Group 1: Peiyuan Li, Qi Qiao, Alena Zheng, Yanqi Liu, Jiaqi Wu

STAT 436 Kris Sankaran

May 6, 2024

Word count: 1525

Dataset / Source Code / Write-Up / Shiny Application

Project Milestone 3: Refinement and Reporting

Introduction

This project, guided by the Design Studio modality, focused on a demographic dataset from Kaggle (Watanabe, 2024) about international students in the US. It aims to show several invaluable facts about international students in the US to the US higher education institutions. The project's goal is to provide insights to the institutions to aid in developing more effective recruitment strategies, support services, and academic programs to meet the needs of international students, to enrich institutions' educative environments, and to strengthen the institutions' global presence and competitiveness. Utilizing the Design Studio approach to convert static data into interactive narratives, the project seeks to make complex information more accessible and appealing to institutions. With the above motivations and goals in mind and inspired by the literature review, the project's key visualization tasks are:

a) Source of funds: revealing how students finance their U.S. education.

b) Countries of origin: mapping where students originate to identify key regions.

c) Field of study: highlighting trends in popular disciplines.

d) Academic levels: detailing enrollment numbers by academic level to reveal preferences.

Literature review

Before delving into the project details, it is important to understand the context. Demographic data, while rich in information, can be challenging to summarize (Riffe et, 2021); however, it provides crucial insights for higher education institutions in the US to examine the recruitment of their international students (Goralski & Tootoonchi, 2015). This understanding forms the basis of this project, which was further enriched by the visualizations we drew inspiration from.

The first inspiration for our visualization designs comes from Gurpreet Saluja's visualization of demographic data on international students in the US (Fig 1, 2) created in 2023. It features a color-coded map that illustrates population distribution and a bar graph that presents data on active students. However, the map (Fig 1) lacks interactivity, limiting the data density that could be presented. Additionally, the bar graph (Fig 2) is restricted to showcasing statistics from a specific point in time and cannot extend over longer periods. Therefore, we created a dynamic choropleth map that displays the geographic origins of undergraduate international students arriving in the U.S. The design is selected for its effectiveness in highlighting the density and distribution of students by country. It can provide a clear visual representation of data across geographic regions and can show data over time.

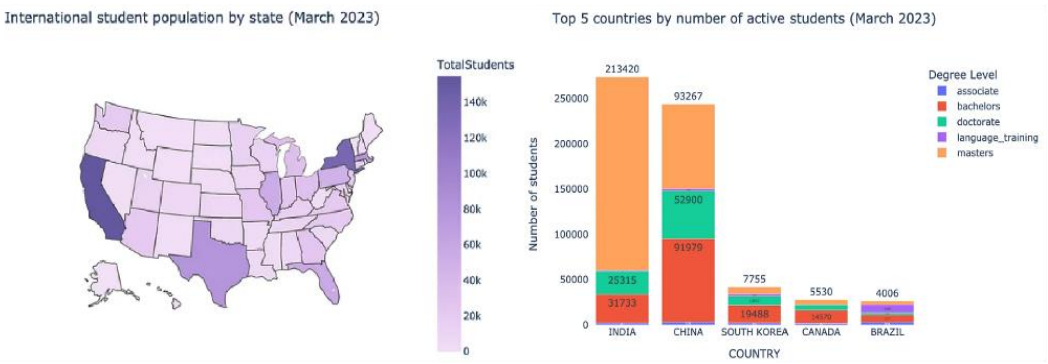

Fig 1, 2. Gurpreet Saluja's visualization of demographic data on international students in the US

The second inspiration comes from the static stacked bar chart for international students in the US among academic levels—Associate, Bachelor's, Master's, and Doctoral (Fig 3). The author uses data from 2012 to 2022 to draw some conclusions, such as an increase of 37% in the number of international science and engineering students, with the number of Master students more than doubling (Fischer, 2024). Similarly, another literature suggests that after the 2008 OPT policy extension being enacted, there is an upward trend in the number of students pursuing STEM degrees at various academic levels post-policy implementation (Amuedo-Dorantes et al., 2019). Consequently, a dynamic stacked bar chart has been implemented to display the numbers of international students across various academic levels over time.

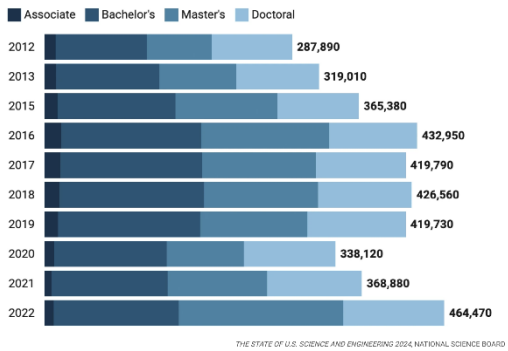

Fig 3. Static stacked bar chart for international students in the US among four academic levels, including Associate, Bachelor's, Master's, and Doctoral

The third inspiration comes from the Global Higher Education Population graph (Fig 4). It employs a pie chart delineated with distinct color assignments for each student's country of origin, with clear labels of the countries' names and the number of students (Goralski &

Tootoonchi, 2015). However, the static graph demonstrates data only in 2020, which does not allow users to make comparisons across different years. In contrast, a dynamic pie chart allows users to select the range of years or a specific year that they are interested in, making it easy to compare the prevalence of each funding source across the years of interest.

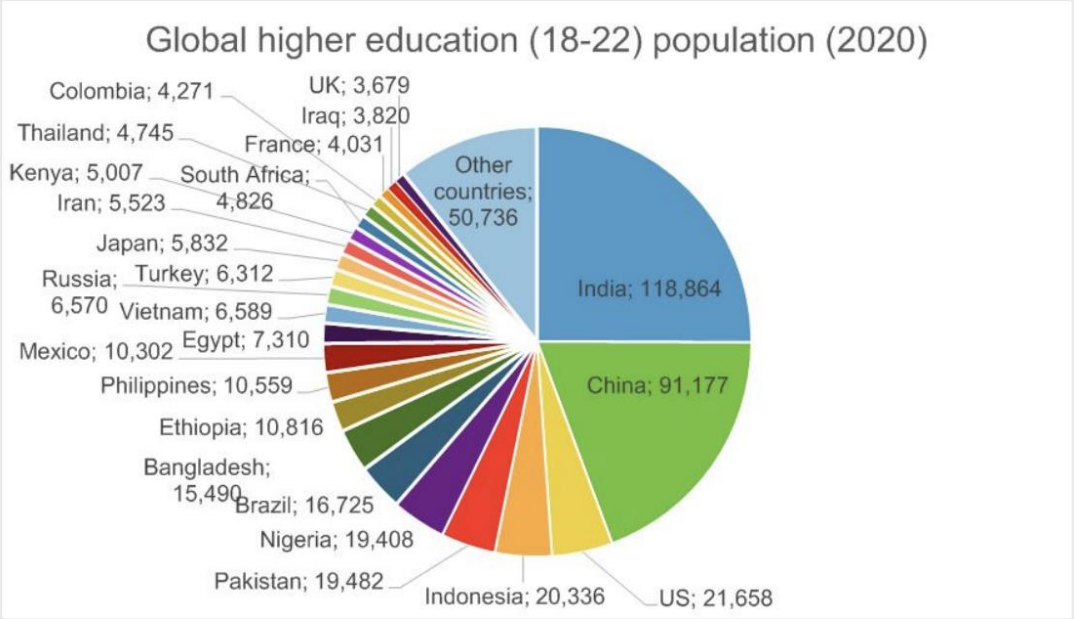

Fig 4. Pie chart graphs the Global Higher Education Population in 2020 with a detailed label for each country

The final inspiration derives from Mode.com (Mode Help, n.d.), highlighting filters' importance in data visualization (Fig 5). These filters allow users to customize their data exploration based on specific criteria, enhancing the analytical depth and personalizing the experience. This functionality is pivotal for conducting thorough analyses of datasets and quickly identifying trends that may be obscured in broader data compilations. Filters also facilitate a detailed examination of the field of study data. By applying these filters, there is an opportunity to substantiate the findings reported by Bound et al. (2021), which state that it is common for

international students in the U.S. to pursue science, technology, engineering, and mathematics (STEM) majors at various academic levels—findings that could significantly influence institutional academic planning and recruitment strategies.

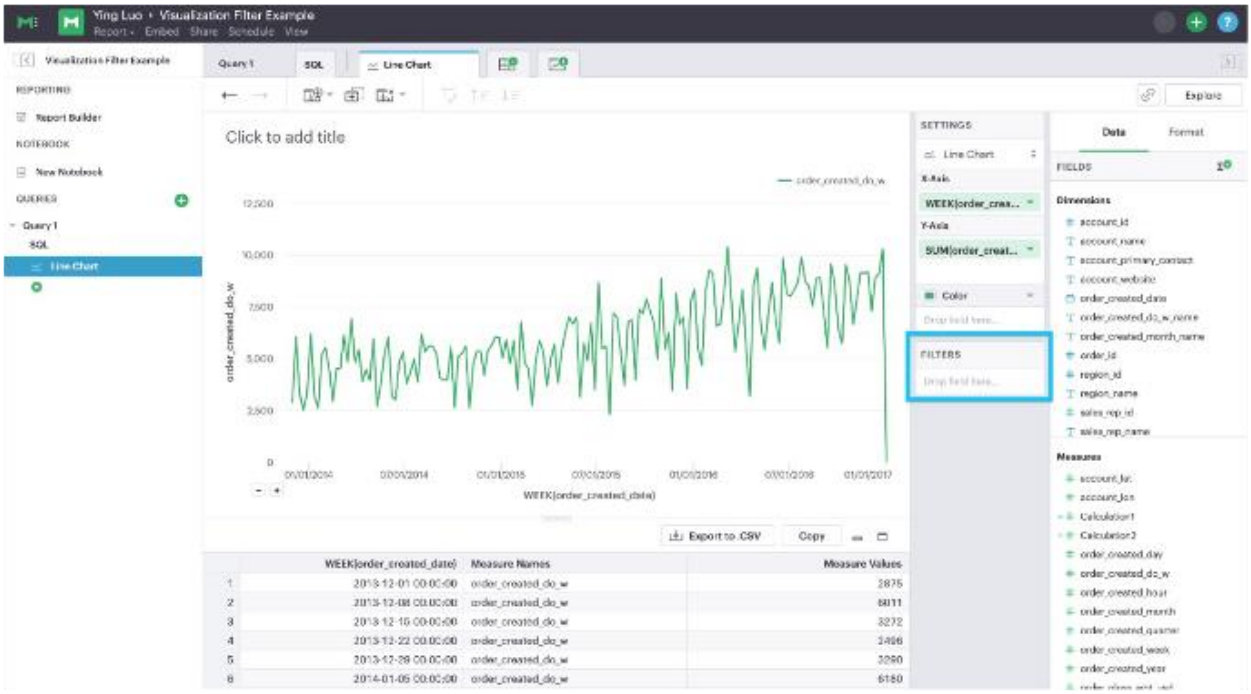

Fig 5. Mode.com's visualization filter panel, highlighted with a blue box for emphasis

Designs

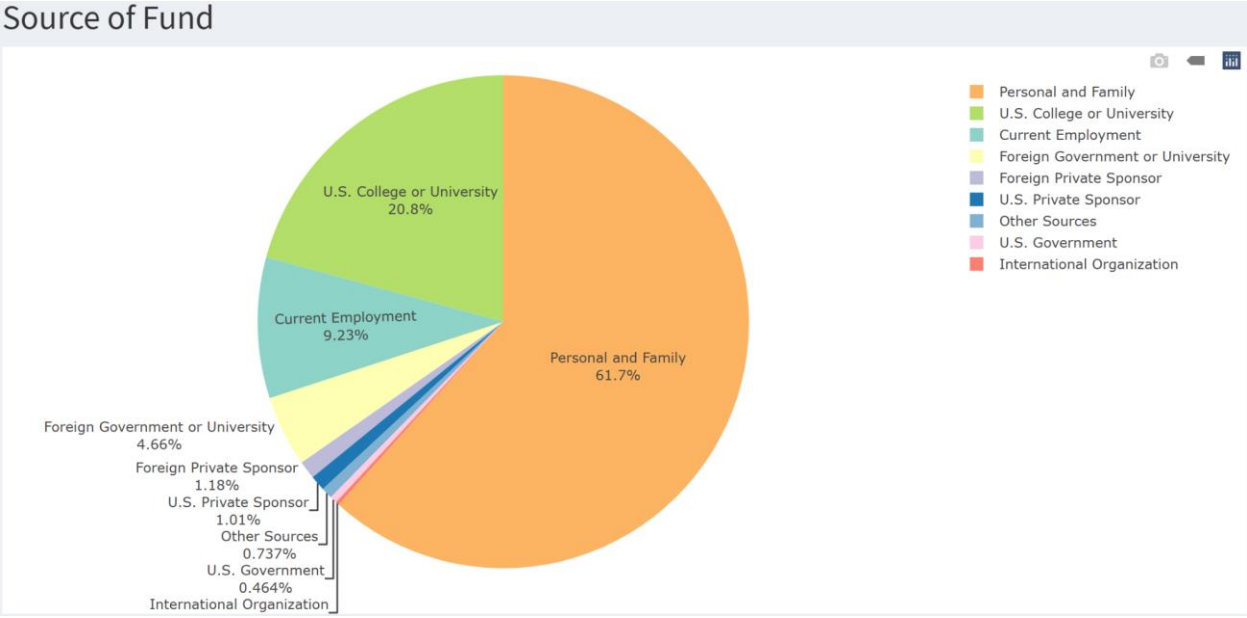

Fig 6. The interactive pie chart shows the source of funding for international students in the US

The first design of the project uses a dynamic pie chart to visualize the distribution of various sources of funding for international students in the US, which categories include Personal and Family, US College or University, Current Employment, etc (Fig 6). It explicitly shows their source of funding with a clear label for each category's proportion. Educational institutions can utilize this chart to better understand international students' source of funding, which enables more targeted financial support and program planning (Bound et al., 2021). Adjusting the years range of the graph can show shifts in funding sources over time due to changes in economic conditions or educational policies. Moreover, the parameters are crucial: '*labels*' are set to funding source names, '*values*' represent the monetary amounts or counts, and *'type'* is specified as '*pie*' to denote the style of the graph. The uses of '*hoverinfo*' and '*textinfo*' can provide detailed data and percentage values on hover and make the visualization be more informative.

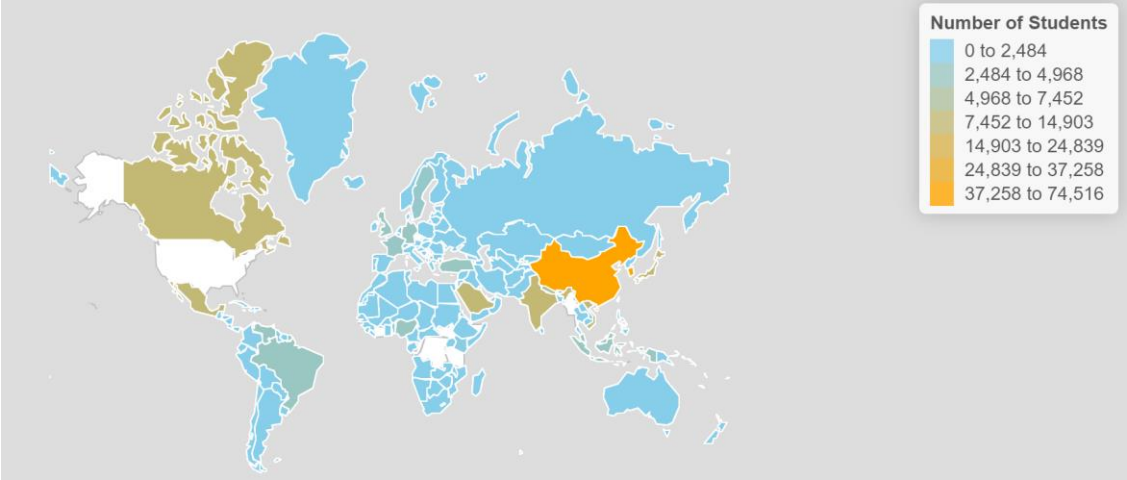

Fig 7. Interactive choropleth map shows the countries origin for international students in the US

The second design is an interactive choropleth map showing each country as origin for international students in the US (Fig 7). By using a constrained base map, the graph integrates features like clicking, zooming, and hovering for detailed information using the R package "*tmap*" combined with "*leaflet*". In this design, dynamic intervals were setup as the legend to visualize the distribution of student numbers worldwide across different periods. Universities and college admission offices can benefit from this visualization by identifying key recruitment areas and understanding regional trends, which can guide their international marketing and outreach strategies (Goralski & Tootoonchi, 2015).

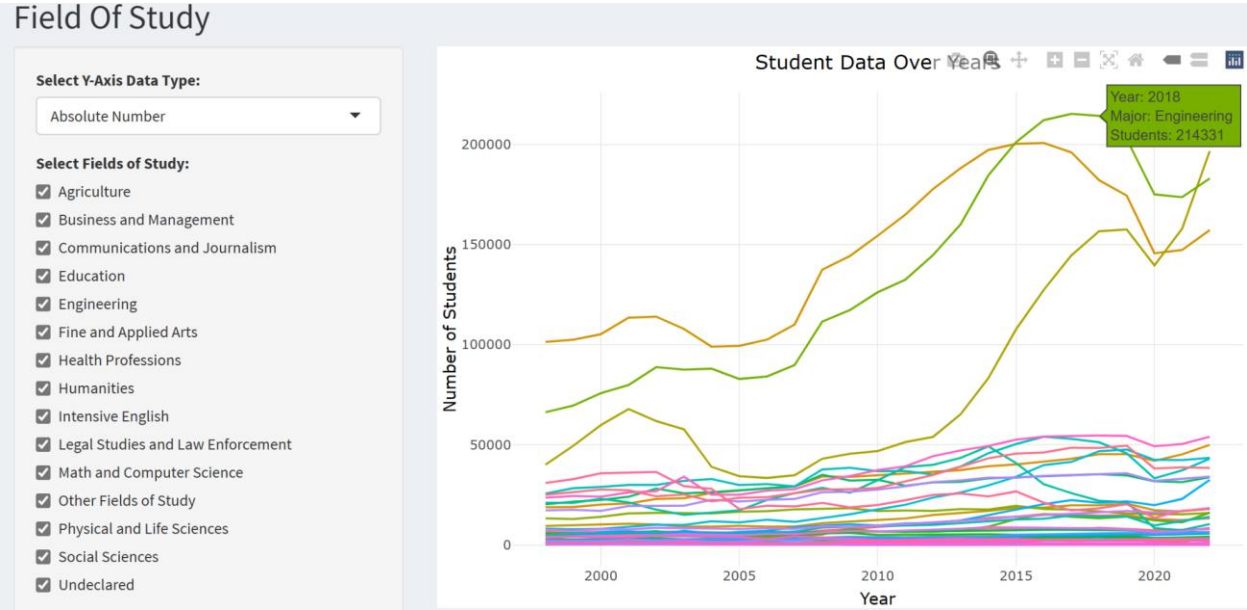

Fig 8. Interactive line graph shows the field of study for international students in the US

The third design is a dynamic line graph showing the field of study for international students in the US over years (Fig 8). The graph explicitly shows15 fields of study in the dataset, including agriculture, business and management, communications and journalism, and education. Here the employment of the line graph aims to show the differences among various study fields while highlighting their students' number changes over time. Users can select the fields of study to display and choose whether to view the data as an actual number or a percentage of total students. The actual number of students in each field can bring essential information for researchers to understand the raw scale of interest or participation. In addition, users can hover their mouse over the graph to view detailed student numbers in different fields of study for each year.

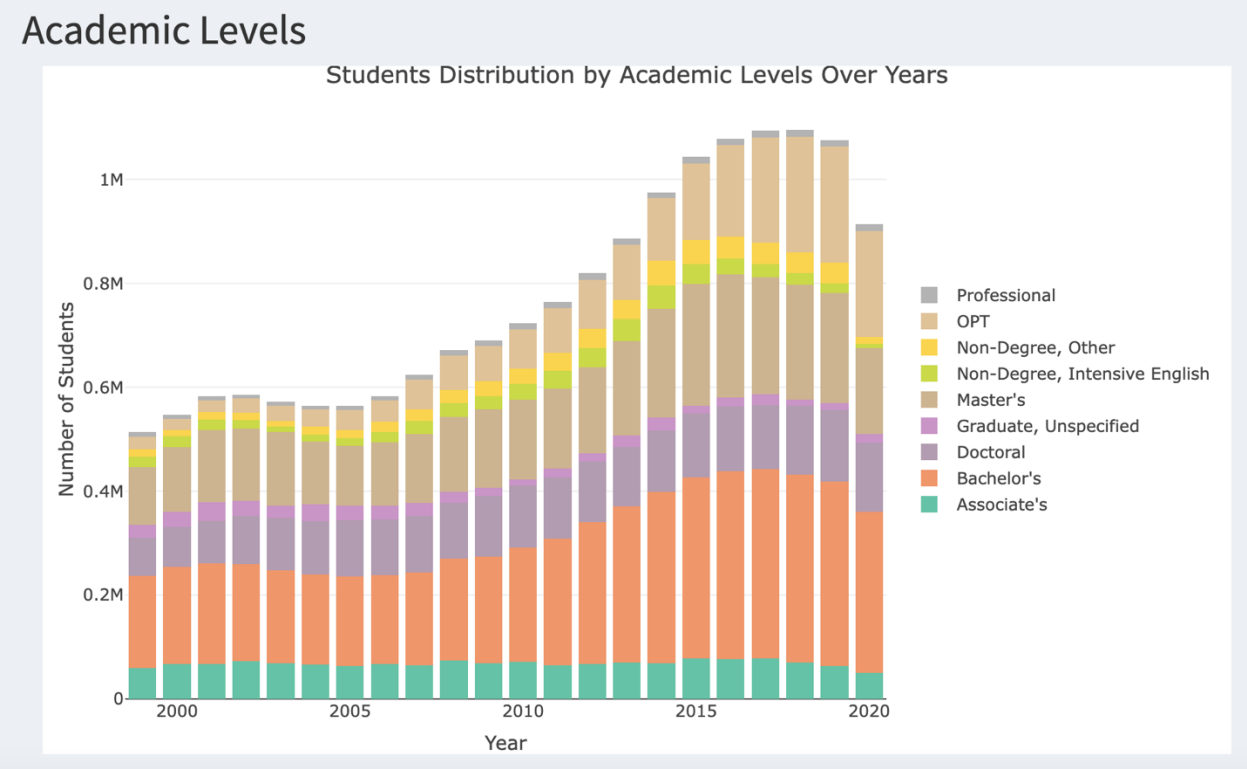

Fig 9. Interactive stacked bar charts show the student distribution by academic level (Students' type) over the years

The last design is a stacked bar chart that is utilized to illustrate the distribution of students across nine academic levels over time, including professional, OPT, non-degree, other, non-Degree, and intensive English, etc (Fig 9). By clicking on the legend, users can select and compare specific academic levels. Similar to our previous design, we also use '*hoverinfo*' and '*textinfo*' to provide detailed data. The use of stacked bars allows for an effective visualization of the growth or decline of each academic category over the years within the total student body, making it easier to compare trends across different fields as the total student population changes significantly over years. Institutions can use the comparisons to plan their future academic programs, to facilitate constructions to meet future students' needs, or to admit less students of certain fields of study due to limited resources (Amuedo-Dorantes et al., 2019).

Conclusion

In this project, the visualizations show international students' data in the US among four features: source of funds, countries of origin, field of study, and academic levels. Several pieces of literature show each feature’s importance to US institutions. Each feature has an interactive chart that shows the students' data over time or in a specific year. Users can obtain clear and detailed demographic data and trends across these features when they’re interested. Therefore, this Design Studio has the potential to serve as a comprehensive tool that enables U.S. educational institutions to optimize their recruitment strategies for international students.

References

Amuedo-Dorantes, C., Furtado, D., & Xu, H. (2019). OPT policy changes and foreign-born STEM talent in the U.S. *Labour Economics, 61*, 101752. https://doi.org/10.1016/j.labeco.2019.101752

Bound, J., Braga, B., Khanna, G., & Turner, S. (2021). The globalization of postsecondary education: The role of international students in the US higher education system. *Journal of Economic Perspectives, 35*(1), 163–184. https://doi.org/10.1257/jep.35.1.163

Fischer, K. (2024, March 20). The Chronicle of Higher Education. Retrieved from https://www.chronicle.com/newsletter/latitudes/2024-03-20

Goralski, M. A., & Tootoonchi, A. (2015). Recruitment of international students to the United States: Implications for institutions of higher education. *International Journal of Education Research, 10*(1), 53+. https://link.gale.com/apps/doc/A417473393/AONE?u=anon~7e0e9aa9&sid=googleScholar&xid=fdc90c07

Mode Help. (n.d.). Visualization filters: Query and analyze data. Retrieved from https://mode.com/help/articles/viz-filters

Riffe, T., Sander, N., & Klüsener, S. (2021). Editorial to the special issue on demographic data visualization: Getting the point across – Reaching the potential of demographic data visualization. *Demographic Research, 44*, 865–878. https://www.jstor.org/stable/27032938

Saluja, G. (2023, July 1). International students in the US: Visualized. *Medium*. Retrieved from https://medium.com/@gsaluja/international-students-in-the-us-visualized-f32aac5eee89

Watanabe, T. (2024, January 10). International student demographics. *Kaggle*. Retrieved from https://www.kaggle.com/datasets/webdevbadger/international-student-demographics