STAT436 Project: U.S. Migration Patterns Report

Group 2: Aarya Deshpande, Corey Johnsen, Jeffrey Wang, Nico Butera, Samuel Pekofsky, and Zeke Jeske

Group 2

2025-05-05

Introduction

Data source

Our project visualizes the U.S. Census Bureau's migration flow data, which contains estimates of county-to-county and state-to-state migration flows using data from the American Community Survey (ACS). The ACS contains questions that ask where people currently live they lived exactly one year ago. In order to have a large enough sample for geographies with small population, the data is then combined over a five-year period of the survey (ACS responses are often aggregated over 3- or 5-year windows). In this report, we focus on the most recent data available, which is from 2016 to 2020. The documentation for these data explains how the data was collected, issues with the data, and what the fields mean.

Research questions

The motivation for visualizing US intranational migration is simple: migration plays a crucial role in understanding social stratification and economic mobility, but all available data are so vast that they are hard to comprehend. Migration data is, therefore, a space that is ripe for a few visually inclined data scientists like ourselves to transform this complexity into clear, intuitive visuals that make the patterns and stories of migration accessible to everyone, even those without the time or expertise to navigate the raw data themselves. Interestingly, while a number of visualizations already exist for international migration, little work has been done using the ACS data on domestic migration.

From this data, we hope to answer the following questions that will serve as the main foundation to our critical Report:

- What are the major migration trends in the US? In particular, we hope to identify whether there are certain states or regions that are experiencing high levels of in or out-migration and possible paths of these migrants.

- Are there local migration trends within certain regions? For example, do we see neighboring states share a high number of migrants or do counties within states have a general pattern?

Data fetching and processing

In the `prep.R` script, we fetch the migration data from the Census Bureau's FTP server (www2.census.gov). The original format is a fixed-width text file. To read it, we must specify the exact column positions of each field we are interested in. We filter the data to include US states and the District of Columbia, removing other territories like Puerto Rico (for which data is incomplete), and save the flows to `ctyxcty.csv` and `statexstate.csv`.

Then, we use the `tigris` package to fetch shapefiles for U.S. counties and states from the Census Bureau. We join the resulting `sf`s with aggregated migration data, resulting in two geospatial data frames: `states`

and `counties`, each of which contains the in-migration, out-migration, and net migration for each state and county, respectively. States has 51 rows (50 states + DC) and counties has 3143 rows. These two data frames are saved as GeoJSON files. We use them for most of our visualizations, but for the more complex visualizations we work with the full network data (`ctyxcty` and `statexstate`).

Here is a peak at `states` and `counties`:

Table 1: Sample of 10 rows from counties

state_code	cty_code	name	name_full	in_flow	out_flow	net_flow	lower_48	geometry
13	125	Glascock	Glascock County	145	177	-32	TRUE	MULTIPOLYGON (((-82.75576 3...
35	009	Curry	Curry County	4206	4050	156	TRUE	MULTIPOLYGON (((-103.7399 3...
09	005	Litchfield	Litchfield County	7052	8731	-1679	TRUE	MULTIPOLYGON (((-73.51795 4...
21	239	Woodford	Woodford County	1602	1703	-101	TRUE	MULTIPOLYGON (((-84.866 38....
21	041	Carroll	Carroll County	361	677	-316	TRUE	MULTIPOLYGON (((-85.33309 3...
28	077	Lawrence	Lawrence County	354	400	-46	TRUE	MULTIPOLYGON (((-90.24519 3...
48	477	Washington	Washington County	2420	2076	344	TRUE	MULTIPOLYGON (((-96.79455 3...
27	007	Beltrami	Beltrami County	4353	2804	1549	TRUE	MULTIPOLYGON (((-95.60286 4...
21	081	Grant	Grant County	1524	1433	91	TRUE	MULTIPOLYGON (((-84.78528 3...
55	043	Grant	Grant County	3644	2823	821	TRUE	MULTIPOLYGON (((-91.15681 4...

Table 2: Sample of 10 rows from states

code	name	in_flow	out_flow	net_flow	lower_48	geometry
27	Minnesota	106560	120310	-13750	TRUE	MULTIPOLYGON (((-89.59206 4...
35	New Mexico	60407	67128	-6721	TRUE	MULTIPOLYGON (((-109.0502 3...
20	Kansas	88462	93711	-5249	TRUE	MULTIPOLYGON (((-102.0517 4...
36	New York	258667	461202	-202535	TRUE	MULTIPOLYGON (((-79.76215 4...
25	Massachusetts	146920	178057	-31137	TRUE	MULTIPOLYGON (((-70.23405 4...
46	South Dakota	28159	29021	-862	TRUE	MULTIPOLYGON (((-104.0579 4...
41	Oregon	139288	114539	24749	TRUE	MULTIPOLYGON (((-123.6647 4...
44	Rhode Island	34239	30410	3829	TRUE	MULTIPOLYGON (((-71.28802 4...
31	Nebraska	47681	52441	-4760	TRUE	MULTIPOLYGON (((-104.0534 4...

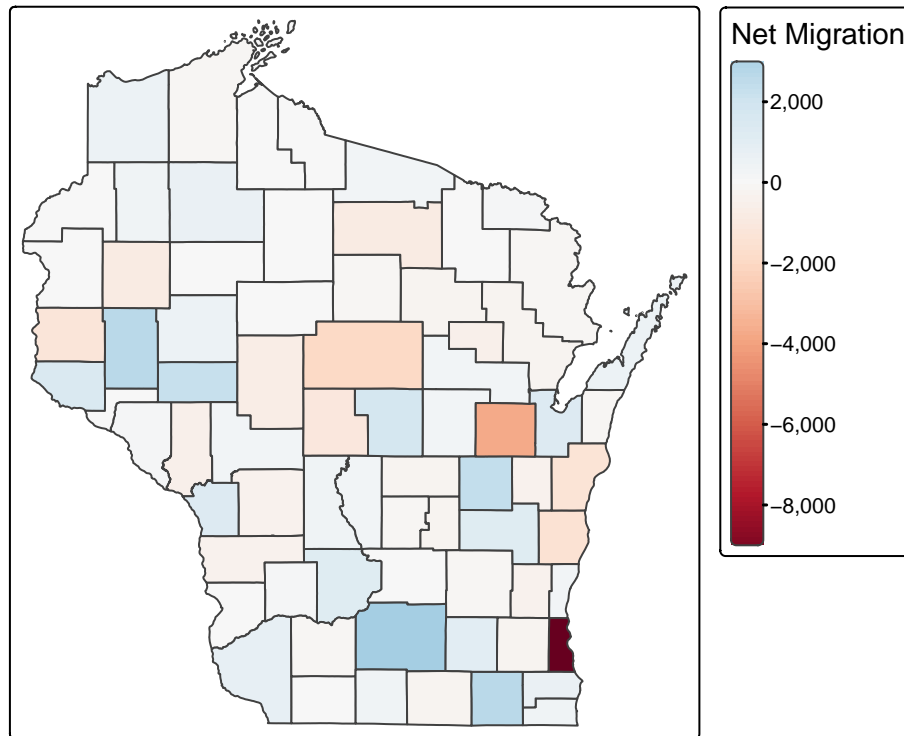| code	name	in_flow	out_flow	net_flow	lower_48	geometry
13	Georgia	278030	247999	30031	TRUE	MULTIPOLYGON (((-81.279393...

Literature Review

Although domestic migration plays a key role in shaping economic mobility and regional development, most public tools built to visualize these patterns remain limited in both scope and usability. The USDA/UW-Madison Net Migration Patterns map offers a comprehensive county-level view of net migration but lacks directionality and user-driven filtering, making it difficult to analyze migration flows in context. Commercial dashboards like Allied's U.S. Migration Report offer some directional insights and interactivity but rely on private data and proprietary pipelines, limiting transparency and reproducibility. In contrast, the ACS provides publicly available, granular migration estimates, but the raw format requires significant preprocessing to extract meaningful patterns. Our work aims to fill that gap by making this complex dataset interpretable through transparent, customizable visual design.

From the visualization literature, we focused on aligning form with task, selecting techniques that match common analytical goals like comparison, ranking, or spatial lookup. While static choropleths are a popular choice for migration data, they often fall short when it comes to representing direction or volume. To address these limitations, we introduced bar charts for ranking, directional network graphs to capture flows, and interactive filtering to support localized exploration. These decisions were guided by principles of mixed-method visualization, which encourage combining multiple views to accommodate different types of questions and reduce interpretive ambiguity.
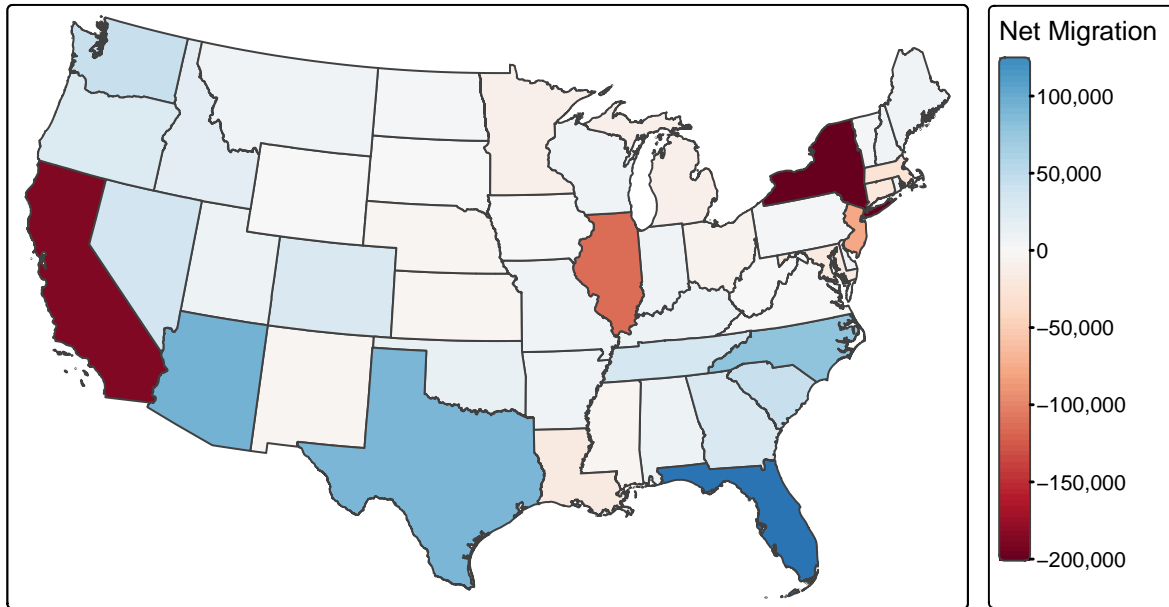
Visualizations

Static plots

Net Migration in Wisconsin Counties

This county-level net migration uses a color gradient, from blue representing net positive inflow of migration to red with net negative outflow while hue intensity represents the inflow/outflow magnitude. Dane County, for example, has a positive 6,000-8,000 net inflow margin (the strongest of all counties). This visualization is inspired by the USDA/UW-Madison migration tool from Milestone 1. This format allows the viewer to easily differentiate between high-inflow and high-outflow counties. Additionally, this approach made it easy to spot clear regional trends—urban counties tend to gain population, while many rural or post-industrial areas see losses. While the map does not show directional flows, it complements our later visualizations by grounding the viewer in the geographic scale of the data.

In the plot below, keep in mind that we're only measuring *domestic* migration; immigration (moves from outside the U.S.) and emigration (moves to outside the U.S.) are not included in these visualizations.
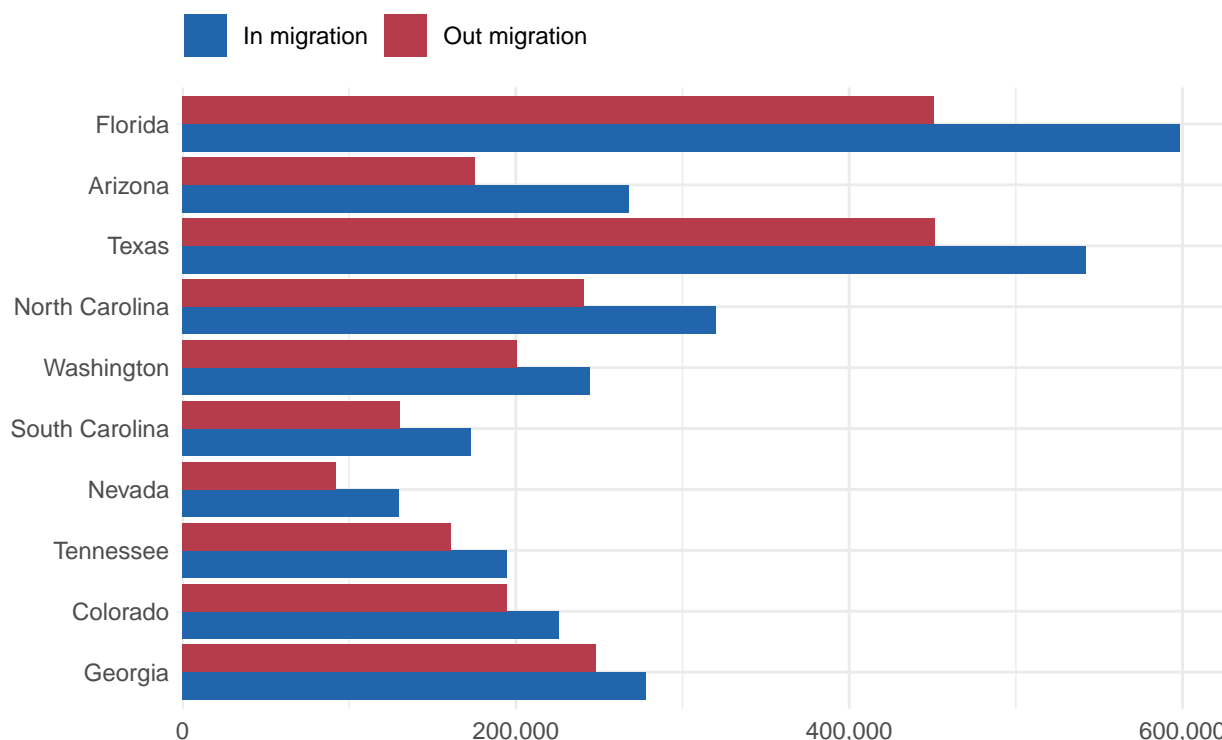
Net Migration by State

Similar to figure 1, this more expansive color-coded state-to-state thematic migration map aligns with the Allied's U.S. Migration Report from Milestone 1, on making accessible state level patterns. Yet again, the color gradient reveals which states experienced significant population gains (blue) or losses (red), effective for audiences like policymakers and researchers for interstate movement. However, it summarizes the in-state version and thus has the tradeoff of lacking county-level movement.

This visualization clearly illustrates how certain states, such as Florida, Texas, and Arizona, consistently draw net positive migration, while others like California, Illinois, and New York see sustained outflows. These results reinforce patterns observed in our county-level map, validating them at a broader geographic resolution. However, this state-level view does lose the granularity of within-state variation, which is something our interactive tools help to recover later in the report.
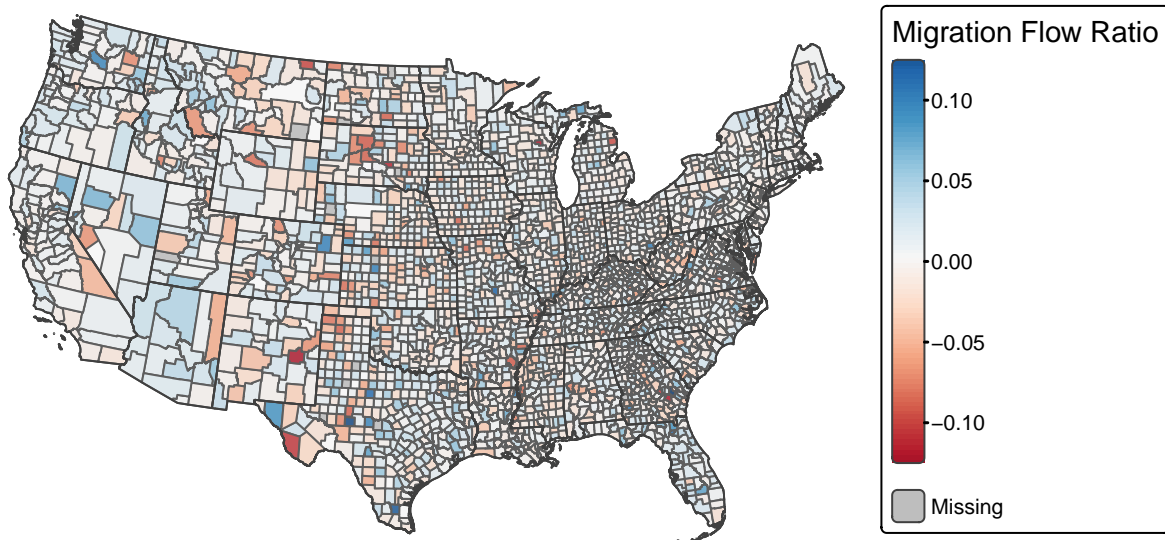
In–migration vs Out–migration by State
Showing only the 10 states with the highest net migration.

■ In migration   ■ Out migration

This visualization ranks the top ten states with the highest net migration to offer a quick comparative and quantitative perspective on migration trends. Unlike previous maps, this bar chart emphasizes magnitude rather than geography to make it easier to quickly identify which states gained the most residents. Here, it is clear that the top three "migration hot" states are Florida, Texas, and Arizona. This ranking is appropriate in real estate and demographic studies to assess state-level attractiveness. However, despite the clear ranking, it lacks spatial context, meaning users cannot see where people are moving from or to. While this plot sacrifices directional or geographic nuance, it complements our other views by offering a focused, high-level snapshot of migration winners.

Population Adjusted Migration Flow Ratio by County

Our final static visualization takes a more normalized approach by mapping net migration as a proportion of county population. Instead of highlighting absolute counts, this ratio, computed as net flow divided by 2021 population, helps reveal where migration is disproportionately high or low relative to the size of the county. However, we filtered out counties with fewer than 1,000 residents or fewer than 100 total movers to avoid distortions from extremely small samples. This view brings out interesting contrasts not immediately visible in raw numbers; for example, small counties with moderate movement can show extreme ratios, suggesting potential demographic volatility or unique local factors. Positive values indicate more people moved in than out, while negative values reflect net outflows. Though the ratio itself isn't always directly interpretable in real-world terms, it's a useful comparative measure across counties. This plot adds a valuable dimension to our report by identifying counties where migration trends are especially outsized relative to population, an important perspective for demographers and local planners alike.

Interactive visualization

After seeing static visualizations with a focus on general US migration trends or specially designed plots for local regions, we present a interactive visualization that showcases a network-based plot overlaid on a US map for the source and destination of migration flows. The network features nodes at every county and the top 3000 edges for migration between two counties with thicker edges indicating more movers. You can select whether to see in, out, or net migration and an individual state to look at (through both sidebar and clicking on the map), which updates the colors of all states and filters the edges to reflect the chosen migration statistic, inspired by the Allied Migration Report.

Hosted Shiny App

This not only enables a board level view of hot migration counties, general paths, and state level information but also a detailed view of each state to see smaller details that are missed in larger trends. The network visuals provide extra context to see the paths people are taking rather than just isolated region level views.
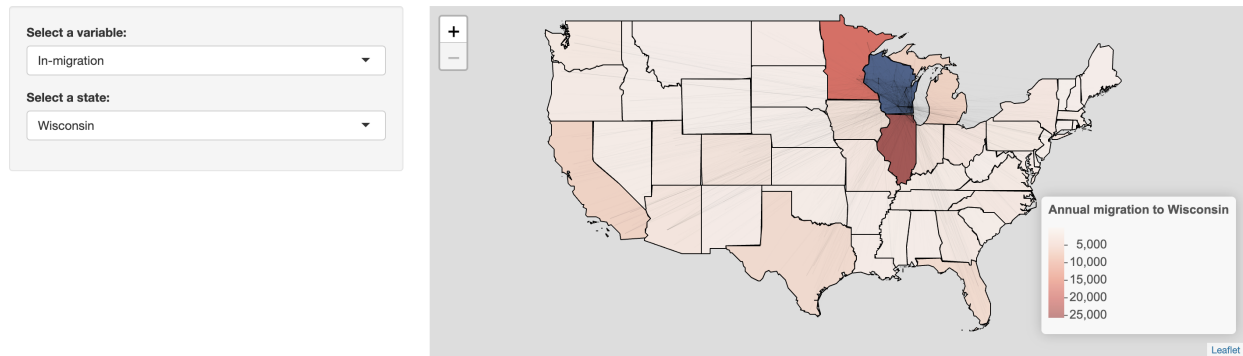
Figure 1: Top In-Migration Sources to Wisconsin

Beyond the network details, the interactivity provides different levels of details and focuses that static plots aren't able to achieve without an overwhelming number of plots. In regards to our main questions, I can see Florida, Texas, New York, and California having the most migration with the network layer showing that large counties are often a major zone for migration (both in and out), suggesting that these states have migration due to their quantity of such counties. In local trends, when filtering my view to a single state, the migration of nearby states are often relatively higher compared to other states (Wisconsin for example has Illinois and Minnesota being been a major migration partner) while sometimes the big three migration states are still major players (ie. Colorado). In the end, this interactive plot complements the previous visuals to complete our answer of migration trends across the US.

Conclusion

Each of our visualizations brought out a different piece of the migration puzzle. From high-level plots to granular network maps, we've tried to make sense of where people are moving, where they're leaving, and how those flows differ across regions. Our original questions focused on both broad national trends and more localized movements and we now feel like we have a solid answer to both. States like Florida, Texas, and Arizona continue to attract new residents, while others like California and Illinois show net losses, a finding consistent across multiple views. On a more local scale, the interactive plot helped reveal regional relationships, like Wisconsin's strong ties with Illinois and Minnesota, which are harder to spot in static plots.

That said, there are still directions we could take this further. County-level flows are inherently noisy, and low-population areas make some plots tricky to interpret without filtering. It would also be interesting to layer in more demographic info, like income or age, to add more context to these patterns. But overall, we think our visualizations succeed in turning something messy and massive into something meaningful and explorable. Hopefully, they provide not just answers, but new questions for others to pursue.