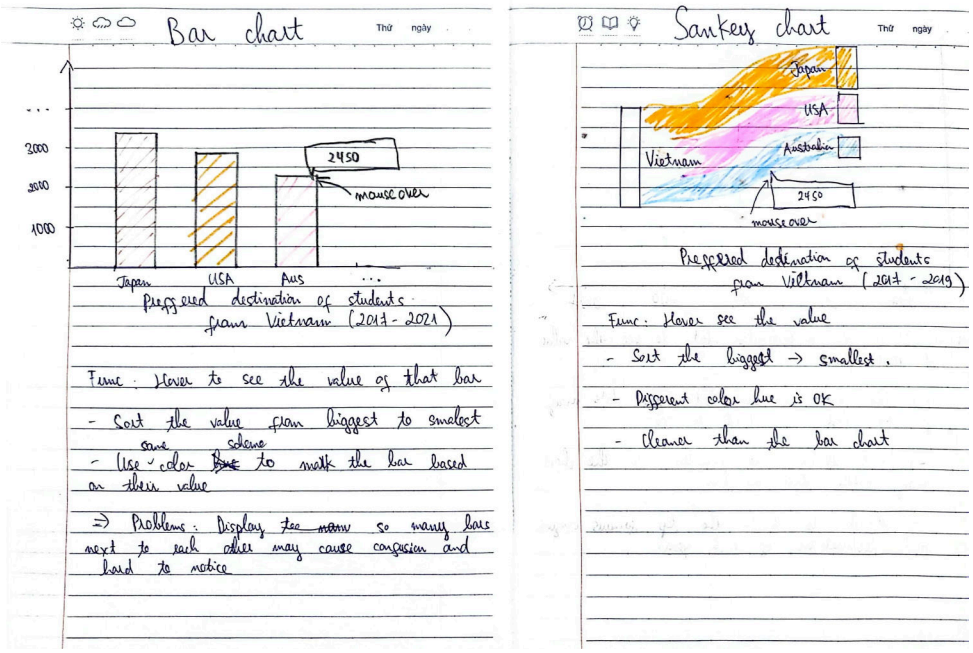


Figure 4: The sketch and quick comparison between bar and Sankey chart

A bar graph is a chart with bars whose lengths represent different categories. These categories, like gender, have no specific order. The graph shows the frequency of each category, with the size of the bar indicating this frequency (*Bar Graph - an overview | ScienceDirect Topics*, no date). With countries represented along the x-axis and the number of students along the y-axis, the bar chart offers a straightforward visual comparison. Each bar provides a clear representation of the number of students from a particular country or the popularity of different destinations, making it easy for users to quickly identify the top choices for students between 2017-2021.

Sankey diagrams are visual tools used to represent the flow of energy, materials, or other resources. Initially used for steam engines, they are now used for modern systems like power plants and for a macroscopic view of global energy use. Besides energy, they are also used to depict the flow of various resources (Lupton and Allwood, 2017). By visualizing the flow of students between countries, the Sankey chart provides a general overview of migration patterns. With nodes representing countries or destinations and the width of the connecting lines indicating the volume of student movement, users

can easily notice the relationships between different locations. However, one major concern with the Sankey chart is that it is difficult to digest for people without a major knowledge background and may require a lengthy introduction (*Sankey Diagram*, no date). So if we intend to visualize our information this way, we must simplify the design of this chart.

When deciding between the bar chart and the Sankey chart for this visualization, the choice depends on the level of detail and insight desired. The bar chart is straightforward and great for a quick comparison, while the Sankey chart gives you a deeper dive into migration patterns. If we want a simple overview of popular destinations or origins, the bar chart is the way to go. But if we aim to show the user a more detailed understanding of how students move between countries, the Sankey chart offers a richer visualization of international student migration and sure is a better choice.

4.3 Visualization 3 (Line chart)

Lastly, when it comes to our final question for the visualization: "How have the numbers changed for each pair of origin-destination over the years?", we both strongly agree that nothing is better than using a line chart to represent those data. After modifying the line chart design in the previous section, a new sketch of the line chart was completed and ready to be used.

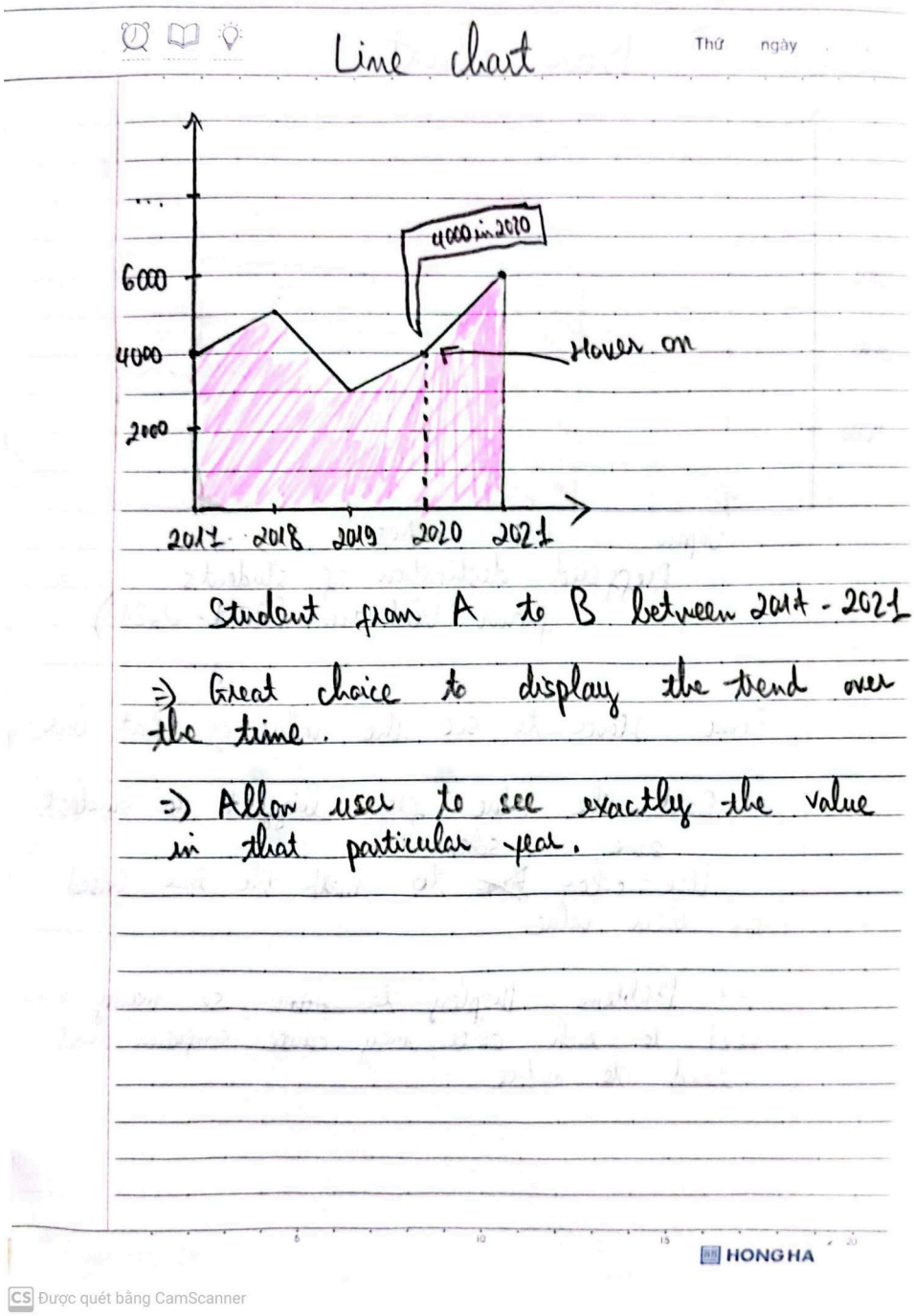


Figure 5: The final sketch of the line chart

A line chart is utilized when there's a need to highlight fluctuations in a variable's values (displayed on the y-axis) corresponding to continuous values of another variable (displayed on the x-axis). The focus on trends over time is conveyed by line segments that consistently progress from left to right, with the direction of the line's slope indicating an increase or decrease (Atlassian, no date). For that reason, when displaying the trend of the students' migrations between 2 countries in a period, we strongly believe that this must be the most suitable chart.

4.4 Overall website design

During our project, we received a lot of feedback from the tutor during each stand-up meeting and also during the discussion time in class. Thus, we rearranged some of the elements in the web page and made a huge transition from our initial design.

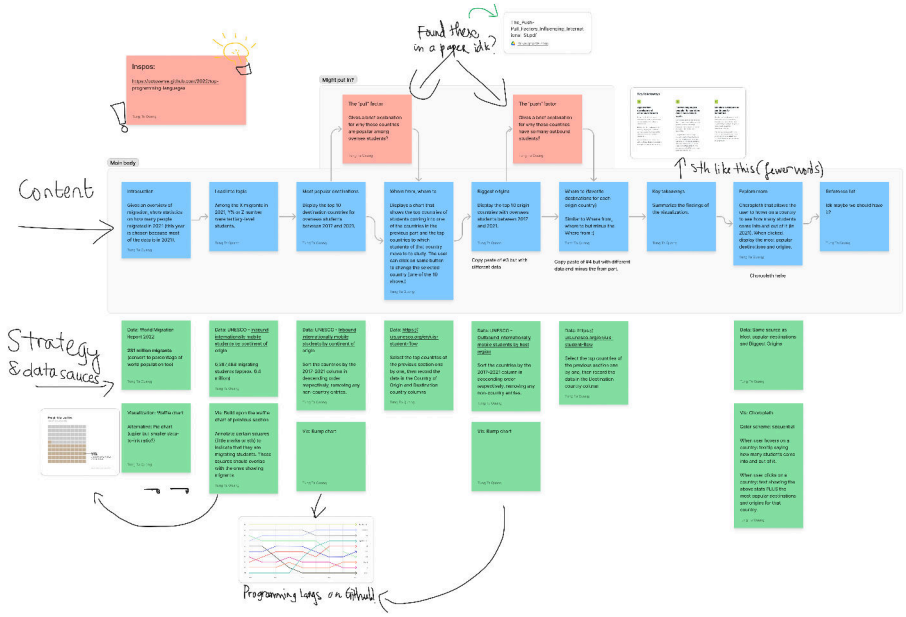


Figure 6: The first design of the website

In our first design, we provided the users with lots of leading pages and the content of the page was lengthy, which caused the total pages of the website to stack up to 9 pages. In this design, we aimed to illustrate all the data we wanted to display and didn't focus much on the user story. Besides that, the flow of the data between each page was not directly relevant to each other, which may confuse the users. However, after stand-up meeting 3, after getting some useful feedback from our tutor, we immediately updated and rearranged our website, trimmed all the unnecessary parts, and picked a new layout.

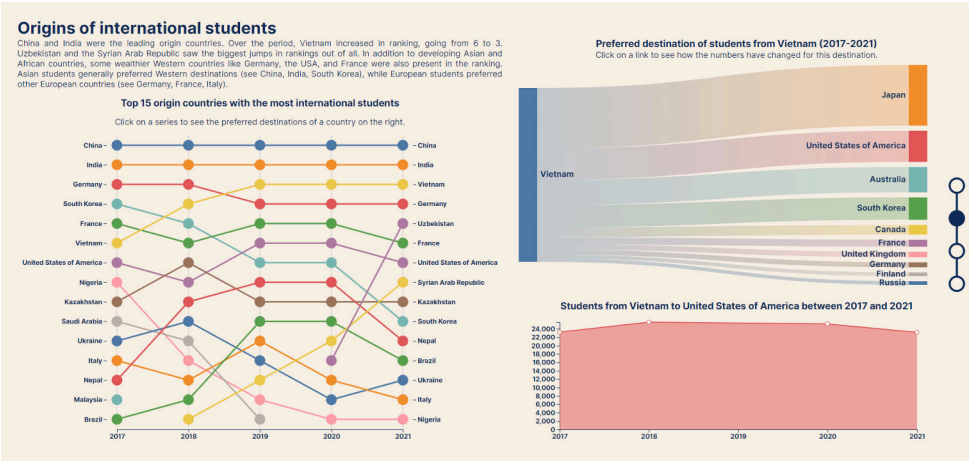


Figure 7: The final design of the website

In the final design of our webpage, the biggest change that we have made is the position of the charts. Separated charts will now be located on the same page in order to pass the interaction of the users

as the value for the next chart. By using this approach, we can enhance the experiences by letting them discover all the data related to the biggest origin or destination at once without any interruption. Moreover, we also cut the website to 4 pages only, which include the introduction page, origins of international students, destinations of international students, and acknowledgments.

In conclusion, we believe that our visualization design effectively presents the data and allows users to explore and understand the complex dynamics of international student migration.

5 User Validation

5.1 Methodology

User validation is a critical component of our project as it ensures that our visualization is not only accurate but also user-friendly and effective. For our research methodology, we have adopted a straightforward approach to gather feedback on our visualization project. Firstly, we have deployed the website on GitHub Pages (https://104222196.github.io/), making it accessible to users. Through this platform, individuals can interact with our visualization easily. Secondly, we have created a Google Forms survey to collect feedback from users regarding their experience with the visualization (https://forms.gle/PhDuuKZNLZekVu1x9).

This survey contains questions covering usability and overall effectiveness of the whole website. We have gathered participants from diverse backgrounds through various channels to ensure a wide range of perspectives. During testing sessions, participants are encouraged to explore the visualization and fill in the form independently, while we observe their responses and note any encountered issues. After collecting responses, we analyze both quantitative and qualitative feedback to have further insights and identify areas for future improvement. Through this approach, we aim to gather valuable feedback and refine our visualization to better serve our users' needs in the future.

5.2 Participants' background

In our report, most of the responses come from the 18-24 age group, making up 92.3% of the total feedback. The remaining responses are from participants aged 16-18 years old. This shows that the majority of our feedback comes from undergraduates, with a smaller portion from high school students.

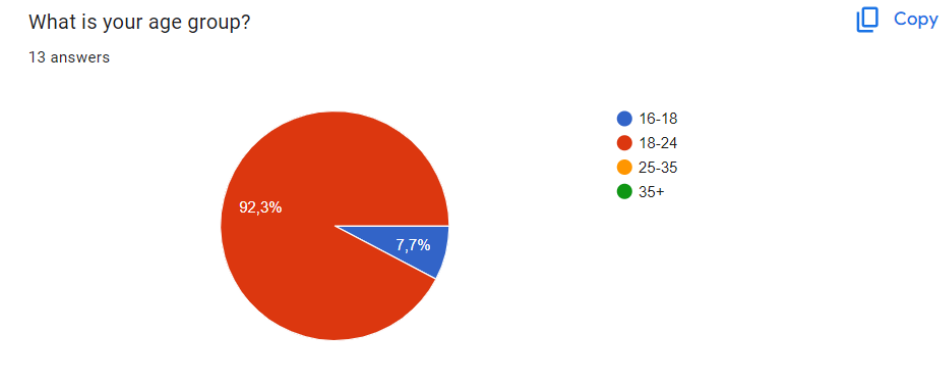


Figure 8: The age group of our survey participants

Additionally, we checked if participants were studying computer science or not. According to the report, there is 76,9% of the participants with Computer Science background. This information helps us understand if there are any differences in feedback based on their academic backgrounds. Those with a computer science background might notice technical aspects of the visualization, while those without

might focus more on how easy it is to use and understand. Analyzing feedback from both groups helps us improve the visualization to meet the needs of all users.

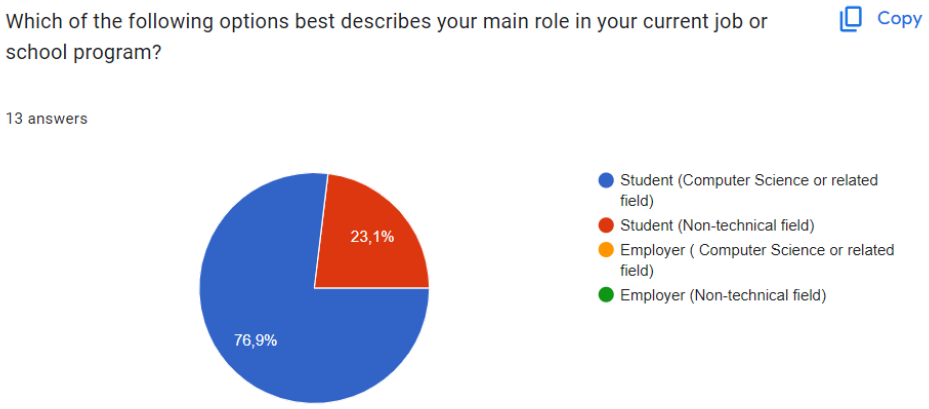


Figure 9: The knowledge background of our survey participants

5.3 Bump Chart effectiveness

When asked about the effectiveness and appeal of the bump chart on the page, an overwhelming majority of participants, 12 out of 13, agreed that the chart is appealing and effectively conveys the required information.

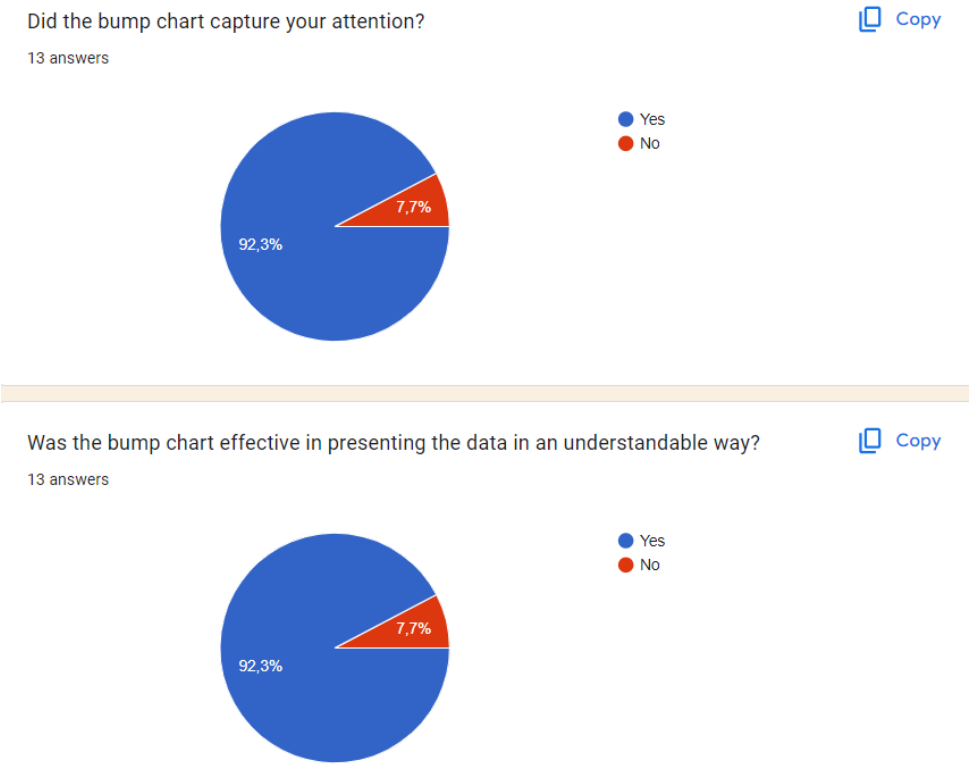


Figure 10: The responses of the bump chart effectiveness

To further demonstrate the effectiveness of the bump chart, we conducted a task-based assessment to ensure users could extract information from the chart successfully. When asked, “In 2018, which ORIGIN country took 7th place in the ranking?” an impressive 12 out of 13 participants correctly identified Kazakhstan. Additionally, we measured the time it took participants to complete the task. Interestingly,

30.8% of participants found it challenging to locate this data on the chart, taking more than 20 seconds to do so. Only one participant managed to complete the task in under 5 seconds, while the majority required between 5 to 20 seconds. Despite some users finding the task more time-consuming, the high accuracy rate highlights the bump chart's effectiveness in conveying the required information.

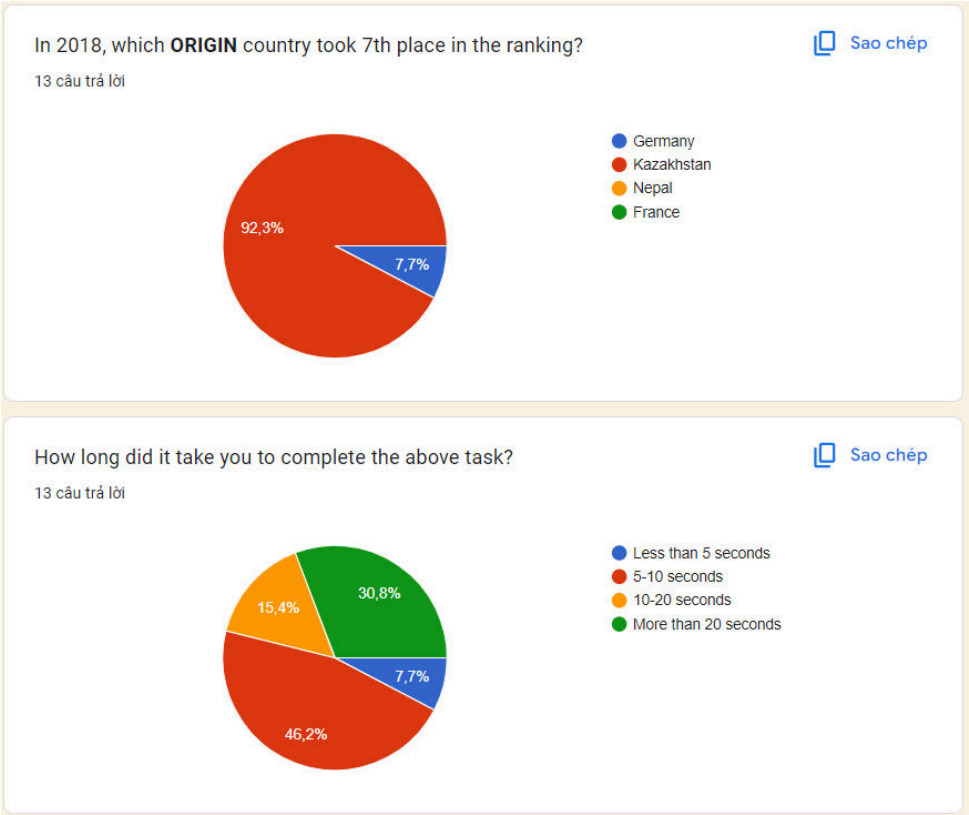


Figure 11: The responses of the first users' task

Lastly, we gathered extensive feedback from users regarding the overall design of the bump chart. Interestingly, the majority of users reported minimal difficulties when interacting with the chart and expressed satisfaction with its current design. However, some users did provide valuable feedback, particularly noting challenges caused by the website's lack of familiarity. They highlighted the need for additional indicators to aid navigation and streamline task completion, especially for users less acquainted with the website's interface. While the general point of view was positive, these insights underscore the importance of enhancing user guidance and accessibility features to improve the overall user experience with the bump chart.

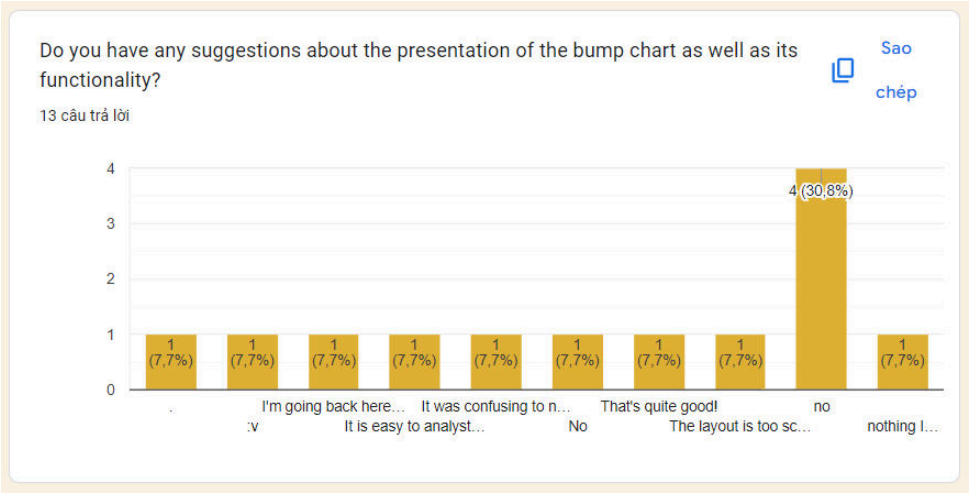


Figure 12: Users' feedbacks on the bump chart

5.4 Sankey chart effectiveness

When evaluating the Sankey chart's effectiveness and attractiveness on the page, 12 out of 13 participants agreed that the chart is visually appealing and all of the participants agreed that the chart effectively communicates the required information, which surprised us because we thought that this should be the most difficult chart for the users.

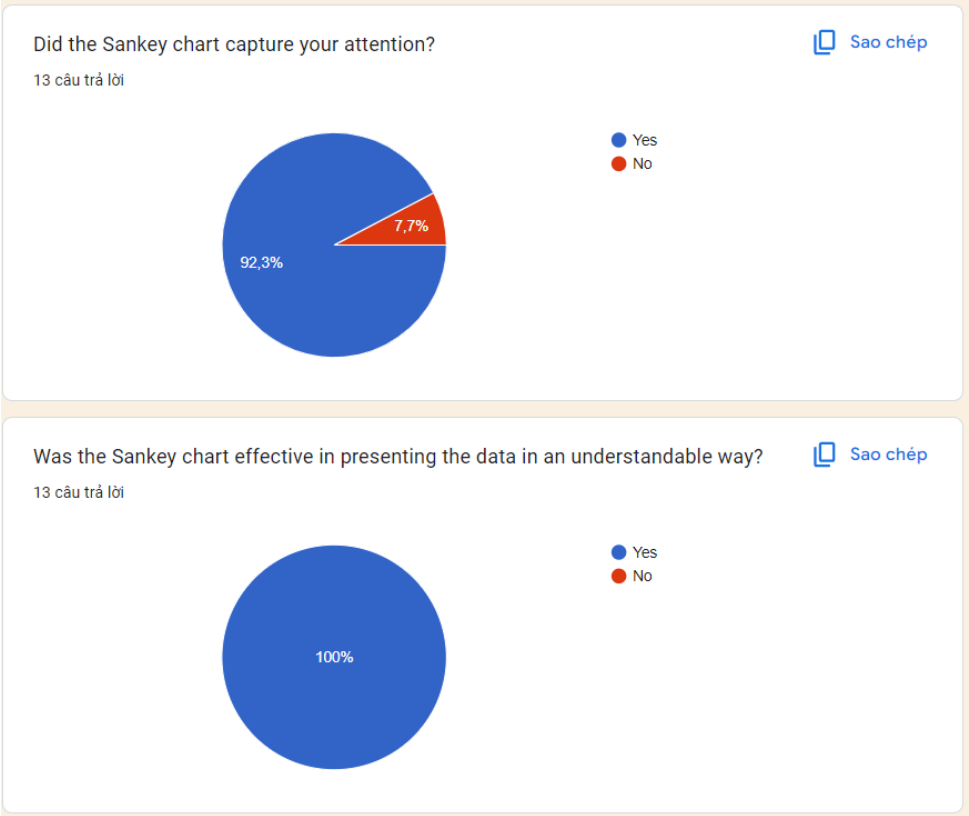


Figure 13: The responses of the Sankey chart effectiveness

To demonstrate the efficiency of the Sankey chart, we carried out an evaluation focused on task completion. We asked participants to identify the main destination for Vietnamese international students from 2017 to 2021. Out of 13 participants, only 10 correctly answered with Japan. We also timed how long it took participants to find this information. Interestingly, 46.2% of participants were able to locate

the information in less than 5 seconds. However, 30.8% took more than 20 seconds, and 23.1% took between 5-10 seconds. Despite the range in completion times, these results highlight the Sankey chart's effectiveness in presenting data, while also indicating that we still need to make future improvements in terms of accuracy and efficiency.

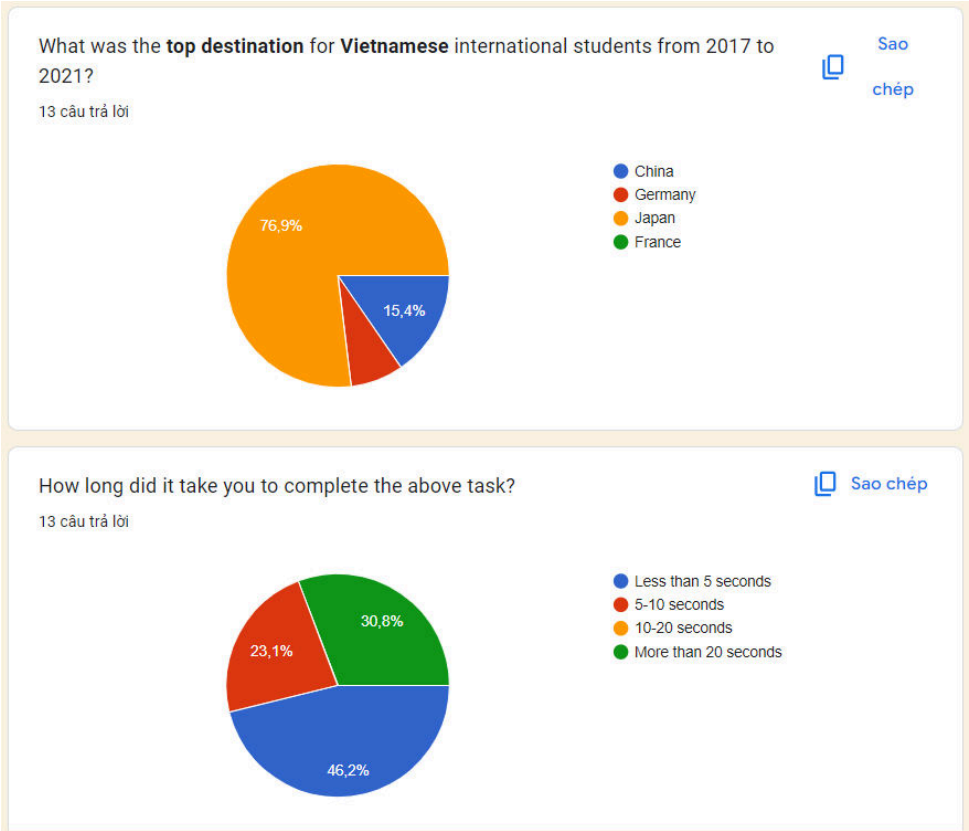


Figure 14: The responses of the second users' task

Finally, we gathered plenty of feedback from users about improving the overall design of the Sankey chart. While most users didn't face many issues with the bump chart and felt there was nothing to change, others provided valuable feedback. They pointed out that because the chart itself is not so popular, they struggled initially. Additionally, some users didn't realize they could interact with the chart. For that reason, we have added some instructions ahead of each chart to make sure that the user knows how to interact with each chart and see the information they want.

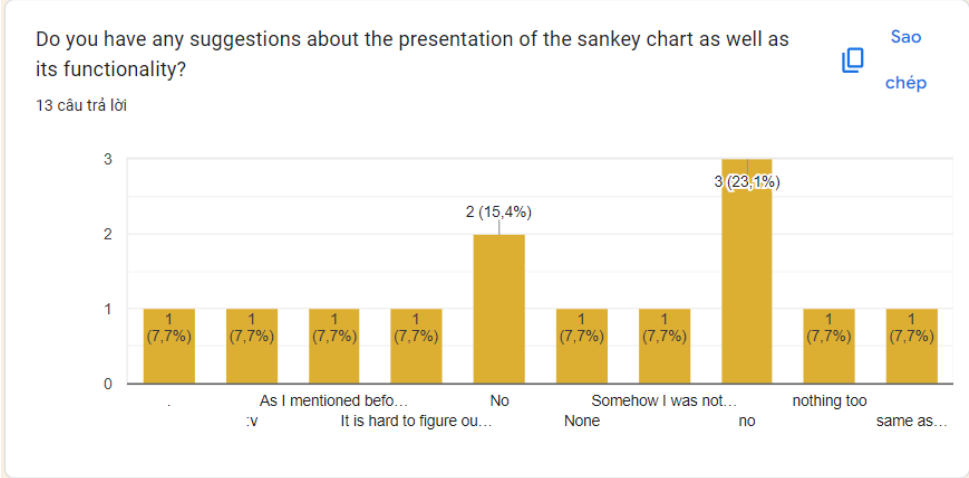


Figure 15: Users' feedbacks on the Sankey chart

5.5 Line chart effectiveness

When users were asked about their experience with the line chart on the page, 10 out of 13 participants found it visually appealing. However, a higher number, 12 out of 13, agreed that the chart effectively provided the required information. This difference in responses suggests that while the chart may not have been universally perceived as visually appealing, it was still considered effective in conveying the necessary data. This could be because we placed them at the bottom right-hand side of the page, which the users often neglect when they first see the page.

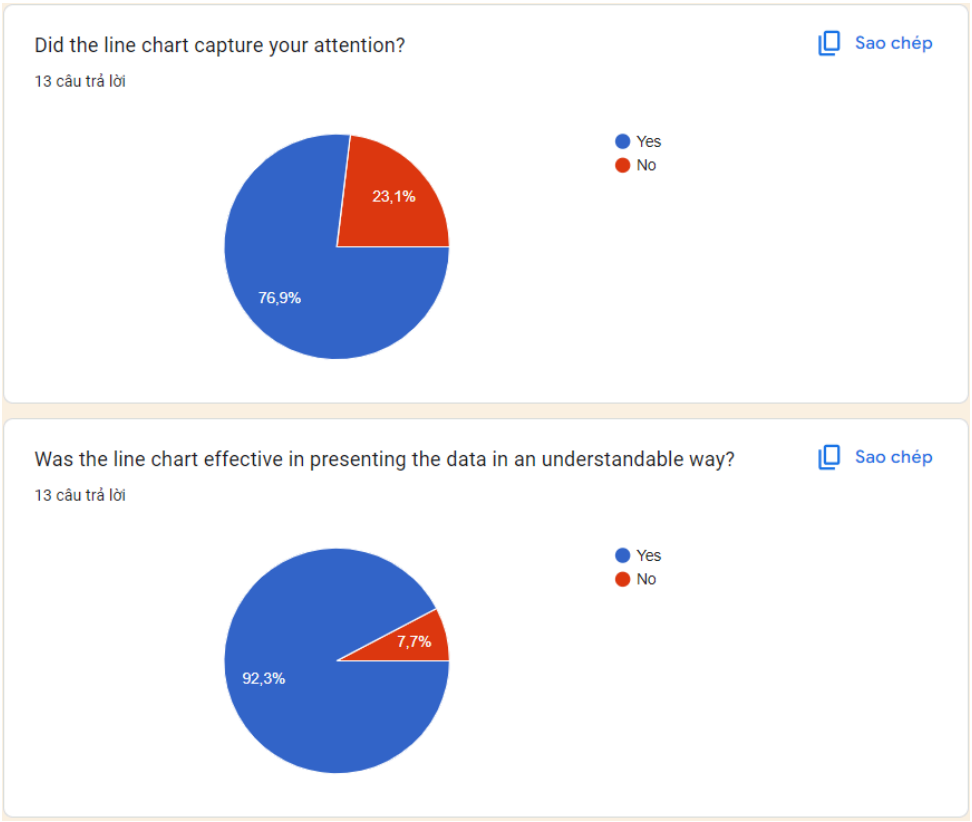


Figure 16: The responses of the line chart effectiveness

To see how well the line chart works, we asked users to find specific info on it. We asked a question, “What was the trend of students from Vietnam going to South Korea from 2017 to 2021?” Surprisingly, 100% of participants got it right. We also timed how long it took them. Surprisingly, over half of them (53,8%) found it easy, taking less than 5 seconds. Only 15,4% of participants took more than 20 seconds to answer, and the rest took between 5 to 20 seconds. This tells us the line chart is not only good at showing info, but it is also easy to use for the users.

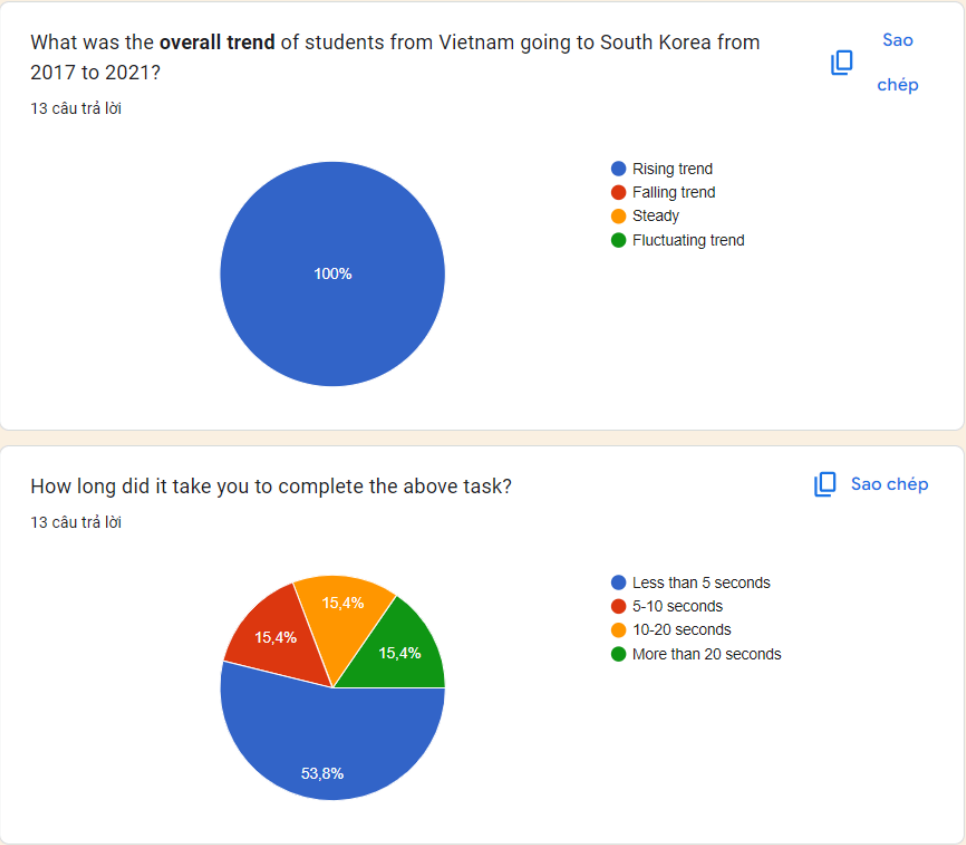


Figure 17: The responses of the third users' task

In our last review, we got a lot of feedback from users about possible improvements to the line chart's design. Interestingly, there weren't many complaints about the chart's design. Users mostly said they were happy, saying they thought the charts were well-made and presented effectively. The fact that there weren't many complaints might be because the chart is clear, simple, and easy to understand, which probably made users have a good experience and feel satisfied overall.

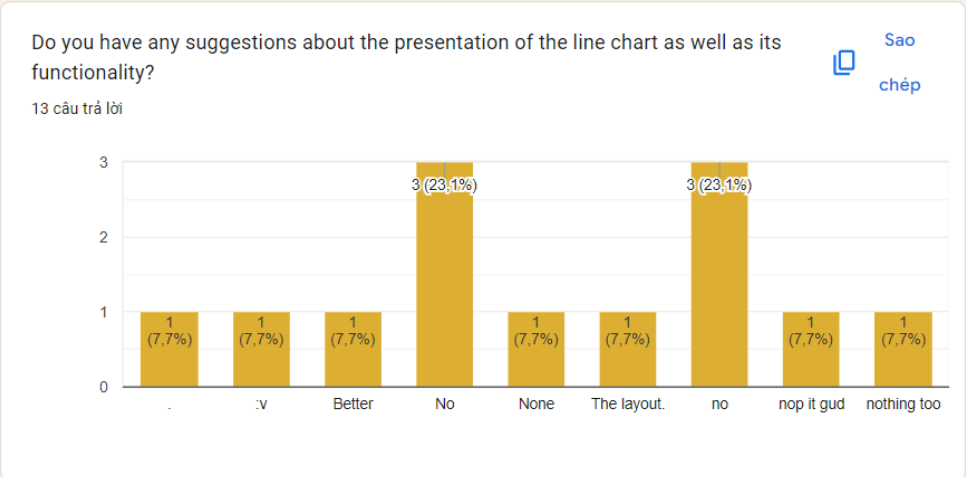


Figure 18: Users' feedbacks on the line chart

5.6 Overall website effectiveness

When asked about the overall website's user-friendliness, 9 out of 13 users found it easy to navigate. However, when it came to their overall comfort using the website, an even higher number, 11 out of 13 users, agreed that they thoroughly enjoyed their experience.

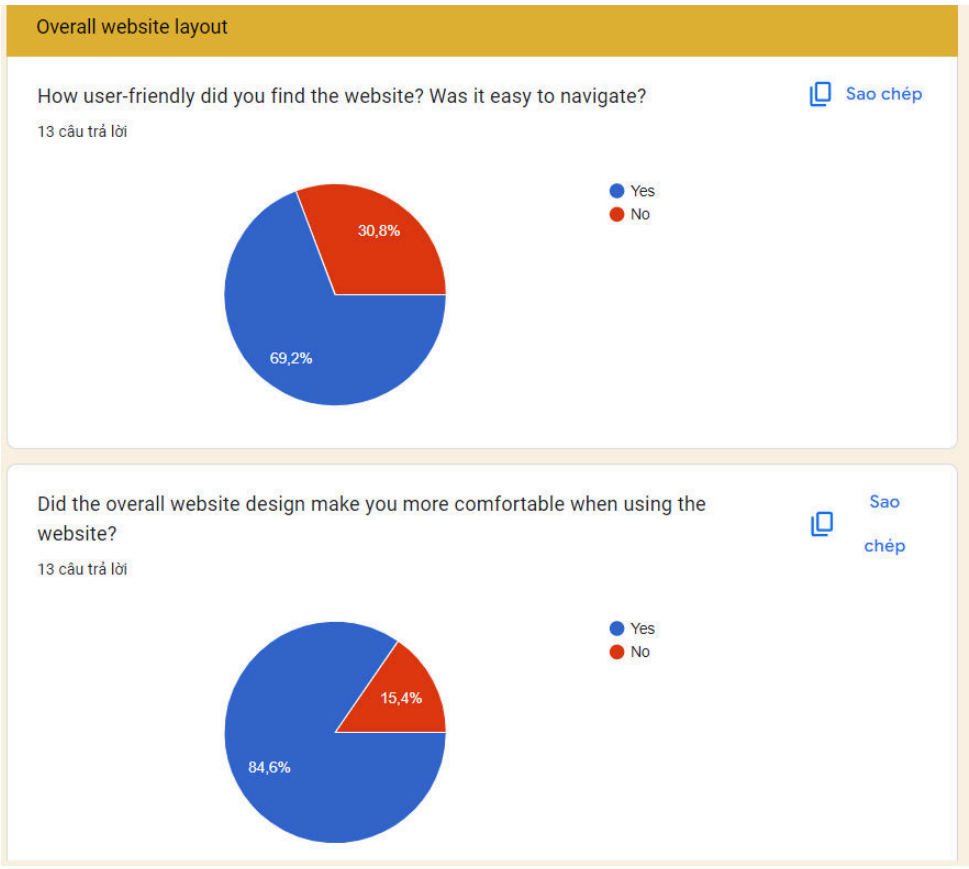


Figure 19: The responses of the overall website effectiveness

At last, users shared some thoughts on how to make the website better. Most didn't have trouble with the bump chart, but some had useful suggestions. They said the layout could be better because it looks messy, and the charts are too close together. Also, they mentioned that the website should work well, especially with snap scrolling. Despite these issues, users liked the fun and interactive charts, which made the experience exciting. Some users mentioned that it might take a bit of time to get used to the charts, but they thought the detailed data was reasonable. These comments give us good ideas for making the website more user-friendly and enjoyable to use.

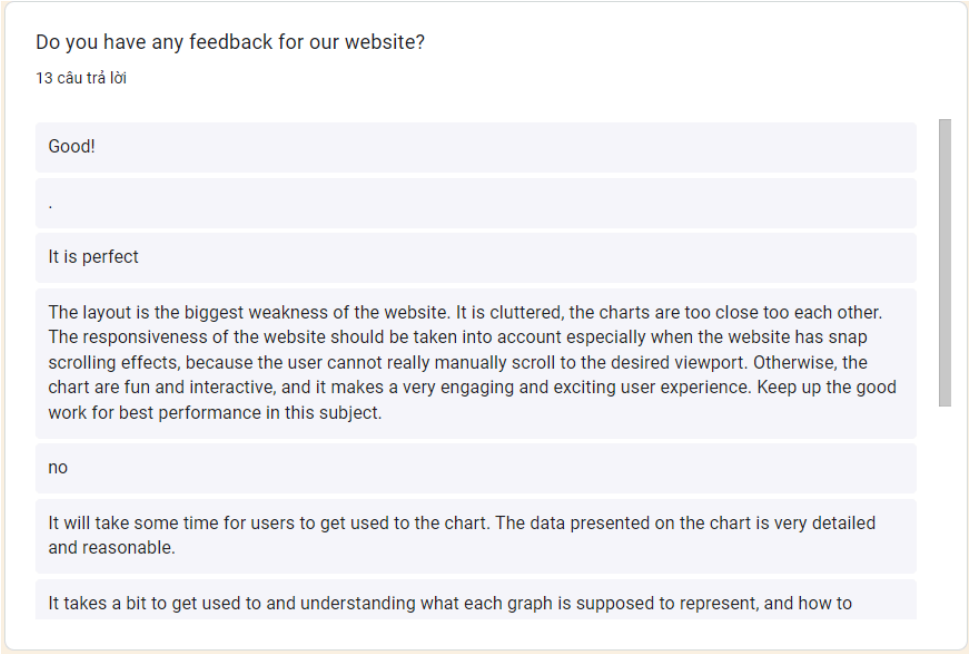


Figure 20: The responses of the overall website effectiveness

In conclusion, the user validation process provided valuable insights into the strengths and areas for improvement of our visualization project. Through feedback gathered from users, we gained a better understanding of how they interact with the charts and navigate the website. These insights will inform our efforts to enhance user experience, ensuring that our visualization remains both informative and user-friendly. Overall, the user validation process served as a crucial step in refining our project to better meet the expectations of our audience.

6 Conclusion

To wrap up, this process book showcases how our project came together, starting from the initial idea to the final visualization. We focused on illustrating migration patterns using D3.js, aiming to make the data both informative and engaging. Throughout the project, we learned a lot about data visualization—how to make data easy to understand and how to build interactive features that keep users interested. Working together, we solved problems, shared ideas, and improved our skills. Overall, this project was a journey of learning and teamwork. We now have a better grasp of how to visualize data effectively and collaborate effectively. Looking ahead, we're excited to apply what we've learned in future projects and keep exploring the world of data visualization.

References

Atlassian (no date) *A Complete Guide to Line Charts*. Available at: https://www.atlassian.com/data/charts/line-chart-complete-guide (Accessed: April 6, 2024)

Bar Graph - an overview | ScienceDirect Topics (no date). Available at: https://www.sciencedirect.com/topics/mathematics/bar-graph (Accessed: April 6, 2024)

Kelleher, C. and Wagener, T. (2011) "Ten guidelines for effective data visualization in scientific publications," *Environmental Modelling and Software - ENVSOFT*, 26, pp. 822–827. Available at: https://doi.org/10.1016/j.envsoft.2010.12.006

Khan, S. (2018) *Data Visualization "Color and Data Visualization "*

Line Graph (Line Chart) - Definition, Types, Sketch, Uses and Example (no date). Available at: https://byjus.com/maths/line-graph/ (Accessed: April 5, 2024)

Lupton, R. C. and Allwood, J. M. (2017) "Hybrid Sankey diagrams: Visual analysis of multidimensional data for understanding resource use," *Resources, Conservation and Recycling*, 124, pp. 141–151. Available at: https://doi.org/10.1016/j.resconrec.2017.05.002

Sankey Diagram (no date). Available at: https://www.fusioncharts.com/ (Accessed: April 6, 2024)

Sridharan, M. and Sridharan, M. A. (2018) *Bump Chart*. Available at: https://thinkinsights.net/data/bump-chart/ (Accessed: April 5, 2024)