

Group1: Alena Zheng Jiaqi Wu Peiyuan Li Qi Qiao Yanqi Liu

Stat436 LEC001

Project Milestone 1: Study Goals and Literature Review

Word Count: 400

### Modality Decision

Our team has selected the Design Studio modality for our data visualization project, a decision inspired by the unique blend of creativity and interactivity that this approach offers. Design Studio stands out for providing a versatile creative environment that encourages the exploration of innovative visualization techniques and ideas, challenging conventional data presentation norms. This modality emphasizes user engagement through interactive interfaces, allowing users to directly interact with data, thereby fostering a deeper understanding and greater interest in the underlying stories. Furthermore, Design Studio leverages dynamic visualization and storytelling techniques to transform static data into compelling narratives, making complex information more accessible and engaging to a broad audience. By prioritizing creative freedom, user participation, and dynamic storytelling, our team aims to create visualizations that not only inform but also inspire and engage users, turning data exploration into an immersive experience.

## Motivations and Goals

Our project, guided by the Design Studio modality, focuses on an international-student-demographics dataset retrieved at <https://www.kaggle.com/datasets/webdevbadger/international-student-demographics/>. We will harness data visualization to serve two primary clients: international student offices in the U.S. (and related departments) and its prospective international students. Our motivation stems from the pivotal role of international students in enriching the cultural and intellectual diversity of U.S. educational institutions and their significant economic contributions. To address the needs of our clients, we aim to create interactive visualizations that illuminate the varied panorama of international education in the U.S.

For educational institutions, our goal is to underscore the value of international students by showcasing their distribution across academic fields, origins, funding sources, and academic levels. This insight is intended to inform recruitment strategies, support services, and program offerings tailored to international students' needs.

For prospective international students, who typically face the daunting task of navigating their academic futures abroad, our visualizations aim to provide a user-friendly platform to explore academic programs by field of study, funding opportunities, and the geographical distribution of students, thereby facilitating informed decision-making and planning.

With the above motivations and goals in mind, we decided our key visualization tasks to be:

- a) field of study preferences: highlighting trends in popular disciplines
- b) funding sources: revealing how students finance their U.S. education
- c) student origins: mapping where students originate to identify key regions
- d) academic level: detailing enrollment by academic level to reveal preferences
- e) overview (genders): displaying gender distribution to support diversity initiatives.
- f) interactive features: enabling personalized exploration by state, field, or origin

Through these visualization tasks, we aim to empower institutions to enhance their international programs and assist students in making informed academic choices, ultimately fostering a more informed and engaged international student community in the US.

## Literature Review

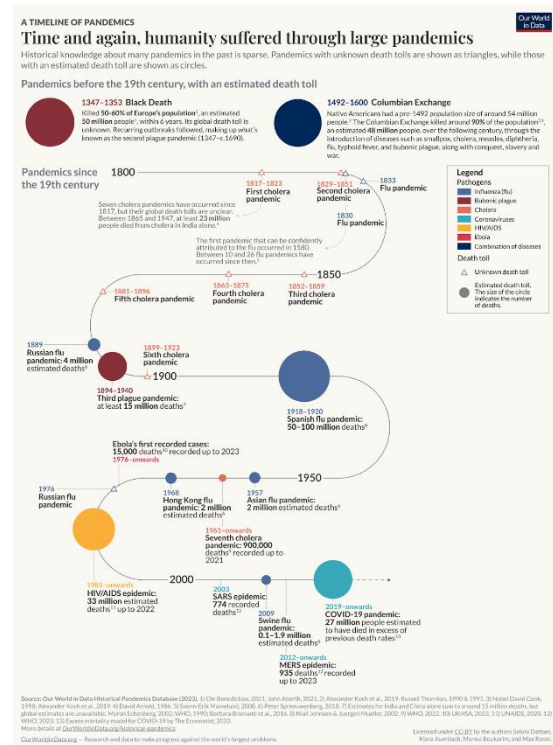
The exploration of data visualization techniques and their applications in diverse fields has become increasingly significant with the advent of big data and interactive technologies. This section reviews relevant literature, tools, and methodologies that underpin our project's approach to creating impactful visualizations for international student demographics.

### Visualization Tools and Interactive Techniques

Gapminder harnesses the Bubbles technique to dynamically represent complex statistical data, where each bubble's size and position illustrate various demographic and economic indicators. This interactive approach provides a clear snapshot of each country's status on these indicators, enabling users to track temporal changes through an interactive timeline. Using color coding to differentiate between categories further enhances intuitive understanding and engagement, making complex information both accessible and captivating (Gapminder.org).



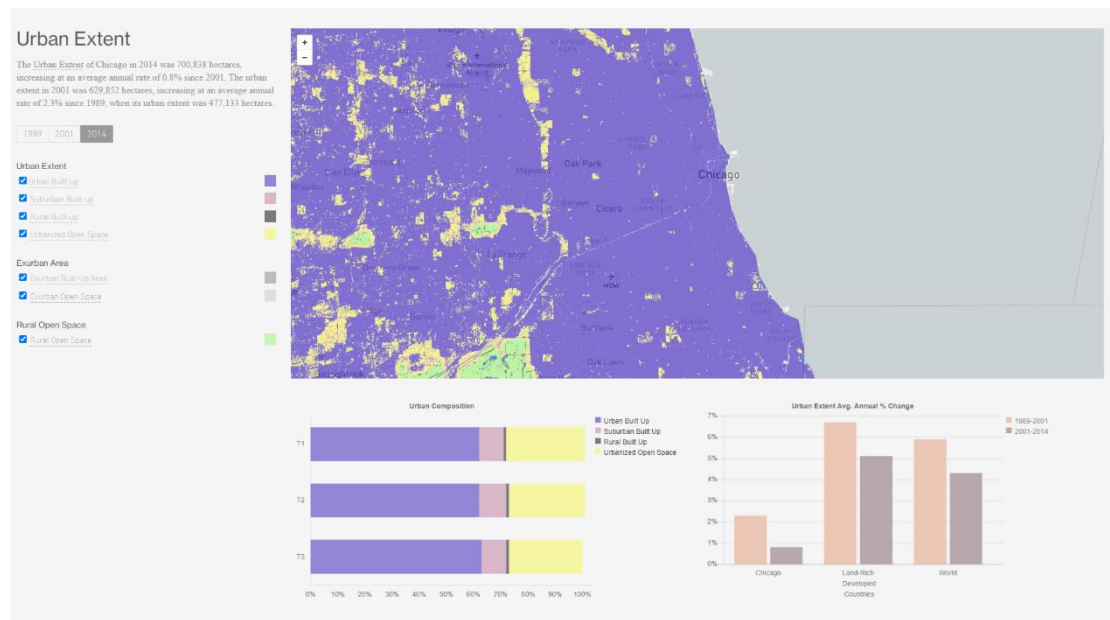
Mode.com emphasizes the significance of filters in data visualization, offering users the ability to tailor data exploration according to specific criteria. This feature promotes a detailed examination of datasets, allowing for the identification of trends that might be obscured within broader data sets. Filters thus play a crucial role in deepening analytical depth and personalizing the data exploration experience (Source3).



The screenshot shows the Tableau Desktop interface. At the top, the user is 'Ying Luo' and the view is 'Visualization Filter Example'. The main workspace displays a line chart titled 'Click to add title'. The chart's X-axis is 'WEEK(order\_created\_date)' and the Y-axis is 'order\_created\_do\_w'. The chart shows a fluctuating green line representing weekly order creation counts over time, from early 2014 to early 2017. To the right of the chart is a 'SETTINGS' panel with tabs for 'Line Chart', 'X-Axis', 'Y-Axis', and 'Color'. Below these is a 'FILTERS' section. At the bottom of the interface is a data table with columns for 'WEEK(order\_created\_date)', 'Measure Names', and 'Measure Values'.

	WEEK(order_created_date)	Measure Names	Measure Values
1	2013-12-01 00:00:00	order_created_do_w	2875
2	2013-12-08 00:00:00	order_created_do_w	6011
3	2013-12-15 00:00:00	order_created_do_w	3272
4	2013-12-22 00:00:00	order_created_do_w	2496
5	2013-12-29 00:00:00	order_created_do_w	3290
6	2014-01-05 00:00:00	order_created_do_w	6180

The Atlas of Urban Expansion's presentation of Chicago employs an innovative combination of maps and bar charts to facilitate a multi-dimensional understanding of urban development. Color coding distinguishes various urban areas, enhancing the map's clarity and usefulness. Interactive features invite users to engage more deeply with the data, transforming passive observation into an active exploration. This approach not only makes the visualization more user-friendly but also significantly enriches the analytical experience by enabling detailed urban expansion analysis (Source6).



## Techniques for Effective Data Visualization

The evolution of data visualization technologies has markedly improved how we engage with and interpret data. At the forefront, Microsoft Power BI distinguishes itself by delivering a comprehensive suite of tools designed for data aggregation, analysis, and visualization. It's celebrated for its user-friendly interface and interactive capabilities, making it a favorite among professionals seeking to democratize data

insights within organizations (Technique1). In parallel, the Shiny package from R has revolutionized the creation of interactive web applications directly from R, allowing for seamless integration with other graphing libraries such as ggplot2 for an enriched data exploration experience (Technique2).

Leaflet for R stands out for mapping and spatial analysis by offering advanced features for interactive maps, making it an indispensable tool for projects requiring nuanced geographical data visualization (Technique3). Similarly, R charts provide comprehensive documentation and support for various R graphing libraries, ensuring users can efficiently leverage Leaflet, ggplot2, and base R for their visualization needs (Technique4).

Plotly and Tableau further enhance the landscape of data visualization in big data contexts. These platforms are lauded for their simplicity and intuitive interfaces, which empower users to derive insights quickly through diverse graphical representations. Their accessibility is particularly beneficial for individuals without programming expertise, enabling business analysts and others to conduct deep dives into data with minimal effort (Source4). Together, these tools and techniques represent a leap forward in making data visualization more interactive, accessible, and insightful, catering to a wide array of analytical needs and user backgrounds.

## Application to the International Student Demographics Project

The reviewed literature and techniques provide a foundation for addressing the visualization needs of our project. By incorporating interactive elements and focusing on clear, communicative graphics, our visualizations will effectively convey the diverse panorama of international education in the U.S. to our target audience. Using tools like Power BI, combined with interactive mapping and customized filters, will enable us to create engaging and informative visualizations that address international students' and U.S. educational institutions' specific interests and needs.