a) A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection

Data Description

In this project, we use two datasets: edges_final.json (edges) and nodes_vfinal4.json (nodes). Edges aggregates data on migration between continents, and nodes contain descriptive information about each continent.

Our datasets can be broken down into three components: a dataset for the migration data on the chord diagram, a dataset of metrics for the spider diagram, and another dataset for the bar graphs. We sourced our data from multiple sources, explained in detail below. Additionally, for clarification, our data is mostly at the continent level. However, due to the difficulty of sourcing data for East Asia and North America on the continent level, we broke that down into countries.

Data for migration (chord diagram),

The source is the nodes file is as follows:

https://www.science.org/doi/full/10.1126/science.1248676?ijkey=ypit4%2Fxi7wo4M&siteid=sci&keytype=ref

The chord diagram mainly uses the edges file, which contains three variables. “Source” is an integer pointing towards the source continent of the migration, and “target” is an integer pointing towards the target continent of the migration. Weight is the weight/size of the migration, where a larger number means a larger migration, which becomes a wider band connecting the source and target countries on the chord diagram. Together, they describe the size and direction of migration movements between continents.

Dataset for the chord diagram

In the dataset for the chord diagram, we used several metrics from the nodes file to describe the sending and receiving countries in migration: healthcare access and quality index, annual CO2 emission, and civil liberty. The sources are listed below:

- Healthcare access and quality index: https://www.healthdata.org/results/country-profiles/haq
- Annual CO2 emission: https://ourworldindata.org/explorers/co2?time=1990..2020&facet=none&country=CHN~USA~OWID_WRL~JPN~PRK~KOR~MNG~CAN~TWN&Gas=CO%E2%82%82&Accounting=Production-based&Fuel=Total&Count=Per+capita
- Civil liberty: https://freedomhouse.org/report/freedom-world

Healthcare access and quality index is an index that is based on amenable mortality and describes the region's effectiveness and timeliness in preventing deaths that could be avoided with sufficient and timely care. The larger the index, the better the region's healthcare access and quality.

Annual CO2 emission is the amount of annual carbon dioxide emission generated by human activities such as burning fossil fuels, measured in billion metric tons. The larger the number, the great the annual CO2 emission of that region.

Civil liberty describes the extent to which citizens are guaranteed freedom including but not limited to freedom of religion, freedom of speech, and freedom of the press. The larger the number, the better the civil liberty.

Together, the spider diagram shows the relative scores of each region's metric, allowing readers to gain a perspective of understanding how each region stands in these measures compared to other regions.

Dataset for the bar graph

Lastly, for the bar graph, we wanted it to serve as an additional diagram giving information about the sending and receiving countries. We still maintained CO2, Healthcare Access and Quality Index, and Civil Liberty index to show the numerical values of them (since they weren't labeled in the spider diagram). We decided to also look at some other factors that indicate the quality of life within the

country. The two outward-facing bar charts are used for better comparison. With the Source country on the left and the Target country on the right. We used the following data source for the data:

- Human development index (HDI): https://hdr.undp.org/data-center/documentation-and-downloads
- Average life expectancy: https://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf
- Total population (millions): https://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf
- GDP per capita: https://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf
- Gender gap index: https://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf

HDI is a compounded index for life expectancy (health), education (knowledge), and per capita income (standard of living). Average life expectancy measures on average how long a resident in the country lives. Total population measures the number of residents that a country has. GDP per capita measures the gross domestic product divided by the total population of the country. Lastly, the gender gap index measures the extent to which a country closes its gender gap, economically and socially. The larger the gender gap index, the smaller the gender gap in the region.

Data Cleaning

After obtaining these data, we aggregated and cleaned the data using a Jupyter Notebook with Python to produce the final datasets that are imported into our index.html file. We first mapped the different countries and continents to corresponding integer values in a Jupyter Notebook.

The original dataset came in a matrix format data frame but in order to create a JSON file just to include the different connections, we were able to decompose it into our edges.JSON file. This file had three columns, Source, Target, and Weight. Since we were only able to find comprehensive data about only countries in North America and East Asia, we broke down the immigration and emigration data into country-level data after filtering down to those specific ones.

Our final version of the nodes.JSON file first only contained 3 columns, Continent, Country (if applicable), and Weight. In this case, the weight was the sum of all the people that immigrated and

emigrated in 2005. This was used to help draw the ring structure of our chord diagram. After we were able to find the different attributes of the country-level data (Civil liberty index, Health index, etc), we joined that with the original nodes data frame on the name of the country. This final version was used in our chord diagram, spider diagram, and bar plots. The nodes dataset contains the aggregated data (from different sources), combined through columns, to be used for the spider diagram and bar graph.

Data Selection Criteria

Overall, our data were selected based on how well they reflect the social, economic, and political status of a region, as well as the availability of information online. For example, we originally wanted to include other variables such as Government expenditure on education (% of GDP) and the gender gap. However, due to a lack of availability of such information online, we were only able to source this data for several regions for several years. It was difficult to deal with the resulting NAs in other rows, so we decided to take these variables out. The goal of this project is to find potential correlations and relationships between migration and the status of a region, so we want to find as much descriptive data on each region as possible, while also making sure that such information can be well-represented for all years and all continents.

b) An overview of your visual design rationale. A good rule of thumb to follow is "every pixel must be justified." Instead of a 100,000-element breakdown, give us an overview of the design decisions you made and the trade-offs inherent in how you displayed the data. This part ought to include a description of the mapping from data to visual elements. Describe marks and channels you employ such as position, color, or shape. Mention any transformations you performed, such as log scales.

Design Rationale

We designed our project around the topic of migration, aiming to explore how migration might be correlated with the social, economic, and political status of each region, and how their strength compares to other countries. To visualize this information, we chose to create three diagrams, each showing a different aspect of migration movements.

The chord diagram aims to show readers how people move from region to region (including within the region itself). The marks are the bands between regions, and the channels include the width of the band (representing the size of the movement), the two sides that each band connects to, the color of each band, and the size of the region on the circumference of the chord diagram. We chose to use the width of the band to represent the relative magnitude of the movements since our main goal is to find a comparison between the movements, which the width of a band can represent. Additionally, the two sides that each band connects represent a sending and receiving country. The color of each band shows the source country of the migration movement, which can be interpreted using the country labels. Lastly, the relative size of each region on the circumference shows how much migration a region receives or sends. The longer the line a region has on the circumference, the more involved it is in migration. Together, these elements tell the readers the amount of migration that is happening in 2005 across different continents, and the direction of this migration, allowing us to spot big trends.

The spider diagram illustrates the relative strength the sending and receiving countries have in the metrics listed. Our graph uses three metrics, as explained in section A. These metrics are representative of the economic health of a region. A region's annual CO2 emission is positively correlated with its GDP growth, implying that the higher the CO2 emission, the more likely that this region has a high GDP growth. Similarly, more developed countries generally have higher healthcare access and quality

index. They also tend to have higher civil liberty. Therefore, these metrics show how well a region is doing economically and socially. Having two countries on the same graph allows the readers to compare the relative economic strength of the two regions. The mark on this graph is the position of the vertices of each country. The channels are the color of the triangles (showing which country it is), and the aligned position of the axis (3 axis in total).

Lastly, the bar graph is an additional resource allowing the readers to compare the strength of the sending and receiving countries.

We decided to use these three visualizations to tell a story about migration across continents, because these visualizations not only show how migration movements happen across the continents but also compare the relative social and economic strength of these regions, thus allowing the readers to get a sense of a potential relationship between migration and how it affects the economic and social status of a country.

The design of our project also went through several stages of change, due to factors such as information availability and aesthetic viability. Originally, we think that migration is the most interesting when we can visualize how each continent's migration changed over the years. Therefore, we wanted to implement graphs that allow the readers to select different years (such as 2000 vs 2005 vs 2010) and compared the direction and size of migration that have changed for the continents over a specific time period. Nevertheless, we soon realized that this creates more challenges in finding complete datasets for each continent across all the years we want to display. An additional challenge is that most of the data are recorded on the country level, but rarely on the continent level, so it created an additional challenge for us if we would like to use these data to describe each continent since we have to search for which country is in which continent, add the respective continent label to each country, and the aggregate the data. Consequently, in order to minimize and eventually eliminate any NAs in our datasets, we chose to go with data in one year first. We eventually chose to use the data from 2005, because the author manually aggregated the data to the continent level for the 2005 data, so we were able to directly use these data for our project.

Moreover, as mentioned above, we were originally planning to use 6 metrics/variables for the spider diagram to generate more descriptive information about each continent. However, due to a lack of data, we decided to go with 3 metrics, which also gives us enough space to create a visualization that gives enough space for each variable and is easy to understand.

Lastly, after finishing up the spider diagram, we soon realized that because our spider diagram only shows the relative score of each country on each metric, it creates difficulty for the readers who would like to know the absolute numbers to understand the scale of each region's impact. To do this, we added a third component of our visualization: a bar graph that shows the absolute numbers that the sending and receiving countries have for each metric (3 in total). This allows the readers to understand not only the relative strength of each country but also their absolute strength and scale, adding more authenticity to our visualization. Through this evolution of rationale and change in design elements, we produced our final visualization on the topic of migration.

c) An overview of your interactive elements and their design rationale. Give us an outline of the design decisions that went into the interaction affordances you added to your visualization. What process did you use to choose the interactions you developed? How did you make them discoverable, usable, and interesting?

We chose the interactive elements so that they are accessible and intuitive for the readers. Overall, we want to tell a story about migration and how it is correlated with the social and economic conditions of different regions. To do this, we adopted several interaction affordances to strengthen the connection between visualizations.

When readers open our website, they should first notice the chord diagram, showing how migration movements happened between regions in 2005. If they **hover over the thick circumference of the circle**, **that country, the migration movements from all other regions would fade**, thus allowing the reader to focus on learning about migration out of the target region. Moreover, as they hover their mouse over the bands, **the band would light up with a white stroke around them**. This helps the readers to better visualize the relative size and direction of the migration movement (especially when there are many movements between continents. **As they click on the band, the spider diagram on the right would show the data of the metrics for the receiving and sending countries**. This allows the reader to then compare the social and economic conditions of these countries.

We made our interactions **discoverable** by visibly changing opacity and having an obvious change in numerical value when a user hovers over the different numbers. However, we didn't want a user to be able to do too much and cause a distraction by only having hovering interactions so when you click on ribbons that have a white outline when hovered, the spider diagram will also change values slowly. When we were researching how to implement a chord diagram with interactivity, it was a very familiar design pattern to allow a user to hover and click on the specific ribbons. Therefore, we made it more usable by adopting that into our design. Our data and presented visuals were **interesting** based on the feedback we heard during demo day from our peers.

d) The story. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

Since 2000, international goods have become much more accessible, but the increasing number of policies and restrictions on migration have added barriers for immigrants and protection for receiving countries (World Migration Report). When migration movements are examined, people often choose to look at intracontinental movements, but we took this a step further by looking at intercontinental movements. We started this project with assumptions and biases, but it was interesting to see our results confirming and contradicting them.

Living in the U.S. and hearing news about migration into the U.S, most of us assume that the U.S. would be a major receiving country in the world. However, as shown on the chord diagram, we discovered that the U.S. was only involved in a small part of the global migration movements. In fact, the majority of migration movements come from East Asia (broken down into countries including China, South Korea, etc) and South Asia. This is mainly because Asia is a large sending country to the world, but as we can see on the graph, many regions such as South Asia also have a significant portion of intracontinental movements, reaching close to 40% of the total migration moments from and to South Asia (World Migration Report). Likewise, around 50% of the migration movement to and from Africa is intracontinental, and almost all movements to and from Europe are intracontinental. Therefore, examining the chord diagram, we realized that the U.S. is not a significant receiving or sending country in the world migration movement, drawing our attention to other continents for analysis.

Furthermore, we also expected to see countries that are mainly sending migrants to other regions have a lower index for the metrics on the spider diagram and the bar graph. However, it turns out that such is not the case, based on our visualization. For example, there is no significant difference between the human development index (HDI) of majority-sending and majority-receiving countries. For example, even though South Korea is a majority-sending country, its HDI (0.86) is not so much different from the U.S, a majority-receiving country (0.9). Based on the World Migration Report, this could be

explained by several factors. First, residents of majority-receiving countries receive remittance from their immigrant relatives abroad, supporting their income. Since HDI is based on education, life expectancy, and per capita income, remittance would also raise the HDI of the receiving countries. Additionally, a common misconception that people have is that immigrants have a negative impact on the receiving countries. Nevertheless, although there might be minor negative impacts on the average wage of the receiving countries, the net impact in most cases is positive, as immigrants in 2005 and onwards were mostly skilled laborers rather than unskilled laborers. Especially between places such as South/Southeast Asia and the oil-rich countries of the Middle East, recruiters are unlikely to select the weakenest or poorest of any group, so immigrants in these areas are more likely skilled laborers who bring positive impacts, economically and socially, to the receiving countries.

It is also important to note that we did see some correlation between certain metrics with migration movements. For example, it seems like most of the majority-receiving countries have a higher healthcare access and quality index compared to the receiving countries. For example, the U.S. is a majority-receiving country, with healthcare access and quality index much higher than that of most majority-sending countries. Similarly, Canada—another majority-receiving country—also has a relatively high healthcare index. Therefore, we did notice that there is a correlation between whether a country is majority-sending or majority-receiving and our metrics, although such a trend is not seen with all metrics.

Overall, we believe that these visualizations not only tell an interesting story about the migration movements back in 2005 but also help readers to clear many misconceptions about migration. When we live in a country for too long, it is easy to believe that the world is only this small, so the chord diagram helps us to see the big picture of how the U.S. and other countries/continents stand in the global migration movement. It is also important to note that migration in general benefits both the receiving and sending countries because of the fluidity of skilled labor. Nowadays, migrations are increasingly based on skill sharing, rather than using unskilled laborers to work in low-barrier

industries. Recognizing this difference, we think it is crucial that these visualizations help to clear the stigma around immigrants.

Member Contribution to the Report

Ruby Li

- Wrote the report
- Helped with coding
- Researched on world migration movements
- Analyzed the visualizations

Aurora Zhang:

Aurora was responsible for finding data on migration, as well as the 8 metrics, which are mostly drawn from different resources. Keep data resources, and integrate them into the one dataset. Aurora produced bar plot and all labels and interaction with it.

Jane Xie

- Wrote the spider diagram and created its interaction
- Wrote the interaction between the chord diagram and spider diagram

Stephanie Zhang

- Referenced and coded the chord diagram
- Helped write the interactivity part
- Transformed and filtered our data from CSV to JSON files

Aurora Zhang: Aurora was responsible for finding data on migration, as well as the 8 metrics, which are mostly drawn from different resources. Keep data resources, and integrate them into the one dataset. Aurora produced bar plot and all labels and interaction with it.

Citation

"World Migration Costs and Benefits of International Migration 2005."

https://publications.iom.int/system/files/pdf/wmr_2005_3.pdf.