Analyzing and Forecasting U.S. Immigration Trends
Group 7: Alan Lin , Trevor Petrin , Ganesh Kumar Ramar , Mingyu Chen
Repository: alanklin/AA-Capstone: Boston College Applied Analytics Capstone Project

For our Applied Analytics Capstone Project, we are interested in pursuing an incredibly relevant topic that has sparked massive discussion in the past decade: immigration. In particular, we'd like to explore and extract patterns from U.S. Customs and Border Protection (CBP) datasets dating back to 2020. Our goal is to investigate and forecast any changes in monthly migrant encounters at U.S. ports of entry (overseen by field offices) or sectors. If we can identify and predict these changes, it may be beneficial and possible to increase or direct their funding to hire more personnel or prepare other measures to withstand the stress of an oncoming surge of migrants. To provide some background information, the CBP is the nation's largest federal law enforcement agency and it currently operates with a budget of $19.6 billion dollars for fiscal year 2024. Their top concern is to keep terrorists and weapons from entering the U.S. while enforcing laws at the border, such as preventing illegal entry, seizing illegal drugs and contraband, and protecting U.S. agriculture from harmful diseases and pests. In many ways, the CBP's work is incredibly important to the security and health of U.S. travel and trade, but recent events and policies have undoubtedly soured public perception on their effectiveness given their sizable budget.

One of the most controversial policies was Title 42, which was adopted by both the Trump and Biden administration to turn away migrants, regardless of asylum-seeking status, on the basis of the public health emergency that was the COVID-19 pandemic. At the time, the Centers for Disease Control and Prevention (CDC) maintained that the order would protect CBP personnel and U.S. residents, as well as prevent the U.S. healthcare system from being overwhelmed by an influx of potential COVID-19 carriers. Over the course of Title 42's lifetime, around 3 million expulsions were recorded, with most migrants being expelled into Mexico or their country of origin. Some were transported by air, but this was rarely done as it was an expensive operation, resulting in many migrants being stranded in foreign lands. Unaccompanied children were initially also expelled, often to dangerous conditions where they could be vulnerable to more violence, kidnapping, and human trafficking. A later federal judge ruling in 2022 made unaccompanied children exempt from Title 42, which caused some families to split prematurely to prevent their children from being turned away immediately. With millions of migrants denied a fundamental human right to seek asylum, there was massive outrage directed towards both the administration and CBP from human rights advocacy groups, immigrant centers, and many more people. In addition, Title 42 did not seem to stem the flow of immigration, a pattern we'd like to further examine. If possible, we could potentially connect the impact of Title 42 on COVID-19 cases, following the outcries from public health officials, medical professionals, and epidemiologists stating that the policy

was not necessary to protect public health. There seems to be little evidence that Title 42 prevented illness and COVID-19 from spreading in the U.S.; instead, the health of migrants were generally of more concern, with many becoming more vulnerable to illness due to poor conditions at the U.S. Mexico border such as inadequate shelter, sanitation, safety, and access to food and water. As you may see, there is a large well of potential insights and knowledge to be drawn from trying to analyze the effects of this controversial policy.
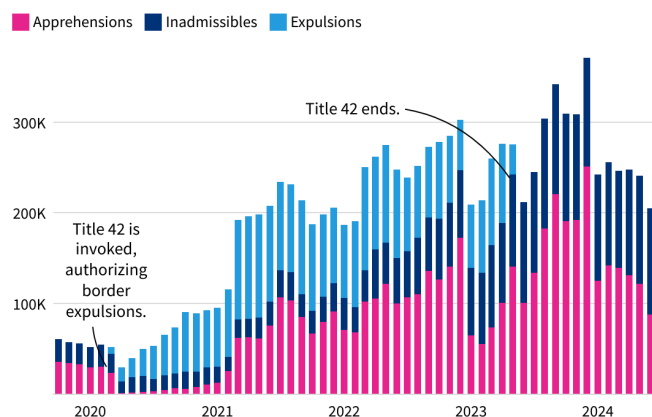
The other main asylum status is Title 8 of the US Code, which is the section containing immigration law outlining processes for handling migrants at the border and deportation if they do not have a legal basis to remain in the country. The main difference between Title 8 (Apprehension/ Inadmissible) and Title 42 (Expulsion) is Title 8 allows migrants more time to lodge an asylum claim than they were afforded under Title 42, which allows for the expulsion of migrants without due process. Other differences include that Title 8 bars migrants from re-entry into the country for a period of time from 5-20 years depending upon the type of detention whereas Title 42 did not give a set period of time before an individual could request reentry. However, if migrants do not claim asylum when they are apprehended when crossing the border, they can be quickly deported without a court hearing. If they claim asylum, though, they must pass a screening with an officer to avoid deportation. These 'credible fear' interviews, designed such that most non-Mexican asylum

Reported border encounters increased from 2020 to 2023.

Monthly nationwide border encounters by category, October 2019– June 2024

Recent data is available up to June 2024.
Source: Customs and Border Protection

USA**FACTS**

seekers are rejected, deport migrants who fail for at least 5 years. Those who pass can either be detained in long term detention facilities or are released with a notice to appear in court for a review of their case. Migrants forfeit their right to stay in the US should they fail to report to their hearing and will be deported if that is the case, however, because of the current backlog of cases, it can take years to get a court date.

Between 2019 and 2024, there were 11 million border encounters nationwide, roughly equivalent to the population of North Carolina, the 9th most populous state, peaking with over 12,000 encounters per day. Observing the rate of increase in reported border encounters, it is clear that US Customs and Border Patrol (USCBP) will not immediately have the funding nor staffing to handle the increased rate of apprehensions. These metrics can be used to anticipate the hiring and resource needs of the department at

the Federal level to handle the continued influx of migrants. These numbers, however, are a good rough estimate for the department's use, as previously stated, there are issues including uncaught undocumented immigrants and multiple reentry attempts by Title 42 expulsions.

Encounters with migrants from some countries rose dramatically in 2021

Migrant encounters at U.S.-Mexico border, by citizenship and fiscal year
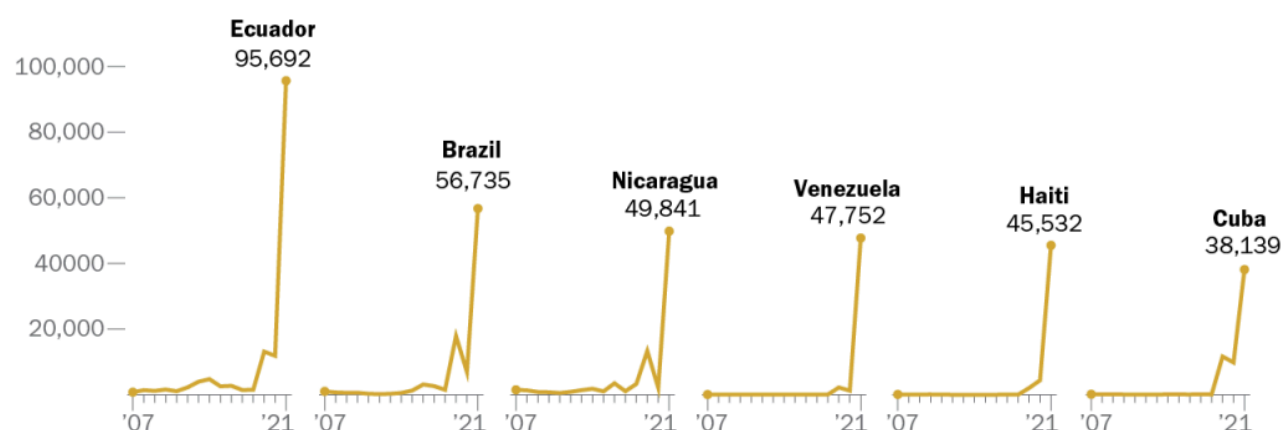
Note: Beginning in fiscal 2020, annual totals combine expulsions and apprehensions into a new category known as encounters. Annual totals before fiscal 2020 include apprehensions only.
Source: U.S. Customs and Border Protection.

PEW RESEARCH CENTER

As for understanding the source of the influx of migration, the data can be used to determine the trends not only for each country, but also for regional migration trends. Beginning in late FY2020-2021, there was a spike in migration from South American countries Ecuador, Nicaragua, Brazil, and Venezuela encountered at the US-Mexico border driving the overall spike in encounters. Additionally, there were noteworthy rises in encounters as well from Haiti and Cuba during that time marking a sharp change in the way migrants came to the US from those regions.
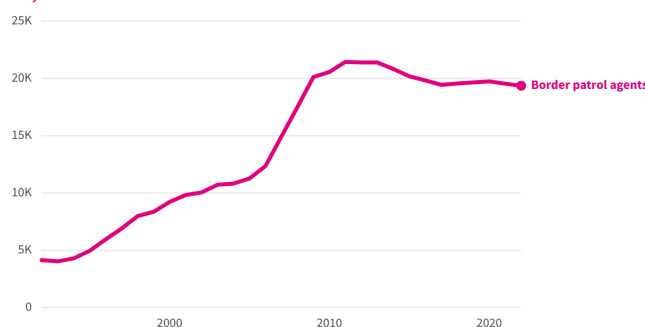
As for the dataset, CBP data was used tracking the encountered migrants' country of origin as well as their family status, and under what immigration code they were held and when they were apprehended. The data was aggregated by month, country of origin, family status, and in two separate datasets by state or CBP

USA**FACTS**   OUR NATION, IN NUMBERS

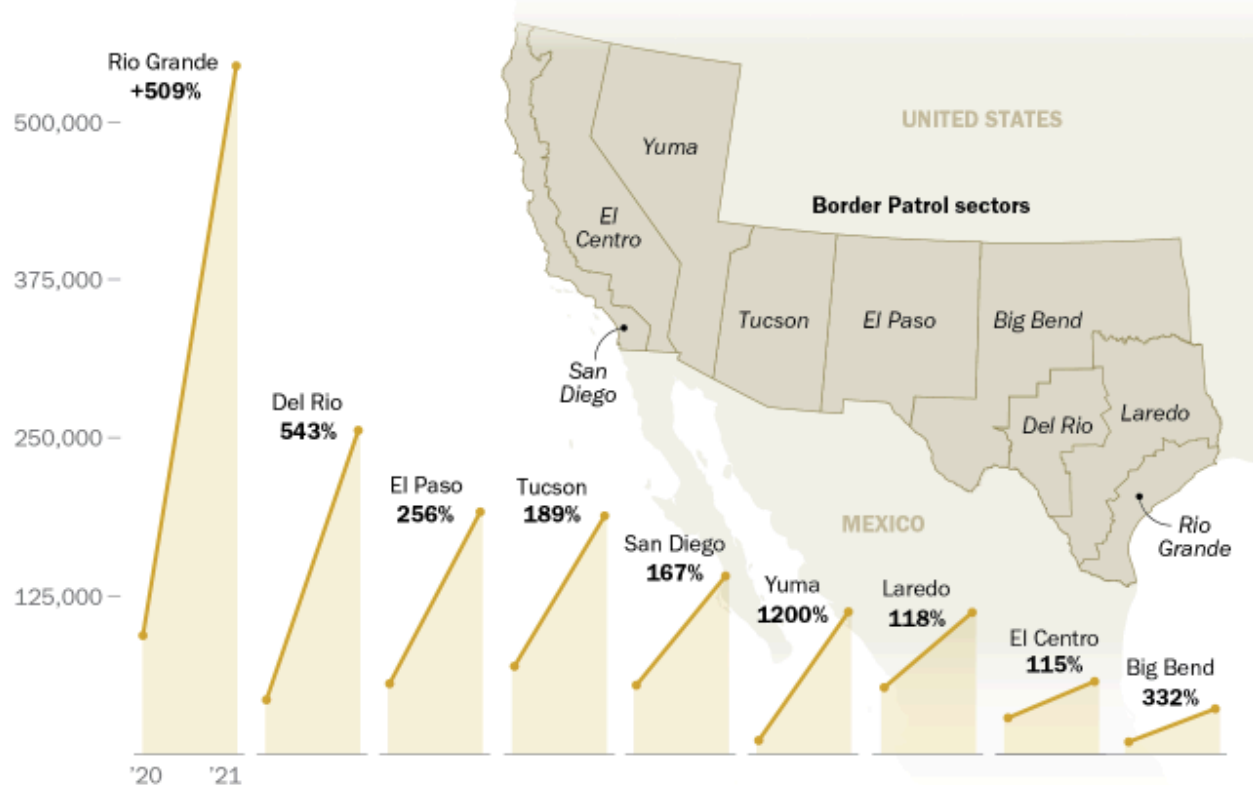**Border patrol agents**
19,357 in 2022

Explore the data and source at usafacts.org © 2019 USAFacts.org. All rights reserved.

region for where they were apprehended. This data can be used to create a forecasting problem concerning predicting the number of migrants overall as well as at each separate region throughout the country to better allow CBP to anticipate personnel and resource allocation needs. Especially with the number of border patrol agents stagnant as shown in the graph above. This problem will be supervised learning for predicting the number of crossings and regression, as the number of migrants is a quantitative variable.

Migrant encounters more than doubled in all nine southwest border sectors in 2021

Change in migrant encounters at U.S.-Mexico border, by sector, fiscal 2020-2021

The main stakeholders for this forecasting problem include the Federal Government and United States taxpayers. The Federal Government must know how to allocate their resources to orderly and legally process migrants they encounter without being overwhelmed by the sheer volume of people. They must understand where to put their funding, agents, and attention in collaboration with state governments to be ready to apprehend the influx of people where they occur on the border. Additionally, it gives insight to the state of other countries investigating the demographics and origin of the migrants coming to the borders, northern and southern. The data can give insight to the trend and seasonality of the arrivals as

well, showing when individuals are deciding to travel and possibly revealing socio-economic trends in their country of origin. As for the US taxpayers, they have a right to know where their money is going in support of the Federal Government and for the state of their borders to understand how the influx of people may impact their daily lives. As can be seen in the graph above, all of the CBP Regions along the Southern Border saw a spike in incoming migrants from FY 2020 to 2021 straining migrant services and overwhelming the CBP. We want to take a data driven approach to forecasting and solving this problem to predict the influx before they arrive.

We are well-aware that migration is a very complex system, making forecasting an incredibly difficult process. Immigration data can be hard to get a hold of in a raw format, which can make it difficult to formulate a question that can be answered with the data that we have. In our searches, we have come across articles mentioning that they obtained exclusive data from a Border Patrol agent such as the number of illegal migrants crossing the Southwest U.S. Border in 2024. Currently, the CBP keeps track of "encounters" which means they don't account for "gotaways", otherwise known as individuals that entered the US without encountering CBP agents. With no formal consequences from expulsion via Title 42, it was reported that 1 in 3 apprehensions were of individuals on at least their second attempt to cross the border. Main offenders of this phenomenon were from the countries of Mexico, Guatemala, Honduras, and El Salvador. As a result, Title 42 appears to have substantially inflated the number of migration encounters. Furthermore, recent news shows that New Mexico saw an increase in illegal immigration from July 2024 to August 2024, primarily due to an executive order that restricted asylum seekers from entering the country until average daily crossings dropped below a certain threshold. Potential months of waiting led many desperate migrants to rely on Mexican smugglers and American contractors to enter the country through other means, namely, enter through a different part of the US that is less guarded. With Mexican cartels supporting these migrants, including attacking and assaulting Border Patrol agents, the lives of these personnel are constantly put under risk, not to mention the unforgiving desert that these migrants are travelling through. If anything is to be learned from the current situation, it is that the U.S. immigration system is rather flawed and requires an incredible amount of restructuring and policy change. Although our project is limited in its application to enact policy reform, we do want to provide more insight into immigration trends and potentially give both migrants and US CBP personnel a chance, no matter how slim, to work through the immigration system without endangering each other's lives.

For the economic value this forecast would provide, the assumptions for this problem include that an unlawful immigrant faces either expulsion outright, costing the US nothing to turn them away, detention, then expulsion, or deportation if CBP misses the migrant at the border. The average cost per migrant deportation is roughly $19,559 (Ingram 4), and the alternative fate if they are not granted asylum is detention before deportation costing $200 per day with an average stay of 44 days (Urbi 3). The

proportion of migrants that are expelled outright was roughly 50% of encounters (USAFacts). This gives the following formula to calculate potential savings per individual migrant who would otherwise be detained:

$$Savings\ per\ Detainee\ = (Deport\ -\ (Stay\ \times\ Detain/Day))$$
$$\$10,759 = (19,559\ -\ (44\ \times\ 200))$$

Substituting the background research into this equation, along with the number of missed encounters estimated at roughly 660,000 along the southern border alone (USAFacts) and projecting the proportion expelled rather than being detained remains the same. Assume with the resources properly allocated CBP can encounter an additional one-tenth of the attempted uncaught illegal crossings:

$$Annual\ Savings\ = ((1\ -\ Expelled)\ \times\ (Savings\ per\ Detainee)\ \times\ (Total\ Missed\ Encounters))\ \div\ 10$$
$$\$355m =\ (1\ -\ 0.5)\ \times\ (10,759)\ \times\ (660,000)\ \div\ 10$$

If proper resource allocation can allow for 1/10 of previously estimated uncaught migrant crossings on the southern border alone to be encountered, the cost savings as opposed to deportation will be roughly $355 million making proper forecasting a key component towards efficient government spending. As was previously mentioned, though, we do not want it lost upon the stakeholders that this is at its core a human issue, where the goal is to be as humane to migrants as possible while ensuring the safety of the American people by keeping the small proportion of criminals and illicit drugs out of the borders. This will also disincentivize immigrants from crossing illegally via smugglers or criminal/terror groups resulting in their funding and entry into the country. America was built off of the backs of immigrants, and the stakeholders must not forget that, but disincentivizing migrants from using human traffickers and other means of crossing illegally is in the interest of national security.

Project Outline:

	Objectives	Deliverables
Week 1	Identify Problem Statement, Gather Dataset & Quick Preliminary Data Investigation, Set-Up Github, Written Report	Written Report, GitHub Repo link
Week 2	Dataset Investigation. IDE. Define Variables	Written Report, Repo Update
Week 3	Set up Pre-processing Pipeline, Train/Test/Validation splits. EDA	Written Report, Repo Push

Week 4	Get data ready to input into the model. Perform preprocessing on dataset splits	Written Report, Repo Push
Week 5	Feature Engineering. Final train/test/validation splits	Written Report, Repo Push
Week 6	Build simple models appropriate to the problem. Model tuning and evaluation	Written Report, Repo Push
Week 7	Build more advanced models. Tune and evaluation	Written Report, Repo Push
Week 8	Build even more advanced models. Tune and evaluation	Written Report, Repo Push
Week 9	Select best and most appropriate performing model	Written Report, Repo Push
Week 10	Improve model by improving data. Perhaps we use the most updated data from CBP/DHS	Written Report, Repo Push
Week 11	Explain Model, Feature Importance, Prediction Outcomes, Identify/Measure Bias	Written Report, Repo Push
Week 12	Package for Model Deployment, Model Monitoring	Written Report, Repo Push
Week 13	Bring everything to a close. Final Discussion	Final Written Report, Repo Push

Brief Dataset Discussion:

We acquired our datasets from the U.S. Customs and Border Protection website, linked here: https://www.cbp.gov/document/stats/nationwide-encounters. There are 4 datasets of interest here. FY22 - FY25 Nationwide Encounters by Area of Responsibility - December, FY20-FY23 Nationwide Encounters by Area of Responsibility, FY22 - FY25 Nationwide Encounters by State - December, and FY20-FY23 Nationwide Encounters by State. We've merged each accompanying pair so that we have data from FY20-FY25 for both Encounters by Sector (AoR) and Encounters by State and removed one copy of FY23 as there was overlap between the two datasets. Looking closer, the Encounters by Sectors dataset includes data for both Field Offices and Sector. Field Offices presumably operate within a given Sector and usually are responsible for multiple ports of entry. The figure down below shows how the US has been split up into 20 sectors. Notice how Texas is split into 4 sectors. As such, while we do have data for each state, we believe that basing our model on Area of Responsibility will prove to be more effective and targeted, i.e. we can narrow down our predictions to Field Office or Sector.

● Border Patrol stations
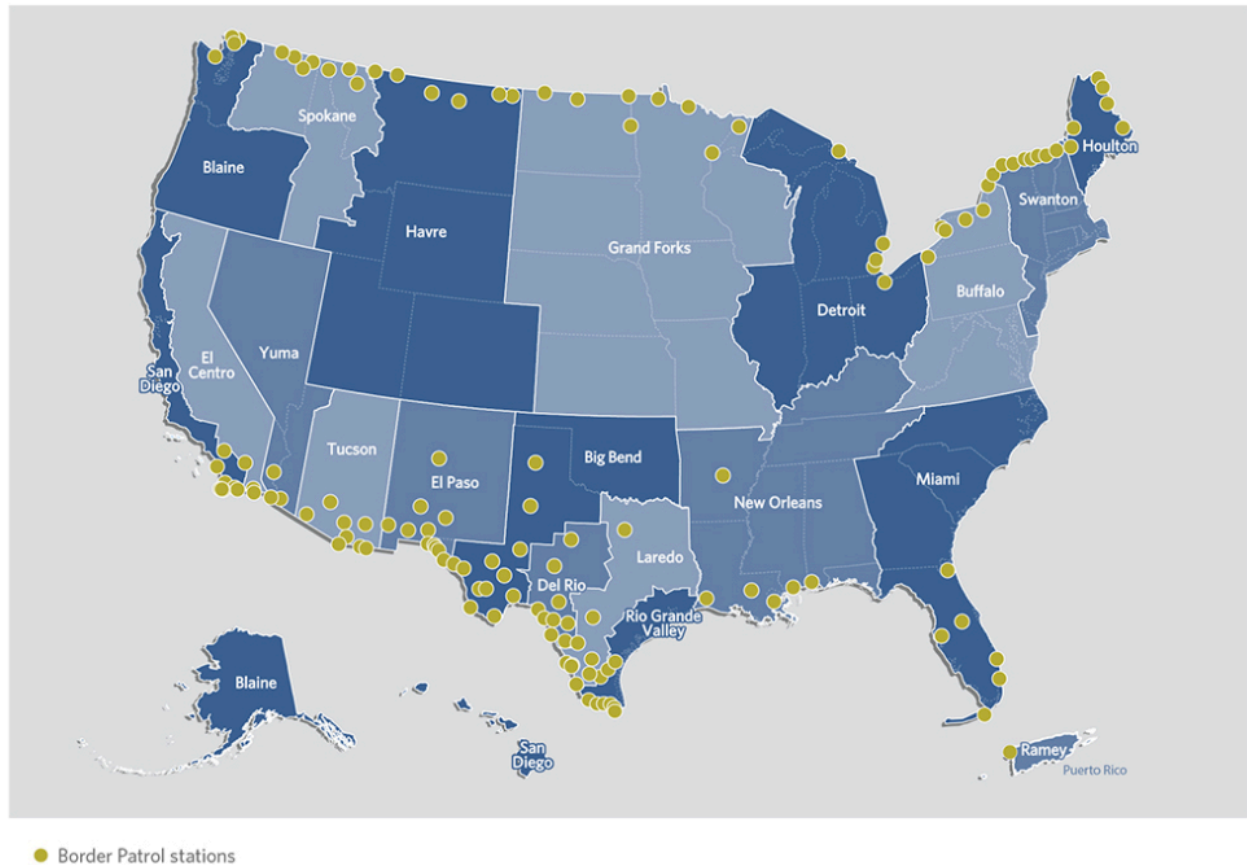
The resulting Encounters by Sector dataset has around 71,000 entries with 12 columns. Variables include Fiscal Year, Month, Land Border, Area of Responsibility, Demographic (basic information), Country of Citizenship, Title of Authority (processed through Title 8 or Title 42), Encounter Type, and Encounter Count. With data between 2020-2025 (and counting!), we should then be able to explore any underlying trends, produce a forecast, and provide suggestions for a given sector/AoR, For example, perhaps we see a decline in the Rio Grande sector, which is then followed by a rise in El Paso due to migrants attempting to cross a different part of the border. With proper anticipation, resources such as manpower, funds, and holding capacity could possibly be effectively distributed to sectors that need more assistance.

While immigration remains a complex and deeply human issue, our findings aim to support policymakers in balancing humane treatment of migrants with the safety and security of U.S. borders. Proper forecasting not only improves resource allocation but can also reduce reliance on costly deportation practices. Moving forward, our project will refine these insights, build predictive models, and offer actionable recommendations to address the multifaceted challenges of immigration.

Sources

ABC News. "What Is Title 8 Immigration Law?" *ABC News*, n.d., https://abcnews.go.com/Politics/title-8-immigration-law/story?id=99226848/.

ABC7 Chicago. "What Is Title 8 Immigration Law? And What Happens When Title 42 Ends?" *ABC7 Chicago*, n.d., https://abc7chicago.com/title-8-immigration-law-42-explained-vs/13234991/.

Border Report. "Illegal Immigration Up 22% in New Mexico." *Border Report*, n.d., https://www.borderreport.com/immigration/border-crime/illegal-immigration-up-22-in-new-mexico/.

Columbia University Mailman School of Public Health. "Voices of September: Asylum Press Release." *Program on Forced Migration and Health*, September 2021, https://www.publichealth.columbia.edu/research/programs/program-forced-migration-health/voices/september-2021-asylum-press-release.

Ford School of Public Policy. "The Impact of Title 42 on Disease Prevention and Migrants' Well-Being." *University of Michigan*, 2023, https://fordschool.umich.edu/news/2023/impact-title-42-disease-prevention-and-migrants-well-being.

Pew Research Center. "What's Happening at the U.S.-Mexico Border in 7 Charts." *Pew Research Center*, n.d., https://www.pewresearch.org/short-reads/2021/11/09/whats-happening-at-the-u-s-mexico-border-in-7-charts/.

The Pew Charitable Trusts. "Immigration Enforcement Along U.S. Borders and at Ports of Entry." *The Pew Charitable Trusts*, February 2015, https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2015/02/immigration-enforcement-along-us-borders-and-at-ports-of-entry.

USAFacts. "Border Patrol Agents." *USAFacts*, n.d., https://usafacts.org/data/topics/security-safety/crime-and-justice/homeland-security/border-patrol-agents/.

USAFacts. "Statistics on Unauthorized US Immigration and US Border Crossings by Year." *USAFacts*, n.d., https://usafacts.org/articles/what-can-the-data-tell-us-about-unauthorized-immigration/.

Washington Office on Latin America (WOLA). "Ending Title 42: Key Resources and Analysis." *WOLA*, n.d., https://www.wola.org/analysis/end-title-42/.

Urbi, Jaden. "This Is How Much It Costs to Detain an Immigrant in the US." CNBC, CNBC, 5 July 2020, www.cnbc.com/2018/06/20/cost-us-immigrant-detention-trump-zero-tolerance-tents-cages.html?msockid=2afada3725666fb01541c9b3244d6e2d.

Ingram, Julia. "Trump's Plan to Deport Millions of Immigrants Would Cost Hundreds of Billions, CBS News Analysis Shows." *CBS News*, CBS Interactive, 17 Oct. 2024, www.cbsnews.com/news/trump-plan-deport-immigrants-cost/.