Data Visualization QMSS 5063, Spring 2019
By: Yihong Liu, Zu Cheng Mark Ong

Looking at Migrant Stock and Industrial Growth
Process Notebook

*Image from: https://www.brainpop.com/socialstudies/culture/immigration/*

Introduction

We are interested in looking at migration stock and industry growth patterns across countries. We will explore the following questions:

1. What is the distribution of migrant flow across countries?
2. How does migration flow evolve over time for each country?
3. How does the value added of different industries evolve over time for each country?
4. What is the relationship between migration flow and value added for each industry?

Initial Project Plan: Plots
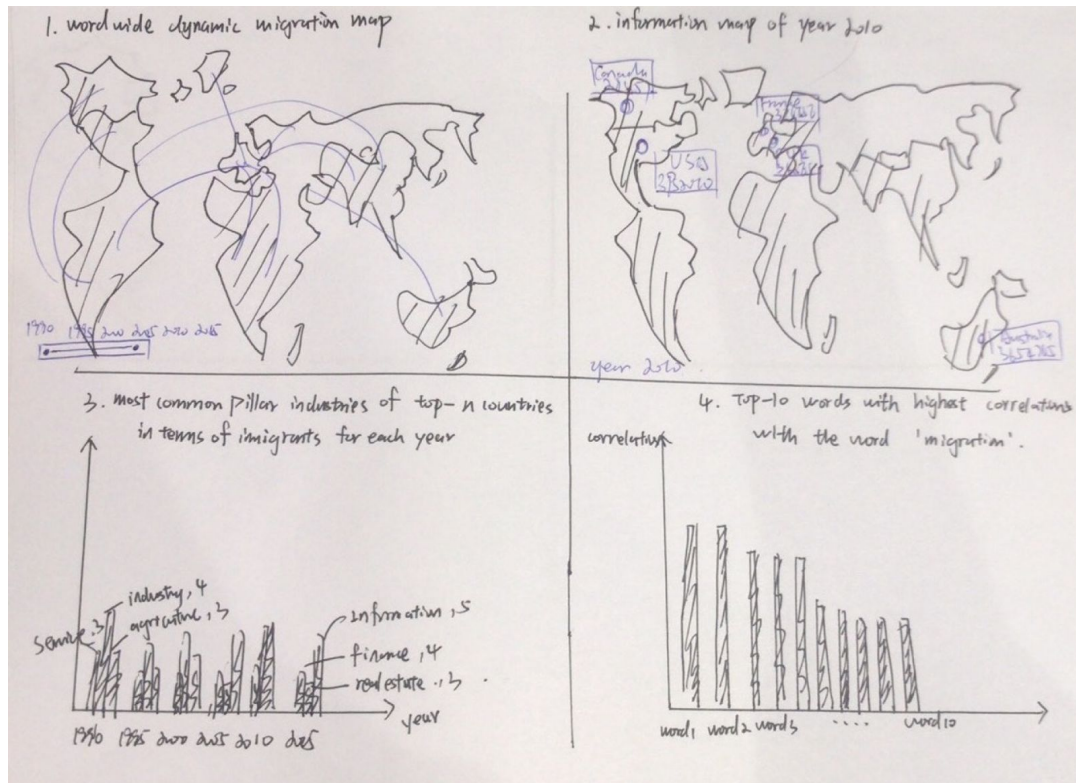
Our initial plan includes the following four plots:

1. A map to show how does migration flow evolved over time (in terms of volume of migration, countries with the major inflow of migrants, countries with the major outflow of migrants, etc). We plan to plot migration data, over a time series, on maps.
2. A map or a line plot to illustrate which countries observed the greatest increase in immigrants over time.
3. A bar plot to show how do the top industries of countries change over time? Specifically, we want to look at the top n industries for each of the top n countries in term of the number of immigrants. We plan to find out what are the common major industries among the countries with the greatest inflow of migrants. We hope to gain insight on relationships (if any) between industry trends and migration flow trends.
4. What do the sentiments towards immigrants change over time? If feasible, we plan to analyze sentiments from twitter data, whether 'migration' is typically mentioned alongside with positive or negative words.

Initial Project Plan: Data

We planned to use the migrant stock and industry value added datasets from the United Nations. Both datasets are available for all UN Members States and other territories in the world for which National Accounts information is available. Also, the industry value added dataset has used the *United Nation 2008 ISIC Industry Standard*, *Rev.4* to break down the total value added by sector, providing a well-defined definition of industries.

For the *International migration stock at mid-year by major region, country or area, 1990 - 2017 (from United Nations, Department of Economic and Social Affairs. Population Division (2017))* dataset, it has collected data of international migrant by age, sex, destination, and origin. Estimates are presented for 1990, 1995, 2000, 2005, 2010, 2015 and 2017.

For the *Value added by activity, 1970 – 2017(from United Nations Statistics Division (UNSD))* dataset, it has collected data contains annual value-added-by-activity (e.g. agriculture, mining, construction, financial, etc) of each country from 1970 to 2017. Value added by activity breaks down the total value added by sector. It could be used to reflect the contribution of industries to the overall economy.

We planned to use text data extracted through the Twitter API If we would carry out sentiment analysis towards immigrants.

Visualization 1. Migrant stock data at country level over time (1990-2017)

We explored several ways to create a moving 'time-series' data (animation-like) that can represent migrant flow over time. However, one drawback is that the visualisation would play out like a video on loop and the user will not be able to look at the worldwide migrant flow for a specific year. Also, the dataset we are using only contains migrant stock numbers and related information for 7 specific years (1990, 1995, 2000, 2005, 2010, 2015, 2017). The time series animation might not turn out to be as impressive as what we had in mind since it will turn to be like a hyperlapse with 7 frames.
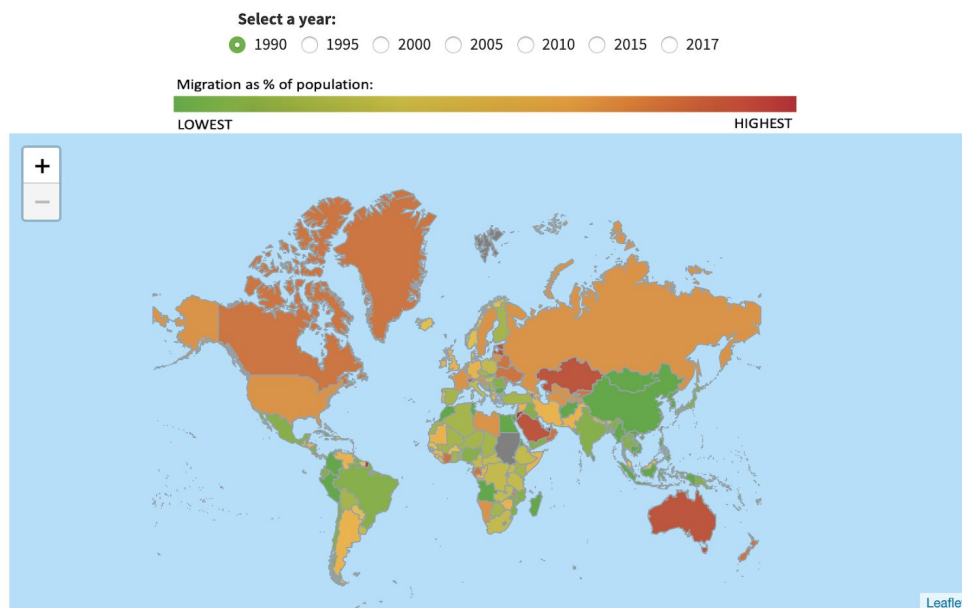
We thus decided to plot a leaflet map and include radio buttons for users to choose a year of interest. The leaflet map will plot the migrant stock data for the selected year and serve as a snapshot of migrant flow at world level. The user can click through 1990 to 2017 to see how the migrant stock numbers evolve over time. Lastly, we decided to include the more relevant information in a pop-up box, which appears when user clicks on the country on the leaflet map. We chose to provide information on the total migrant stock, top 3 migrant sources, and the percentage of migrant stock relative to the country's population.
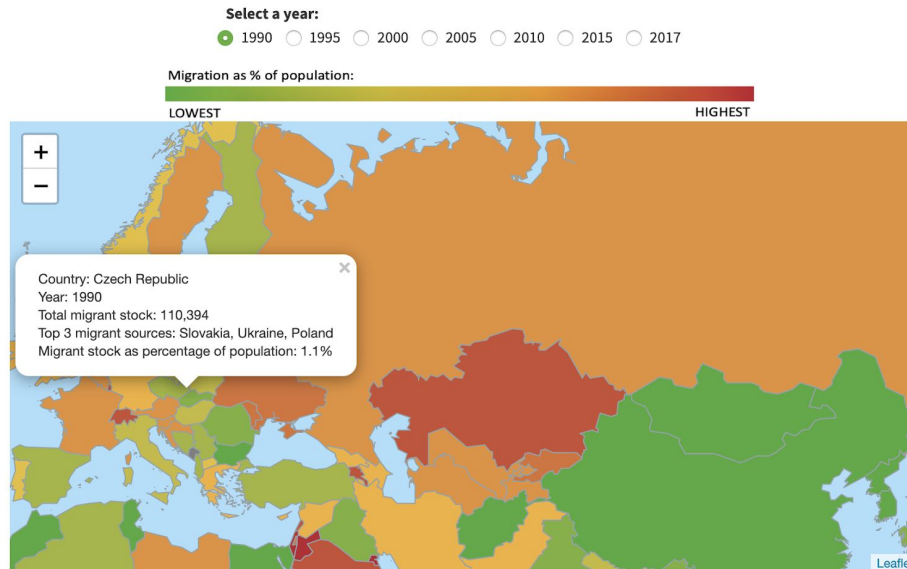
To make sure that the 'fill' colours differ in a more obvious way among countries on the map, we chose to use a 3-colour gradient instead of a 2-colour gradient. We decided on an intuitive set of colours where green represents the lowest percentage of migrant relative to population, yellow represents the mid-level, and dark red presents the highest percentages.

Lastly, we wanted to design the leaflet map in an overall aesthetically pleasing way. This includes intentionally choosing pastel tones for the 'fill' colours, and adding a pastel blue leaflet background to represent the sea. We thought it would look better than using the default leaflet theme which leaves the map background grey in colour.

Visualization 2. Migration trend across time (1990-2017)

We wanted to explore how migration stock numbers change across the years, at country level. We thought it would be good to provide population numbers together with the migrant stock numbers since the first visualization briefly looked at percentage of migrant stock relative to country's population. Also, we decided to group the migrant stock by gender with the intention of finding more insightful trends. We chose to use plotly in shiny because we liked the pop-up containing X and Y values when users hover their cursor on the data points. At the same time, we found a way to set the default plotly mode as 'compare data on hover'. We thought it was useful to show all the data value (all 6 lines) for the same year when the user hovers over any one of the data point because it would make the graph trends more digestible.
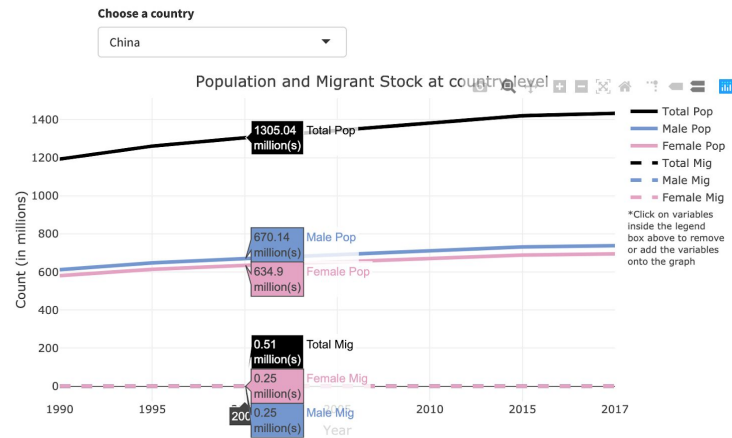
At the same time, we chose plotly because users can select and deselect variables in the legend box to remove/add them onto the plot. This is useful because the migrant stock numbers and population numbers are sometimes in very different order of magnitude (especially for countries with very low percentage of migrant stock relative to population, think China and India). When all six lines are plotted together, it is sometimes difficult to see fluctuations in specific years since range on the y-axis has to extend across a large set of values to plot both the population variables and migrant stock variables. By deselecting either the population or migrant stock variables, the slope and fluctuations can be made more obvious to the human eye.
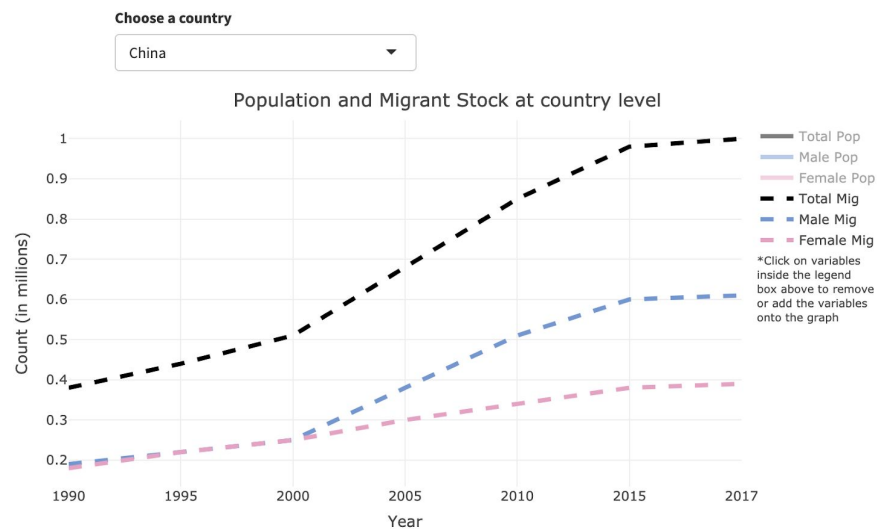
Visualization 3. Industry value added across time (1990-2017)

We want to explore how do the value-added of industries change over time for each country. During the process, we realized that compared to the bar plot (our initial plan), the line chart could perform better if we want to make a comparison between industries and areas.

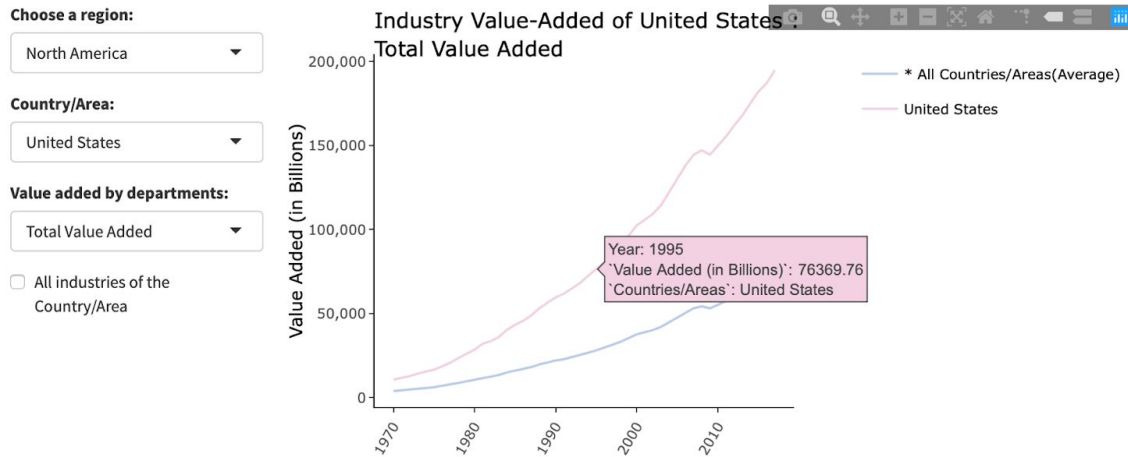In addition to that, since we didn't provide a map plot for the industry data, so to make the comparison within regions easier, we designed an interactive plot with hierarchical selections. From the sidebar, the user could select the area and the industry of interest, the default plot would show the value added of the industry for all countries in that region. The comparison could be shown clearly.

Choose a region: North America

Country/Area: * All Countries/Areas(Sum)

Value added by departments: Total Value Added

☐ All industries of the Country/Area

Industry Value-Added of All countries in North America : Total Value Added

Legend:
- * All Countries/Areas(Average)
- * All Countries/Areas(Sum)
- Bermuda
- Canada
- United States

Year: 2012
`Value Added (in Billions)`: 59696.86
`Countries/Areas`: * All Countries/Areas(Average)

Value added equals the difference between an industry's gross output and the cost of its intermediate inputs. See the definition by U.S. Bureau of Economic Analysis (BEA).

Show 5 entries   Search: [   ]

Industry Valued Added Across Time

	Countries/Areas	Year	Industry	Value Added (in Billions)
7592	Bermuda	2017	Total Value Added	64.99
7600	Bermuda	2016	Total Value Added	63.25
7608	Bermuda	2015	Total Value Added	60.9
7616	Bermuda	2014	Total Value Added	58.98
7624	Bermuda	2013	Total Value Added	58.89

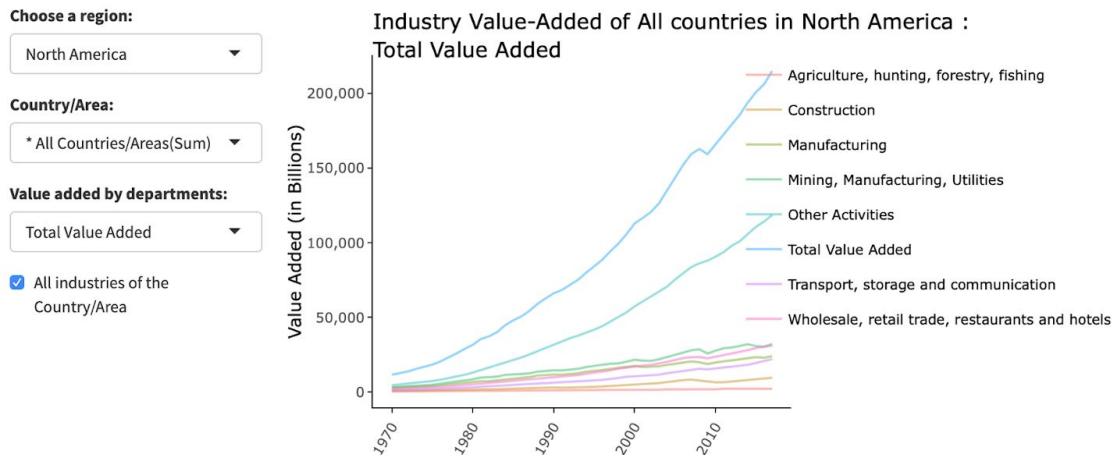Showing 1 to 5 of 240 entries   Previous 1 2 3 4 5 … 48 Next

And by selecting a specific country, country-level trends are shown accompanied by the regional average one.

Industry Value-Added of United States : Total Value Added

Choose a region: North America
Country/Area: United States
Value added by departments: Total Value Added
☐ All industries of the Country/Area

Year: 1995
`Value Added (in Billions)`: 76369.76
`Countries/Areas`: United States

* All Countries/Areas(Average)
United States

Value added equals the difference between an industry's gross output and the cost of its intermediate inputs. See the definition by U.S. Bureau of Economic Analysis (BEA).

Then we reconsidered the problem from the point of the users, thinking that they might also want to see the comparison of different industries in an area. Thus, we further provided a checkbox of 'All industry of the Country/Area'. By clicking that, the value added of different industries of one area could be shown in one plot.

Industry Value-Added of All countries in North America : Total Value Added

Choose a region: North America
Country/Area: * All Countries/Areas(Sum)
Value added by departments: Total Value Added
☑ All industries of the Country/Area

Agriculture, hunting, forestry, fishing
Construction
Manufacturing
Mining, Manufacturing, Utilities
Other Activities
Total Value Added
Transport, storage and communication
Wholesale, retail trade, restaurants and hotels

Value added equals the difference between an industry's gross output and the cost of its intermediate inputs. See the definition by U.S. Bureau of Economic Analysis (BEA).

Visualization 4. Describing the relationship between migrant population and industry value added

For the fourth visualization, we initially planned to do a sentiment analysis of migration. However, during the process of doing the project, we reached to an agreement that instead of adding a new section to our website, we'd rather go deep into the story and make a visualization

that could vividly show the relationship between migration and industry development from multiple dimensions. So, we designed a scatter plot to describe the relationship. In addition to the geography and industry options, we also wanted to provide options for population types, to enable flexibility selection on migration data too. Thus, we added two more options, the gender type, and the population type, on the top of the graph.

When we checked the plots for each country, we found that for many of the developed countries, both the absolute value of migrant increment and industry value added have been increasing over time. We thought it might be better to provide options to see the relationship of their growth rates. By clicking '5-year Growth rate' options, data of 5-year growth rate for both variables are presented.

Hosting on Shiny

Initially, we explored the the idea of using a github.io page to hold all of our visualisation plots. However, due to technical issues with putting up plots that rely on Shiny's interactivity (radio

buttons, slider, etc) onto a github.io page, we eventually decided on using an R Shiny dashboard and hosting it on shinyapps.io.

The site can be accessed via: https://yvonneleoo.shinyapps.io/datavizfinal/

Final data sources used

We haven't added any new data source during the process. The visualisations are built using data obtained from the original sources as planned:

- UN data on total international migrant stock
- UN data on international migrant stock by destination and origin
- UN data on total population and international migrant stock by gender
- UN data on value added by activity, 1970 - 2017
- UN 2008 ISIC industry standard international standard industrial classification of all economic activities (ISIC), rev.4

Conclusions

Our data exploration led us to several conclusions:

1. There is an overall North-South divide in the number of migrants.
2. Population and migrant stock number of a country tend to move in tandem over time.
3. From the graphs we could see that for most countries/areas in the world, Other Activities (ISIC J-P) contributes most to the total value added. According to the United Nation 2008 ISIC Industry Standard, it includes the following industries:
    - Information and communication (ISIC J)
    - Financial and insurance activities (ISIC K)
    - Real estate activities (ISIC L)
    - Professional, scientific and technical activities (ISIC M)
    - Administrative and support service activities (ISIC N)
    - Public administration and defence; compulsory social security (ISIC O)
    - Education (ISIC P)
3. For many countries, a strong positive correlation between migration increment and industry value added in absolute value have been shown. But when we view the relationship of their 5-year growth rates, the positive relationship becomes less significant. It may indicate that given more available data, we could further explore whether there is a time lap between migration increment and industry developments, and whether there exists a causal effect.