CS171 Final Project Process Book

Dream Seekers

Leader: Catharine Wu (21425273, jwu@g.harvard.edu) +9254514517
Group Member 1: Zhaodong Chen (60940029, zchen01@college.harvard.edu) +18083842696
Group Member 2: Kidist Alemu (41424133, kalemu@college.harvard.edu) +2676014482
Zoom link:
https://harvard.zoom.us/j/7490397286?pwd=TlFyaXdKRDRwcGJrRHdqaTVuZVZDQT09

Week 8 & 9

Team Agreement

- Each team member will be responsible for writing part of the code, but all team members should be involved with the technical aspects of the project. All code should be inspected on a regular basis and documented well.
- All team members should share updates and opinions on weekly meetings, final design decisions will be discussed among all members; fair compromises should be made when necessary.
- Individual tasks should be made clear during weekly meetings. Work hours should be split as evenly as possible (actual task output may differ based on an individual's ability / previous experience). This ensures not only fairness but also learning opportunities for everyone. We will keep each other accountable so that one person does not work too much / too little.
- We will use a Git workflow to aid our progress as a team and help us split up the work of coding. Code changes will be committed/pushed frequently.
- Work will not necessarily be done together in person, but good communication via Facebook Messenger is expected in a timely manner. Work may be done remotely as long as collaboration and communication are done well.
- Team meetings will be held on TBD via Zoom. Team members can schedule additional meetings if needed via Messages.

Signatures: Zhaodong Chen, Kidist Alemu, Catharine Wu
Date:10/23/20

Topic: International student experience in the U.S.
Motivation: All three of us have international heritage
Brainstorm Ideas:

1. Percent of international students from different countries at U.S. universities
    a. Time series of the percentage
    b. Type of universities, e.g. public vs. private, liberal arts colleges, etc.

    c. Social/economic background
    d. Where they end up(if there is data on that)
2. Individual/country-level data, testimonials...
3. Application process/timeline
4. Metrics for Harvard International Student Data
    - Year
    - Origin
    - Number of students admitted from a country
    - Number of students that committed to Harvard
5. Initial Visualization Ideas
    1. Parallel coordinates
    2. Radar plots
    3. Map

Questions:

1. How does the number/percentage of international students change over time?
2. Where do international students come from?
3. What type of universities do international students tend to choose?
4. What are the popular majors for international students?
5. What's the typical source of funding?
6. What are the social/economic backgrounds of international students?
7. Do international students stay in the U.S./go back to their home countries after graduation?
8. How's the application process like? (timeline, fee, …)
9. What's the distribution of international students at Harvard?
10. What socioeconomic and demographic inequities are observed from the trend of admission?

Related Work/Data Source:

1. https://opendoorsdata.org/data/international-students/
2. https://educationdata.org/international-student-enrollment-statistics
3. https://www.iie.org/Why-IIE/Announcements/2019/11/Number-of-International-Students-in-the-United-States-Hits-All-Time-High
4. https://www.migrationpolicy.org/article/international-students-united-states-2017
5. https://www.statista.com/chart/20010/international-enrollment-in-higher-education/

Abstract: The motivation of our project is that all three of us have an international heritage with two of us being international students. We are interested in exploring the composition of international students in U.S. universities, e.g. what's the percentage for different origin countries, how the statistics have been changing over the years, and social/economic background. We also want to learn those students' application process and experience living and studying abroad. The datasets we have in mind are as follows: 1. The percentage of students from different foreign countries for different types of universities and overall. 2. Demographic information on international students such as age, family income, etc. 3. Where they end up(if there is data on that), i.e. if they decide to stay in the U.S. after graduation. 4.

Students' application experience: timeline, fee, etc. We plan to acquire the data from the College Board, IIE, Statista, international offices, and other education and news sources.

Main Question: What is the journey of international students to the US?
Audience: Our primary audience is anyone in the education field and students interested in applying to the US, but this may also extend to a broader audience.

Data Cleanup: We'll focus on places of origin, distribution in various institutions, applicant backgrounds, and application process. We plan to mainly use the formatted csvs from opendoordata.org, which are relatively clean. But we do expect some missing values, which as of now would be left as it is. The opendoordata.org datasets are industry-standard data on international students used by various international offices so we believe that it is clean enough for our visualizations.

Week 10

You will submit your dataset(s) that you are planning to use for your project. The data should be cleaned as much as possible at this point. Each team member will then create individual sketches of visualizations using a pen and paper that might answer your questions. In addition, you can create a few exploratory visualizations in Tableau. As a team, you will then decide which of these visualizations and insights you plan to pursue in your project. Finally, you will create an initial storyboard of the data story that you plan to tell. Please note, we expect you to have at least one **novel** visualization in your final implementation. Check the "Technical Requirements" in the final project rubric for details.

Here are the previous design sprint instructions that may guide you through the process:

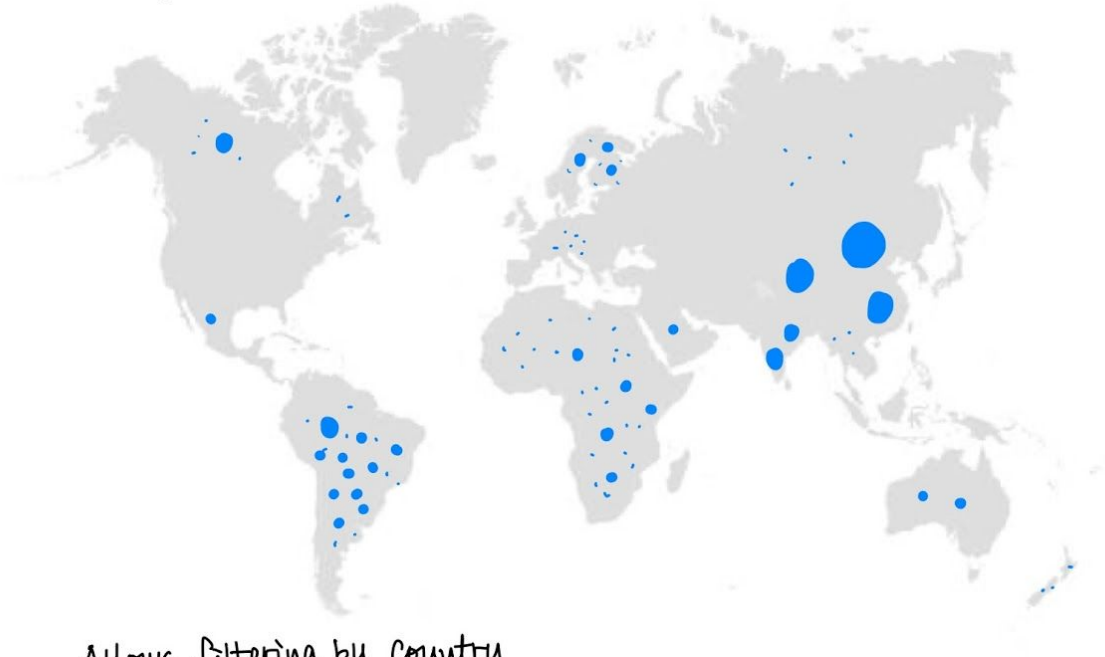
- Sketch
- Decide
- Storytelling

Please check the final project overview for more information about the final project.
Please note that using Tableau is not required, it is optional but encouraged.

Sketch

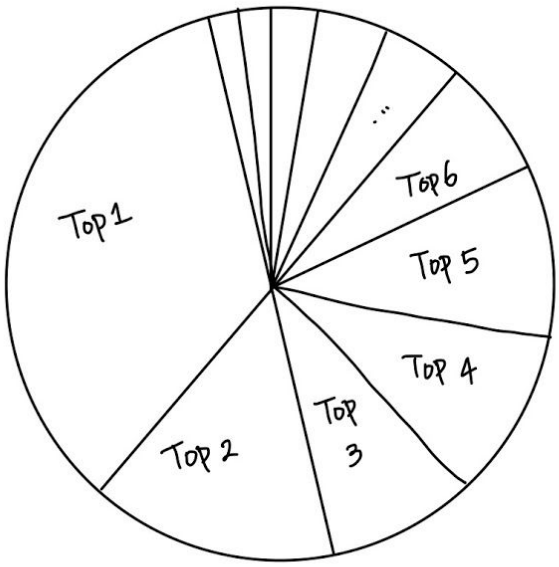
Catharine's sketches

① # of international students by place of origin.



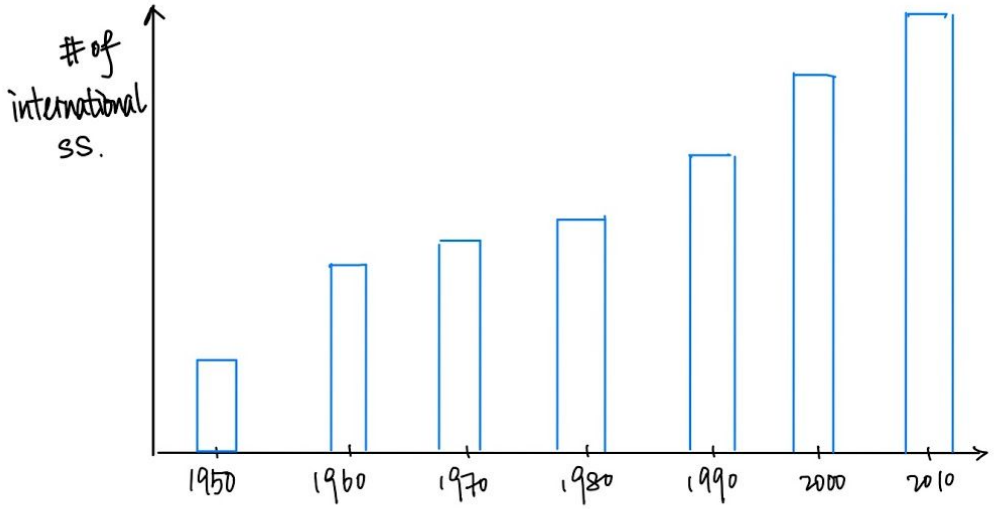
Allows filtering by country

② Percentage for each country

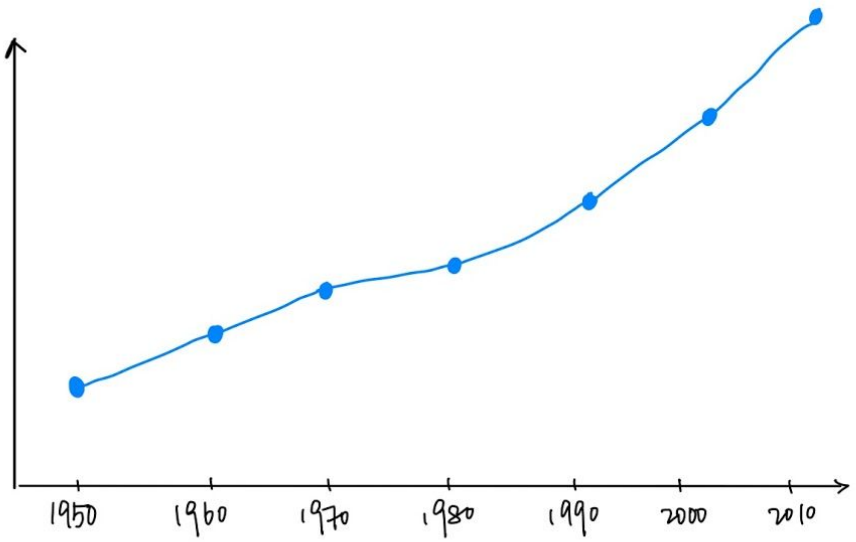


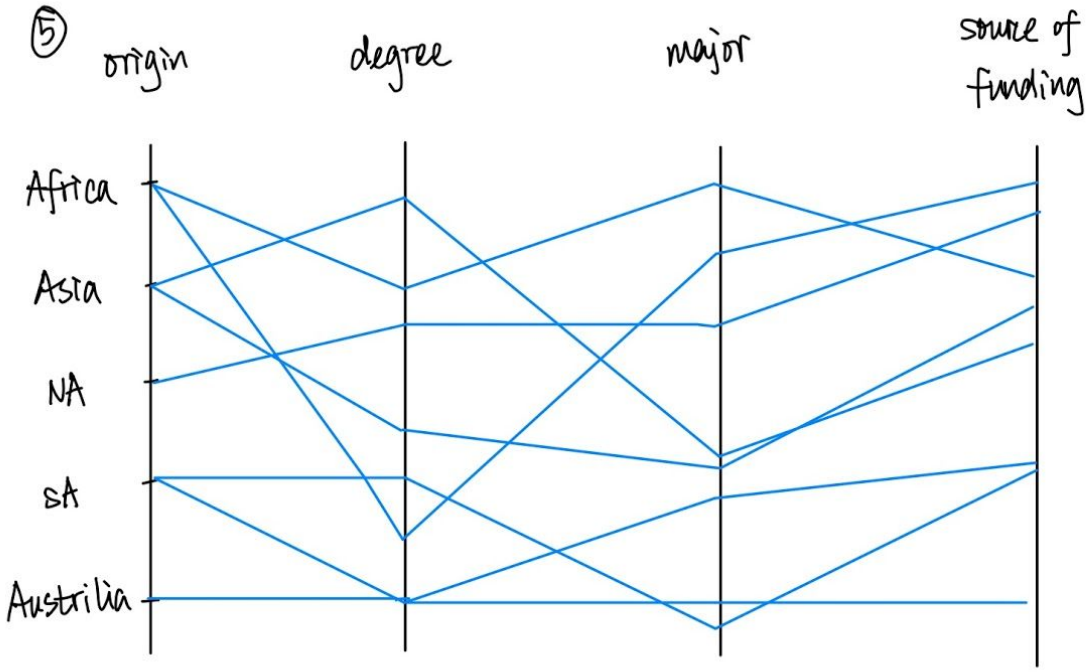
changes over time: (Allow filtering by year)

③ By number



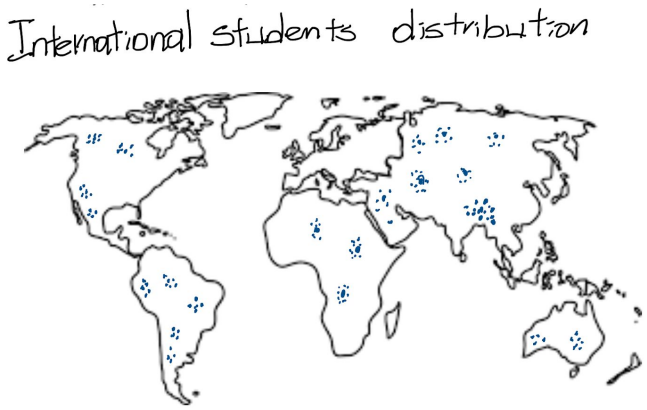
④ By percentage (potentially by country too)





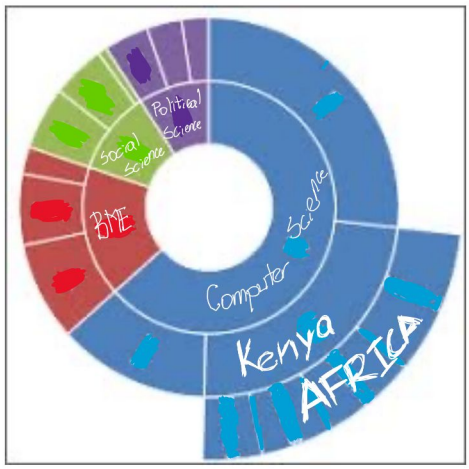
Kidist's Sketches

- 1. World Map: Countries, with dots of students with hover function of directed line from origin to destination



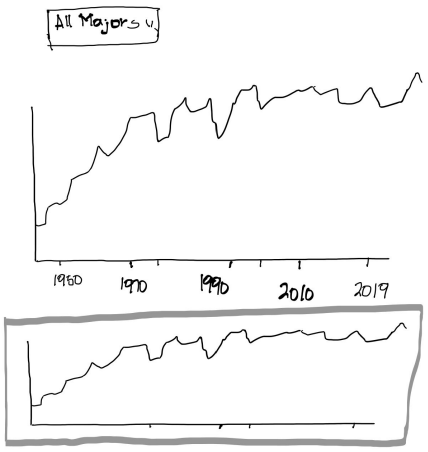
- 2. Sequence sunburst: continent, country, spilt by concentration
    - https://bl.ocks.org/kerryrodden/766f8f6d31f645c39f488a0befa1e3c8

Common Major pursued by internationals across the countries



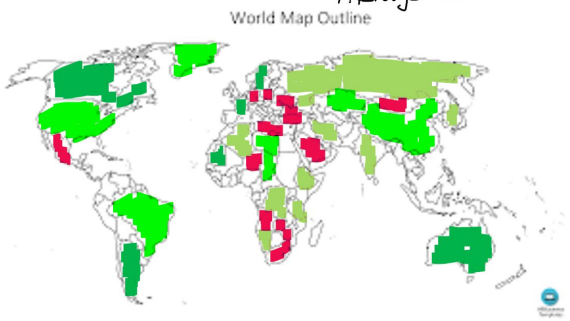
- 3. Trends in concentration choice

Trend of concentrations



4. Average annual income vs Average cost of college attendance

Average annual income Vs
Average cost of college applications



5. College Application Time Series

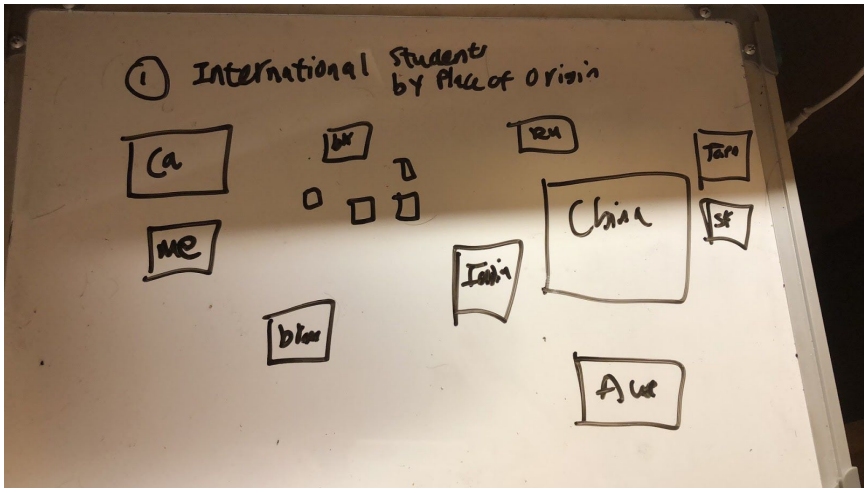
Inspiration : https://github.com/marmelab/EventDrops

Even time Series

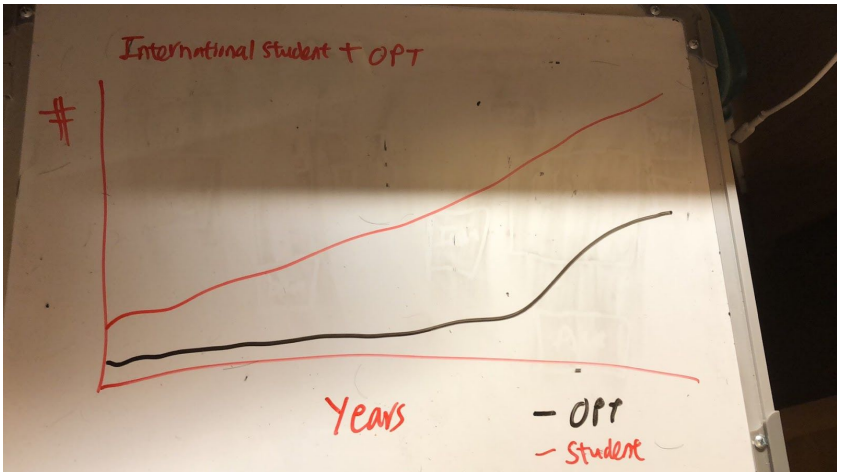
Month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Cost			Subject Test		Decisions come out!					SATs for early Apps		SATs for regular Apps

Zhao's Sketches

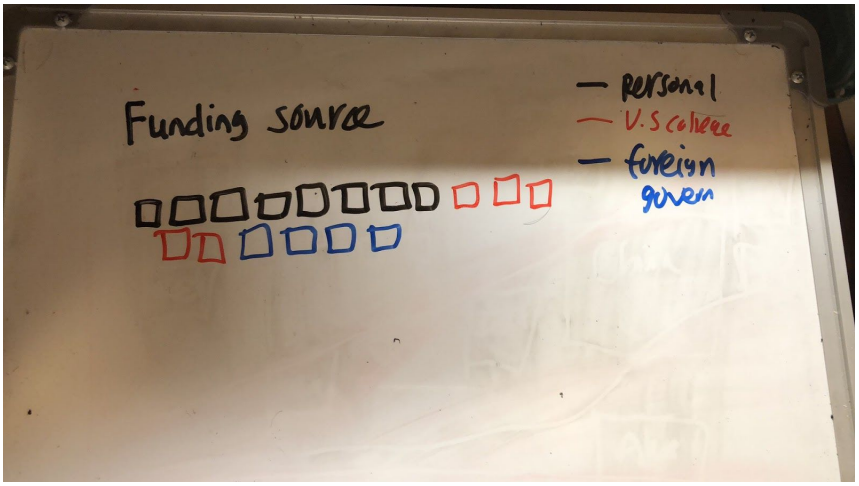
1



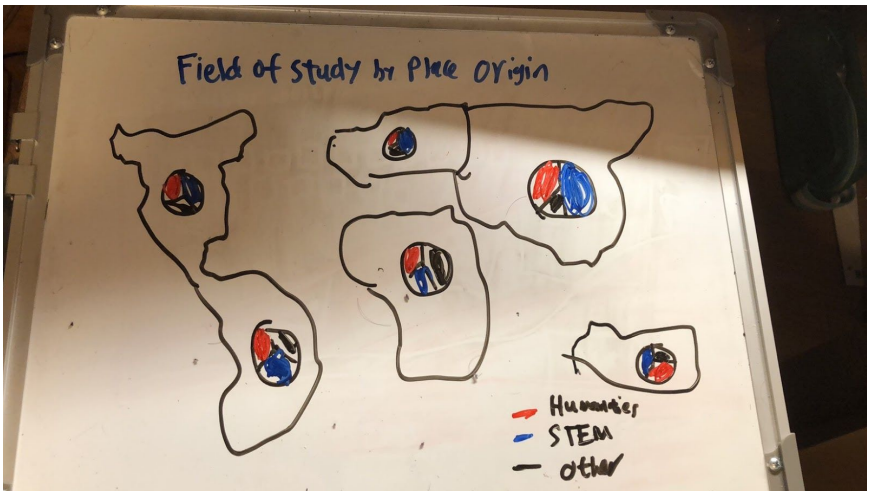
2



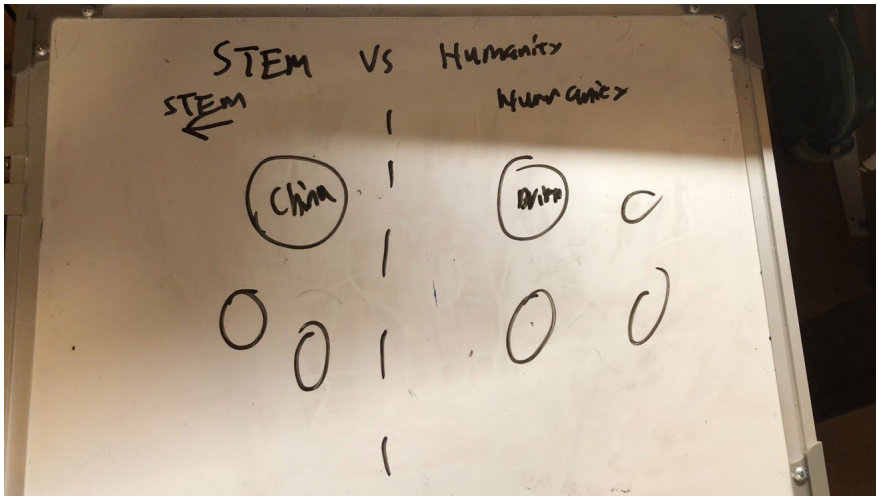
3



4



5



Final decision on Sketches:

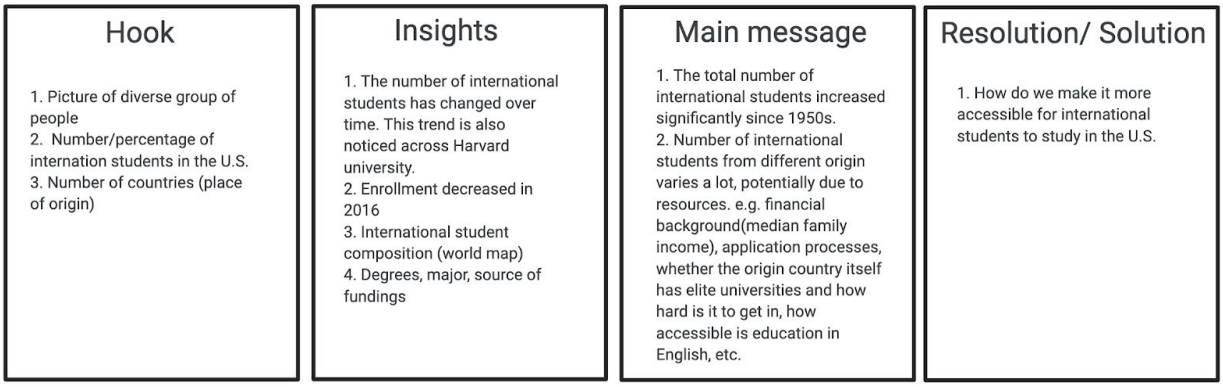
1. Map
    a. Combining **Catharine's 1** and **Kidist's 1** - having a world map with international student populations. Add animation of directed lines moving from the origin country to destination(the U.S.), when hovering over an origin country, both the origin and the U.S. are highlighted with a line flowing from origin to destination.
2. Bar chart/line chart indicating number/percentage of international students over the year.
    a. Catharine's 3 & 4
3. More detailed information
    a. Major, source of funding, degree, etc. both general & specific: **Catharine's 5, Kidist's 2, and Zhao's 3**
    b. OPT: **Zhao's 2**

Questions we want to answer:

Where are the students coming from? Proportion?
What is the Major/Concentration breakdown based on countries?

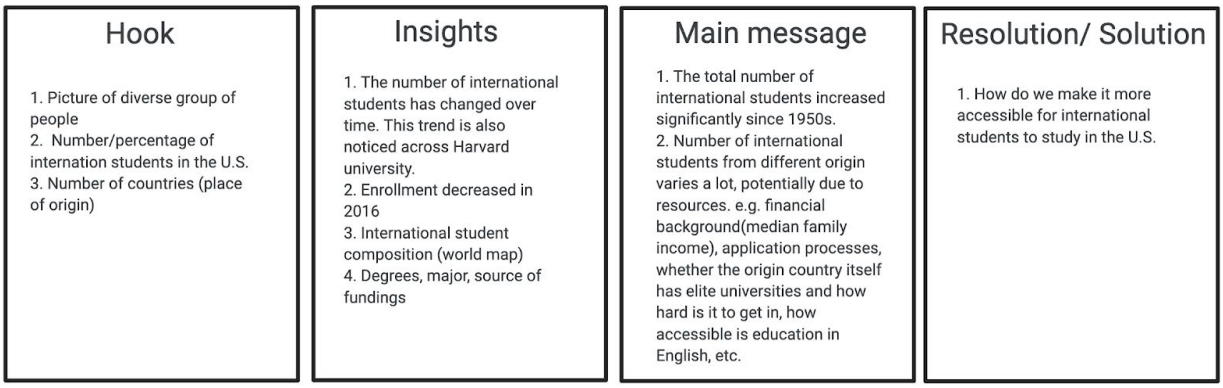
What is the funding source breakdown based on sources?
Population change of International students over time?
How could we boost international students from countries with few international students? (find the channels of actions)
What is the international student breakdown in Harvard?
How representative is Harvard compared to the overall picture in the US for International students?

Jamboard:
https://jamboard.google.com/d/13sh3c1FDIM67xrBfiswJi_2ioHN3ywOJIRqOaRInulU/edit?usp=sharing



Week 11

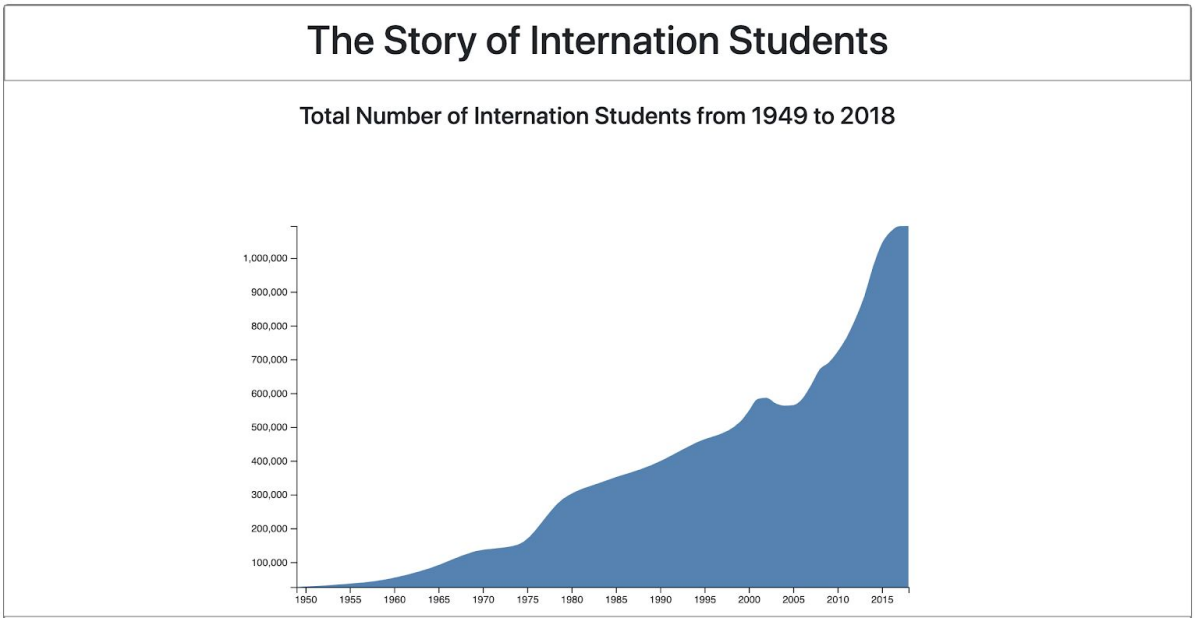
Storyboard

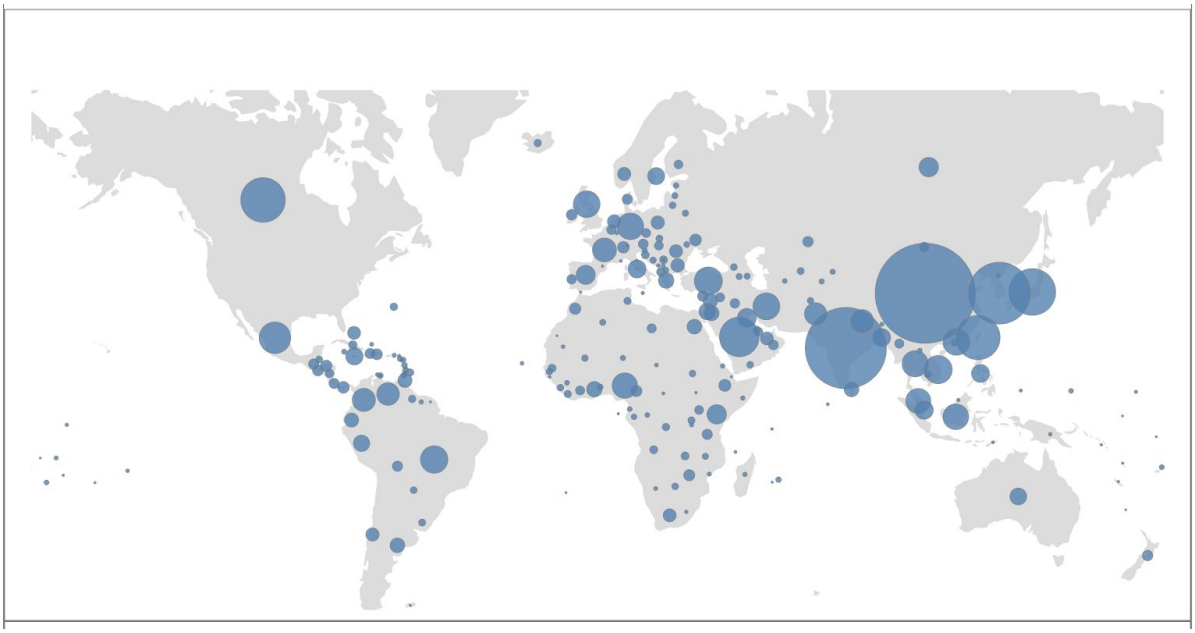


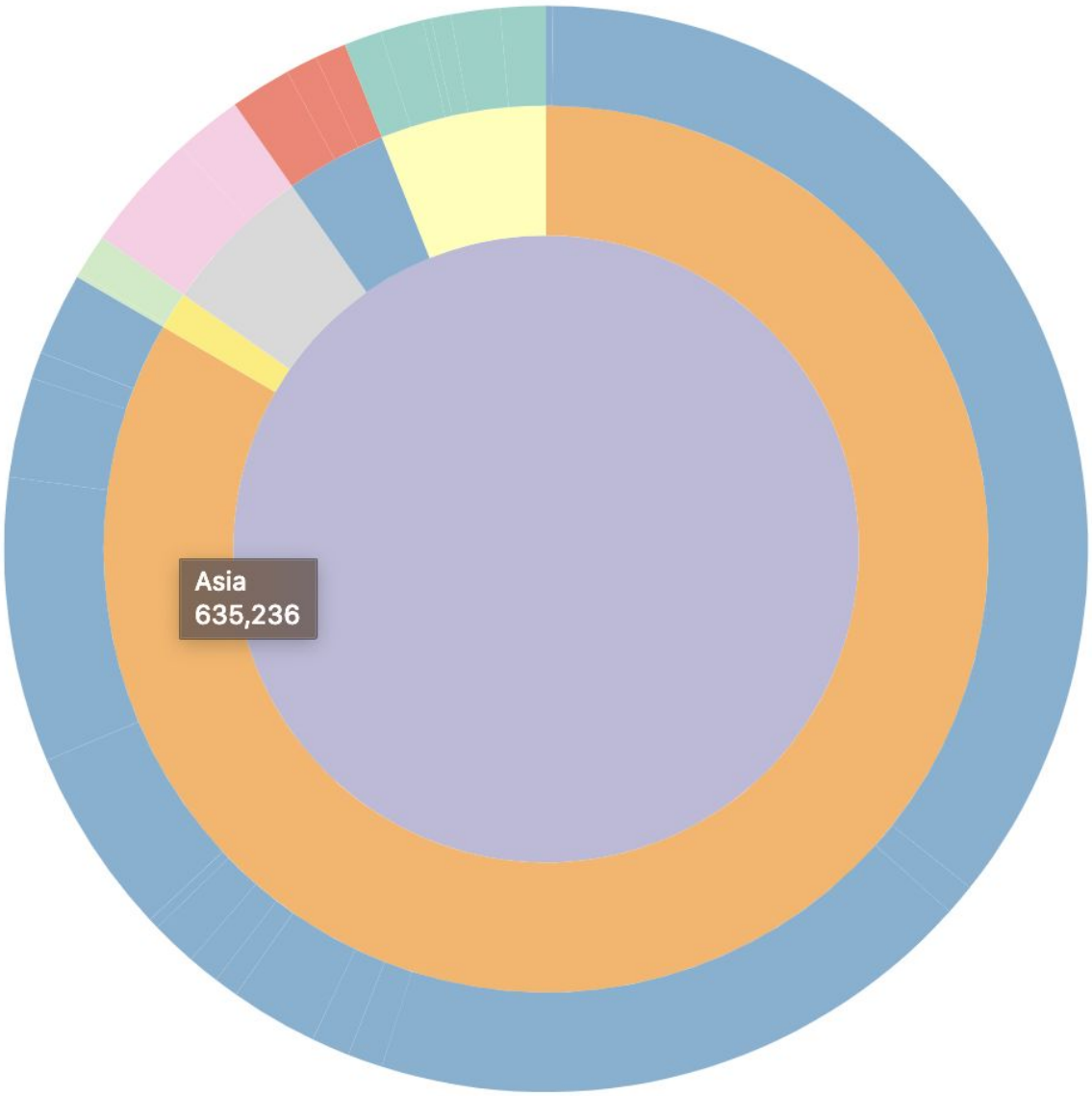
V1 Prototype Status

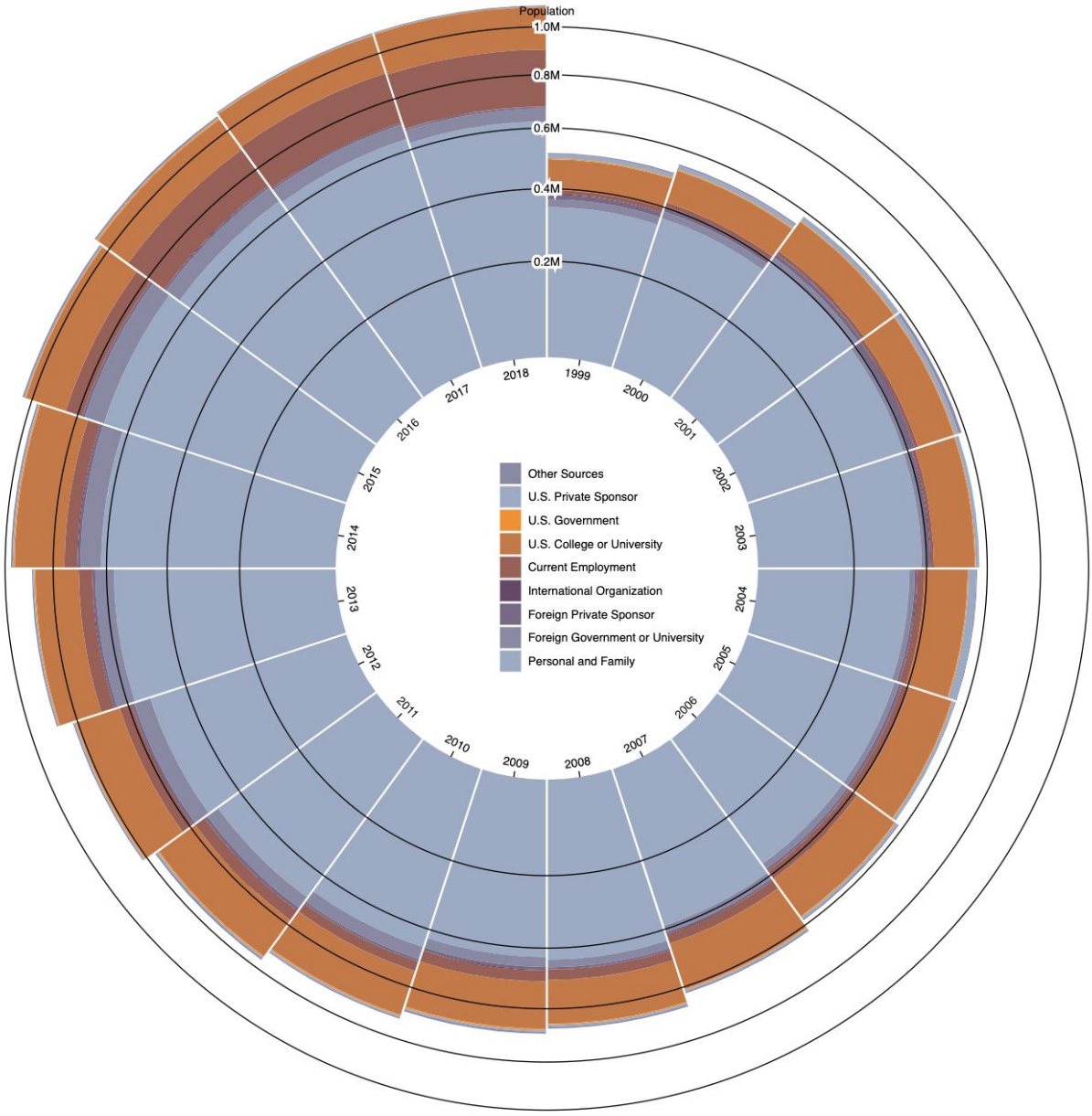
	Visualization	Assignee	V1 Prototype status	To do
1	Area/Line chart showing the change in number of international students over the years	Catharine	Area chart created	1.Add brush to allow filtering by year for map. 2.Make it a line chart and add animation to show increase in motion.
2	Map of distribution	Catharine	Bubble map created	1.Add filters: by year, region, etc. 2.Add tooltip to access more info by hoverover
3	Sunburst	Kidist	Sunburst Created	1. Add zoomable feature
4	Income Distribution	Zhaodong	Radial Stacked Bar Chart	Implement color choices and finetune the visualization, adding hover over

Visualizations






Asia
635,236



Data Processing

1. *Place of Origin* for number of international students per year: The dataset itself had a lot of rows/columns explaining the features, which needs to be removed. Also matched each country to their regions. The missing values are left as it is.
2. *Map data*:
    a. Geographical information was found here: https://www.kaggle.com/eidanch/counties-geographic-coordinates
    b. The difficult part was to combine this dataset with Place of Origin, a lot of the country names don't match or are missing, so I had to input them by hand.
3. Sunburst:
    a. Data source: *Census-Fields-of-Study-by-Place-of-Origin*

    b. Processing: Took the average for each concentration, group data based on continent, convert into nest json structure.
4. Income Distribution:
    a. Data source:https://opendoorsdata.org/infographic/primary-source-of-funding-of-international-students/
    b. Originally we were interested in replicating the pie graph used in the original data source. However, we found troubles in expressing the time dimension to add more clarification. Filtering does not do the trick since the overall total would be different. Proportionality loses information if we are not certain on the total comparison. So I was inspired by Mike Bostock's Radial Stacked Bar Chart since it is useful to show cyclical data which ours are. It is a creative solution that solves the year by year comparison as well shows the increase between the year and the proportion within the year. We have not finalized our color choices yet so the color of the prototype might not be best according to the class lecture but we will improve on that.

Webpage design and structure

For V1 prototype, we each created at least one visualization. Currently they are not integral, but the structure of the webpage (setup in html) can be found in the html file here: https://drive.google.com/drive/folders/19WPCR-OiVb0V1LNjHIITsFowfLlPVyVe?usp=sharing

---

Week 12

Page one: Image / video of international students

Page two: Line chart

Page three: Quote on the growth of international students in higher education institutions

Todo:

- Labels for axis
- Increase font size of the numbers on the axis
- Data points on the line that are hoverable for every point on the xaxis (we did this on one of the homeworks)

- Explanation of the visualization: Text appearing with description of the data that is being visualized.
    - Example:
    - “As of 2019, there are over 1 million international students in the US”
    - “On average **823,459** students from all over the world set on the journey to the US seeking higher education.”

Page three: Map

- Animation for the data points appearing in sequence
- Description of the visualization:
    - Highlight how different countries have varied number of int students
    - Add small blurb about the features that are possible with the visualization

Quote: Intl students from South America and Africa are disproportionately under admitted into higher education institutions

Page four: Sunburst

Page five: Quote-- According to --- the main funding source for higher education are personal expenditures. This

Page six: Radial

Page seven: Median income distribution map

Page Eight: Quote --- Lack of support for low income intl students

Page Nine: To attract more students, schools needs to provide more help.

Color scheme: https://colorbrewer2.org/#type=diverging&scheme=RdBu&n=9

['#b2182b','#d6604d','#f4a582','#fddbc7','#f7f7f7','#d1e5f0','#92c5de','#4393c3','#2166ac']

Font: Poppins --- Headings

Montserrat --- Body Text

Week 13

Test

Tester Name: Sarah Mokthar
Tester Email: smokhtar@mit.edu

General Observations from the think-aloud study:

After walking through the webpage, the tester found the visualizations to be very effective in communicating the intended message which was clear. The tester also found the visualizations were easy to understand and interactable. The tester liked the quotes in between the visualizations as it helped drive/introduce the main message intended to be extracted from the visualizations.
Map: not clear if origins are from
Hover detail was not obvious.

Sunburst: hover text cropped
Hierarchy is not clear
Very clear and complies a lot of data together

Radial graph: other source coloring blends in with the background
The number represented is not clear

Map:
Unclear what the number of students shows
Color are clear

Thoughts after observing other data story:
Our storytelling could set the scene more by telling the story of the international students
Both of our testers have international backgrounds so they were able to understand it right away
It might be harder for viewers who do not understand the international student process

Our visualizations are effective and innovative compared to some of the other ones
The right of amount of visualizations
The choice of quotes to separate the sections is effective

What does the tester like about your data story?

The test found the story line to be clear and nicely flowing. The transitions from one visualization to the next was clear and the flow was nicely put.

The tester found the visualizations to be interesting and intriguing. The sunburst in particular was nice to explore.
The maps are very clear and the radial chart makes it easy to understand the changes in quantity over the years and the distribution of funding across different sources.

What improvements does the tester point out?

- It took a bit of time to realize the text on the first map
- The exploratory options in the visualizations are not explicit
- Text formatting in the hovers of the visualizations
- The layers were not clear in sunburst
- Have clearer explanations on the exploratory options

Was the intended key message clear to the tester? Why or why not?

Yes it was. The tester was able to understand and deduce that there is a general trend in increase of international students in the US but there is a problem in equity for students that are from developing countries in Africa and South America.

Did the tester get your next steps or call to action? Why or why not?

Yes they did**.** The tester was able to watch the video linked to the next video and click on the links for additional information and next step actions.

Tester Name: Yiwen Wang
Tester Email:
Notes:

- Overall feedback: story is clear, visualizations look good
- General:
    - The bold texts are not distinguishable
    - Better color scheme
- Line chart: Add more explanations, animation looks good
- Map 1:
    - Background color/image for the U.S.
    - Add an arrow or little plane on connection
- Sunburst:
    - remove the question mark or change to one with no background color
    - The explanatory texts are not centered
    - Texts overlapping & too big
- Radial:
    - Numbers showing in hover over don’t make sense
    - Change to a categorical color scheme
- Map 2:
    - Round numbers in hover over, add units

- Legend for bubble size

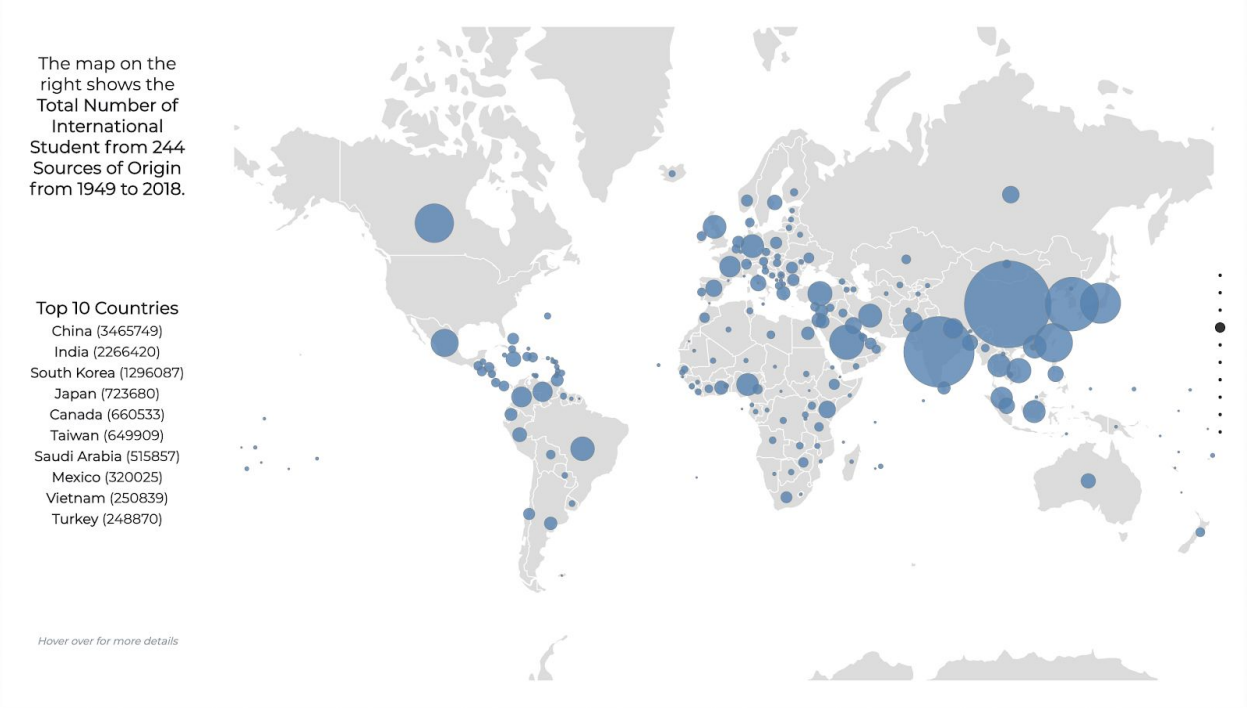
Improvements on the Visualizations

Map 1

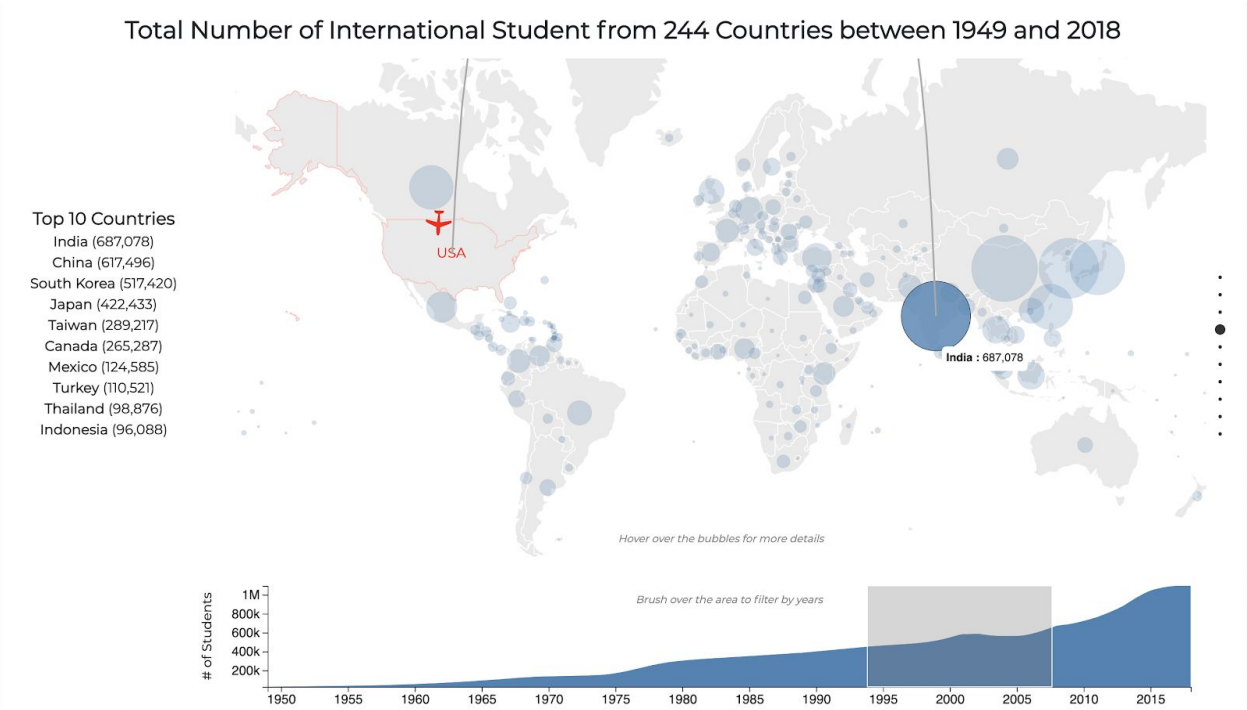
Changes made:

- Add area chart to allow filter by years
- Add a plane that flies from the origin country to the U.S.

Before



After

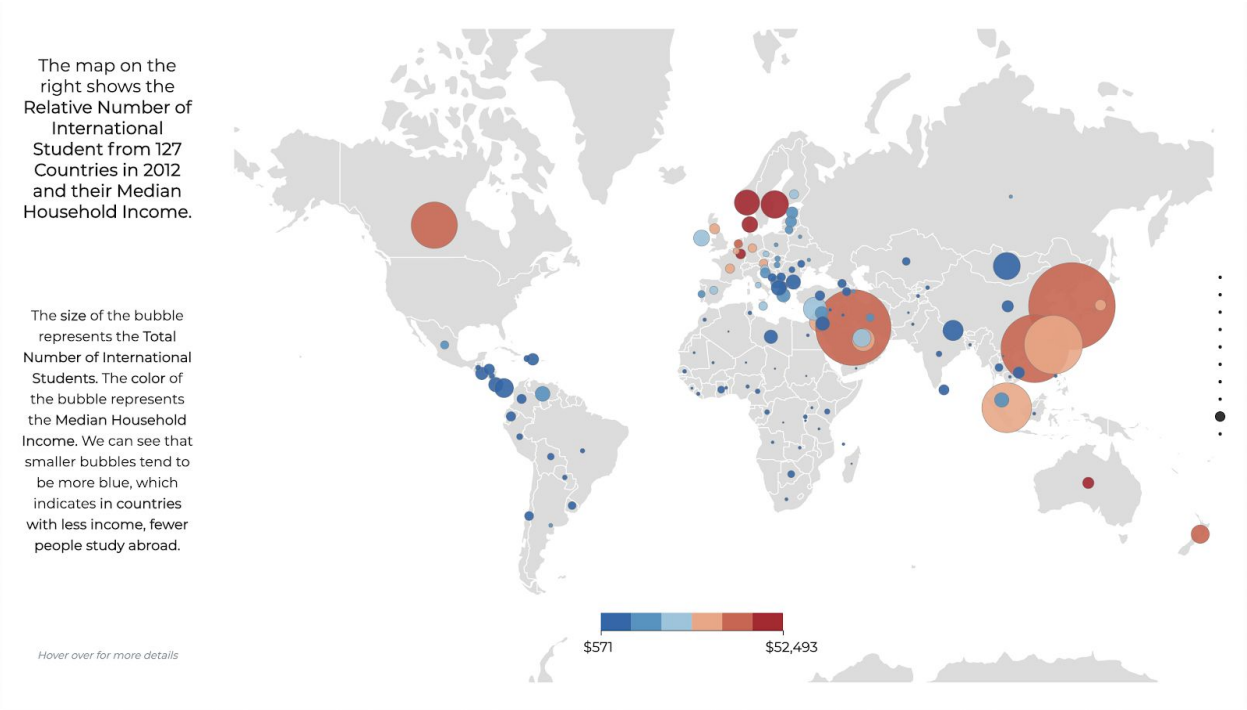


Map 2

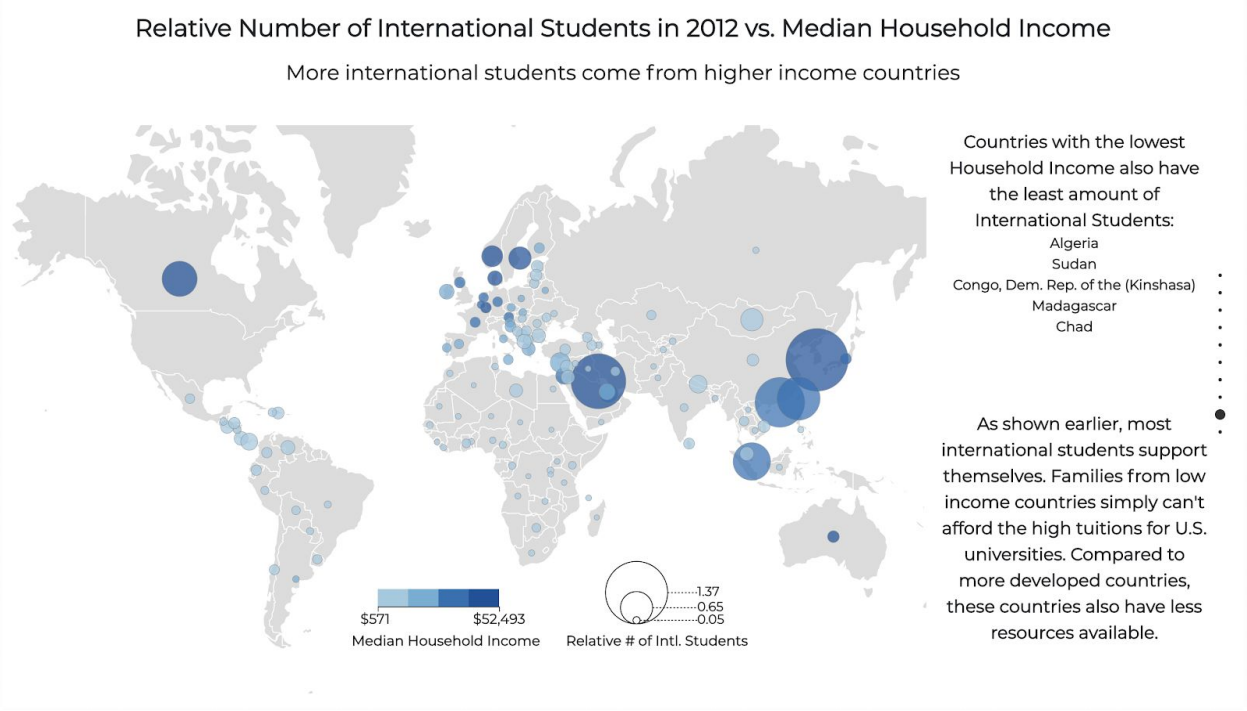
Changes made:

- Changed to a unified diverging color scheme
- Added legends for circle sizes
- Added more explanations

Before



After



Radial

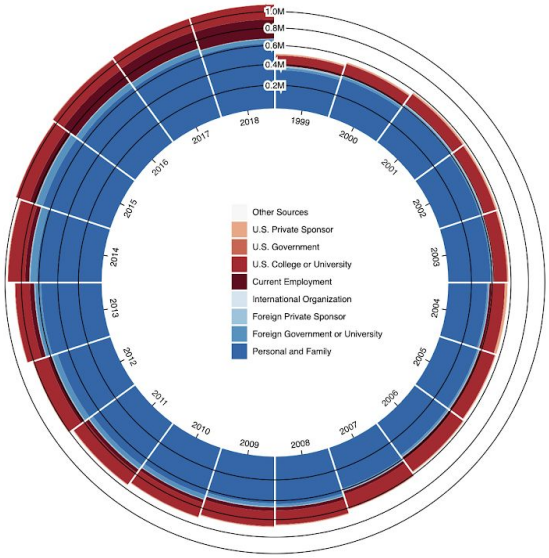
Changes Made:

- Adding Section Heading
- Improved hover information with details Consolidation of the smaller forms funding source into other sources.

- Creating more visible color contrasts while keeping the red (U.S base) and blue (foreign country based) gradient theme intact Before

Before

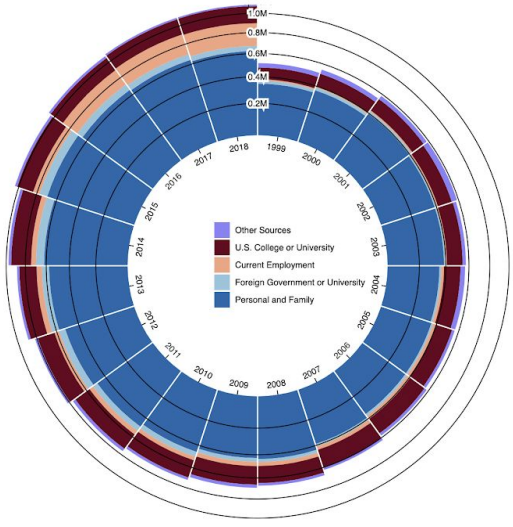
From 1999 to 2018, there has been a continuous growth in the total amount of students studying in the United States. We could see that the education funding for a majority of these international students came from their personal or family. Another big source of funding is from the college/university which account for around the 16-20% mark throughout the different years. One of the trends that is interesting to see is the growth of employment as funding for education. It went from several percent to a huge jump to 10-20% in recent years.



After

The Various Funding Sources of International Students Over the Year

From 1999 to 2018, there has been a continuous growth in the total amount of students studying in the United States. We could see that the education funding for a majority of these international students came from their personal or family. Another big source of funding is from the college/university which account for around the 16-20% mark throughout the different years. One of the trends that is interesting to see is the growth of employment as funding for education. It went from several percent to a huge jump to 10-20% in recent years.



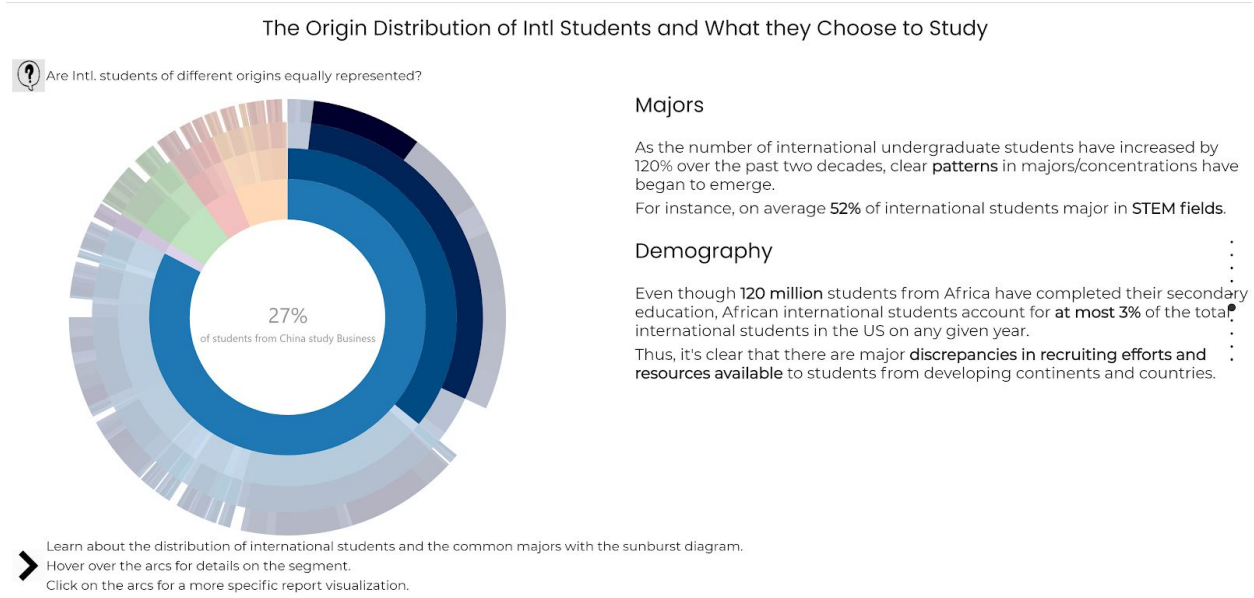
Sunburst

Changes Made:

★ Add Layer indicator on hover

★ Text formatting to eliminate text overlap

Before



After