CS 171 Final Project Process Book

Immigration Trends

Sierra Katow & Kristina Hu
TF: Romain Vuillemot

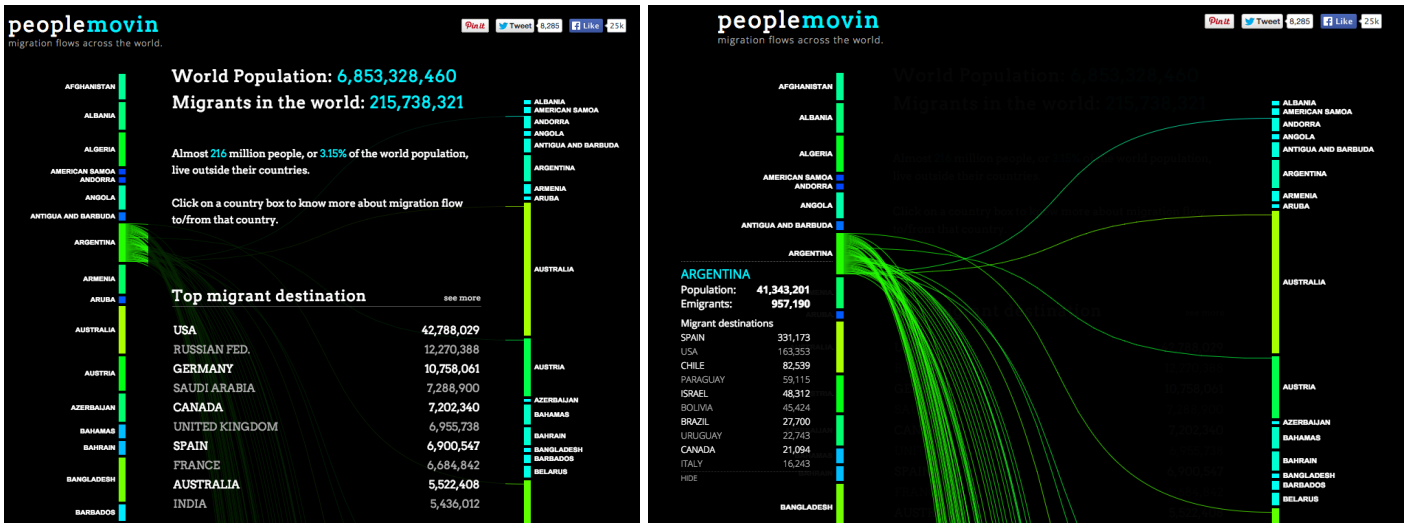
Overview and Motivation

As the descendants of immigrants who came to the United States, we are fascinated by the stories of our ancestors and how they fit in with the larger movement of human migration. Global migration trends influence not only country demographics, but also culture, economy, and religion. In today's age of increased mobility and informational flow, where are people coming from and going to? This is what we set out to discover.

Kristina and Sierra have also explored this topic independently. During Kristina's semester abroad in Argentina, Senegal, and India, she wrote a major research paper on Chinese immigration trends to developing countries. Sierra is a secondary concentrator in History and is currently taking two history seminars that examine the perspectives of immigrants and minorities of the U.S., one of which is perhaps the only Asian American history class offered at the college.

Related Work

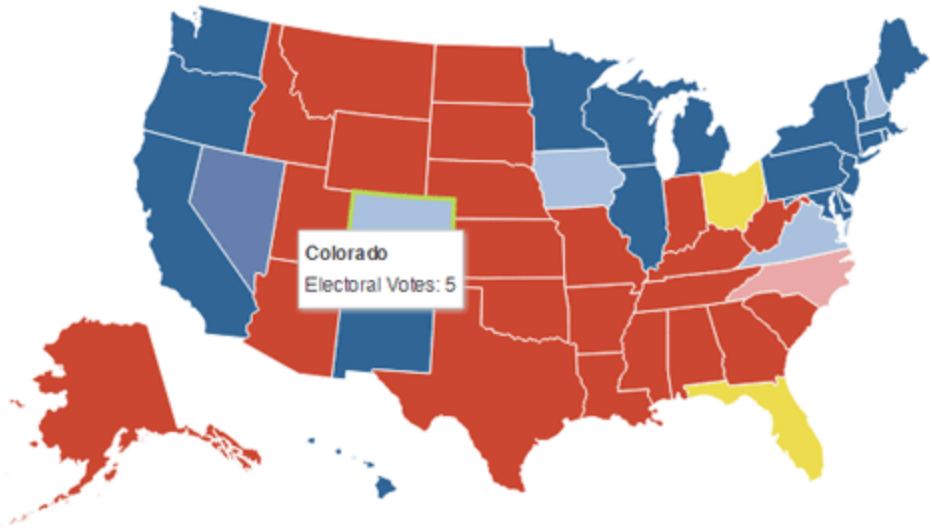
We were inspired by the visually pleasing and intuitive design of http://peoplemov.in/, a visualization website that captures migration data from country to country. The design makes use of arcs to illustrate movement of people from one country to another, an idea we implemented in our final design (arc flows). We also utilized a similar color scheme: black background with bright, easy-to-discern colors.



Peoplemovin, a data visualization illustrating migrant flows across the globe.

One thing we did not like about the Peoplemovin design is its organization of countries. Although the alphabetical ordering makes it easy to reference and find countries, the visualization tells you nothing about the extent of geographic movement of these migrants. Do people tend to migrate to nearby countries or far away?

To answer this question, we looked to the D3 Datamaps library, a lightweight library that includes both US and world maps for use in design. Given the highly geographical nature of our project data, we decided early on that we wanted to use D3 Datamaps as the framework for our landing page visualization, which illustrates at-a-glance information about migration between countries.



An example of a D3 Datamaps map object of the United States. Borders are easily discernable.

Questions

Below are the primary questions we hope to answer with our visualization:

1. For each country, what are the top destinations for outgoing emigrants and top sources of incoming immigrants?
2. Do immigration patterns differ between men and women?
3. Do other factors may influence the trends we see in migration over the past 5 decades -- in particular, life expectancy rates, infant mortality rates, or GDP levels?

Answering these questions will help us learn about global migration patterns from 1960-2000, and may help us discern why different groups, nationalities, or ethnicities of people move and for what reasons. For instance - do people tend to migrate to nearby countries or countries far away? What is the relationship between life expectancy rates, infant mortality rates, and GDP levels with migration flows? The quantitative data have important qualitative consequences for our understanding of international trade, population growth, development, politics and more.

Initially, we hoped to explore the more inferential/subjective side of human migration: for instance, what historical, social, political, or other factors influence the trends we see in migration over the past 5 decades? As we progressed in our implementation, we realized that we had way too much data to analyze to answer very specific questions. In addition, one can say that historical, social, and political factors that drive people to migrate are subjective and overly qualitative in nature. After all, the best visualizations are the ones that allow the user to explore, discover, and communicate new insights or information on their own.

Data

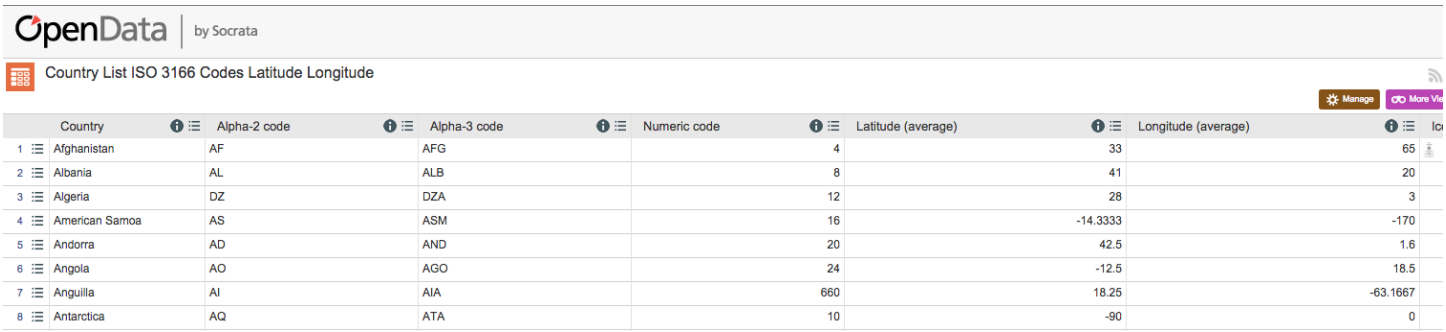
Because we used data from so many sources, there was substantial data processing/cleanup - both manually and via script.

Initially, we wanted to use data from the OECD's International Migration Database, a database that gives inflows and outflows of foreign population by nationality for the years 2000 - 2012 for all OECD countries. Users can filter data by country or year, but we were not able to find time-series data for other migration characteristics, such as gender, labor force, education level, etc.

We ended up using the **World Bank Global Bilateral Migration Database**, which gives global matrices of bilateral migrant stocks spanning the period 1960-2000, disaggregated by gender and based primarily on the foreign-born concept. Over one thousand census and population register records are combined to construct decennial matrices corresponding to the last five completed census rounds. More specifically, this database gives number of migrants from country A (origin) to country B (destination) for every possible combination of countries A and B by gender for years 1960, 1970, 1980, 1990, and 2000.

- This solves problem # about including time as a dimension, ***AND includes all countries***! Users can filter data by country, year, ***and gender***.
- Unfortunately, the dataset did not include time-series data for other country characteristics, such as GDP, life expectancy, infant mortality rate, etc. Therefore, we decided to supplement our migration data with data from the
    - World Bank GDP Database: a database that gives the GDP in USD (adjusted for inflation) of each country
    - World Bank Life Expectancy Database, a database that gives the life expectancy in years (both male and female) for people born in each country
    - World Bank Infant Mortality Database, a database that gives the infant mortality rate (per 1,000 live births)

For geographic data, we used the Country List ISO 3166 Codes Latitude Longitude open-source database from opendata.socrata.com, which matches countries to their alpha-2, alpha-3, and numeric codes, as well as their average latitude and longitude coordinates.



OpenData | by Socrata

Country List ISO 3166 Codes Latitude Longitude

Manage | More Vie

	Country	Alpha-2 code	Alpha-3 code	Numeric code	Latitude (average)	Longitude (average)	lo
1	Afghanistan	AF	AFG	4	33	65	
2	Albania	AL	ALB	8	41	20	
3	Algeria	DZ	DZA	12	28	3	
4	American Samoa	AS	ASM	16	-14.3333	-170	
5	Andorra	AD	AND	20	42.5	1.6	
6	Angola	AO	AGO	24	-12.5	18.5	
7	Anguilla	AI	AIA	660	18.25	-63.1667	
8	Antarctica	AQ	ATA	10	-90	0	

After downloading all of the complete datasets, we manually formatted them into .CSV and .JSON files that could be read and interpreted by our scripts.

After manual data processing, we stored the migration data in aggregate_data.csv, a file which contains the entire dataset and has the following header items: ***Origin, Gender, Destination, Count, Year***

	A	B	C	D	E	F	G	H
1	Origin	Gender	Destination	1960	1970	1980	1990	2000
2	Afghanistan	Female	Afghanistan	0	0	0	0	0
3	Afghanistan	Female	Albania	0	0	0	0	0
4	Afghanistan	Female	Algeria	22	8	6	5	4
5	Afghanistan	Female	American Sa	0	0	0	0	0
6	Afghanistan	Female	Andorra	0	0	0	2	3
7	Afghanistan	Female	Angola	0	0	0	0	0
8	Afghanistan	Female	Anguilla	0	0	0	0	0
9	Afghanistan	Female	Antigua and	0	0	0	0	0
10	Afghanistan	Female	Argentina	2	4	6	20	0
11	Afghanistan	Female	Armenia	0	0	0	0	4
12	Afghanistan	Female	Aruba	0	0	0	0	0
13	Afghanistan	Female	Australia	16	31	350	1236	4449
14	Afghanistan	Female	Austria	19	64	105	149	854
15	Afghanistan	Female	Azerbaijan	0	0	0	0	0
16	Afghanistan	Female	Bahamas, Th	0	0	0	0	0
17	Afghanistan	Female	Bahrain	22	32	2694	6227	10054
18	Afghanistan	Female	Bangladesh	49	40	25	12	13
19	Afghanistan	Female	Barbados	0	0	0	0	0
20	Afghanistan	Female	Belarus	0	0	0	0	0
21	Afghanistan	Female	Belgium	7	34	15	120	111
22	Afghanistan	Female	Belize	0	0	0	0	1
23	Afghanistan	Female	Benin	0	0	0	0	0
24	Afghanistan	Female	Bermuda	0	0	1	0	2
25	Afghanistan	Female	Bhutan	0	0	0	0	0
26	Afghanistan	Female	Bolivia	0	0	0	0	0

For GDP, life expectancy and infant mortality data, the data were formatted and stored in separate files with the following header items: ***Country, 1960, 1970, 1980, 1990, 2000***

A	B	C	D	E	F
Country	1960	1970	1980	1990	2000
ABW					1873452514
AND		78617570	446377495	1028989439	1133644295
AFG	537777812	1748886596	3641723447		
AGO				10260193361	9129594819
ALB				2101624963	3686649387
ARB		30910527987	3.36534E+11	4.4386E+11	7.26013E+11
ARE			43598747467	50701443696	1.04337E+11
ARG		31584210365	76961923742	1.41352E+11	2.84204E+11
ARM				2256838858	1911563665

Our GDP dataset. Countries are stored as 3-letter country codes.

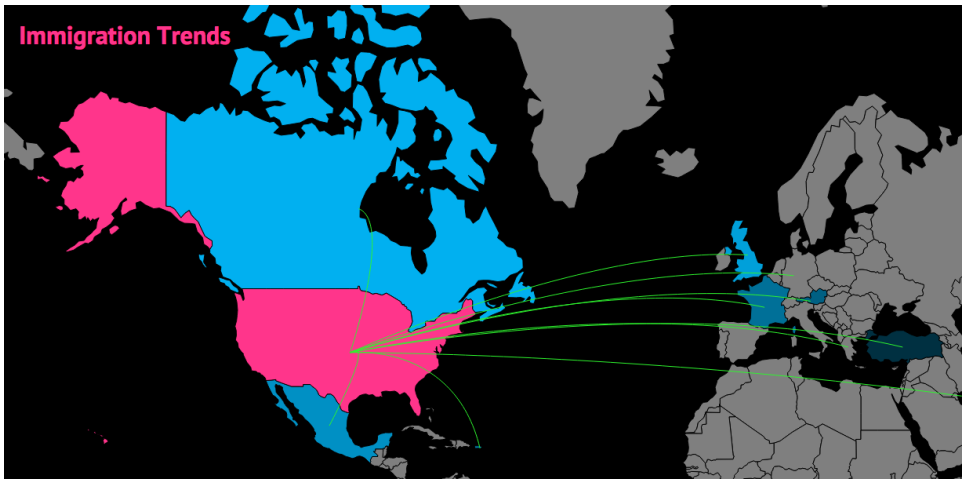
One challenge we ran into was not having complete data for all years or countries. Similarly, both the World Bank and D3.Datamaps use ISO-3155-1 alpha-3 country codes to represent countries, but these did not always match up with each other, so we had to manually reconcile differences between country codes across datasets from separate sources. Furthermore, some countries have become obsolete due to international political reasons -- we decided to remove those from our dataset.

Exploratory Data Analysis and Design Evolution

We received the following preliminary feedback from Romain about our initial proposal submission:

1. Time is an important component of migration patterns. You may want to include some timeline or anything that can let the user understand when migration flows started, peaked, ended, etc.
2. Flows are also crucial, and visually speaking they may be represented as connections between countries (using links, arrows, etc..). I am aware this is not your primary goal, but you put a lot of expectations on a map which may not fully do justice to the complex data-set you have. You may want to find a better way to show countries.
3. You seem to be using a line/area chart for education level, but this are categorical data thus best represented as bar chart
4. Finally, using some reactive interaction (like in HW3) would allow the user to query the data in a very flexible way. This is listed in your optional feature but I think this important for you to consider.

Our landing page started out as a world map (WorldView) that highlighted a country and generated outgoing arcs to that country's top ten emigrant destinations. The selected country is highlighted pink on mouseover. Destination countries are colored different shades of blue, with brighter shades indicating larger migrant populations and darker shades indicating smaller migrant populations.

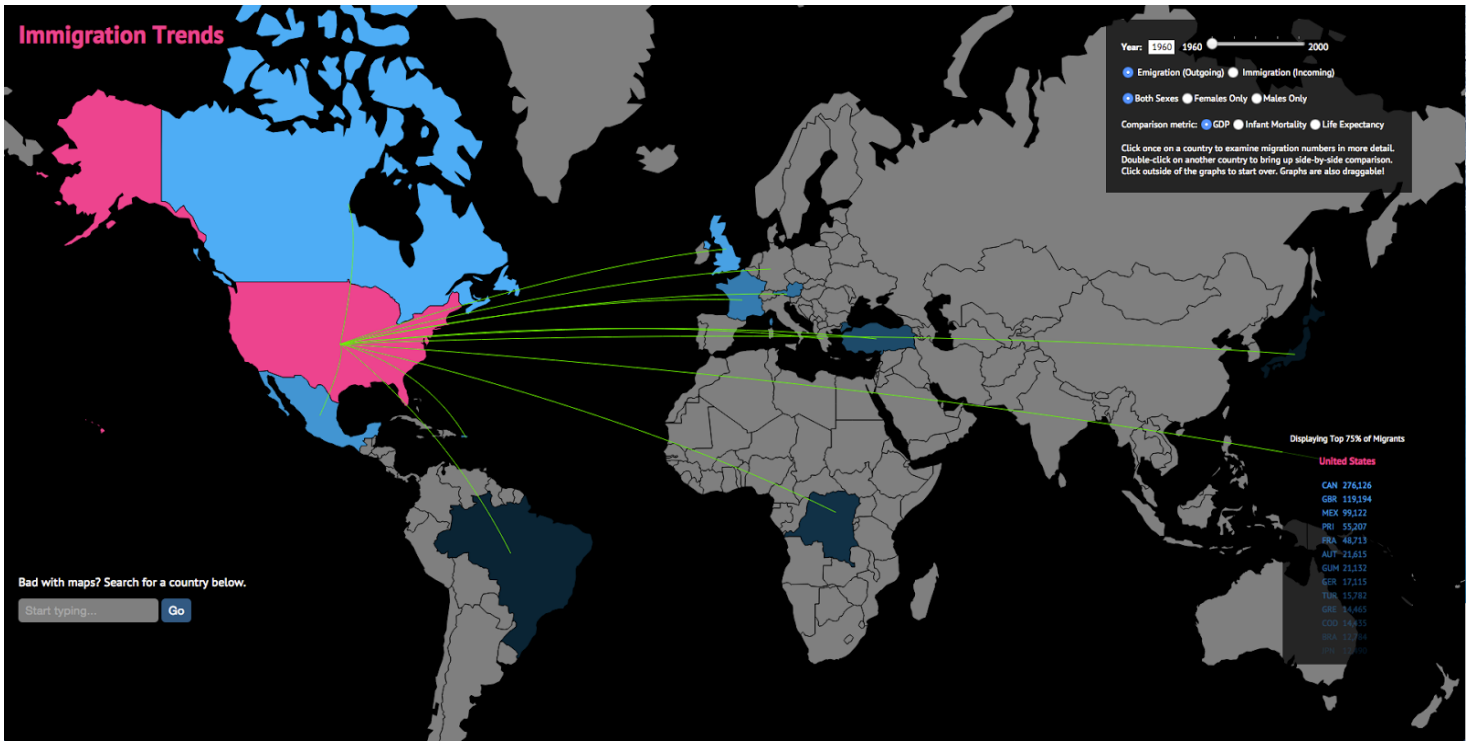


We quickly noticed one major issue: smaller countries are much harder to find on a map layout than larger countries. Furthermore, we needed a way to filter the data and display more quantitative data, whether through pop-up bar graphs or otherwise -- or select certain pairs of countries to analyze.

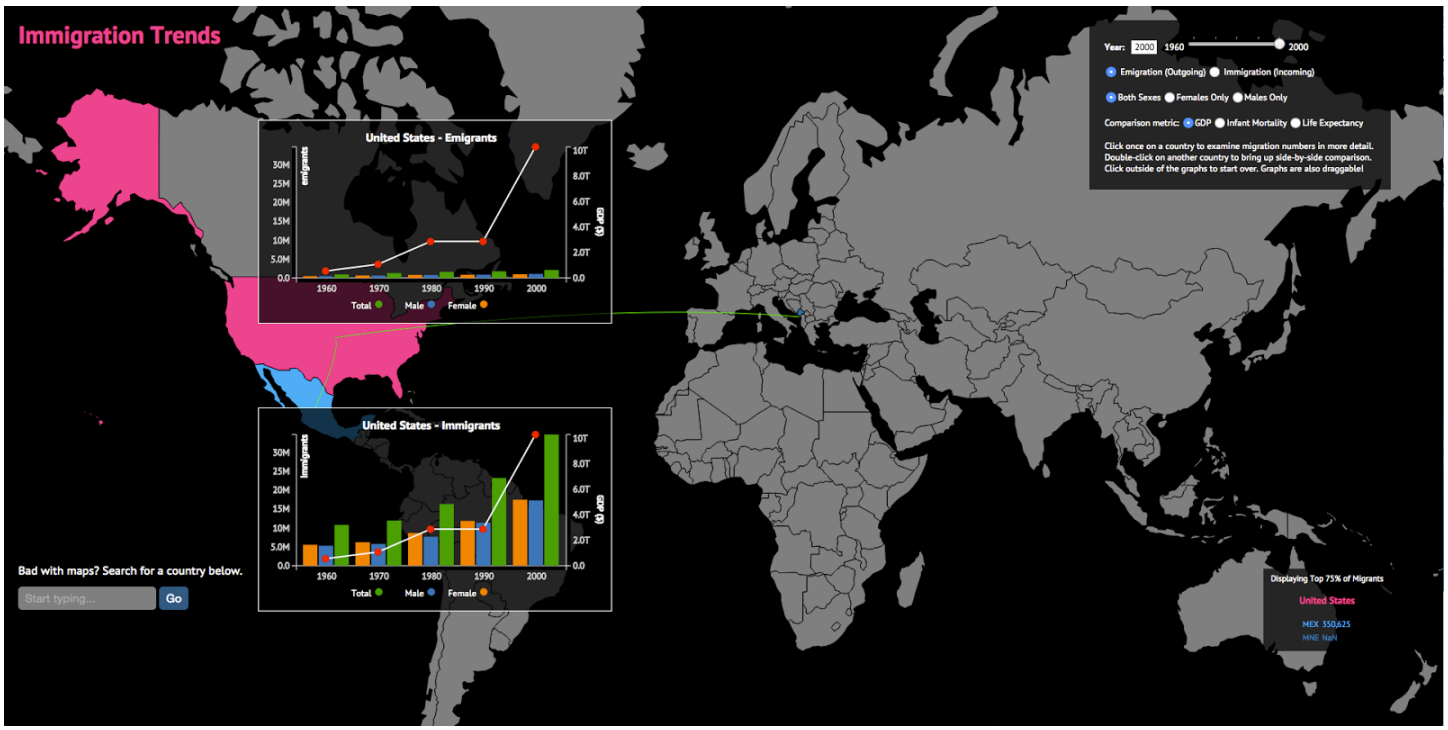
In addition, we needed to figure out how to integrate these different ideas -- how many views did we need, and how would we display them? We decided to keep the world map in permanent view, because it is useful for the user to reference. We originally also wanted to show each country's top 10 migrant destinations/sources, but realized through our discussions with Romain that not all countries should be treated the same way, because this top 10 may only account for 10 or 20% of a country's immigrants. Therefore, we decided use a 75% cutoff.

Lastly, we decided to implement additional comparison metrics such as GDP, infant mortality, and life expectancy into our graph because migration numbers mean nothing without context -- even when compared against itself over time. We decided to overlay this information on top of our migration data for easy viewing.

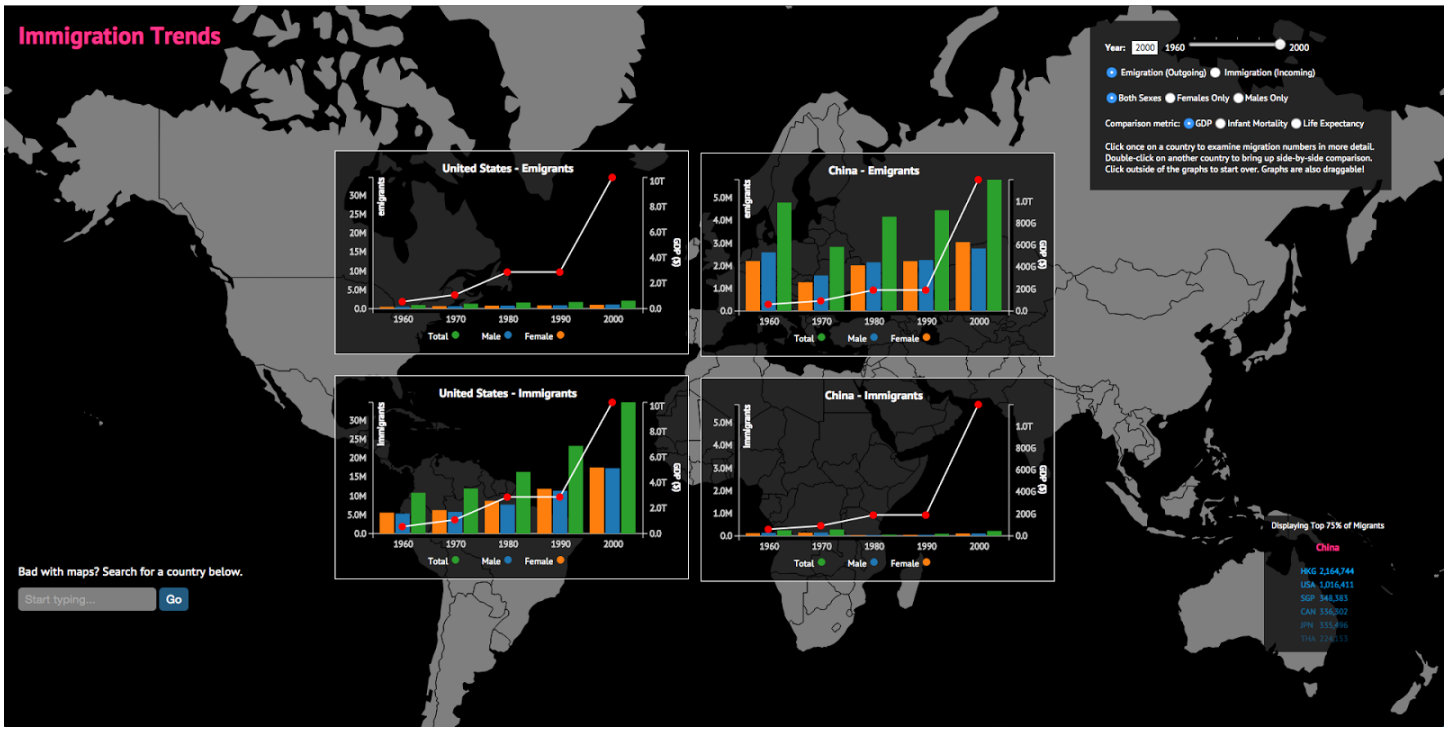
We finally decided on the following three views:
WorldView: (the landing page with the world map -- arcs bloom upon country mouseover)



CountryView: (pops up bar graphs when a country is clicked)

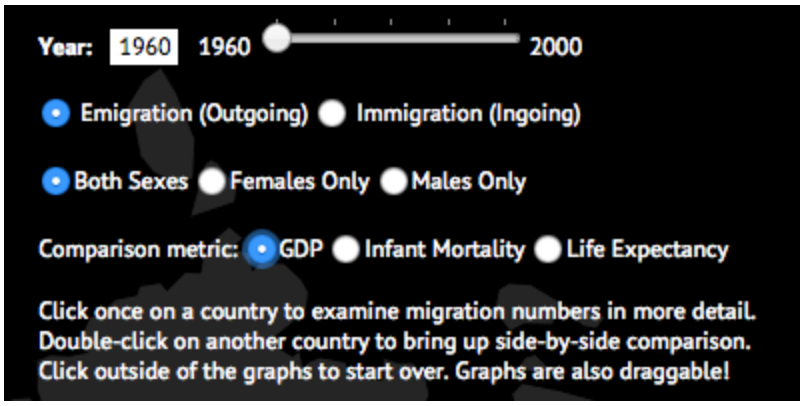


TwoCountryView: (pops up another set of bar graphs when a second country is double-clicked)



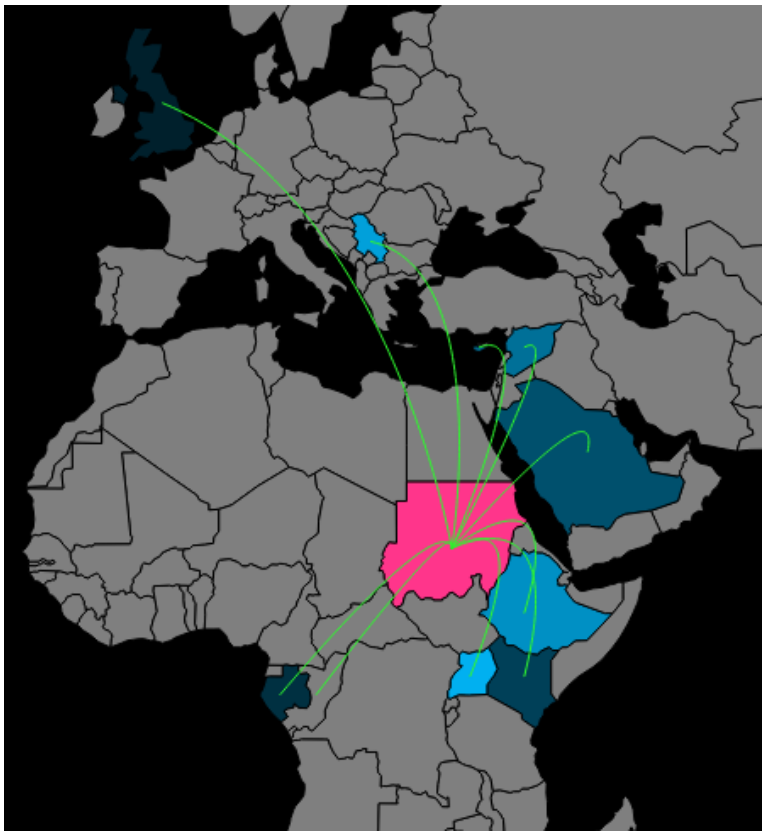
Implementation

For our WorldView, users will have the option to toggle between *ingoing (immigration)* and *outgoing* (emigration) data, as well as for different years and by gender.

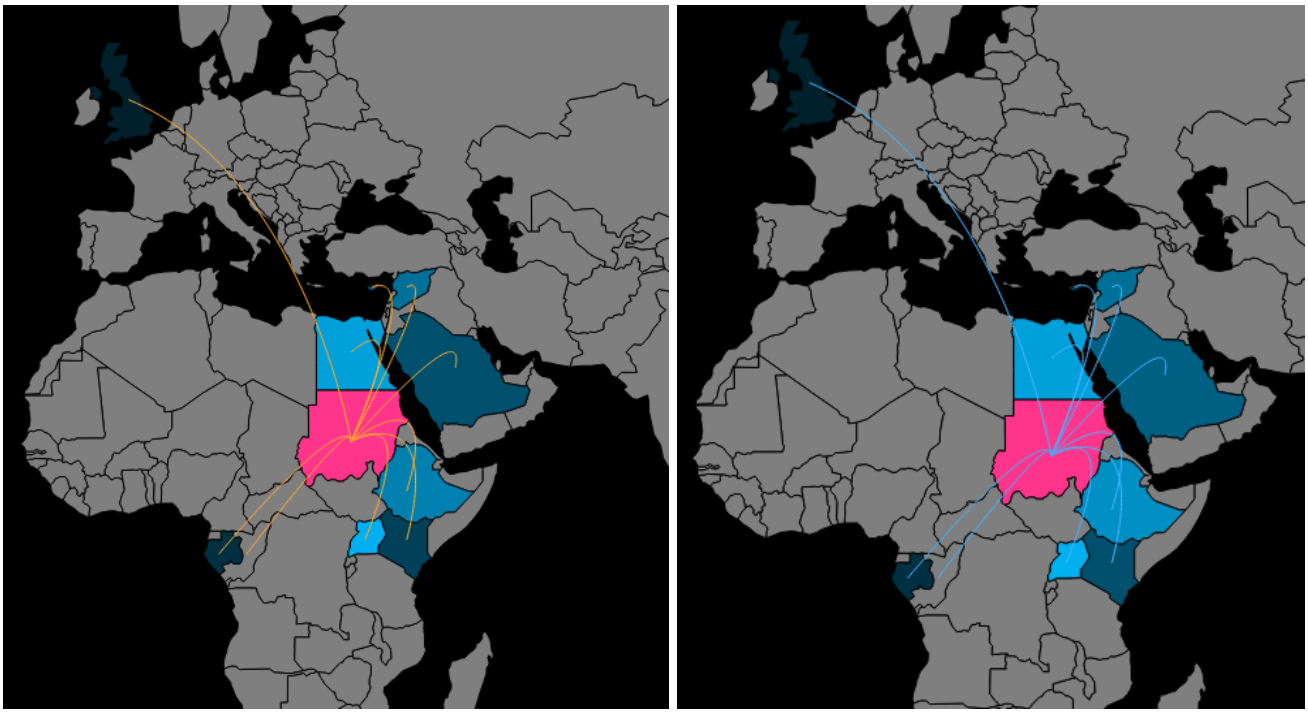


Our legend appears in the upper-right hand corner of the landing page. Users can toggle between different years, genders, and emigration/immigration.

When the *Emigration* radio button is selected and the user hovers over a country, outgoing arcs are drawn from the country to the top destination countries that account for 75% of the emigrants out of the hovered country. When the *Immigration* radio button is selected and the user hovers over a country, incoming arcs are drawn from the country's top immigrant sources that account for 75% of the immigrants to the hovered country.



When the user hovers over a country, arcs will bloom into or out of that country based on whether or not the Emigration or Immigration radio button in the legend is selected.

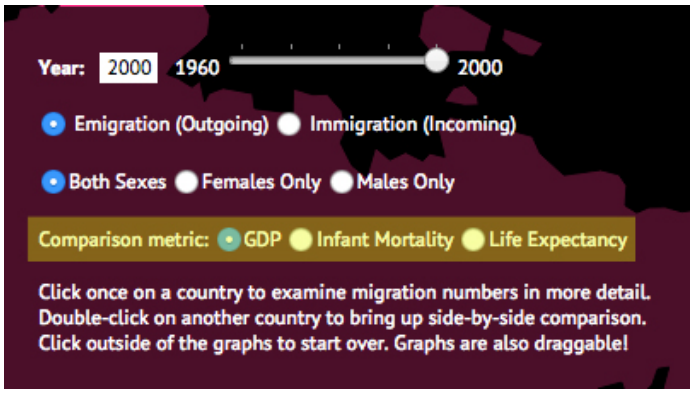


Left: female data. Right: male data.

A table in the lower right-hand-corner displays numerical data giving the top 75% of migrant population's destinations/sources. Furthermore, these are color-coded in a gradient that matches the color of the country on the map. As stated previously, brighter countries indicate a larger migrant population to or from that country, and darker countries indicate a smaller migrant population to or from that country. They fade into darkness to show they are less significant to that particular migration subset.

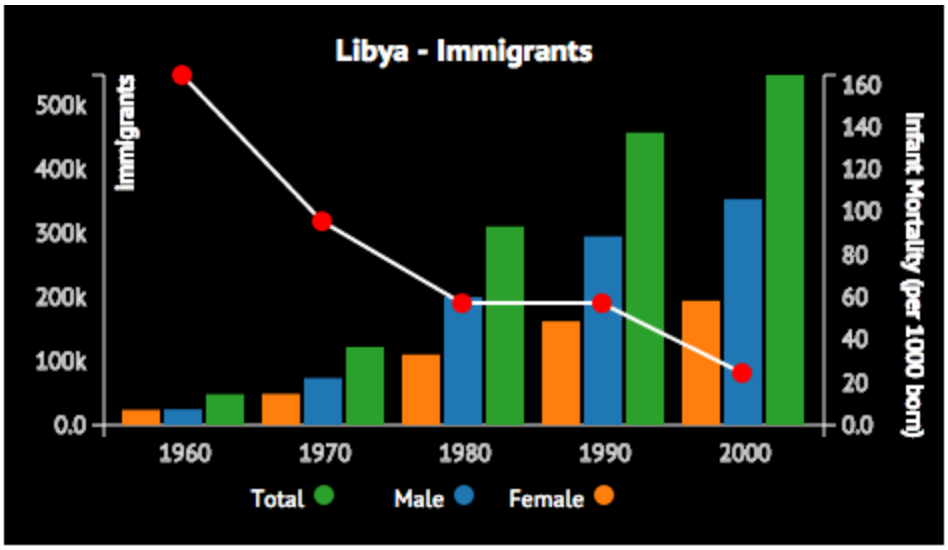
Users can also analyze specific countries. Upon clicking a country once, two charts appear on top of the world map displaying the total number of emigrants and immigrants to or from that country in the years 1960, 1970, 1980, 1990, and 2000 and separated by Male, Female, and Total. These are generated so that they share the same y-axis scale for easy comparison.

In addition, users have the option of selecting a Comparison Metric from the original legend. When either GDP, Infant Mortality, or Life Expectancy is selected, the bar graphs will update to show an overlaid line chart with data corresponding to the requested metric.



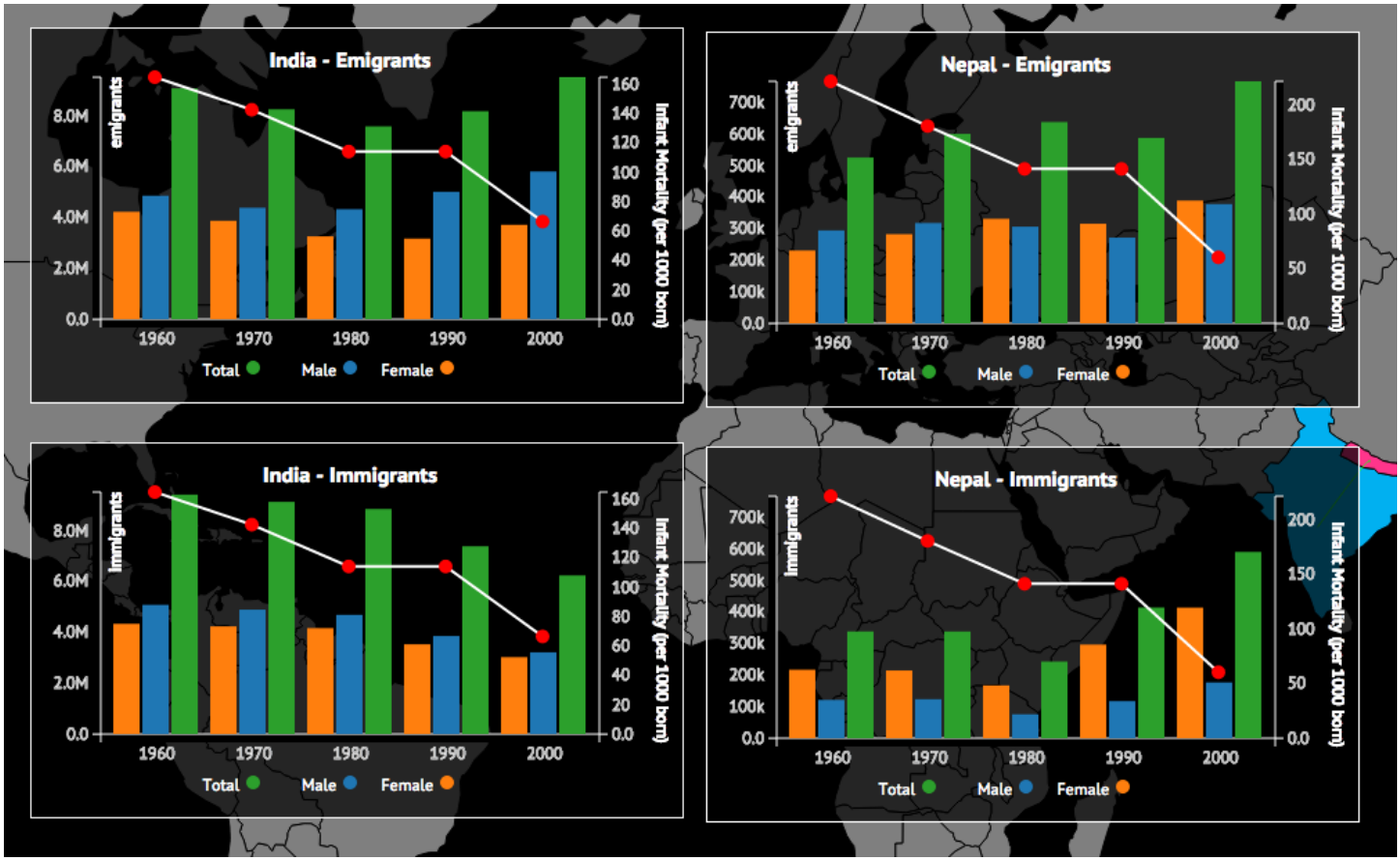
The main legend gives three options for comparison metrics: GDP, infant mortality and life expectancy.

United States and China immigration and emigration data with overlaid line charts. GDP is represented in USD and Infant mortality rates are in deaths per 1,000 live births.



Libya immigration data with infant mortality data overlaid. One can see that as infant mortality declined over the past half-century, immigration to the country skyrocketed.

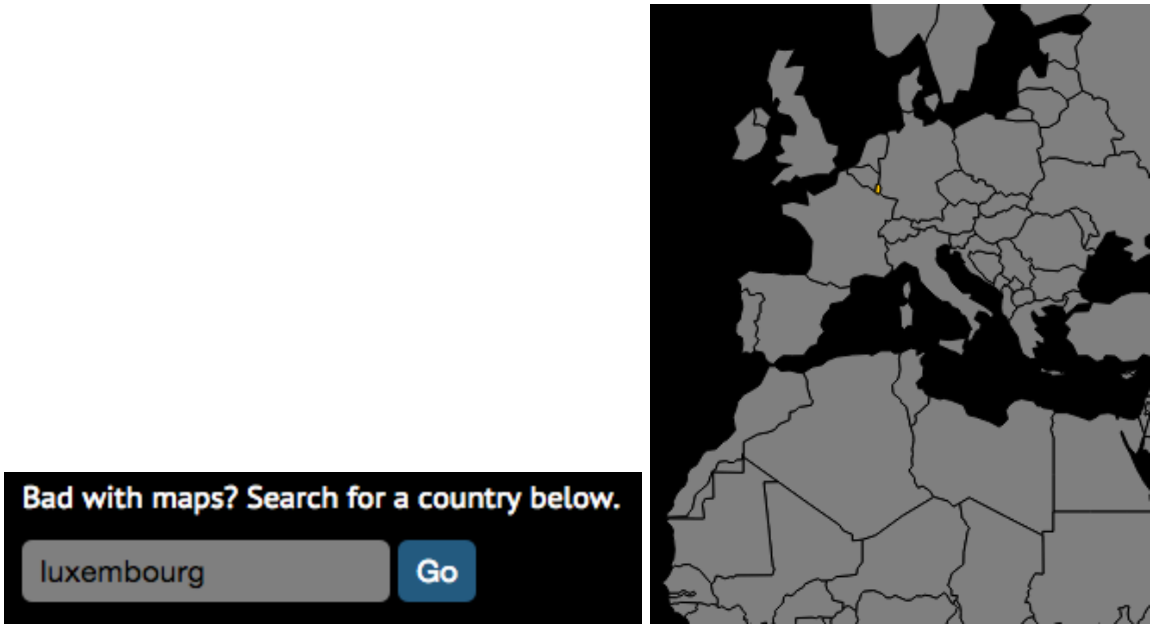
The user also has the ability to view two countries side-by-side -- click once on a country to examine migration numbers in more detail. Double-click on another country to bring up side-by-side comparison. Click outside of the graphs to start over. Graphs are also draggable!



Two-country view with India and Nepal selected.

In the event that a country is difficult to find on the map for the user, there is also a search bar in the lower-left hand corner of the world map which will highlight that country on the world map when entered. You can begin typing in the name of the country and the search bar will give auto-complete suggestions of possible country names. We implemented this using

Typeahead.js in conjunction with our list of country names from **Datamaps**. When the country name is typed in, then you press the “Go” button and the correct country will flash yellow on the map to make it better locatable.

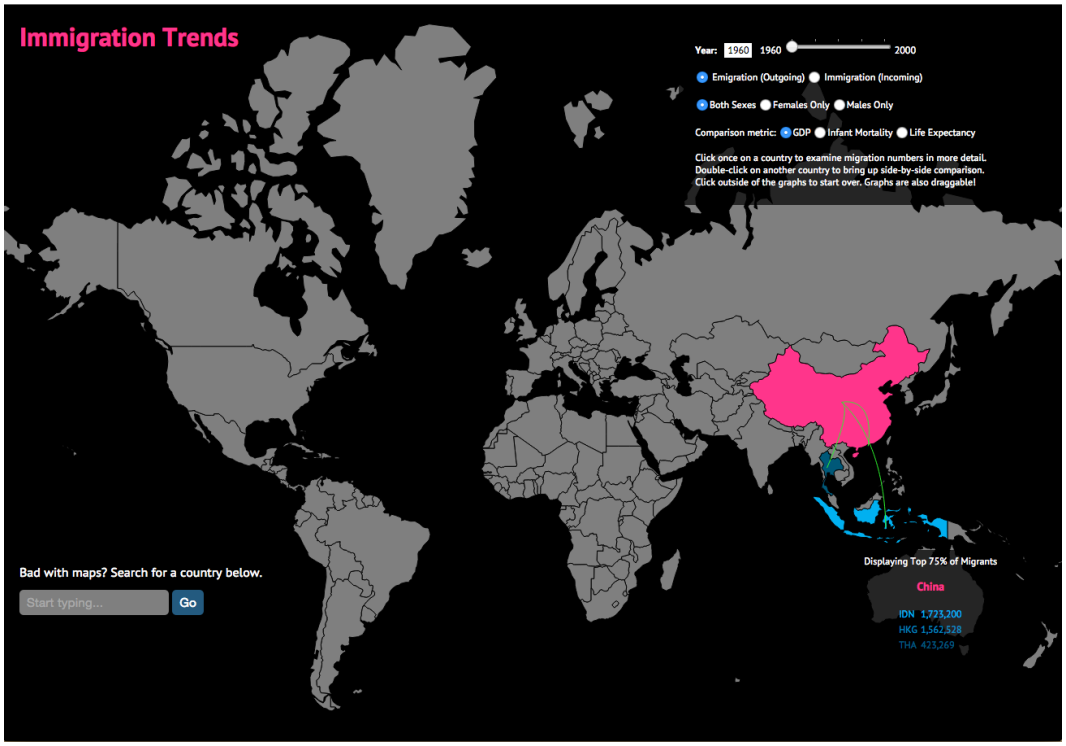


Entering “Luxembourg” into the lower-left-hand search bar makes the country flash yellow several times on the screen so you can find it.

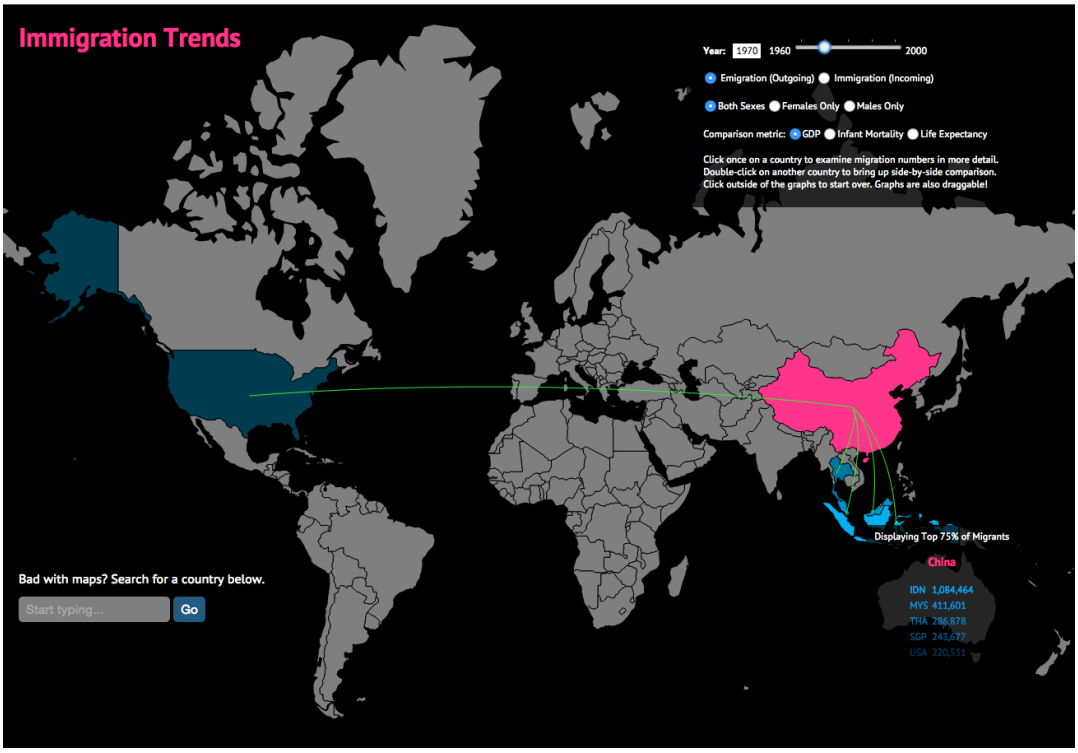
Evaluation

Our visualization reveals interesting trends about human migration flows over the past half-century. Below are just a few:

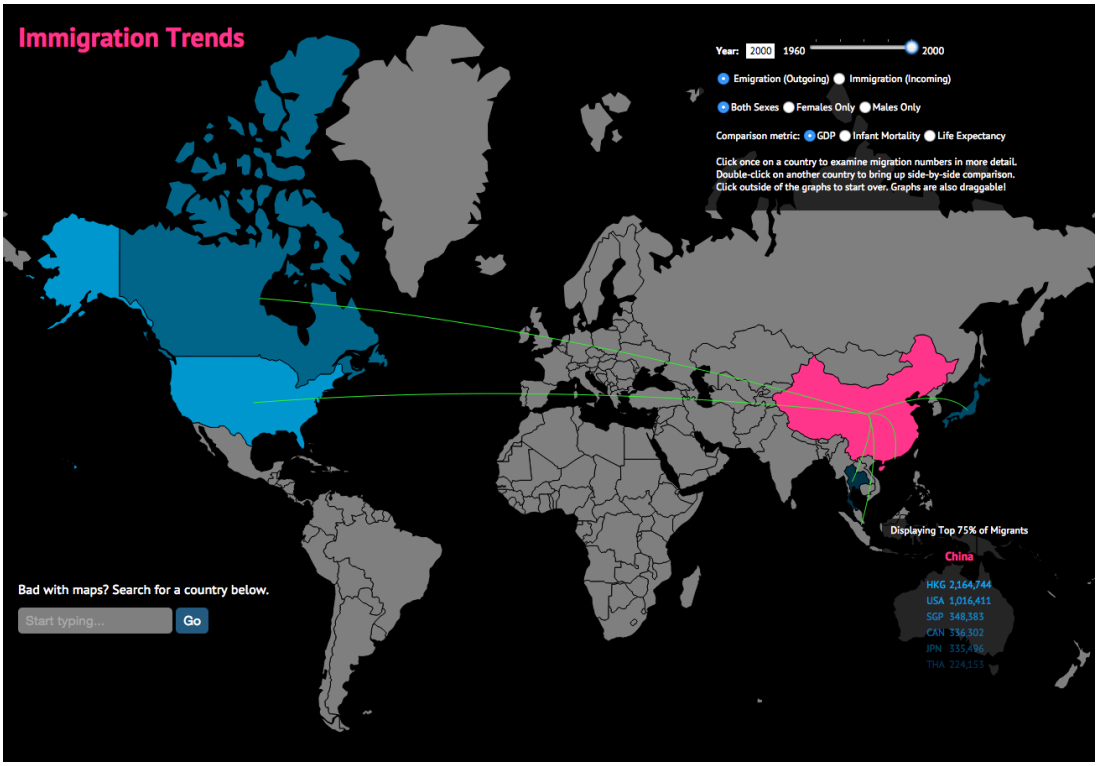
In general, we see that most countries have experienced an increase in migration (whether incoming or outgoing) since 1960. Furthermore, people nowadays have the ability to migrate to almost anywhere in the world, thanks to improved transportation, technology, health and flow of ideas. For example, let’s use our visualization tools to learn about China.



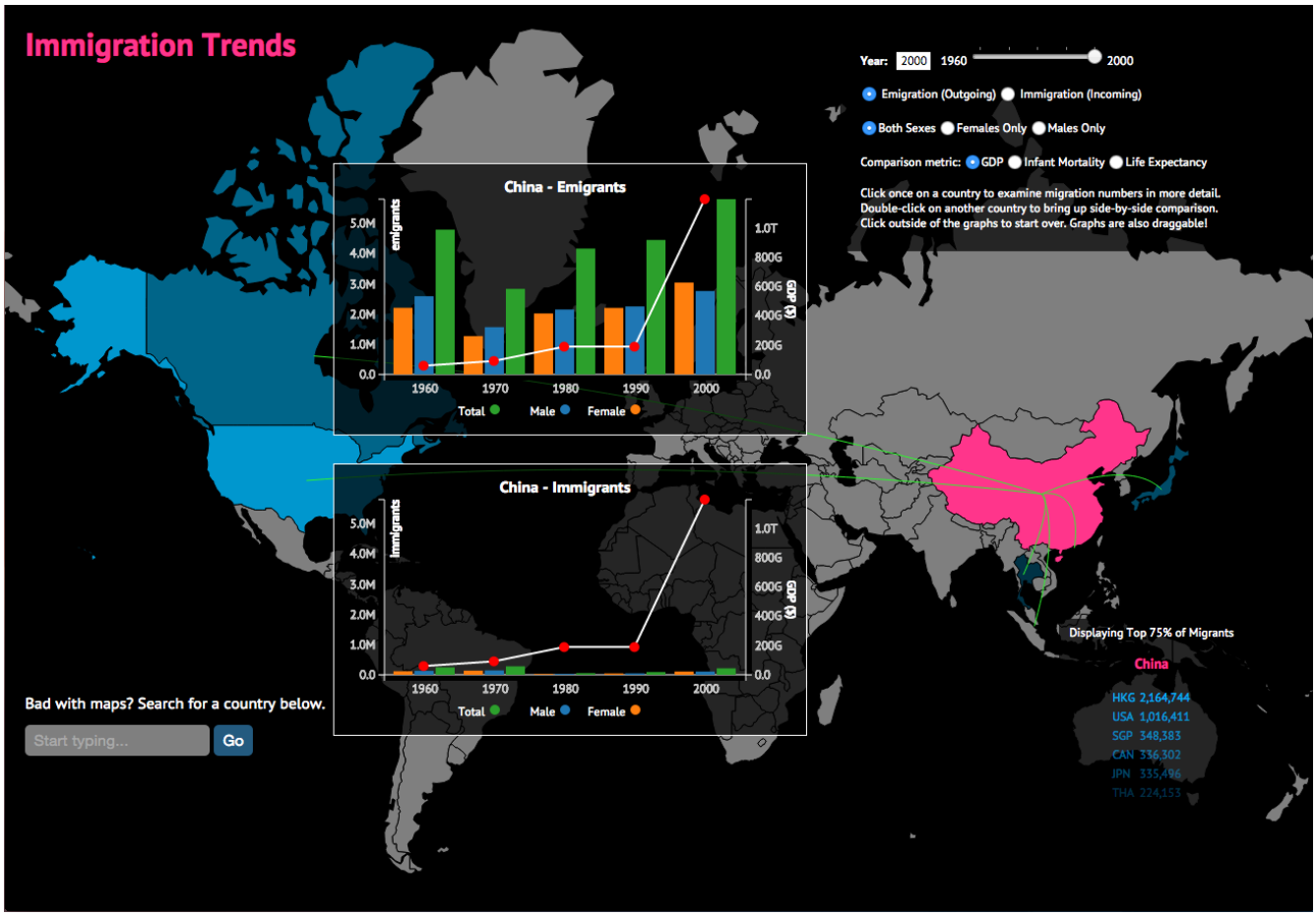
China emigration data, 1960. Over 75% of Chinese emigrants travel to Hong Kong, Thailand or the Philippines.



China emigration data, 1970. The United States, Singapore and Malaysia have been added to this list.

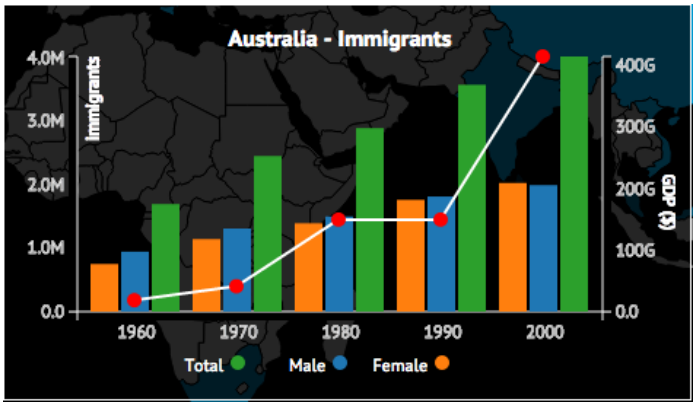


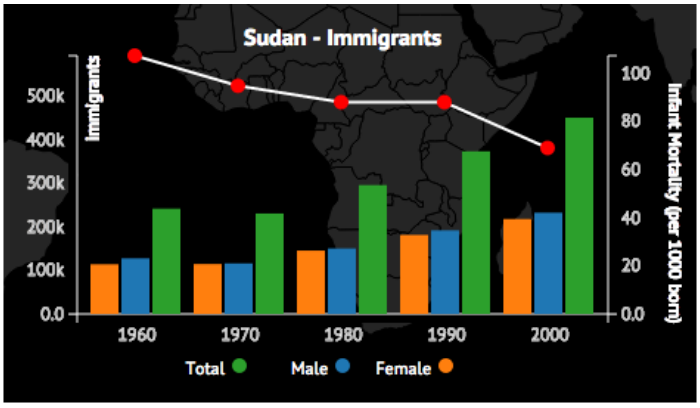
China emigration data, 2000. As you can see, emigrants are moving further and further away from China. Nowadays, the top emigration destinations are Hong Kong, USA, Singapore, Canada, Japan and Thailand. Why is this the case?

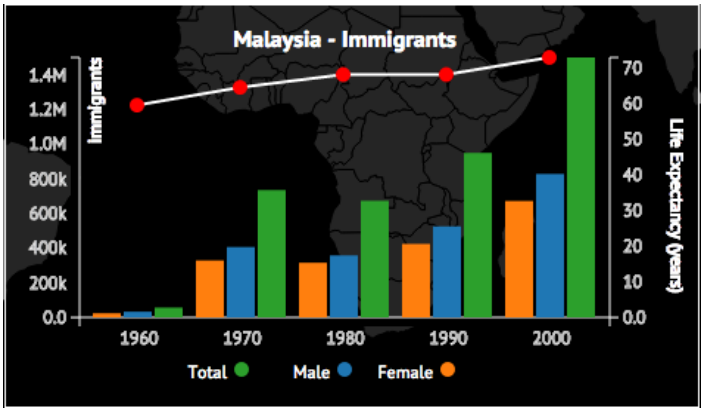


Upon clicking on China, the user observes that emigrants leaving China far outnumber immigrants to China. This is an interesting observation, especially given that China has grown tremendously over the past 50 years and is now a world economic superpower, as is indicated by the GDP data shown on the line chart. One possible explanation is that as the Chinese become wealthier, many seek business or educational/family-raising opportunities overseas, particularly in other developed countries.

In general, we notice that immigration levels increase with life expectancy and GDP and decrease with infant mortality rates. Furthermore, GDP levels and life expectancies have been increasing over time, with infant mortality rates on the decline.

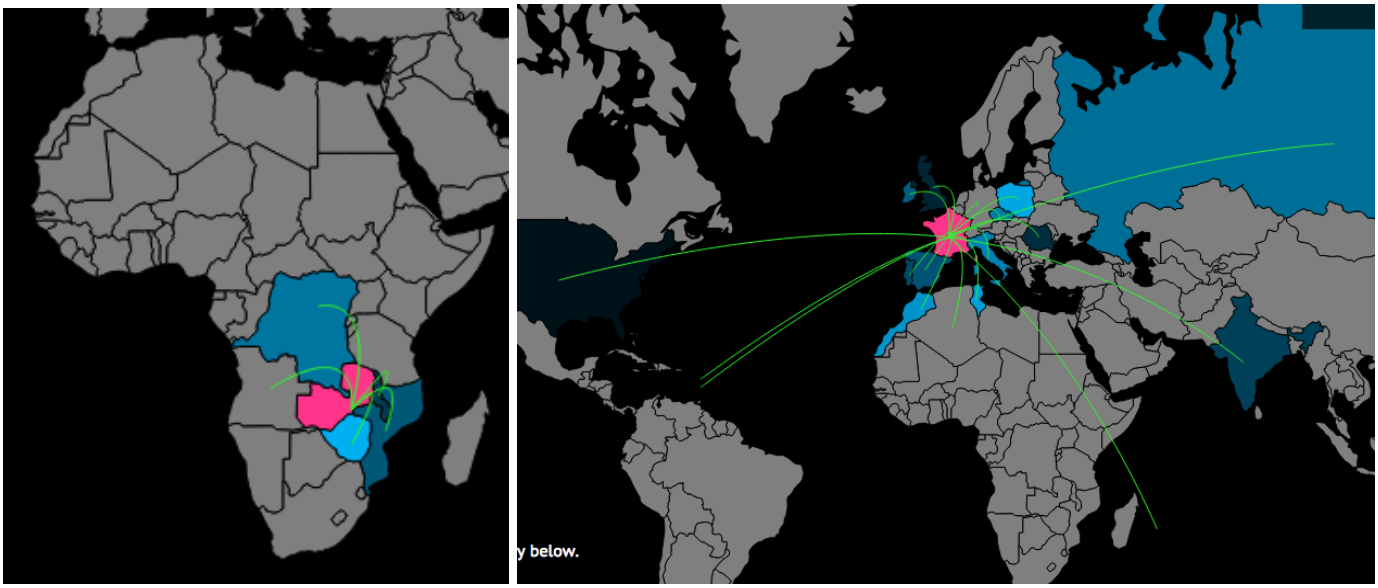


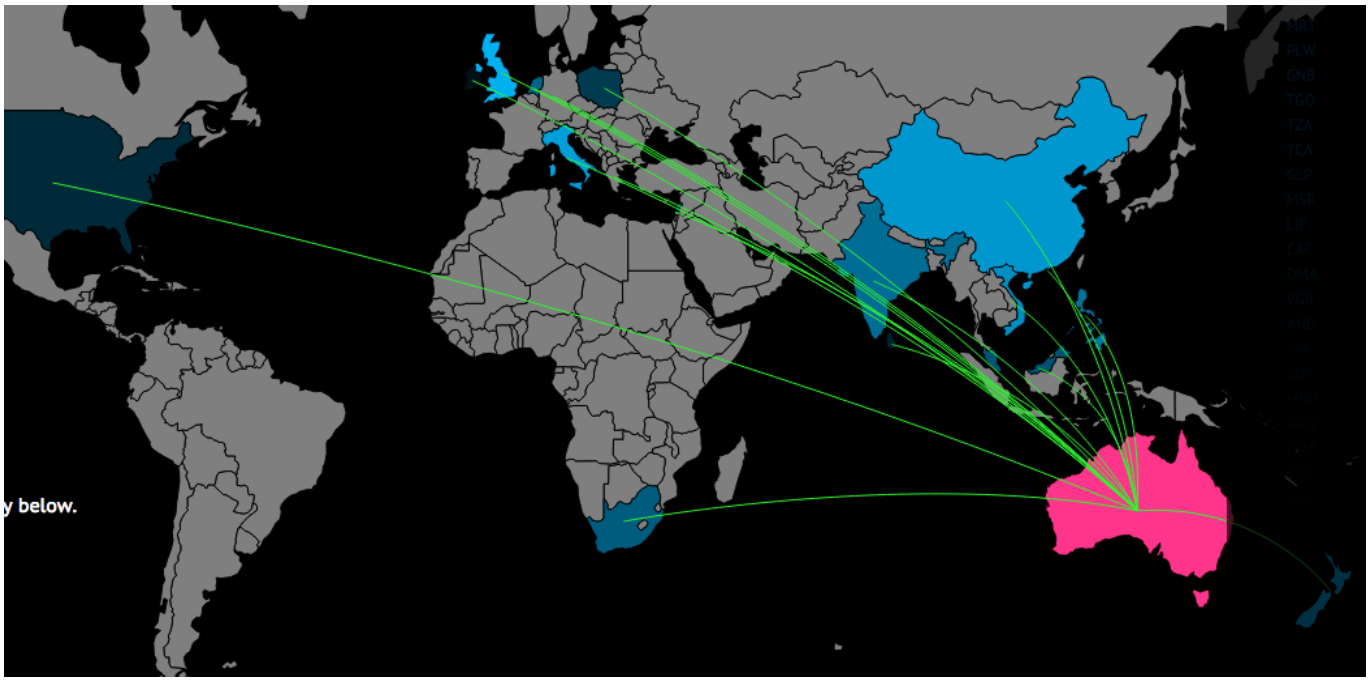




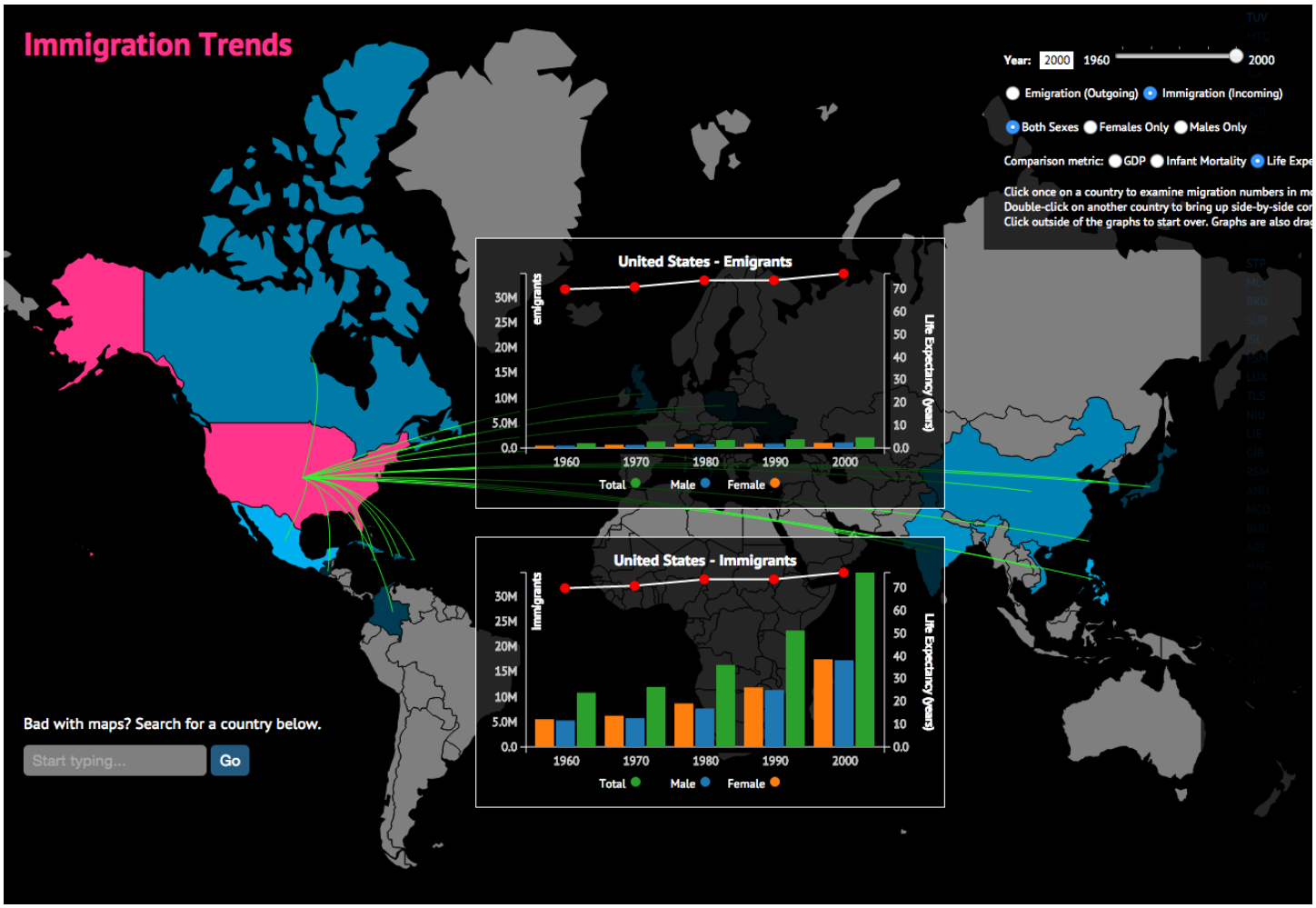
We also notice that migration from developing to developed countries is the fastest growing component of international migration. Consider Taiwan:

Upon the same vein, we also observe that immigrants to African countries are largely from other countries in Africa, whereas immigrants to North America, Europe and Australia come from all over the world:





The United States has and continues to be the most important migrant destination in the world, with over 30 million immigrants in 2000 alone:



Finally, although most migrants are still male, the proportion of female migrants has increased noticeably between 1960 and 2000.