Samuel LeBlanc

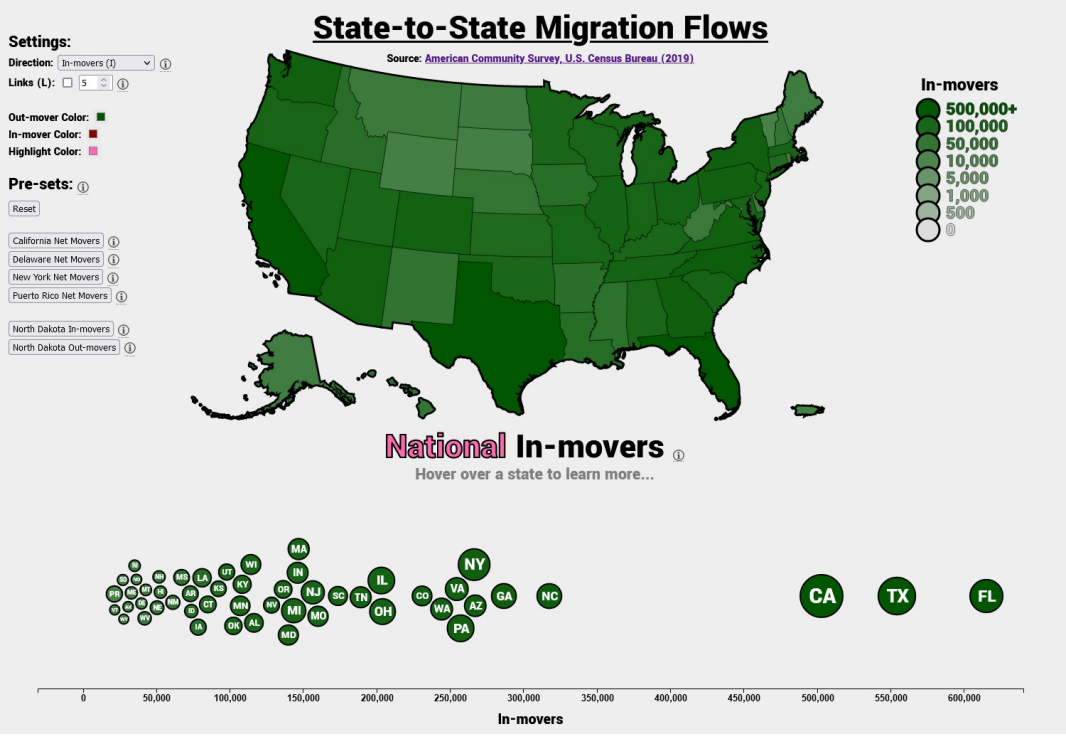
12/16/2022

Data Viz Final Project

Link: https://samleblanc.github.io/Migration-Map/

My initial intention for this project was to create a visual representation of the intersection of the climate crisis and interstate migration. However, due to a lack of expertise in the area of climate modeling and insufficient time, I was forced to narrow my focus to simply interstate migration. As an undergraduate in mathematics, I have always been intrigued by networks and this project was a great chance to illustrate the complex and symmetrical network of interstate migration which holds notable public policy implications for the future. My visualization is a two-part (and I hate to use this word, dashboard) which is made up of a choropleth and a beeswarm chart. The two pieces work simultaneously to display the same migration data with an extra degree of perspective that one chart could not provide alone. The visualization is made with D3, a combination of v4 and v6.

Note, I have been struggling with screen dimensions and resizing. I have attached an image that contains all of the elements. If you are missing an element such as the legend or the x-axis, you may need to adjust the browser zoom level.



The main data for this project comes from the *State-to-State Migration Flows* (SSMF) dataset which is a product of the American Community Survey powered by the U.S. Census Bureau. Similar to other data sets from the Census Bureau, the SSMF data are both extremely clean and well-documented, allowing for straight-out-of-the-box analysis. As the name implies, the SSMF provides the total number of migrants (in each direction) between all pairs of states across the United States. The SSMF utilizes the survey machine that is the American Community Survey, sampling 3.5 million households per year on a rolling basis. The American Community Survey and the Puerto Rico Community Survey inquire survey participants (including those in group quarters) regarding their previous residence one year prior. By tabulating the current state of residence crossed with the state of residence one year ago, state-to-state migration flows are calculated. These state-to-state flows are available in full starting in 2005 (the first year of full implementation of the survey), and in part as early as 1999.

The data are provided in Microsoft Excel spreadsheets on the Census Bureau website, here. The rows of each table denote the state of current residence, while the columns denote the state of previous residence in the preceding year. In 2010, the tables were augmented to include estimates of the population aged one year and over in each state, non-movers, and those migrating from abroad. It is important to note that the SSMF data are based on a sample and thus are subject to sampling variability. The degree of variability in estimates derived from sampling is represented by the margin of error. For the sake of clarity, in my visualization, I have excluded the margin of error and displayed only the central value of the distribution. Non-sampling errors, which are not accounted for in the SSMF tables, must also be taken into account when interpreting ACS estimates. Further information regarding sample size and measures of data quality can be found in the American Community Survey Methodology section.

This project also required the use of several supplemental datasets, all of which are obtainable from the Census Bureau. The first of these is a shapefile containing all of the U.S. states and territories. To facilitate the processing of the shapefile, I utilized the web version of Mapshaper, which I described in detail in my Data Visualization Tutorial. Additionally, I also collected state population totals published as part of the 2020 Census data, which I accessed via the Census API.

As noted earlier, data from the Census Bureau is both clean and easily accessible. As a result, minimal pre-processing was necessary. The steps that were taken for pre-processing included converting the

data from its wide format to a long format and joining to the state centroids, which were calculated by taking the average of the latitude and longitude extremes for each state from the shapefile. The finalized dataframe can be found in the file *data/state_migration.csv.*

All the elements of the visualization were created in D3. D3 (Data-Driven Documents) is a JavaScript library for creating interactive data visualizations in web browsers, combining powerful visualization components and a data-driven approach to DOM manipulation. It is used for producing dynamic and interactive data visualizations in web browsers using SVG, HTML, and CSS. This project required significant problem-solving in comparison to my previous experiences with D3, with the most challenging aspect being the implementation of a Voronoi simulation on the back end to ensure the circles in the beeswarm chart did not overlap.

For this project, I borrowed parts all or parts of two other D3 scripts. The first is a custom projection that combined Albers USA with Albers Puerto Rico. I found this script on Observable and it is kept and cited in *projection.js.* In addition, I also used another Observable file to create the links between states and make them bend and look nice.

Personally, I am a huge fan of reader-driven visualizations, especially when I know that the audience is data-literate. To avoid overwhelming viewers with what is already a busy visualization, I opted to let them explore and discover the stories hidden beneath the surface. Still, I wanted to show viewers how these stories may arise, so I provided a handful of pre- settings maps with descriptions of an example of a story that I took away from the data. Human migration is an immensely complex and personal topic, and thus there is an endless supply of diverse and unique stories that can be told. With such a complex network of data, I found it hard to come away with a single story. But the more I worked with the data the more I was inspired by the small stories that appeared. These are what I aim to highlight with the pre-set settings. By inviting readers to uncover their own stories, I aim to make this visualization a classic tale of “show and not tell”.