

# Visual Factoring of Historical Immigration Flows in the USA

Ruoxuan Cheng, Yitian Fan, Sanchuan Jin, Yu-Ru Lin,  
University of Pittsburgh

## Abstract

In this paper, we present a prototypical system for visually factoring historical immigration flows in U.S., with a goal to reveal multiple immigration factors and their interactions simultaneously. Our visualization system contains three aspects of immigration, which are historical trends, immigration flows and immigrants' details, involving the area chart, the choropleth map and the bipartite chart. In particular, we design a novel factor-driven bipartite chart for presenting both immigration flows and detailed information to get more informative representations. We cover different aspects of the data to help users get a step-by-step understanding of the immigration of U.S.. Based on the preliminary user study, we demonstrate the utility of our visual design in assisting users to gain more in-depth understanding about the complex factors (e.g., geographic, ethnic, etc.) in the immigration issues. This design can be potentially used for discovering other social issues which involve complex factors and interactions.

**Keywords:** immigration; immigrants; visualization; visual factoring; USA

**doi:** 10.9776/16578

**Copyright:** Copyright is held by the authors.

**Contact:** ruoxuan.cheng@gmail.com, yurulin@pitt.edu

## 1 Introduction and Related Work

Immigration to the United States is a complex demographic phenomenon. From the beginning, it has been a major source for population growth and culture change. It is important to have immigration data for further analysis in history research and society analysis.

Since 1820s, the data has been kept well in Department of Homeland Security (DHS); however, the current visualizations in DHS are not abundant, and much visual work about it is not sound enough (? , ?). These shortcomings inspired us to develop an on-line visualization system to factor the migration flows and to cover various aspects of the data.

Traditional methods focused on the immigration trend, but the immigration factors also play crucial roles in the study of American immigration (? , ?). There are some previous studies that focus on the flow-base visualizations (? , ?). The main purpose of a flow-base visual work is to use flow maps to discover the relationship between immigration data and spatial information (? , ?). As we try to find a more informative and intuitive representation for the flow visualization, we refer to some previous work about employing data mining to represent migration flows and interaction to show the data (? , ? , ?). We also explored some brand-new visualizations which generate a better representation of spatial flows (? , ? , ?).

When building our own system, we try to cover most of the necessary aspects of an immigration completion, which are “who, where and how”. In our dataset, these three aspects are historical trends, migration flows and immigrants' details. We use the trends to know the historical situation; factor the flows to connect them with geographic location; and factor the immigrants' details, such as age, gender, occupation and so on, to make them as flow-like factors, which is an innovation in our work.

Our design also involves a cognitive order, which can help our users explore immigration step by step. We treat trend-flow-details as a suitable step for understanding the immigration completion. Different visual encoding and visual methods, as well as interactive designs, are used in our system. The most innovative part in the system is our use of bipartite charts to represent both the migration flows and immigrants' details. Comparing our designs with others, we see that although the chord charts produced good examples (? , ?), our bipartite chart shows a representation of the flow that is more clear and intuitive.

In the last part, we introduce the preliminary user study, discuss the advantages and disadvantages of our system and present the future plans.

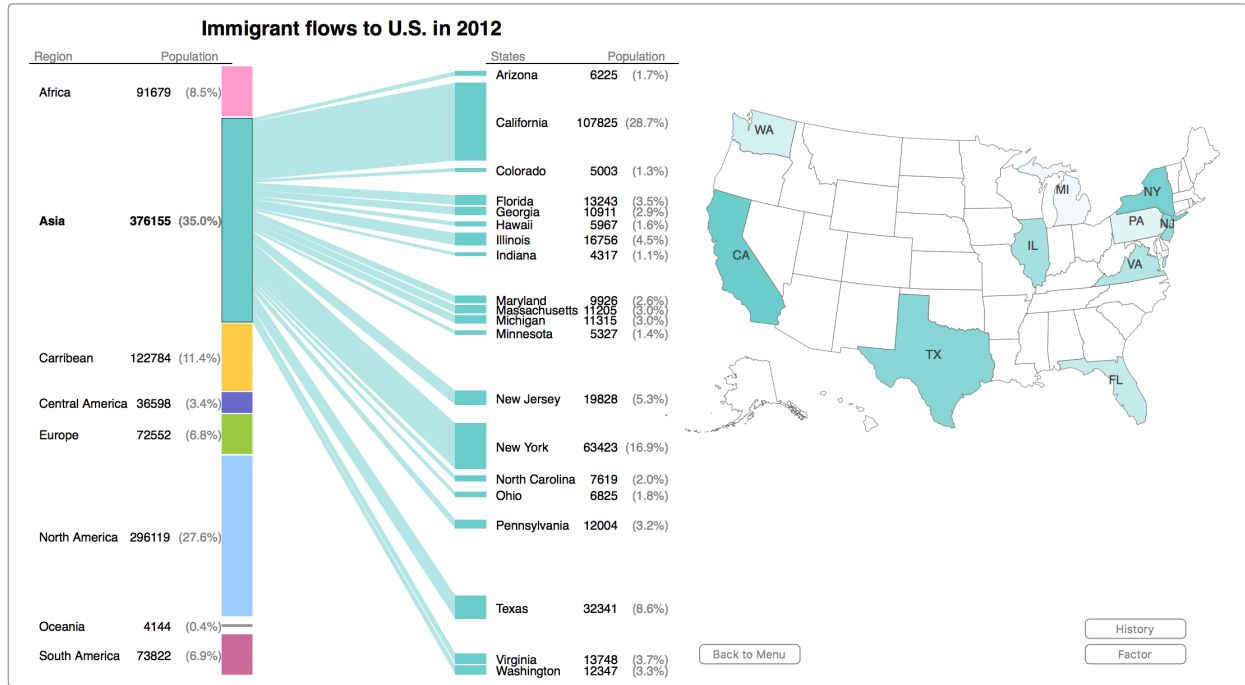


Figure 1: Immigration flows to USA in 2012 with the bipartite chart and the choropleth map

## 2 Visualization Design and Layout

### 2.1 Data Description

We use the immigration dataset from Department of Homeland Security (DHS). The data is across 180 years (1820s - 2000s), including the total immigrants' counts as historical trends, flow information (the places of departure and destination), and details of the immigrants, such as age, sex, occupation and so on.

### 2.2 Layout Design

#### 2.2.1 Main Page

The main page includes the system name and a brief introduction as instructions for new users. The three buttons below the introduction can lead users to different visualization designs. There is also a preview function in the main page. When users hover different buttons, they can see a related graph, which is the according design of each part. This function aims to help users overview and select the part which attracts them most when they don't have a target.

#### 2.2.2 Visualizing Historical Trends

In this part, we use the area chart to show the historical trends of the immigration to the USA. The historical trends are visualized by two key words: (a) region and (b) format. There are seven choices for the region part: six different regions and the total. The format part contains number and percentage in which data can be shown by different views. When showing the trend, we also use different colours to represent the main countries which contribute more counts in a certain region. For further exploration, users can click the buttons on the left and right to change regions and formats for comparisons.

In Figure 2, the example display of this part, the historical trend of Asia as the region with its main countries is shown both by number and percentage, which can help users get a precise understanding by different views. For example, from figure 1, users might lose information by just seeing the left number chart; however, by viewing the right chart, they will surprisingly find that Asia played an important role between

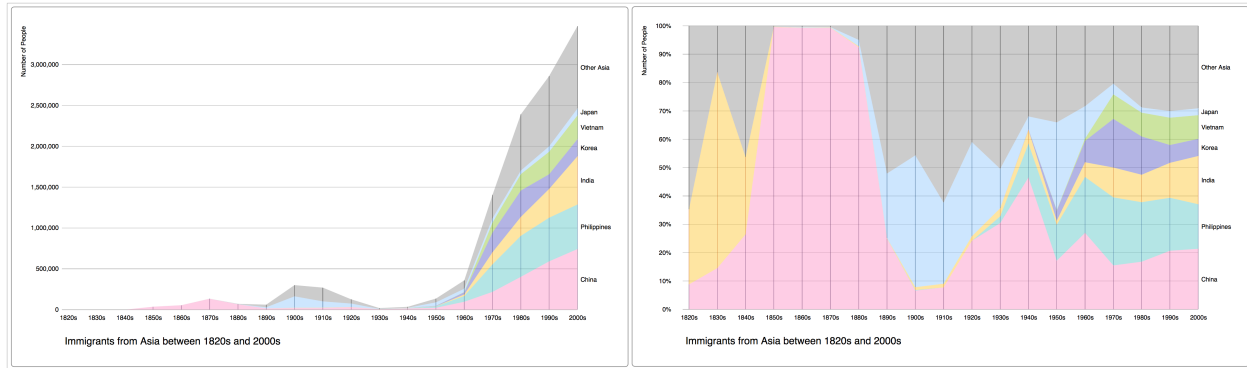


Figure 2: Immigrants from Asia (1820s-2000s) (Left: By Number/Right: By Percentage)

1850s and 1880s. Users can be inspired by the underlying facts they found for further explorations, which may lead them to find the reasons that might cause that high percentage in this period.

### 2.2.3 Factoring the Flows

We use the bipartite chart and the choropleth map together in this part. The bipartite chart visualizes the movement of immigration from different continents to according U.S. states. We produce visualizations based on three key words: (a) year, (b) continents, and (c) states. The left column of the bipartite chart shows the continents in the world; the right column shows the main states in USA; and the links represent the immigration flows from continents to states. The height of each bar represents the number/percentage of the immigrants. The choropleth map located right besides the bipartite chart is to help users understand the geography positions of the states as well as the population density at the same time. Figure 1 is an overview of how we factorize the migration flows in our system.

We use the bipartite chart to visualize the migration flows as an innovation. The bipartite chart allows users to view the flows from both directions and combine all the information together, such as the directions, the population density for both the whole flow and individual flows, the comparison of the flows for both origins and destinations, and the specific names and numbers. Moreover, the bipartite chart combines with the choropleth map which can show geography locations. Thus, the users are able to view the movement, the density, and the geographic information in the same figure, as well as changing for comparisons by hovering. In original data, location is an abstract factor; also, movement and population can't be understood together directly just by numbers; so we believe our design will largely improve the understanding of migration flows for our users.

### 2.2.4 Factoring immigrants' detailed information

This chart is the last design of our system. Assuming that users have already had a general idea of immigration trends and flows, they want to know more about the kinds of the immigrants who form those trends or flows. We choose the bipartite chart here to factor the details to visualizations. We gather three key words in this design: (a) year, (b) regions, and (c) immigrant' types. The design in figure 3 visualizes the percentage of occupation in each continent in year 2012, as well as how much percentage of each continent for a certain occupation (e.g. management and related). Users can click buttons on the left to change the region between continents and U.S. geographic regions; also, users can click the buttons on the right to select the factor type in gender, age, marital status, occupation, and broad class.

The data of the immigrant' detailed information in this part is not really flow data. But we use the bipartite chart again as we found it is helpful for representing these data. Through our study, the story we want to tell is people from different continents or to different regions are going to have these detailed information for their own. We think the data has similar features with flow data, so the bipartite chart that treats them as flow-like visualizations performs better than other traditional methods, such as bar charts or chord charts. In figure 3, users can view the exploration by locations or by different details, which can help

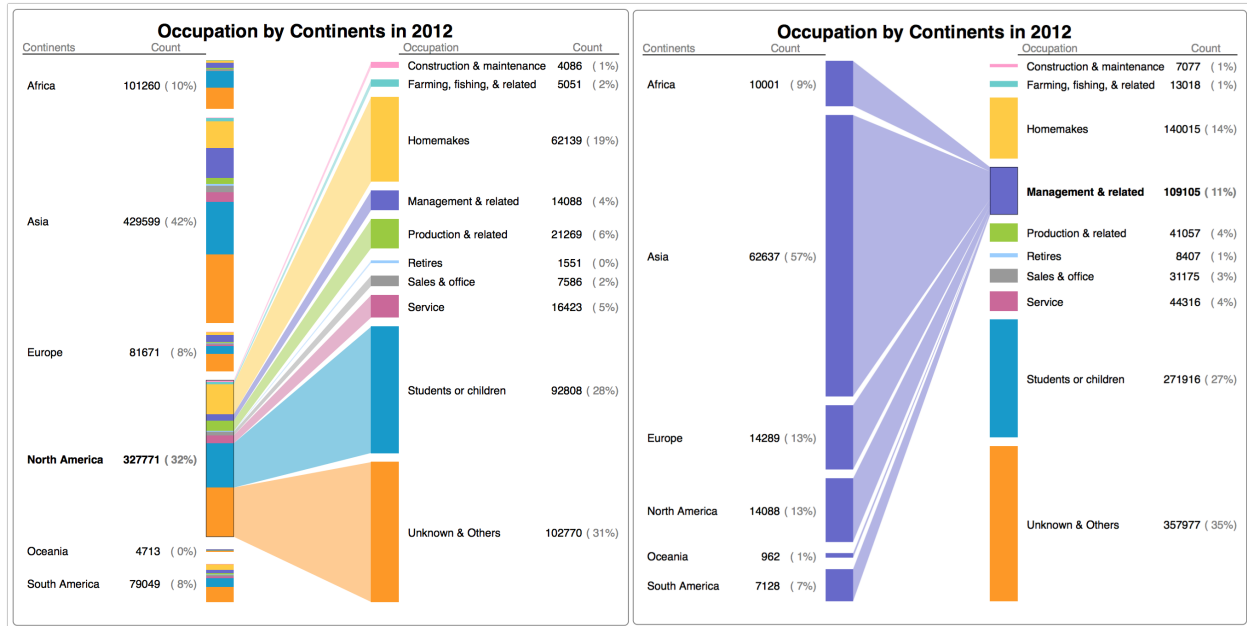


Figure 3: Immigration situation of occupation by continents to U.S. in 2012

them know the relations based on various aspects. Also, users can easily compare the data among different details. The bipartite charts have a good performance here in telling the story about the immigrants.

### 3 Preliminary User Study

To evaluate the feedback to our system, we have conducted a Preliminary User Study with five evaluation questions and three testing questions. Our participants included 21 candidates in total, with 13 males and 8 females by gender; and 5 undergraduates, 13 masters and 3 PhD students with education background. Their self-evaluation of immigration knowledge is 2.76 out of a score 5. The five questions are related to five evaluation criteria, which are functionality, effectiveness, efficiency, usability and usefulness. By evaluating the users' feedback, we found that most of them feel this system can inspire their thinking and is helpful for them. This shows our system is doing well in in-depth understanding about complex factors and has potential for discovering social issues. However, the relative lower scores in functionality and usability parts show what we need to improve. For example, some users expect more interaction functions in the flow part, such as playing with map. Also, users expressed they were willing to see more different kinds of charts in the system. We will work on our system in the future based on the users' feedback.

### 4 Discussion and Future Work

The main advantages of our system are as below: 1. It is a whole system with different types of data, which can give users an overall understanding of immigration in the USA. 2. The three parts of our design are step by step. Users can follow them to get a better exploration, from the overall trend to flows, then to the detailed information. 3. We choose suitable types to visualize different data. Area charts are good to visualize trend and proportional comparisons. Combining bipartite charts and choropleth map together can give users a clear mind of flow data. Also, bipartite charts are informative for visualizing immigrants' details as users can combine the original places or destinations and the detailed information as a whole. 4. Our interaction design is easy for users to master and direct for them to understand. Users can interact with our system without confusion and additional guidance.

According to the feedback in our preliminary user study, there are still some limitations that we plan to address in the future. First, although we design the system as a whole and make them interact with each other, some users may not be engaged in it because they do not have a general sense of immigration

background. We will try to give a tighter relation between our different designs and make it easier for users to consider them as a whole. Second, for the historical trend part, we will add more zooming in and zooming out designs for users to enjoy the visual exploration. The later years contain more data than earlier ones. Therefore, by using zooming design, we can show more data for some parts such as later years and connect them with the other flow and factor designs. Third, the map interaction part needs to be improved. The map will show when users click on the left side. We will add some functions to show the left side when clicking on the map. Also, we will work on the map to let it show more information and choose some other methods to show additional information.

In this paper, we have presented a system for immigration data visualization by integrating several visualization techniques into a new design. Our system aims to help people get a better understanding of immigration in the United States and help them explore more of the data. In the future, we will keep improving our system and making it more informative.

## 5 Acknowledgements

This work is part of the research supported from NSF grant #1423697 and the CRDF at the University of Pittsburgh. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.