

INFO 5100 Final Project

Data Visualization on Migration in USA

Professor: David Mimno

TA: Nathaniel Smith (ncs77)

Team Members: Lu Chen(lc856), Jiayi Tao (jt753)

Part A. Description of work assignment - Final Status Update

Jiayi Tao (jt753):

- Improved interface design
- Added description and instructions
- Added features to enrich the contents and interactions

Lu Chen (lc856):

- Finished modifying data sets
- Made the slide bar to change the year to next one automatically
- Improved the interactions

As a Whole Group:

- Worked together to try to add more features to enrich the content and interactions
- Worked together to solve problems and fix bugs

Part B. Description of Project:

We created an interactive visualization of migration data for each state that allows users to find the trend and detail information about the migration in USA.

We showed the migration trend from four aspects - the total population of each state in each year (line chart), the migration among the selected state with other states (map), and the inflow components (pie chart).

Part C. Description of migration data:

Our main data files contains all states inflow and outflow migration according to the tax return population from 1990 to 2010, obtained at the following link:

<https://www.irs.gov/uac/soi-tax-stats-migration-data>

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

Since each csv represents the outflow or inflow migration data of one state in one year, each state have two data file in one year. And there are 21 years' data. So there are about 2000 data files. Each csv shows the state information, the inflow or outflow between other states. So firstly, we need to clean up the data files using Python.

So firstly, we create two-dimensions tables for each year, showing the state-state migration number. Then we read each file to abstract the state code and id, year, and the migration number between this state and others. If this file is representing the outflow, the state-state migration number in the new created table number need to subtract the state-state migration number showing in this file. If this file is representing the inflow, the state-state migration number in the new created table number need to add the state-state migration number showing in this file. So the new tables represent the state-state net migration number. It takes a long time because the data files do not have the same format.

And we also create a table to get the total population of each states and each year in new creating tables.

The processing data code is in the /data/processing.

The “Zip Code Latitude Longitude City State County CSV” file was downloaded from <https://www.gaslampmedia.com/download-zip-code-latitude-longitude-city-state-county-csv/>.

This file was used in conjunction with a US map to determine the latitude and longitude coordinates of cities.

Part D. A description of the mapping from data to visual elements:

At very first, we decided to use a simple line chart to show the overall population of the United States from year 1990 to 2010, showing trend of an intense population growth within the 20 years.

Below the line chart is a slide bar together with a big USA map. In order to compare inflow and outflow migration population in a straightforward way, we use circle colored in two color, green and purple to represent population inflow and outflow of a certain state. The radius of the circle is positive correlated with the population migrated. Also, to compare between the years difference for migration patterns, we add slide bar from year 1990 to 2010. Another feature is to differentiate the population of all states on the map, we color it different color intensity from yellow to red also positively correlated to its population number.

Below the map we decided to add a preview map to help audience better locate the state it chosen on the main map. Also, it also reasons for the color on the following pie chart. In this map, the color is divided by four geographical regions, southern, northeastern, midwestern, and

western. When audience interacts with either inflow or outflow pie chart, it will highlight the state in green, indicating inflow, or purple, indicating outflow.

Since the population of each state compose of the migration between it with other all states, we want to know the inflow and outflow compositions of each state, especially the percentage of migration from each other state. So we choose to create two pie charts to show the inflow and outflow composition of the selected state on main map by user. However, there are too much states to show in one pie chart. So if the color scale is represent one state, it will be too weak to discriminate the compositions. So we divide it to 4 geographically regions. And each one have one color to represent it. So the pie chart will only have 4 color scale. But we also want to see the migration between two states. So the pie chart has one slice for each state. And since the slice may be too small to show the detail information, we want our users to hover on one slice to show the specific information of this state.

Part E. The story. What does your visualization tell us? What was surprising about it?

Immigration is a hot topic these days in United States, yet migration of people within the country have also brought crucial impact to the United States' economy, politics and culture. Americans have already seen this with the Western expansion, the movement of Southern to Northern cities and the migration from northeastern to southern cities such as Florida.

These charts records domestic migration data for the United State since 1990, based on year to year address changes reported on individual income tax return filed with IRS. For every state on the map, we break down the migration population in two ways. You can now see two views for

each state: on the preview map together with the pie chart, you can have the basic idea of this state popular migration destination and other states migrator by regions.

Also, you can also see the positive related dot flow, representing migration population on the main map from its source state to destination state. In addition, it is quite obvious in the map and line chart that the total population of the United States and each state is increasingly growing. Yet, migration patterns among states change constantly. One good example in case is that common belief that California should be one of the most popular destinations for other states migrators. However, surprisingly, shown from the data visualization, California is not as popular destination, and Californians are moving out into nearly every state.

.