ROBIN: An Interactive Visualization System and Instructional Tool to Democratize United States Domestic Migration Data

Alexander Bendeck
Georgia Institute of Technology
abendeck3@gatech.edu

Clio Andris
Georgia Institute of Technology
clio@gatech.edu

John Stasko
Georgia Institute of Technology
john.stasko@cc.gatech.edu

Abstract

Migration scholars in the United States (U.S.) study how migration patterns within the country relate to characteristics of migrant origins and destinations, such as political leanings or educational attainment. However, few tools exist for experts to visualize these relationships and easily share insights with students or other interested members of the general public. This data thus remains largely inaccessible to potentially interested non-experts. In this work, we present a system called ROBIN *for visualizing U.S. county-to-county migration data in conjunction with other county-level attributes, designed to be used by experts as a communication aid or by non-experts as an exploratory tool. User studies with migration domain experts and non-experts show promising results in our efforts to support and engage both user groups.*

Keywords: Geographic Visualization, Flow Mapping, Design Studies, Migration.

1. Introduction

Domestic (or internal) migration is the process of individuals and households moving from one address to another address within a country. The relationship between migration and the characteristics of migrant origins and destinations is an area of historical and continued interest for researchers in the domain (Grigg, 1977; Ravenstein, 1889). For instance, a recent analysis (Liu et al., 2019) found that people tend to move between counties in the United States (U.S.) whose populations have similar political preferences. At the same time, the topics of migration, elections, and demographics are also of interest to more general audiences who may seek to know whether and how their communities are changing (Heer et al., 2007). This interest is growing given the impact that migration can have on lives and communities. For example, internal migration within the U.S. has resulted in new voting outcomes (Robinson and Noriega, 2010), impacted local and regional economies (Harrison, 2017), and caused cities to grow rapidly over time (Frey, 2010). As a result, U.S. migration experts such as professors and Census Bureau employees are often asked to share their knowledge with journalists, students, or the public.

Visualization can be a powerful technique for quickly and effectively conveying data insights (Fekete et al., 2008), even to consumers who lack extensive knowledge about the data domain (Böttinger et al., 2020). We argue that this is especially the case for migration patterns and related demographic data, which many citizens likely already have some intuition about. While a handful of migration visualization systems support analyzing demographic properties of origin and destination populations (Boyandin et al., 2011; Guo et al., 2006; Scheidl et al., 2021), these systems are not designed to be interpretable by non-experts. Designing such a system would support the education and engagement of the broader public with respect to data about their communities, thus employing visualization as a vehicle for social good.

In this paper, we design an interactive visualization system, ROBIN[1], for exploring U.S. migration data. Our system consolidates county-level data on attributes including election outcomes, household income, and education levels with county-to-county migration data to enable holistic exploratory analysis. ROBIN is meant to be usable by domain experts as a communication aid and by non-expert users to explore the data themselves.

[1] The system is named after the American robin (*T. migratorius*), a migratory bird found throughout the United States.

Our aim is to support engagement of students and the broader public with respect to publicly-available data about important societal phenomena.

1.1. Research Questions

Through our development and evaluation of the ROBIN system, we contribute answers to the following research questions (RQs):

RQ1: Can we support both domain expert and non-expert analysts of U.S. migration data in a single interface? Prior work has not sought to support both of these user groups. We aim to provide a system which domain experts can use as a communication and teaching aid, and which non-experts can use as an exploratory tool. This will also help facilitate knowledge transfer from experts to non-experts.

RQ2: How can we intuitively show relationships between U.S. migration and other attributes? Some expert-focused tools can display flows and other attributes of geographic data concurrently, but rely on multiple linked views or arrow-based designs subject to visual clutter. We seek to design a novel approach that is intuitive for both experts and non-experts.

RQ3: How can we flexibly and intuitively support querying of geographic data, either by individual geographic unit or multiple units at once? In existing flow visualization tools, users are often restricted to analyzing flows entering or exiting one geographic unit (e.g., county) at a time (Egan-Robertson et al., 2023; United States Census Bureau, 2021) or all at once (Koylu et al., 2022). We aim to enable both big-picture and more local takeaways by additionally showcasing movement between groups of counties that share common attributes (e.g., voting patterns).

2. Background & Related Work

2.1. Migration Studies

The movement of people within a country is called *domestic* migration. Factors such as level of urbanization (urban, rural, suburban, etc.) of origins and destinations have been found to impact U.S. domestic migration patterns. Migrants from rural areas migrate to more urbanized areas (Golding and Winkler, 2020), yielding high population decline. Education attainment level, age, and race are also frequently cited as factors that induce migration; non-white migrants tend to move more than their white counterparts, and those with a bachelor's degree or higher tend to move more as well (Ambinakudige and Parisi, 2017). Regarding age, young adults under 35 are the most mobile, while retirees also tend to migrate (Plane, 1992). Migration trends are additionally driven by income and unemployment rates (Rayer and Brown, 2001) as well as political affiliation (Liu et al., 2019). Geographic regions are also important to study in this context. For instance, the U.S. South has attracted many migrants in recent decades, due to warmer climate and more affordable housing (Ambinakudige and Parisi, 2017).

2.2. Migration Vis Techniques & Systems

Prior work in origin-destination flow mapping has established design principles, frameworks, and layouts for both static maps (Jenny et al., 2017, 2018) and interactive systems (Guo, 2009; Koylu et al., 2022; Schöttler et al., 2021; Wood et al., 2017; Yang et al., 2016) which have been used for migration visualization. In this space, node-link diagrams are the most common representation used to visualize flows, with entities or locales shown as nodes and flows shown as links. Line width is very often used to represent flow volume, and arrowheads are commonly employed at the end of flow lines to indicate flow direction (Jenny et al., 2018). Some systems rely strongly on interactive features in their designs, such as to create dynamic force-directed node-link diagrams (Jenny et al., 2017). Migration networks in particular are sometimes visualized as node-edge sociograms, i.e., where city nodes are connected by migration flow edges that have no spatial reference (Liu et al., 2018). Origin-destination matrices are a key technique for visualizing flow without using lines (Wood et al., 2017) and have been employed both as a standalone view or in an interactive pairing with a geospatial map view (Yang et al., 2016). The above techniques focus on designing visual representations of flows to address challenges such as visual clutter, rather than producing tools to help analyze any specific dataset.

A few interactive tools have been built specifically to visualize U.S. migration. The *U.S. Migration Flowmapper* by Stephen and Jenny (2017) displays migration flows by state and county, superimposing flow arrows on top of a choropleth map where color encodes population density. Somewhat uniquely, the U.S. Census' *Census Flows Mapper* (United States Census Bureau, 2021) and the UW-Madison *Net Migration by Decade* mapper (Egan-Robertson et al., 2023) allow users to visualize in and out migration by county without explicit visual links, instead using only a choropleth map where color encodes migration magnitude. These tools were designed for anyone to be able to use, and we based our main design loosely on them for this reason. However, these tools also lack the functionality to flexibly integrate demographic data and flow data for concurrent exploration.

Some more powerful tools allow visualization of flow data over time or with other attributes (Boyandin et al., 2011; Guo et al., 2006; Scheidl et al., 2021). Flowstrates (Boyandin et al., 2011) uses a side-by-side map layout for exploration and comparison of flows over time. The VIS-STAMP approach (Guo et al., 2006) helps analysts investigate multivariate, spatial, and temporal flows using multiple representations including reorderable matrices and geographic small multiples. VisMiFlow (Scheidl et al., 2021) employs visualizations such as line and bar charts interactively linked with a geospatial map view to support analysis of multivariate flow networks. Besides the fact that none of these approaches utilize U.S. migration data, a main gap is that they are designed exclusively for expert use. While useful for expert analysis, these systems consist of complex views and interactions and thus lack affordances for teaching or engaging non-experts.

3. The ROBIN System

3.1. Development Methodology

The visualization design study methodology (Sedlmair et al., 2012) is a theory which helps visualization researchers build tools for domain experts in a specific field. It consists of an iterative, 9-step framework for user-centered visualization system design. Our approach followed a particular variation of the visualization design study methodology known as the "data-first" approach (Oppermann and Munzner, 2020). The data-first approach starts with a dataset, rather than starting with pre-established domain expert collaborators who provide their own dataset. While we followed all stages of the data-first methodology, in this paper we recount a subset of these stages (with stage names from the paper in italics): We first *acquired* the publicly-available datasets from multiple sources (Section 3.2). We then *elicited* preliminary design requirements, leveraging one authors' migration domain expertise. After identifying other domain expert collaborators as well as non-expert system users (the *winnow* and *cast* phases), we *designed* and *implemented* a first version of the system. We iteratively revised the system with feedback from our collaborators – cycling through the *elicit*, *design*, and *implement* phases again – to produce the final design requirements (Section 3.3) and ultimately the final system (Section 3.4), before evaluating the system through a preliminary *deployment* in a lab setting (Section 4). (We envision conducting a more naturalistic, real-world deployment as future work.) Finally, we *reflected* on the implications, takeaways, and limitations of our work (Section 5).

3.2. Data

We use **county-to-county migration data** in the U.S. from 2010 to 2019, available online from the Internal Revenue Service or IRS. This data has been used previously to study U.S. migration flows and the demographics of origin and destination locales (Liu et al., 2019). In addition to the raw county-to-county flows, we compute *net* migration between every pair of counties. For instance, if County A sends 100 migrants to County B and County B sends 80 migrants to County A, then the net migration is 20 migrants from County A to County B. We also compute the *migration efficiency* metric, which intuitively describes the "one-sidedness" of migration flows. We use the definition of migration efficiency from prior literature (Galle and Williams, 1972), where it is defined as the net migration between two places divided by the total, raw migration between the places, multiplied by 100.

Finally, we integrate publicly-available data on county-level metrics which are known to influence where Americans migrate within the country (Liu et al., 2019). We will hereafter refer to these metrics as county *attributes*: **2020 and 2016 presidential election results** from Massachusetts Institute of Technology; **median household income** estimates from the 2019 American Community Survey or ACS; **educational attainment** (percent of adults with a bachelor's degree or higher) estimates, also from the ACS; and **urban-rural classification** (i.e., how urban vs. rural each county is) from the National Center for Health Statistics.

The supplemental material for this article includes more details and citations about all data described in this subsection, as well as key data preprocessing steps.

3.3. Design Requirements

To leverage our dataset of migration flows and county-level attributes, we identified two analytic question types to support. First are "one-to-many" (or "many-to-one") migration questions, where the analyst is interested in the migration flows entering or exiting one particular county. An example would be "Which counties receive the most migrants from Fulton County, Georgia, and how did these destination counties vote in 2020?" On the other hand, "many-to-many" migration questions can be asked about migration flows entering or exiting a set of counties which share a common attribute. Analysts can thus investigate migration patterns for a group of counties that is defined by, for instance, voting in the 2020 presidential election. An example question like this would be "Where do migrants from counties that voted over 60% Democratic in 2020 move to?"

Through a review of relevant literature (see Section 2) and brainstorming sessions among the authors (one of whom has migration expertise), we considered our dataset, intended users, and analytic question types above to formulate our design requirements: Visually show relationships between U.S. migration and other attributes (**DR1**); support both "one-to-many" and "many-to-many" migration queries (**DR2**); and prioritize flexibility, learnability, and ease of use (**DR3**).

3.4. System Interface

In support of **DR2**, we conceptualized two different "modes" for ROBIN to enable users to ask the two types of analytic queries outlined above. **Individual county mode** is for when the analyst is interested in "one-to-many" (or "many-to-one") analysis, while **attribute mode** enables "many-to-many" analysis. An accompanying supplemental video illustrates each mode along with all interface components described below.

To help users answer both question types, we designed our system's user interface (UI) with *dynamic query* components. Dynamic querying (Shneiderman, 1994) allows users to manipulate input widgets such as numerical entry boxes, sliders, and dropdown menus to set query parameters, and then the resulting data matching those constraints is reflected in the interface. This poses distinct benefits for our design requirements. We posited that users could manipulate widgets to precise values if they had specific questions, but would also be able to easily explore how changing the query parameters affects the results. This flexibility supports diverse styles of inquiry (**DR3**). Dynamic query controls are also relatively common across the Web (Manko, 2022), making them already familiar to untrained users.

With the above considerations in mind, we designed the system's UI (Figure 1) with four main components.

3.4.1. Sidebar The **sidebar** allows the user to manipulate dynamic query widgets and change visual encodings on the map (**DR1**). The **color dropdown menu** (Figure 1-C) sets the color encoding for the migrant origin or destination counties on the map. Color options include the county attributes from Section 3.2, plus volume or efficiency of migration. The **county view option** (Figure 1-D) gives two options for viewing migrant source or destination counties on the map. The county glyphs can either by displayed as physical county shapes or shown as "flow circles" whose size represents the magnitude of the migration from or to those counties. With the "flow circles" option, each county glyph on the map to visually encodes both migration magnitude and an attribute value simultaneously, using size and color respectively (**DR1**). The **show arrows** setting (Figure 1-E) shows or hides all individual arrows on the map that link a selected county to its source or destination counties (in individual county mode).

The dynamic query settings are also integrated with descriptive text to serve as a semantic summary for the current state of the system (**DR3**). The specific widgets housed here are slightly different depending on whether the user is in individual county mode or attribute mode. The two persistent widgets regardless of mode are the **migration flow option** (Figure 1-A), which specifies the direction and type of migration (e.g., migration into or net migration out of a county of interest) and **number of top counties** input (Figure 1-B), indicating the number of migrant origin or destination counties to show.

Users can enter **attribute mode** using the **mode selector** radio buttons (Figure 2-A). In attribute mode, the user chooses an **attribute of interest** (Figure 2-B) for selecting counties (**DR2**). The **attribute value selector** (Figure 2-C) changes based on the attribute of interest, but includes numeric inputs to specify cutoffs for Presidential voting, median income, and percent of residents with a college degree, or a dropdown for urban classification. The **sidebar map** (Figure 2-D) highlights the selected counties that match the allowed attribute values (e.g., "suburban") for the chosen attribute (e.g., urban classification) in pink (**DR1**). The sidebar map always shows a single, large arrow to indicate the direction of the migration (as shown in Figure 2-D).

3.4.2. Main Map The **main map** occupies most of the UI and displays the counties returned by the current query parameters; i.e., it shows counties which are either large migrant sources or destinations with respect to the selected county (in individual county mode) or counties (in attribute mode). Only the counties with the largest or most efficient migration flows, up to the number specified in the sidebar, have glyphs shown on the map.

3.4.3. Linked Scatterplot The **linked scatterplot** displays one point per county on the main map. Position along the y-axis of the scatterplot encodes the number of migrants moving to or from each county (based on the query), and position along the x-axis is based on the value of the county attribute currently chosen in the color dropdown. The point color also matches the color of the corresponding counties on the map. Hovering over a county on the main map or on the scatterplot causes a tooltip to pop up on the map above the county, giving details-on-demand for the migration flows entering or exiting that individual county (**DR2**). Hovering over a county on the map highlights the corresponding point in the scatterplot in pink.

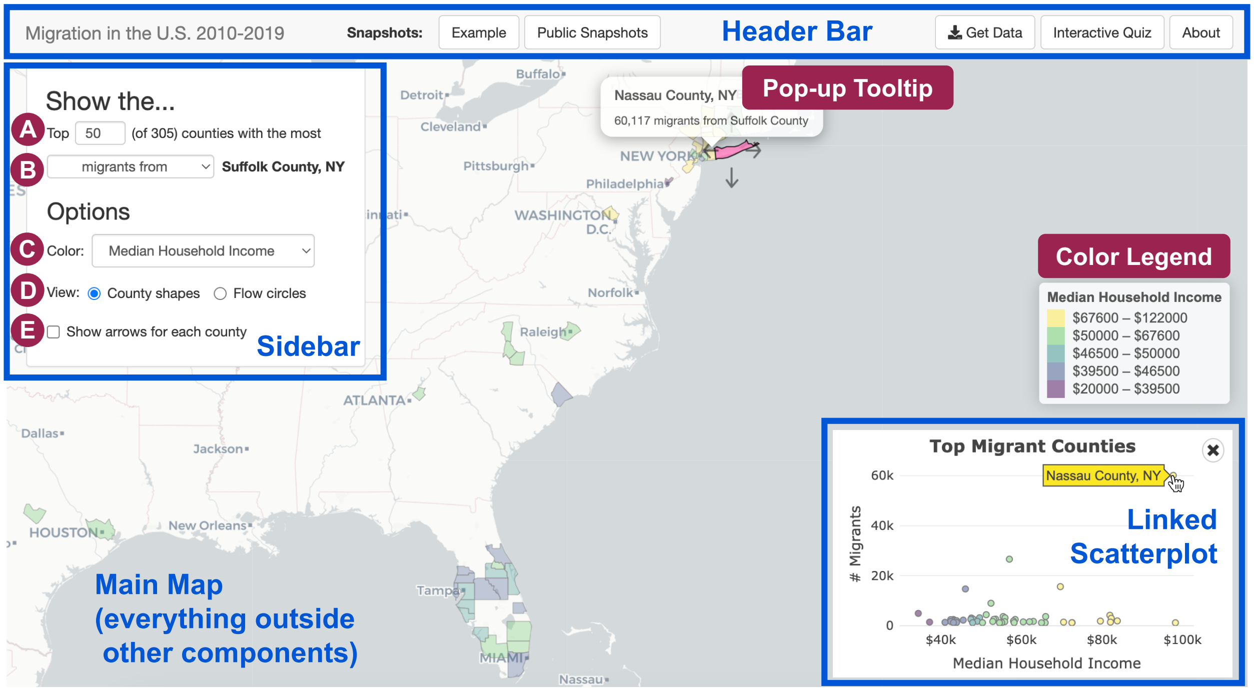

Figure 1. The system in individual county mode after clicking on Suffolk County, New York. The header bar is at the top of the UI; the sidebar is on the left-hand side; the linked scatterplot is in the bottom-right corner; and the main map occupies the rest of the UI. The sidebar houses several dynamic query options: (A) the number of top counties input; (B) the migration flow option, where the user can choose the migration direction and either raw or net migration; (C) the color dropdown menu; (D) the view option; and (E) the show arrows setting. Also note the color legend and a pop-up tooltip showing details for Nassau County, NY which is being hovered on the scatterplot.

3.4.4. Header Bar The **header bar** contains utility buttons. The About button triggers a pop-up with details about the tool and data. To the left of the About button is the button to load the Interactive Quiz (see Section 3.4.5). The Get Data button is to download migration data for the currently specified dynamic query in CSV format. Finally, the Public Snapshots button opens the snapshots panel and the Example button demonstrates the snapshots feature (see Section 3.4.6).

3.4.5. Interactive Quiz We supplement a brief text-based introduction of the system's features with an **interactive quiz** (shown on startup, and thereafter only on demand) meant to engage users in data exploration while learning how to use the system (**DR3**). In the quiz, we aimed to provide a gamified walkthrough focused on analysis questions. Users receive immediate feedback for each question, serving as a check on their understanding of both the data and the system features. With the quiz feature, we sought to exploit previously identified benefits of "game-y" visualizations (Diakopoulos et al., 2011).

3.4.6. Public Snapshots The **public snapshots** feature lets users save the current system view with a textual comment (e.g., describing why the view is interesting). Each saved "snapshot" is added to a library of saved discoveries which can be loaded by other users. This is similar to features in visualization systems designed for asynchronous collaboration, such as ManyEyes (Viegas et al., 2007) and SenseUs (Heer et al., 2007). The feature could be used in an educational setting for domain experts (likely professors) to share insights with students, or for students to share their personal findings with each other. Ideally, having a library of exemplar findings will help make the system easier to understand for all users (**DR3**). The **snapshots library panel** (Figure 2-E) is hidden by default, but can be toggled using a button in the header bar.

3.5. Usage Scenario

Population loss, partially due to net out-migration, has been an issue in many rural U.S. communities. ROBIN can help illustrate where migrants leaving rural areas often move to. To demonstrate the **snapshots**

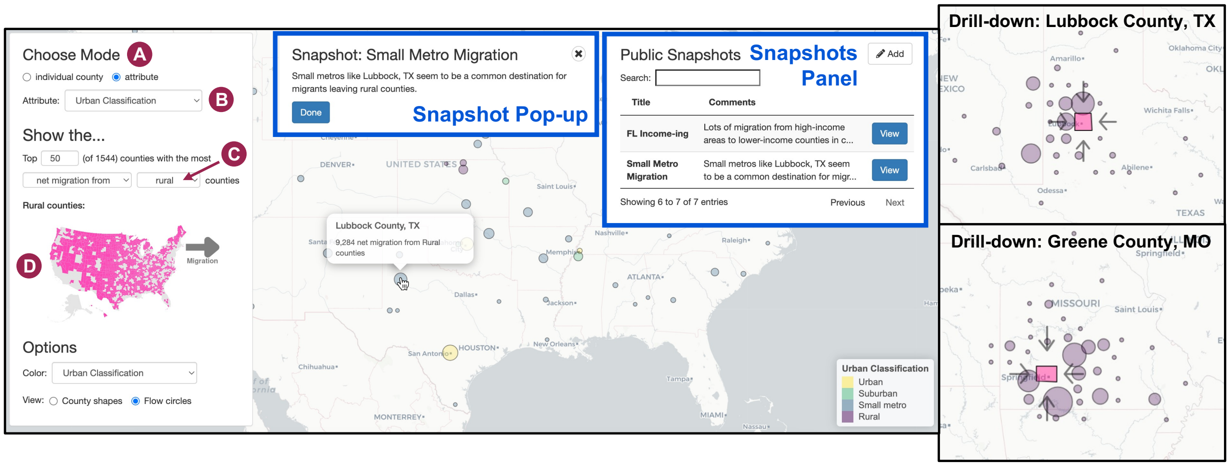

Figure 2. The system in attribute mode. ***Left:*** **Top 50 counties with the most net migration from rural counties. The mode selector (A) is set to attribute mode. The attribute of interest (B) is Urban Classification. Accordingly, the attribute selector (C) allows the user to select any classification – in this case, rural. The sidebar map (D) shows the location of these rural counties. This view is loaded from the snapshots panel, and a snapshot popup contains a blurb about the view.** ***Top right:*** **The drill-down view for Lubbock County, TX. The circles show rural counties that sent migrants to Lubbock.** ***Bottom right:*** **The drill-down view for Greene County, MO showing a similar phenomenon.**

feature, we can load a prior user's saved view that is relevant to this topic. Clicking on the Public Snapshots button in the header bar reveals the **snapshots panel** (Figure 2-E). Clicking on the View button next to the entry called "Small Metro Migration" will load the query settings and populate the map appropriately, and also show a user-submitted text snippet about the view (Figure 2-F). From the sidebar, we can see that the view shows the top 50 counties with the most net migration from rural counties, and the destination counties are colored by their urban classification as well (Figure 2, left). Many of the destination counties are small metro counties. One such destination is Lubbock County, TX. Within attribute mode, there is also a **drill-down feature** to view details on why a source or destination county is present on the map (**DR3**). For instance, clicking on Lubbock County takes us into the drill-down view, showing Lubbock attracting migrants from nearby rural areas (Figure 2, top right). Exiting Lubbock and drilling down into Greene County, MO reveals a similar story (Figure 2, bottom right). This demonstrates a broad mechanism for rural population loss: rural residents make relatively short moves to nearby small metro areas. (The supplemental video shows this scenario in detail, plus an additional scenario.)

3.6. Implementation

ROBIN is implemented in the `R` statistical programming language, using the `Shiny` library for developing web applications and leveraging the `Leaflet` library for mapping. ROBIN is deployed publicly online[2] and the code is released on GitHub[3].

4. User Studies

We conduct two sets of user studies: one set with migration domain experts, and another with non-experts. Below we describe our basic approach and summarize key takeaways from these studies. The recruiting plan and interview guide, as submitted to our university's Institutional Review Board (IRB), are also provided as supplemental material for more detail.

4.1. Approach

For the non-expert user studies, we recruited 17 non-expert participants (hereafter referred to as N1-N17) who were university students taking an introductory data visualization course taught by one of the authors. The participants were 11 males and 6 females between the ages of 19 and 22, recruited through solicitations sent via course announcement channels. These study sessions were in-person and each lasted around 45 minutes. Sessions followed a semi-structured protocol with tasks including taking the interactive quiz, opening the snapshots, and conducting free exploration. Each session ended with a semi-structured debriefing interview about overall impressions and reflections. We note that participants

[2]https://alexanderbendeck.shinyapps.io/robin-migration/
[3]https://github.com/AlexanderBendeck/robin-migration/

used a preliminary system version which was largely identical to the final system presented above, besides small tweaks made based on feedback.

For the domain expert studies, we recruited 9 experts (hereafter E1-E9) as participants, including 8 professors of geography, sociology, or a related field at various U.S. universities and 1 U.S. Census Bureau employee. We first recruited experts via targeted emails to relevant individuals identified by one of the authors, and then used a snowballing process to find additional participants based on prior participants' recommendations. We recruited 8 male experts and 1 female expert; 6 experts had over 20 years of experience working in their field and 2 others had at least 10 years. We conducted the expert study sessions remotely via Zoom, and each lasted around 30 minutes. The study protocol was similar to that for non-experts, though differed slightly due to time constraints, the remote nature of the sessions, and the different questions we had for the expert users. Each session again concluded with a debriefing interview and had an additional short survey focused on our research questions (see Table 1).

4.2. Findings

RQ1: We *can* support both domain experts and non-experts in one interface, especially by leveraging narrative design components. Non-expert participants overall enjoyed using the tool, saying that they *"like the interface"* (N15) and that it is *"pretty intuitive to use"* (N1). In particular, participants said they *"really like"* the interactive quiz (N14) and it was *"really well done for teaching how to use the tool"* (N10). Regarding the public snapshots, 9 out of 17 participants expressed a strong interest in reading snapshots and using the feature as a *"shortcut to finding insights"* (N10). During free exploration in individual county mode, 13 of 17 non-experts clicked counties that were personally relevant, which we saw as an indication that the dataset is appropriate and engaging for non-experts.

Expert participants also had positive impressions of the tool overall (see Table 1 Q1), with E1 describing it as *"really nice software"*. Encouragingly, experts saw the tool as a way to communicate with non-experts in multiple settings (Table 1 Q2). The experts who were university professors were greatly interested in the system as a teaching tool that would be *"really useful to students to explore concepts"* (E4). E2 particularly liked the interactive quiz feature *"to get them [students] sort of seeing [...] what the possibilities are"*. E5 identified ROBIN's potential as a visual aid for communicating with other non-expert audiences like reporters: *"I talk to media people a lot, and for them to be able to look at something like this while I was talking to them would be another thing that would be a use of it for me."* Several experts also mentioned that they could use the tool in their research (Table 1 Q3), perhaps as what E3 described as a *"hypothesis generator"* to search for interesting findings which could then be investigated further. E2 envisioned using the tool to *"look at patterns before I would [...] do more advanced analysis"*.

RQ2: Our glyph-based geographic visualization design helps show data relationships, but textual annotations are needed for non-experts. All expert participants reported that the tool helped illustrate relationships between migration and other attributes (Table 1 Q4). This was noted as a novel contribution: *"I don't know that I've seen anything like this before [. . .] to kind of characterize the Republican versus Democrat split of counties and then to cross that by migration"* (E8). After selecting Cook County, Illinois (home to Chicago) in individual county mode, E5 was able to easily identify an insight by looking at efficiencies, demonstrating the design's effectiveness at elucidating phenomena in the data: *"Boy, you can really see the contrast between the flows in and out. Flows in are mostly from big urban areas. The flows out are mostly to the recreational areas or retirement areas."* For the non-expert participants, the glyph-based design was not always clear by itself. The ability to read the "Show the..." sidebar text as a natural language sentence was an important aid for non-experts to understand the semantics of the glyph-based view on the map. Feedback indicated that reading the text *"helps a lot"* (N3) and *"summed up, like, things really well"* (N17).

RQ3: Dynamic query widgets are useful and flexible, but can be overwhelming for complex queries. The dynamic query widgets proved to be powerful and useful for answering queries, especially in individual county mode where clicking a county provides an intuitive entry point to each view. During free exploration in individual county mode, 15 out of 17 non-expert participants felt they could answer their questions about the dataset. Likewise, expert participants rarely reported being confused in individual county mode (Table 1 Q5). However, in attribute mode, the more complex and numerous query options posed a challenge. A few non-experts indicated that the multitude of options at times seemed *"overwhelming"* (N1), especially in attribute mode. Along these lines, N13 drew a contrast between our system and data-driven articles designed for casual consumption: *"websites that show, like, some [...] demographic tool or infographic tool, it would have less of the filters"*. Several experts also indicated that attribute mode was *"more confusing"* (E6) than individual county mode. This was backed up

Question	*Relevant RQs*	*Avg. Score (out of 5)*
Q1: I enjoyed using the tool.	**RQ1**	4.8
Q2: The tool would be useful for me in an educational setting.	**RQ1**	4.75
Q3: The tool would be useful for me in my work or research.	**RQ1**	4.4
Q4: The tool clearly shows relationships between migration & other data.	**RQ2**	4.8
Q5: I was often confused or disoriented in individual county mode.	**RQ3**	2.0
Q6: I was often confused or disoriented in attribute mode.	**RQ3**	2.8

Table 1. Results of the expert survey.

by the expert survey data as well (Table 1 Q6). However, experts envisioned they could effectively use attribute mode after a bit more practice and noted that a steeper learning curve for powerful features is not unexpected. E3 reported that attribute mode has *"... a little bit more learning curve, but I wouldn't necessarily want you to change anything. [...] Sometimes it should take a little while to learn"*. This sentiment was echoed by E8: *"It takes a minute to, sometimes, know what you're doing and how it's impacting it, but I'd much rather have that – the need for like a minute or two of orientation – and then [...] there's a lot of material to work with"*.

5. Discussion

5.1. Designing for Education and Engagement

A key aspect of this project is using visualization to increase the visibility of publicly-available migration and demographic data. We note that the datasets we utilize are all publicly accessible from government agencies and universities, and the U.S. Census Bureau has recently made some of this data visible through online mapping tools. However, it can still be difficult for those without domain knowledge to explore and learn from this raw data. Additionally, multivariate datasets with a geographic component often require the use of more advanced visualization and analysis tools. Yet we argue that such data should be easy to explore because of the impact that migration can have on livelihoods and communities. We hope that publicly releasing this tool and working with professors to deploy our system in the classroom will help promote accessibility to this data and a greater interest in it. In this way, our work is a small but important step towards democratizing publicly available geographic data through education and public engagement. Given the recognized potential for visualization to engender positive social impact (Syeda et al., 2020), we aim to broadly inspire additional work in this space.

Our experience in this project can also help inform future work in visualization that specifically aims to educate and engage the general public. In particular, we utilized design elements of narrative visualization in our system, such as the interactive quiz and snapshots. These features were received well, and we believe designing tools to resemble interactive, data-driven articles can help increase system usability and social impact. Segel and Heer (2010) outline a bevy of narrative visualization elements which could potentially be included in visualization systems. In this work, we already used *captions/headlines* (descriptive sidebar text) as well as *stimulating default views* (snapshots). We could additionally have used a more *tacit tutorial* compared to our explicit interactive quiz, or a *"checklist"* or *progress bar* to further guide users and gamify the experience of learning the tool.

As we continue to publicly deploy the ROBIN system, we are looking ahead to anticipate some intended broader impacts of the tool for public engagement. In the past, researchers have developed collaborative visualization platforms such as ManyEyes (Viegas et al., 2007) and SenseUs (Heer et al., 2007), the latter of which focused on U.S. Census data. These platforms enabled "social data analysis" and promoted discussion by citizens about publicly available data. We hope that ROBIN can serve in a similar capacity, while additionally serving as a substrate for knowledge transfer from domain experts to non-experts. Furthermore, given the important topics of migration, population growth, and government funding at various administrative levels to public discourse in the U.S., we believe our tool has the potential to spur meaningful discussions about public policy. Citizens could discuss both the potential causes of migration (e.g., policy changes or economic factors), and impacts on election outcomes (Robinson and Noriega, 2010) and urban growth (Frey, 2010), adding their unique perspective to the discussion through the public snapshots feature.

5.2. Limitations and Future Work

ROBIN currently supports visualizing migration between U.S. counties. A few experts had interest in other units of analysis such as states (E7) or metro areas (E8 & E9). However, we decided to leave this for future work. It would not be too difficult to support these different units in the U.S., since migration data and other unit-level attribute data (e.g., income) are obtainable without much trouble for U.S. states and metro areas. Supporting these units thus represents a logical next step for ROBIN. Regarding confusion about the dynamic query options in attribute mode (see Section 4.2), we plan to investigate the use of large language models and other techniques to support natural language querying. Another system limitation is a lack of support for temporal querying to show changes in migration flows over time. While it would be quick to simply add an option to filter the data by year, a more nuanced task analysis and design ideation phase for temporal features is warranted. We also acknowledge that the current focus on U.S. domestic migration data limits the applicability of ROBIN in other scenarios. Beyond the potential challenge of acquiring reliable international migration data at various geographic levels, we would likely need to address scalability issues with these larger datasets and more complex queries in order to fully support international analyses.

We also acknowledge methodological limitations. The sample size of our user studies (17 non-experts and 9 experts) is quite small. Although the non-expert sample consists of students from different backgrounds, they all attend the same U.S. university and are taking a visualization course. We should also study the reactions of non-expert users of various ages, geographic regions, and levels of visualization experience. Perhaps most importantly, we have not yet conducted a long-term evaluation of ROBIN in a real-world setting, and a more naturalistic evaluation is warranted in the future. Given our expert collaborators' expressed interest in using the system in the classroom and even for research, we could potentially conduct a study over several months to assess these use cases. Such an evaluation would be particularly necessary to truly assess the effectiveness of the system at, for instance, supporting learning or facilitating knowledge transfer from experts to students with the snapshots feature. Given that our evaluations to date have been mostly qualitative, we also plan to identify effective quantitative metrics for evaluating the system over a longer period of time.

6. Conclusion

Datasets about migration patterns within the United States (U.S.) are often utilized by migration researchers alongside data about characteristics of localities (e.g., political leanings) to study how the country is changing. However, existing tools to visually explore such data are not designed to be interpretable by non-experts, leaving this vital data largely inaccessible for the public. In this work, we present the ROBIN system designed for exploratory analysis of county-to-county U.S. migration flows alongside other relevant county-level attributes. We find that supporting both domain experts and non-experts in a single visualization interface is feasible by leveraging narrative design features. Our glyph-based geographic visualization design is reasonably effective for representing relationships between migration data and other attributes. The tool's dynamic query widgets enable flexible querying, though can overwhelm users (especially non-experts) in complex scenarios. Finally, we reflect on key aspects of designing for non-expert education and engagement.

References

Ambinakudige, S., & Parisi, D. (2017). A spatiotemporal analysis of inter-county migration patterns in the united states. *Applied Spatial Analysis and Policy*, *10*, 121–137.

Böttinger, M., Kostis, H.-N., Velez-Rojas, M., Rheingans, P., & Ynnerman, A. (2020). Reflections on visualization for broad audiences. *Foundations of Data Visualization*, 297–305.

Boyandin, I., Bertini, E., Bak, P., & Lalanne, D. (2011). Flowstrates: An approach for visual exploration of temporal origin-destination data. *Computer Graphics Forum*, *30*(3), 971–980.

Diakopoulos, N., Kivran-Swaine, F., & Naaman, M. (2011). Playable data: Characterizing the design space of game-y infographics. *Proc. CHI*, 1717–1726.

Egan-Robertson, D., Curtis, K. J., Winkler, R. L., Johnson, K. M., & Bourbeau, C. (2023). Age-specific net migration estimates for US counties, 1950-2020.

Fekete, J.-D., Van Wijk, J. J., Stasko, J. T., & North, C. (2008). The value of information visualization. *Information Visualization: Human-Centered Issues and Perspectives*, 1–18.

Frey, W. H. (2010). Population and migration. *State of Metropolitan America*, 36–49.

Galle, O. R., & Williams, M. W. (1972). Metropolitan migration efficiency. *Demography*, *9*, 655–664.

Golding, S. A., & Winkler, R. L. (2020). Tracking urbanization and exurbs: Migration across the rural–urban continuum, 1990–2016. *Popul. Res. Policy Rev.*, *39*, 835–859.

Grigg, D. B. (1977). Eg ravenstein and the "laws of migration". *Journal of Historical Geography*, *3*(1), 41–54.

Guo, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Trans. Visual Comput. Graphics*, *15*(6), 1041–1048.

Guo, D., Chen, J., MacEachren, A. M., & Liao, K. (2006). A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Trans. Visual Comput. Graphics*, *12*(6), 1461–1474.

Harrison, J. A. (2017). Rust belt boomerang: The pull of place in moving back to a legacy city. *City & Community*, *16*(3), 263–283.

Heer, J., Viégas, F. B., & Wattenberg, M. (2007). Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. *Proc. CHI*, 1029–1038.

Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E., & Jenny, H. (2017). Force-directed layout of origin-destination flow maps. *International Journal of GIS*, *31*(8), 1521–1540.

Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E., & Jenny, H. (2018). Design principles for origin-destination flow maps. *Cartography and GIS*, *45*(1), 62–75.

Koylu, C., Tian, G., & Windsor, M. (2022). Flowmapper.org: A web-based framework for designing origin–destination flow maps. *Journal of Maps*, 1–9.

Liu, X., Andris, C., & Desmarais, B. A. (2019). Migration and political polarization in the US: An analysis of the county-level migration network. *PloS one*, *14*(11), e0225405.

Liu, X., Hollister, R., & Andris, C. (2018). Wealthy hubs and poor chains: Constellations in the US urban migration system. *Workshop on Agent-Based Models and Complexity Science (GIScience)*, 73–86.

Manko, B. A. (2022). Teaching user-friendly web design: A case study on zillow.com in the real estate industry. *Journal of Information Technology Teaching Cases*, *12*(1), 35–42.

Oppermann, M., & Munzner, T. (2020). Data-first visualization design studies. *Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, 74–80.

Plane, D. A. (1992). Age-composition change and the geographical dynamics of interregional migration in the u.s. *Ann. Am. Assoc. Geogr.*, *82*(1), 64–85.

Ravenstein, E. G. (1889). The laws of migration. *J. R. Stat. Soc.*, *52*(2), 241–305.

Rayer, S., & Brown, D. L. (2001). Geographic diversity of inter-county migration in the united states, 1980–1995. *Popul. Res. Policy Rev.*, *20*, 229–252.

Robinson, T., & Noriega, S. (2010). Voter migration as a source of electoral change in the rocky mountain west. *Political Geography*, *29*(1), 28–39.

Scheidl, A., Leite, R. A., & Miksch, S. (2021). Vismiflow: Visual analytics to support citizen migration understanding over time and space. *EuroVis Short Papers*.

Schöttler, S., Yang, Y., Pfister, H., & Bach, B. (2021). Visualizing and interacting with geospatial networks: A survey and design space. *Computer Graphics Forum*, *40*(6), 5–33.

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visual Comput. Graphics*, *18*(12), 2431–2440.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Trans. Visual Comput. Graphics*, *16*(6), 1139–1148.

Shneiderman, B. (1994). Dynamic queries for visual information seeking. *IEEE Software*, *11*(6), 70–77.

Stephen, D. M., & Jenny, B. (2017). Automated layout of origin–destination flow maps: US county-to-county migration 2009–2013. *Journal of Maps*, *13*(1), 46–55.

Syeda, U. H., Murali, P., Roe, L., Berkey, B., & Borkin, M. A. (2020). Design study "lite" methodology: Expediting design studies and enabling the synergy of visualization pedagogy and social good. *Proc. CHI*, 1–13.

United States Census Bureau. (2021). Internal migration in the U.S. - exploring the census flows mapper [https://www.census.gov/programs-surveys/sis/resources/visualizations/flows-map.html].

Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., & McKeon, M. (2007). Manyeyes: A site for visualization at internet scale. *IEEE Trans. Visual Comput. Graphics*, *13*(6), 1121–1128.

Wood, J., Dykes, J., & Slingsby, A. (2017). Visualization of origins, destinations and flows with od maps. *Landmarks in Mapping*, 343–362.

Yang, Y., Dwyer, T., Goodwin, S., & Marriott, K. (2016). Many-to-many geographically-embedded flow visualisation: An evaluation. *IEEE Trans. Visual Comput. Graphics*, *23*(1), 411–420.