Process Book

Edit New Page

Jump to bottom

Kevin Reilly edited this page 12 seconds ago · 19 revisions

The Ebb and Flow of US States

Overview and Motivation

Each year, roughly 40 million Americans move once. In 2016, Utah saw the most population growth, much of which was from out of state residents flocking to the state for reasons generally obscured from analysts. Beneath the often cited economic indicators, such as gross product and consumer spending, are the shifting tectonics of population migrations patterns. A great deal of domestic discourse resides on conversations dominated by these migration patters, their fundamental movers, and talk of decisive policy action to enable drivers of industrial growth and innovation.

The overarching idea of this visualization is to provide viewers with a wide lens view of these migration patterns that are often talked about only in glimpses in the public square. If the user can see where people are going to and coming from, and potentially see these patterns linked with economic data showing the various driving forces, then this could enable consumers of the dataset to have a more rounded understanding of the interweaving complexities across the spectrum driving people to unproot their lives and migrate across the globe.

US Census Bureau

The US Census Bureau collects an enormous volume of data from across the United States. The census bureau website provides this data to anyone with web access at no cost, and even provides some clunky and difficult to use interfaces for exploring the available data sets. The objective of this project was to explore this data set and practice the visualization techniques learned un the Data Visualization course at the University of Utah. While exploring the general migration dataset, the goal will be to provide some insight into some questions about migration patterns

Questions

Initially we targeted some basic questions that we knew a State-level migration level flow visualization, coupled with economic data, that spans more than a decade would answer. The questions are as follows:

- Which states have the fastest growing populations?
- For each state, where are people most likely to move to?
- For each state, where are people most likely to move from?
- Which states experience the largest influx of migrants?
- Which states experience the largest exodus of migrants?
- How do these patterns change over time?

As a stretch, we hoped the project may shine some light on the following questions:

- What factors in a state are the most likely to increase emmigration/immigration?
- Are there multiple factors and causes for different regions?

When first drawing up the canvas and analyzing the various data points available with D3, the scale of the data points in relation to other metrics increasingly became important. California would see a large exodus, for instance, but as a proportion of its population, is this really significant? Or is there possibly a larger *proportion* of people leaving some of the smaller states? It was evident that derived metrics would be important in getting a clearer picture of the data. At this point, additions to the data model were made to include population growth, flow as a percentage of the states population, and total migration in and out of the state was added. These metrics help provide better insight into the patterns present, and also drew up more questions. For the states that were seeing a large exodus but healthy population growth, for example, what was causing that large population growth? Was it international migrations? Was it large births, or low death rates? These are not questions we would be able scope into the project.

As the project continued to evolve, it became evident that a aggregate view of the data collected would only provide a bird's eye view of the patterns. The visualization in the final product does answer the first five questions. Seeing that Florida, Texas, and California see the largest influx of migrants from other states, it became clear that the demographics of migrants would be an important data consideration for understanding motive. Aggregate job numbers, taxes, and potential income each prioritize differently as a driving characteristic for college graduates and retirees.

Finally, the patterns themselves proved to be perhaps too broad. Texas and California are large state, with large population centers mostly clustered around the urban centers. Questions about how these urban centers played a role are central to understand any economic or culutral motivations, and would indeed be a great addition to the visualization. County level data would be a great next step, and would provide a cleaner insight into the patterns already observed. A quick toggle on the left side of the heat map from one dataset to the other would certainly be a great way for the user to interact between the two views and compare state and county level data.

Data

The datasets that we will be using are maintained by the US government. The US government census bureau keeps well curated data for various various purposes across the federal agencies and is regarded as accurate. State to State migration pattern data can be found here

The US Buereau of Economic Development can be found here

This page contains data for different economic indicators for each state every year. The data is separated into segregated spreadsheets for each indicator and year. The indicators include GDP, copnsumer spending, personal income, cost of living, etc.

Data Processing

The data is structured, so cleanup wasn't too difficult, but the format needed to be adjusted. Currently all tables are saved as Excel spreadsheets and the formats were modified. The interstate migration data tables have recurring headers to aid the reader as they scroll through the document which will need to be removed. All tables have titles and footnotes that will need to be removed as well. Data extraction and cleanup was done in Python as follows:

- Download the spreadsheets for all the years in an automated way using the python requests module
- Convert each spreadsheet to a CSV file
- Clean up each spreadsheet into a standard CSV format with headers and values
- Convert the CSV to JSON for consumption by the project

While cleaning the data, a few derived quantities will also be calculated. Among them are:

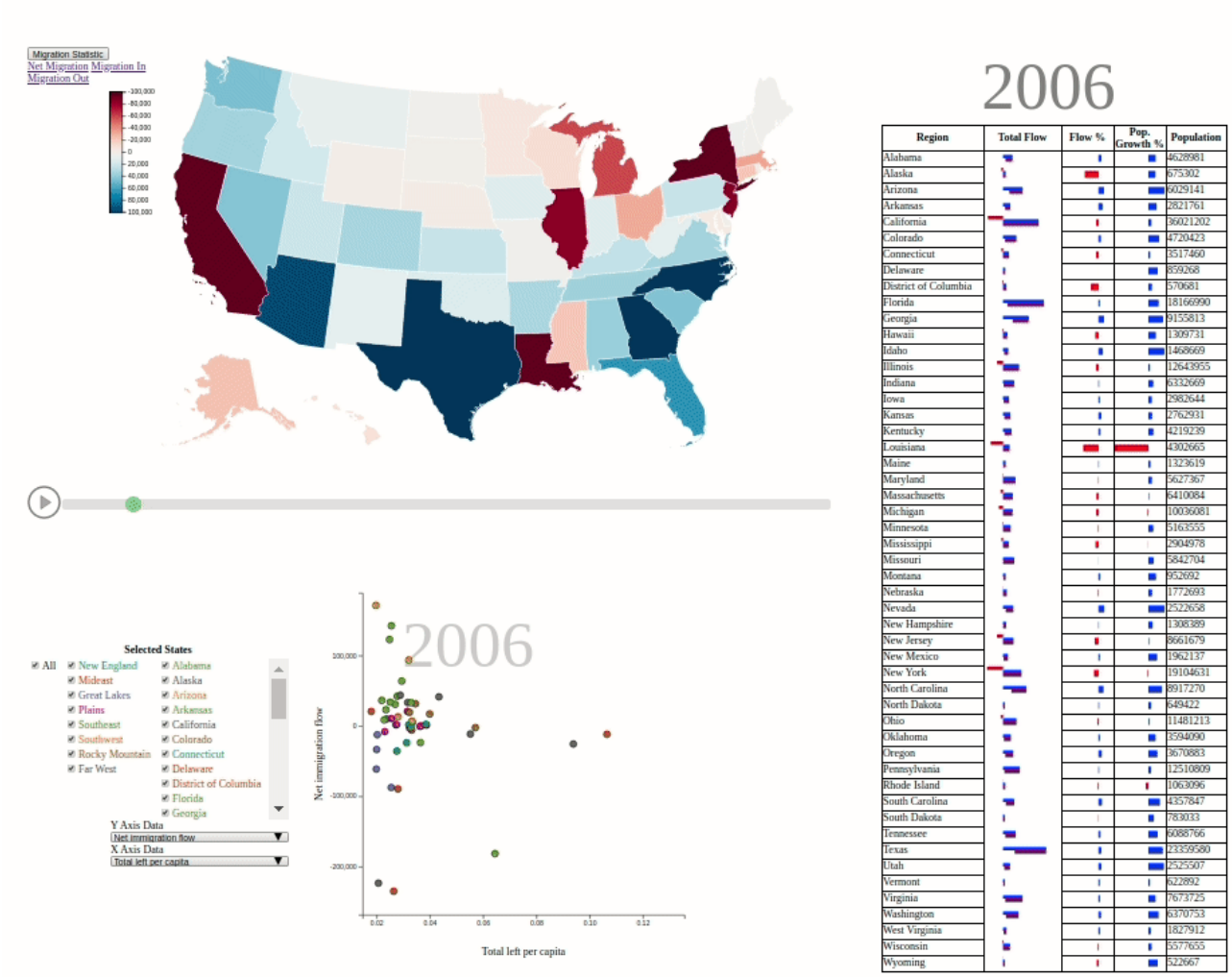
- For each state, the likelihood of moving to every other state for each year
- For each state, the likelihood of an immigrant being from every other state
- Overall likelihood of a foreign resident moving to a state (broken down by nations or

regions)

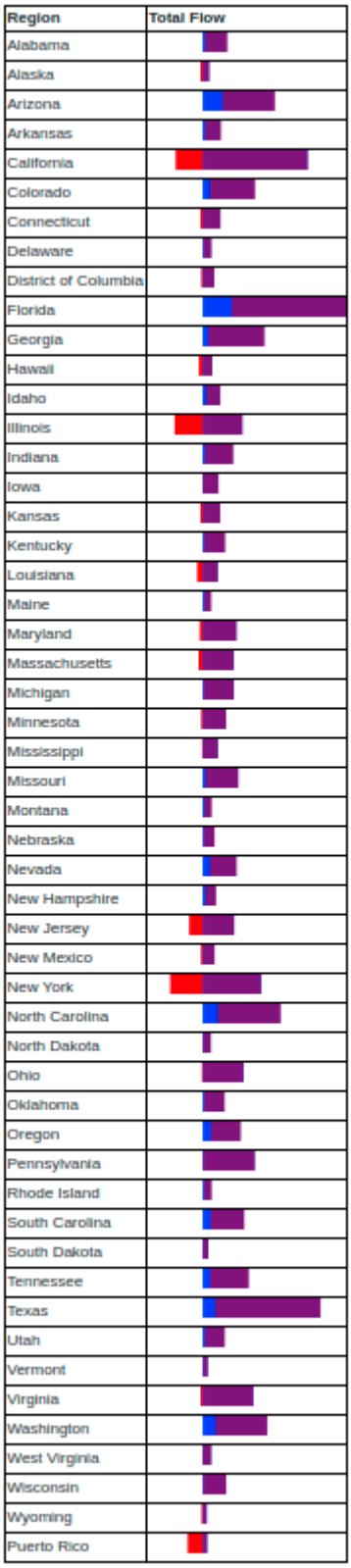
- Overall likelihood of a foreign resident moving out of each state
- R correlation for net immigration rate and different economic indicators (broad brush)

Exploratory Analysis & Design Evolution

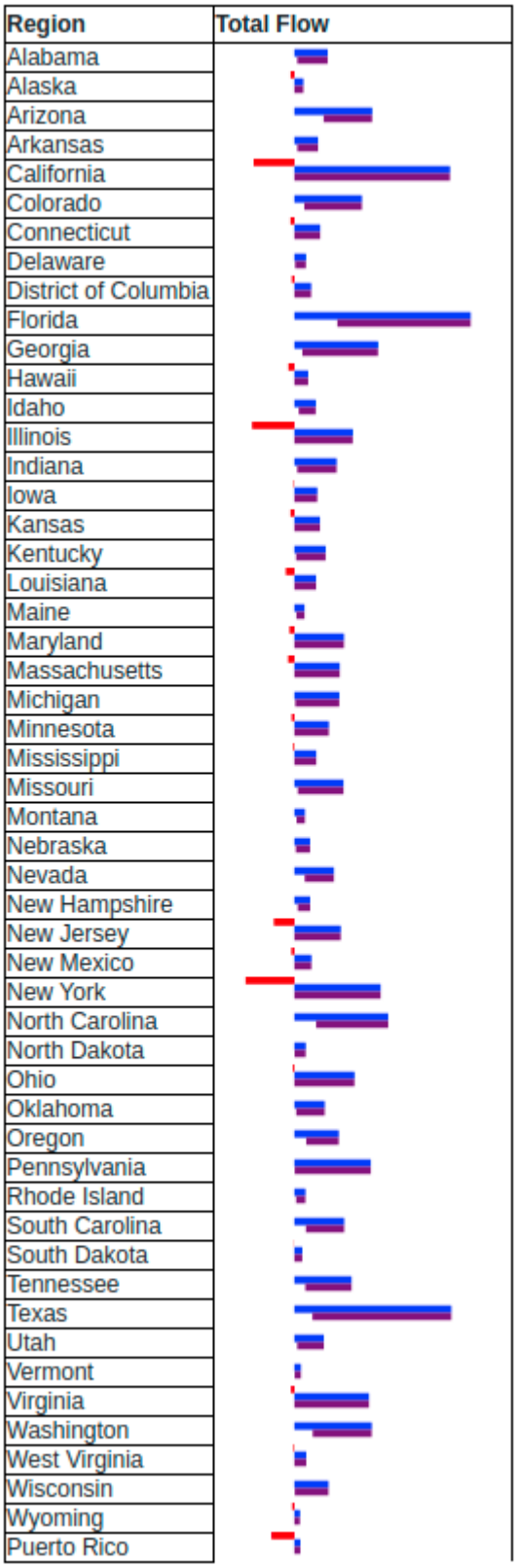
In the first phase of development, the data was extracted and transformed from the census bureau website. During this process, it was realized that much of the data is needed was readily available. There were two proposals for visualizing the migration patterns: the heat map and the chord diagram. After building the heat map, and allowing the user to select between different states and toggle outflow/inflow metrics, it quickly became evident that the chord diagram would not be able to encode the geospatial information, and would otherwise be redundant. Additionally, encoding puerto rico and DC would be expensive. As a result, the decision would be to leave Puerto Rico and DC out of the overall goal unless there was time at the end of the project to display them better.



Different techniques were attempted at trying to display the different migration statitics in a visual format alongside eachother. Since the total flow was related to total inflow and outflow, placing them in different columns seemed to draw them apart. They needed to be together.


Region
Total Flow
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Puerto Rico

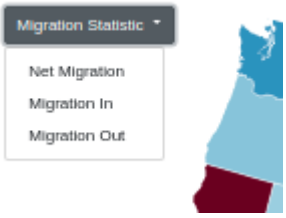
How to encode all three? Placing them on top of eachother would obscure whatever was layered behind. They were expiremently placed on top of eachother as rectangles, with the outflow starting from the end of the inflow metric, which combined would show the net. This seems to show the relationship well, and when the outflow bleeds into the negative, a new red rectangle would show up that goes into the negative. While initially difficult to intepret, the positive attribution from one metric to the next feels like a win.



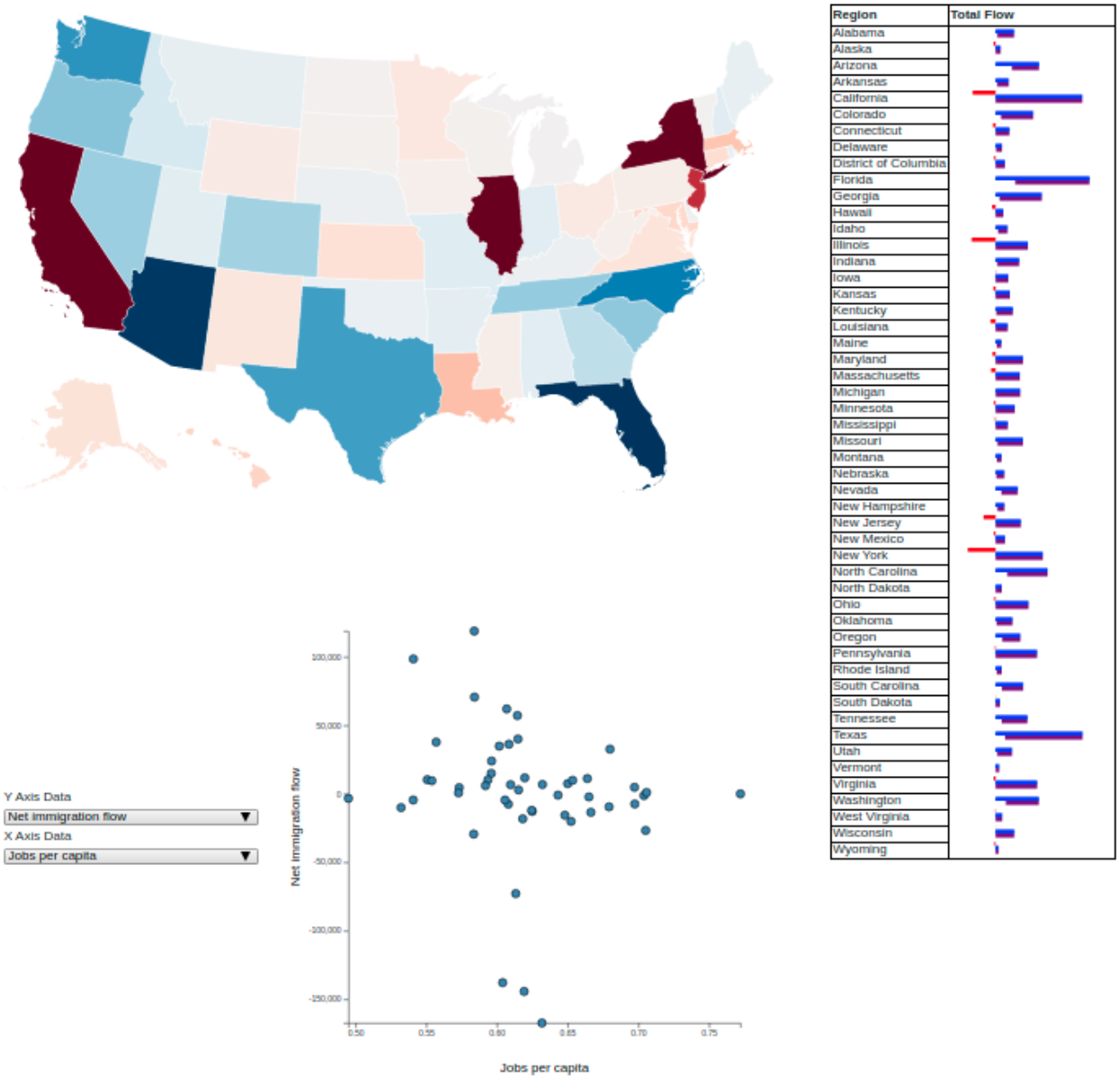
Having these metric next to eachother proved valuable, as you could see the total migration in and out metrics right next to eachother and view the relationship the with the net value to see the proportion of people moving in and out. This is instantly noticeable to the viewer once they recognize the data they are looking at. This was a neat way to show all three in parallel.

Region	Total Flow
Alabama	
Alaska	
Arizona	
Arkansas	
California	
Colorado	
Connecticut	
Delaware	
District of Columbia	
Florida	
Georgia	
Hawaii	
Idaho	
Illinois	
Indiana	
Iowa	
Kansas	
Kentucky	
Louisiana	
Maine	
Maryland	
Massachusetts	
Michigan	
Minnesota	
Mississippi	
Missouri	
Montana	
Nebraska	
Nevada	
New Hampshire	
New Jersey	
New Mexico	
New York	
North Carolina	
North Dakota	
Ohio	
Oklahoma	
Oregon	
Pennsylvania	
Rhode Island	
South Carolina	
South Dakota	
Tennessee	
Texas	
Utah	
Vermont	
Virginia	
Washington	
West Virginia	
Wisconsin	
Wyoming	
Puerto Rico	

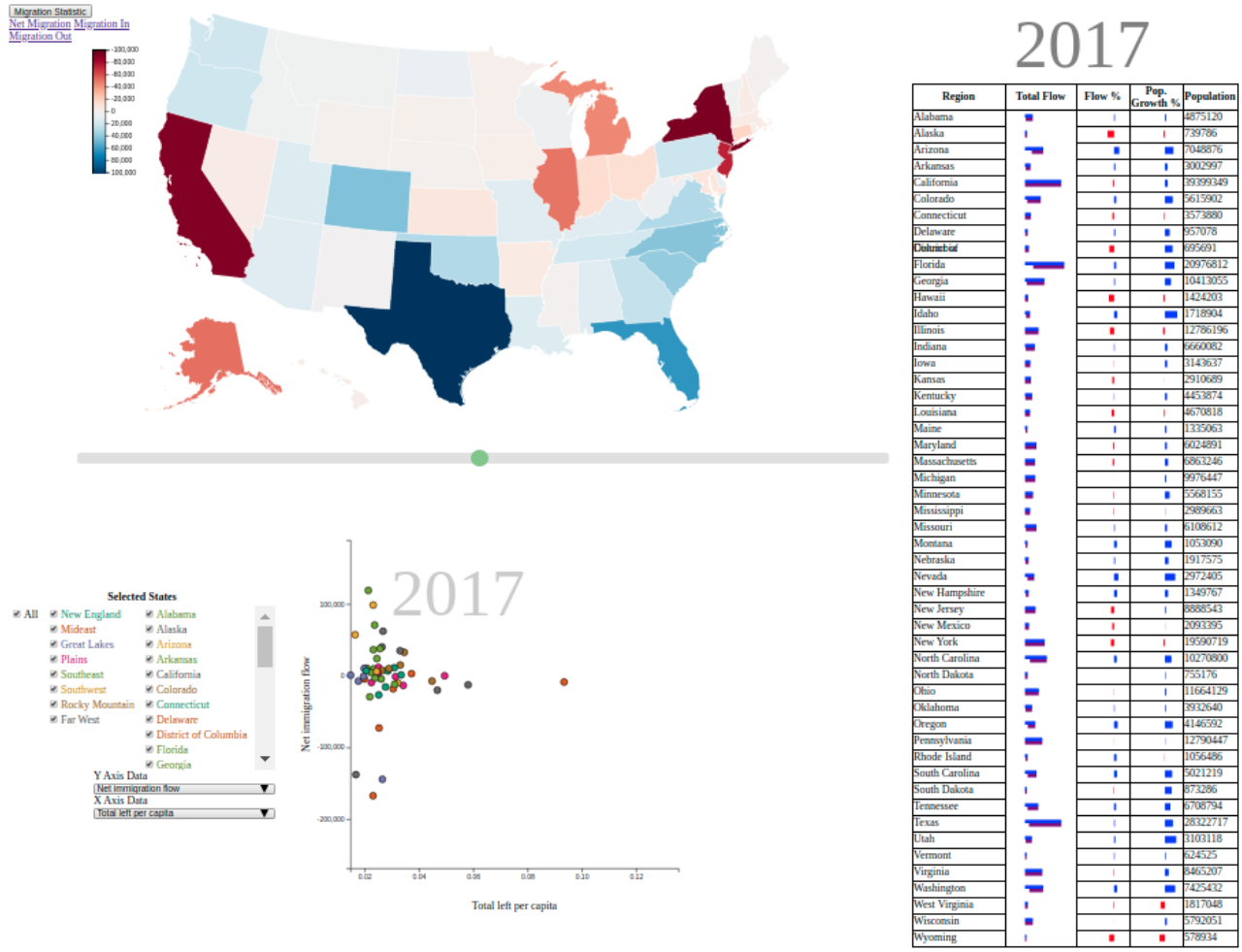
It seemed useful to be able to switch between the different metrics to visualize all that data at once, however any regional groupings would still be difficult to get clued in on. What we wanted to see if there were clear patterns of growth or migration that was visible on the heat map. We needed some way to toggle between the different migration statistics, and that is where the toggle at the left was born.



Adding the economic data was next, and choosing a way to encode it and interact with the heat map and table was an important step. Some of the interactions wouldn't come until later, since the views themselves needed finished first. Once the economic data was available and we could switch through the different metrics, it became clear some derived metrics would be just as important in this case as in the heat map, so those were added. With the table, heat map and scatterplot all in place, picking the important interactions was the last piece.



Beeing able to see where the selected states were in the scatterplot and heat map was important. Also, the heat map could potentially encode more data than was initially planned. After adding some derived metrics to the table that enables the users to see the metrics all next to eachother as they shift from year to year, allowing the user to select which data point to encode in the heat map was next. This allows the visualization of GDP data, flow data, and migration data alongside eachother with regional relationships intact.

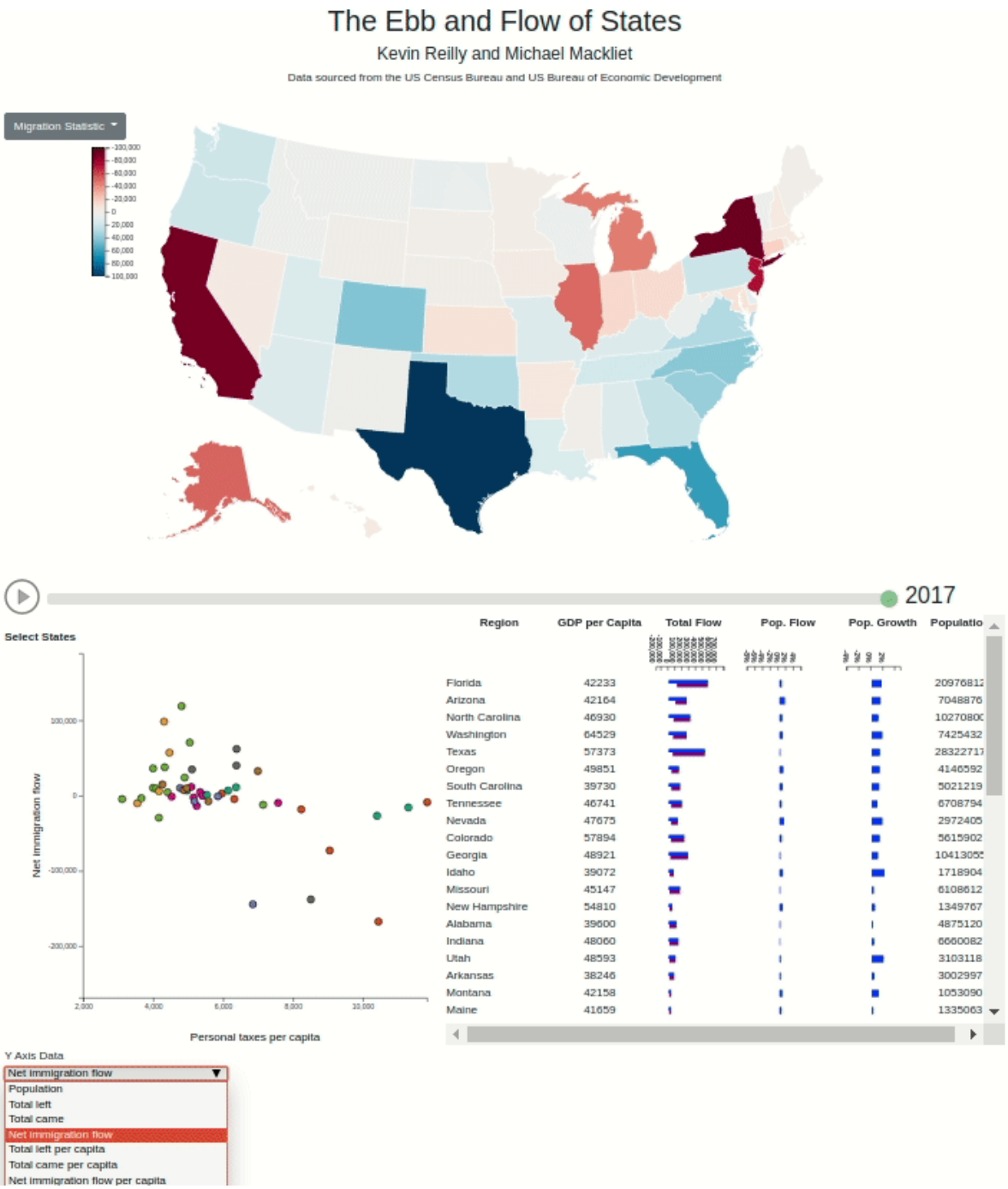


The play button was a helpful piece and allowed the temporal evolution of the visualization to provide key insights into some of the data. First, we were able to find a bug in the way the data was extracted from the dataset. At some point in the extraction, one of the column patterns didn't fit the rest and the population data was off by one column, showing a clear shift around year 2011. This was easily corrected and no other data issues have since been discovered.

After that was fixed, the user could see a clear shift in income and taxes in aggregate in 2009, but no clear shift in migration patterns were shown. So while the recession did have a clear impact on incomes, people didn't seem to move anymore or less than they usually did. Even the total migration metrics didn't show a large shift of any kind. This may indicate that factors other than primarily economic are at play, which is a keen insight. Whether it's mostly life events such as graduation or retirement is yet to be seen, but a large macro economic event didn't seem to be as big of a hit as say a smaller one like Hurricane Katrina was. This does show that the animation is effective, but this could also likely be seen with a simple scatterplot over time as well.

Final Implementation Example

The final pieces were the hardest and required the most work. The regional selection was placed in a drop down hover state to remove it when it wasn't needed. This allowed to user to focus on what was important when needed. What's missing here is the ability to filter the table by this criteria, which would be a huge gain, allowing a little more focus when needed. The table axis was added, as well as the ability to sort the table. The biggest piece that seems to bring the whole piece together is being able to select the metric the heat map will display. By clicking on the table header, you not only sort the table, but change the encoding on the map to match the metric selected. So not only can the user see the migration data metrics as they evolved over time in a geospatial context, but flow data as a percent of stat population is available, as well as population growth with respect to previous years. Finally, the table will update the order based on what was selected so the user can observe how often states will bounce around in the rankings depending on what factors are involved that year. Factors may include the shale oil boom in North Dakota, the migration of recently retired households, natural disasters, or the near constant migration out of Alaska to the continental US.



Evaluation

The current release is a fairly well working model that allows quick interaction between different data metrics and their temporal evolution across the U.S. There are some sticking points, still. Highlighting the state selected in the table, or allowing the filter from the scatterplot to filter the table down a bit, would help the user focus on what they want to. Encoding another map, of county level data, down and allowing the user to switch between these, would also create great anlytical potential. And the color fill bug, when selecting a metric with a larger domain, could easily be fixed. Additionally, adding metrics for birth rate, death rate, or international migration might help fill the gap between migration flows and

population growth.

As a product created in typescript for the browser by a two person team with little front-end dev experience, the total functionality is impressive. Some of the best insights that were gleaned were the huge migration out of Louisiana after Katrina, but the relative GDP increase during that timespan.
This shows that migration factors may often times be inversely correlated, as a natural disaster in a major population zone would see that population move to neighboring states for shelter. That GDP also increased shows that migration and GDP aren't necessarily correlated, as GDP can correlate with an increase in economic inefficiency, i.e. broken windows.

Suprisingly, Utah continues to see major population growth, in part due to migrations from neighboring California, Nevada, Washington, and Nevada. This, perhaps coupled with high birth rates, could potentially be part of the reason the state witnessed major GDP growth, when compared to the nation as a whole. The mountain states, and North Dakota, all share this same feature.

Should county level data be included, it would be interesting to see if mose of the migration from California / New York are from the major metropolitan areas. Whether the flows from these centers follow the same patterns as the states is yet to be seen, though geographical factors seem to be apparent. Most flows from individual states happen to neighboring states, with the exception of Florida, Texas, and Arizona. How age plays a role in this could give some insight on the nations retirement patterns.

###Authors

- Michael Mackliet m.mackliet@gmail.com
- Kevin Reilly reilly.w.kevin@gmail.com

+ Add a custom footer

Pages 3

Find a Page…

Home

Process Book

Proposal

+ Add a custom sidebar

Clone this wiki locally

```
https://github.com/mackliet/gun_data_visualization.wiki.git
```