Mario Moreno

Data Visualization

January 13, 2019

**Data Visualization Proposal**

In the midst of a raging national debate on migration and refugees, plenty has been written about the changing profile of undocumented migrants reaching the US-Mexico border. While traditionally, Mexican migrants have been the predominant demographic group seeking to enter the United States through the southern border, that dynamic has shifted in the last decade. Central American migrants originating in the Northern Triangle countries of El Salvador, Guatemala, and El Salvador have become the predominant demographic of undocumented migrants seeking to enter the United States.

This wave of Central American migration has shed a light on human rights issues all along the Central America-Mexico-US corridor. Experts, for example, attribute the increase in Central American migration to people fleeing grave human rights violations associated with endemic violence, lack of economic opportunities, family reunification, and in some cases, deteriorating environmental conditions in Northern Triangle communities. In traversing Mexico, human rights advocates highlight abuses suffered by these travelling migrants due to Mexican transnational criminal organizations and corrupt government officials. Once they reach the US border, immigrant service providers note how a badly broken American immigration system fails to adequately welcome, process, and account for migrants.

The scope and urgency of this crisis are generally well understood, but few people to date have used data to explore a range of different dynamics involved with the recent surge in Central American migration. Although we know that violence and lack of opportunity are two key drivers in the recent trend, there’s little publicly available data that can help us understand the scope of the problem in these countries. While we know that Mexico has taken on a larger role in stopping and deporting Central American migrants before they reach the US-Mexico border, the extent of those efforts and the experiences migrants have with Mexican authorities haven’t been analyzed using data. Though we understand that the United States has provided foreign aid to Central America in order to combat corruption, violence, and more, the effectiveness of that aid has yet to be mapped to changes in country.

My data visualization project is an attempt to further elucidate those dynamics using available data sources. Using biannual survey data for El Salvador, Guatemala, and Honduras, I’ll attempt to show how citizens perceptions of violence, economic opportunities, and other factors have changed in the last decade. Through annual surveys collected from Central American migrants deported by US and Mexican forces, I’ll endeavor to explore what challenges Central American migrants faced in their journey northward, as well as dive deeper into the factors that drove them to migrate. Lastly, by relying on US foreign aid data, I’ll try to understand if American aid has shifted over time to account for the drivers of Central American migration. I

**The data**

Due to issues of data availability and post 2009 changes in the EMIF Sur surveys, I’ll limit my analysis to 2010 – 2016.

LAPOP Opinion Surveys

LAPOP is a biannual survey conducted in every Latin American country by Vanderbilt University researchers. Questions on the survey include what the respondent identifies as the most pressing national issue, effectiveness of local social services, social involvement, opinions on how to address crime, corruption, and more.

There are roughly 1500 rows in the datasets associated with each survey, with each row representing a respondent. The columns include demographic information and answers to the survey questions which, depending on the survey, are captured in nearly 100 columns.

Since I’ll be looking at the 2010-2016 timeframe, which would include surveys for all three countries for 2010, 2012, 2014, and 2016. That would be a total of 12 datasets, which means I’d wind up with a global dataset of ~18,000 rows. I’ve yet to determine how I’ll handle differences in questions asked between years or countries.

LAPOP API: <https://www.vanderbilt.edu/lapop/raw-data.php>

EMIF-SUR Surveys

EMIF-SUR surveys are yearly surveys of deported Central American and Mexican migrants conducted by a coalition of researchers affiliated with Mexican and Guatemalan universities, as well as Mexican government agencies. Questions on the survey include educational attainment, place of origin, employment and salary prior to emigrating, port of entry, cities crossed in Mexico on their journey, and more.

There are two surveys I will be sampling: surveys of Central American migrants deported from Mexico, and surveys of Central American migrants deported from the United States. The surveys are divided by country (There is one dataset for Honduran migrants deported from Mexico, another one for Guatemalan migrants deported from the United States, and so on.) I’ll be analyzing surveys that match the LAPOP years (2010, 2012, 2014, 2016), which means I’ll be working 24 separate datasets. Each one of these datasets is roughly 3000 rows and 100 columns, where each row represents a respondent and each column an answer to a question.

EMIF-Sur portal: <https://www.colef.mx/emif/dwn/dwn.php>

US Foreign Aid Data

In order to determine aid flows to these three countries, I will combine data from the USAID Aid Explorer API and the Security Assistance Monitor database. The USAID Explorer API tracks a range of US foreign assistance packages to Central America, though it’s often criticized for not including military aid. While that aid is typically not disclosed by the US government, I will rely on a dataset created by the Security Assistance Monitor, which tracks US military and defense aid abroad, for this information.

The USAID Explorer data is 3,219 rows and 47 columns. Each row is an aid project implemented in one of the three Central American countries during the LAPOP years (2010, 2012, 2014, 2016). The columns include the aid project, country recipient, US implementing agency, sector, funding account, fiscal year, and amount. The Security Assistance Monitor data. The Security Assistance Monitor data will require further cleaning, but it will roughly include about 150 rows and 7 columns. Each row is a military aid program funded by the United States in the three Central American countries, and the columns include year, amount, and program details.

Security Aid Totals: <http://securityassistance.org/content/military-aid-download>

USAID Foreign Aid Explorer: <https://explorer.usaid.gov/query>

Other data sources

Time permitting, I will attempt to use ACS data to document US migration patterns for Central Americans residing in the US, UNHCR data to track trends of refugee and asylum admissions into the US and Mexico over time, World bank data provided to these three nations, reported crimes and food insecurity indicators in these three countries, and more.