TIME SERIES ANALYSIS ON U.S. IMMIGRATION DATA

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Big Data Analytics

---

by

Elizabeth Fabio

Spring 2022

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Elizabeth Fabio:

Time Series Analysis on U.S. Immigration Data

Xialu Liu, Chair
Department of Management Information Systems

Aaron Elkins
Department of Management Information Systems

Benjamin Gonzalez O'Brien
Department of Political Science

Approval Date

DEDICATION

This thesis is dedicated to my fiancé, Memphis Huff V, who has supported me constantly throughout graduate school. He is the reason why I finally pursued my master's degree and why I now have something to show for it. I'd also like to dedicate my thesis to my parents, Roy and Carmen Fabio. They helped shaped me to be the person I am today.

ABSTRACT OF THE THESIS

Time Series Analysis on U.S. Immigration Data
by
Elizabeth Fabio
Master of Science in Big Data Analytics
San Diego State University, 2022

U.S. immigration data is analyzed using intervention event analysis and time series clustering. Most research is limited to immigration data in several major migration eras or only focus on immigrants with certain skills that were favored or disfavored by the U.S. This study conducts a comprehensive analysis of publicly available immigration data from 1820 to 2020. We examine impacts of political and economic events that have reshaped the U.S., compare their impacts on legal and undocumented immigration, and further investigate the significant migration patterns from sending countries. Upon understanding the trends and outliers in time series data, immigration flow can be better managed. With these goals in mind, U.S. immigration is assumed to follow a domestic politics and globalization approach. Domestic politics is best explained as a neutral arena for societal interest. Globalization is the international economic factors and social drivers that have a major influence on society.

Based on the intervention analysis, many events had a temporary negative impact. International economic events such as the Great Depression had a significant impact. Political events did have a significant effect depending on what the policy was directed at. For instance, some policies, including the Immigration Reform Act and Control Act of 1986 (IRCA), may have only affected border enforcement but not visa issuances and naturalization.

It is interesting to discover that the immigration-related policies, not only in the U.S. but also in sending countries, play a critical role in the flow of immigration. Hierarchical clustering algorithm was adopted to group these sending countries and to help better understand the dynamics of immigrants from different places. Both shape-based distance (SBD) and dynamic time warping (DTW) were used to measure the distance of time series data between countries because of their higher degree of accuracy when compared to other measures. As a result, two clusters were created for easier interpretation. After all the parameters were adjusted, the various clustering methods still resulted in similar prototypes. This study successfully provided in-depth analysis on political and economic events in time series data and used time series clustering to understand sending countries immigration history and behavior.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Dr. Xialu Liu, who has invested her time and support throughout this entire process. Her consistent encouragement and guidance over the past year helped set this thesis in the right direction. She always made herself available whenever there were any concerns or when I felt overwhelmed.

I would also like to express my gratitude to the rest of my thesis committee, Dr. Aaron Elkins and Dr. Benjamin Gonzales O'Brien, who has contributed insightful recommendations from my thesis proposal defense. These recommendations provided me with confidence in my analysis and results.

Lastly, I want to give credit to my fiancé, Memphis Huff V, who proofread my thesis. He made sure that I did not overcomplicate my analysis and that I was clear and succinct with my conclusion.

CHAPTER 1

INTRODUCTION

The history of mankind is the history of migration (Rystad, 1992). Many studies have delved into its anthropology and analyzed its outcomes and trends based on varying immigration policies. The melting pot of the U.S., for example, is a good study case for its rich immigration history. Sequeira et al. (2019) reviews the long-term effects of immigration by focusing on the Age of Mass Migration (1850-1913), a time period with the largest wave of immigration in U.S. history. Other studies focus more on immigration policy, such as Meyers's (2000) comparative analysis on different immigration policy approaches. This report investigates the history and policy making by performing a time series analysis on U.S. immigration variables to determine if U.S. immigration follows both a domestic and globalization approach.

Meyers (2000) defines the domestic politics model as a neutral arena for societal interests. This means interest groups and parties can either collaborate for a compromise between different ideals or lobby for their policy proposal to be passed. In this “society-centered” approach, policies are meant to represent the general public where socioeconomic factors are identified as the cause for shaping immigration policy (Meyers, 2000). This links to globalization where international economic and social drivers also influence policy making. This means the mobility of workers, capital, and ideas are welcomed especially from certain states that are deemed favorable. With the understanding of domestic politics and globalization, the goal is to detect trends and outliers from immigration data to monitor the effects of newly implemented immigration policies and of economic events. From monitoring these trends and outliers, immigration flow can be better managed in the U.S.

To accomplish this goal, RStudio software is used to run statistical computation with the R programming language. Various packages are uploaded to perform intervention event analysis on the number of lawful permanent residence (LPR) status granted each year and the number of annual and monthly apprehensions. The LPR status is used to quantify the flow of legal immigration, and the apprehension count is treated as a proxy for the flow of undocumented immigration given that the total number of undocumented border crossings is not available. There are models that estimates the unauthorized flow into the U.S., which vary according to the methodology adopted. Hence, in this study, we use LPR status and apprehension data which are available and accessible to measure the flow of legal and illegal immigration, respectively.

Next, a clustering analysis is conducted on sending countries to determine their behaviors and to find similarities and dissimilarities with each other. The aim is to perform hierarchical clustering on apprehension, LPR, nonimmigrant, and naturalization time series data using shape-based distance (SBD) and dynamic time warping (DTW) algorithms on RStudio. These algorithms are selected because they have similar performance scores, and both outperform other models (Paparrizos & Gravano, 2015). The biggest difference between DTW and SBD is their computational cost (Sardá-Espinosa, 2019). SBD may be a preferable model to use since it has a low computation cost (Sardá-Espinosa, 2019). This is critical when working with sizable data.

This study accomplishes the goal described earlier by answering the following research questions:

1. What political and economic factors significantly impact U.S. immigration?
2. How does the implementation of U.S. immigration policies and the occurrence of economic events change immigration in the short term and in the long term?
3. What are significant emigration trends from sending countries?

This study uniquely contributes to current U.S. immigration research in the following ways. First, it provides a comparative analysis between legal and illegal U.S. immigration flow. Past studies only show temporal comparisons between immigration visa types and between skilled and unskilled immigrants (Abramitzky & Boustan, 2017; Beine et al., 2016). Secondly, statistical models and intervention analysis are utilized to study the impact of political and economic factors on immigration. Most papers only focus on a qualitative analysis where the conclusion is often an agreement or disagreement that social and economic drivers affect immigration. This study not only quantifies the impacts of these

factors but also determines its longevity. Lastly, a clustering analysis of sending countries immigration behavior to the U.S. is provided for potential research.

CHAPTER 2

LITERATURE REVIEW

2.1 Importance of Immigration

Immigrants are an integral part of the U.S. economy. They make up 17.4 percent of the labor force as of 2019 and reflect a large share of workers that are important in many industries (U.S. Bureau of Labor Statistics, & U.S. Department of Labor, 2020). This is heavily prominent in jobs that do not require college degrees and may be the least enticing to work in (Sherman et al., 2019). Based on Figure 2.1, service and production jobs have more immigrants in the workforce than U.S. citizens (U.S. Bureau of Labor Statistics, & U.S. Department of Labor, 2020). Immigrants are also more willing to relocate due to labor shortages in the market (Sherman et al., 2019). They help improve the labor market efficiency by filling jobs quickly that could have remained stagnant. Immigrants save the economy roughly five to ten billion dollars annually (Sherman et al., 2019).

Another key contribution that immigrants provide to the economy is that they help support the aging U.S. population (Sherman et al., 2019). Currently, the U.S. population growth rate is at an all-time low of 0.4 percent and expected to decrease every year (Sherman et al., 2019). This can lead to a decline in the workforce and eventually the economy. The immigrant population helps bolster the economy despite the declining population growth. In a 2015 consensus study report from the National Academy of Sciences (NAS), the importance of immigrants in the workforce is emphasized:

> The high employment levels for the least educated immigrants indicates that employer demand for low-skilled labor remains high... owing to the aging of Baby Boomers; higher educational attainment among the U.S.-born; and a fertility rate below the replacement rate for the U.S.-born. (National Academies of Sciences, Engineering, and Medicine et al., 2016)

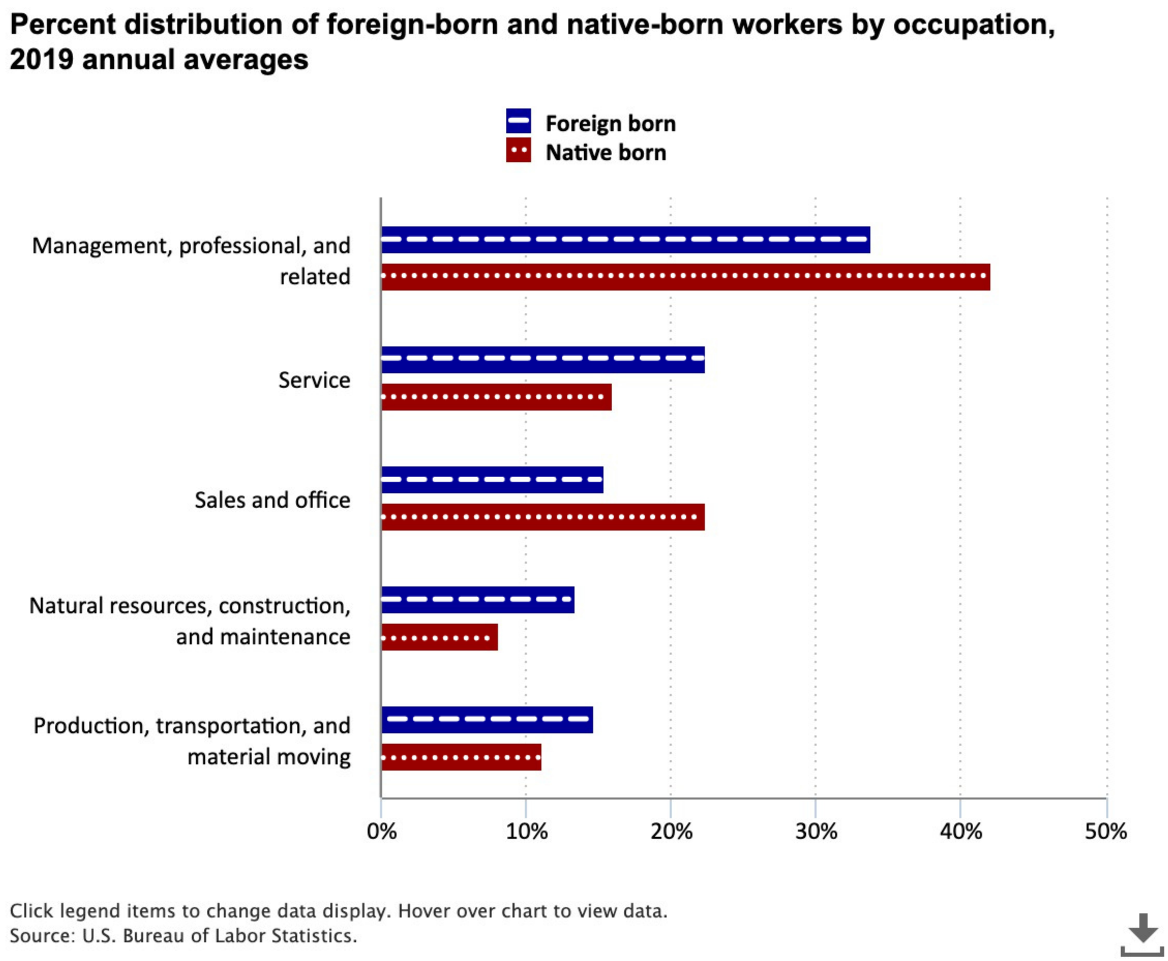


Figure 2.1. Percent distribution of foreign-born and native-born by occupation.

In addition to immigrants providing economic stability, their children can also make the same case. A 2005 study, which analyzed the results of the 1965 Immigration Reform Act, showed that first-generation immigrants were able to attain more education than those whose parents were born in the U.S. (Card, 2005). Since education is much more valued in today's labor market, it is important for them not only to surpass their parents' education level but make a large stride to keep up with society (National Academies of Sciences, Engineering, and Medicine et al., 2016).

2.2 Reasons for Immigrating

People have different reasons for immigrating to the U.S. that can be identified from migration trends. In Figure 2.2, a simplified decision tree shows the three main drivers for migration (Castelli, 2018). These three drivers are macro-, meso-, and micro-factors (Castelli, 2018).

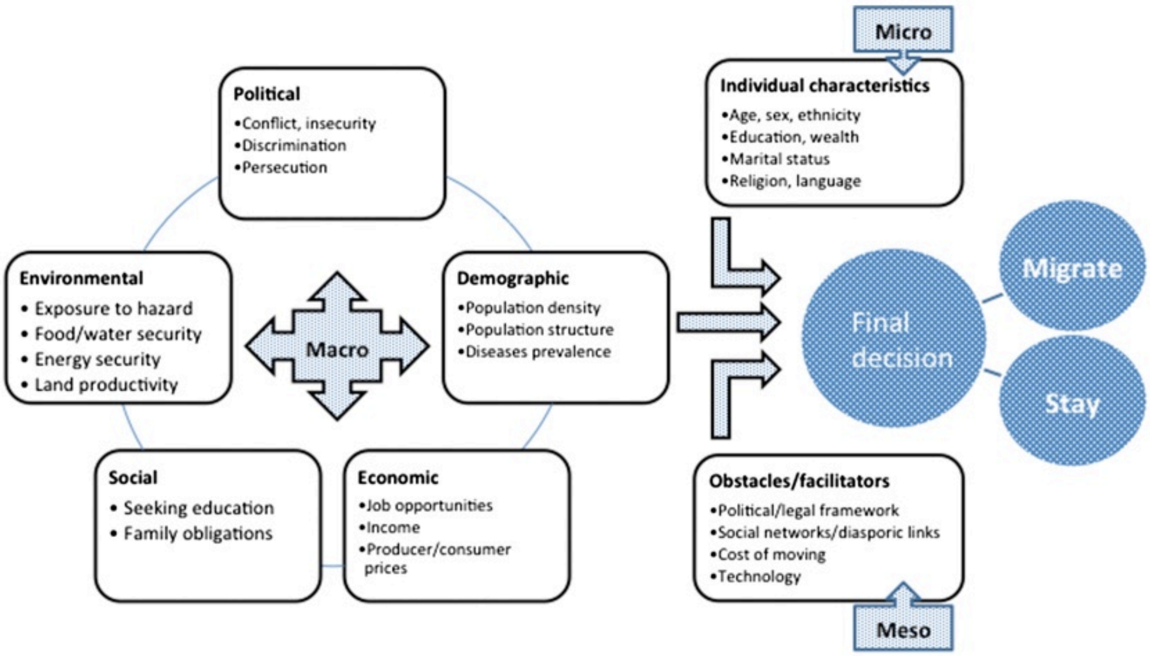


Figure 2.2. Complex drivers of migration.

Macro-factors are major events that can be noticed easily in migration trends and are independent of individual cases (Castelli, 2018). However, other drivers of migration may not be as obvious. For meso-factors, it is more closely related to the individual but still may be out of their control. Examples of this include links to an ethnic group or religious community, land acquisition, advancement of communication technology, and diasporic ties (Castelli, 2018). As for micro-factors, these are all due to the individual's characteristics and attitude (Castelli, 2018). For example, migrating because of a change in marital status can be factors in making the final decision. In this study, political and economic macro-factors are being analyzed.

2.3 Major Events in U.S. Immigration History

Figure 2.3 chronologizes the events in U.S. immigration history that are selected for intervention analysis. These events are further explained in the following sections.

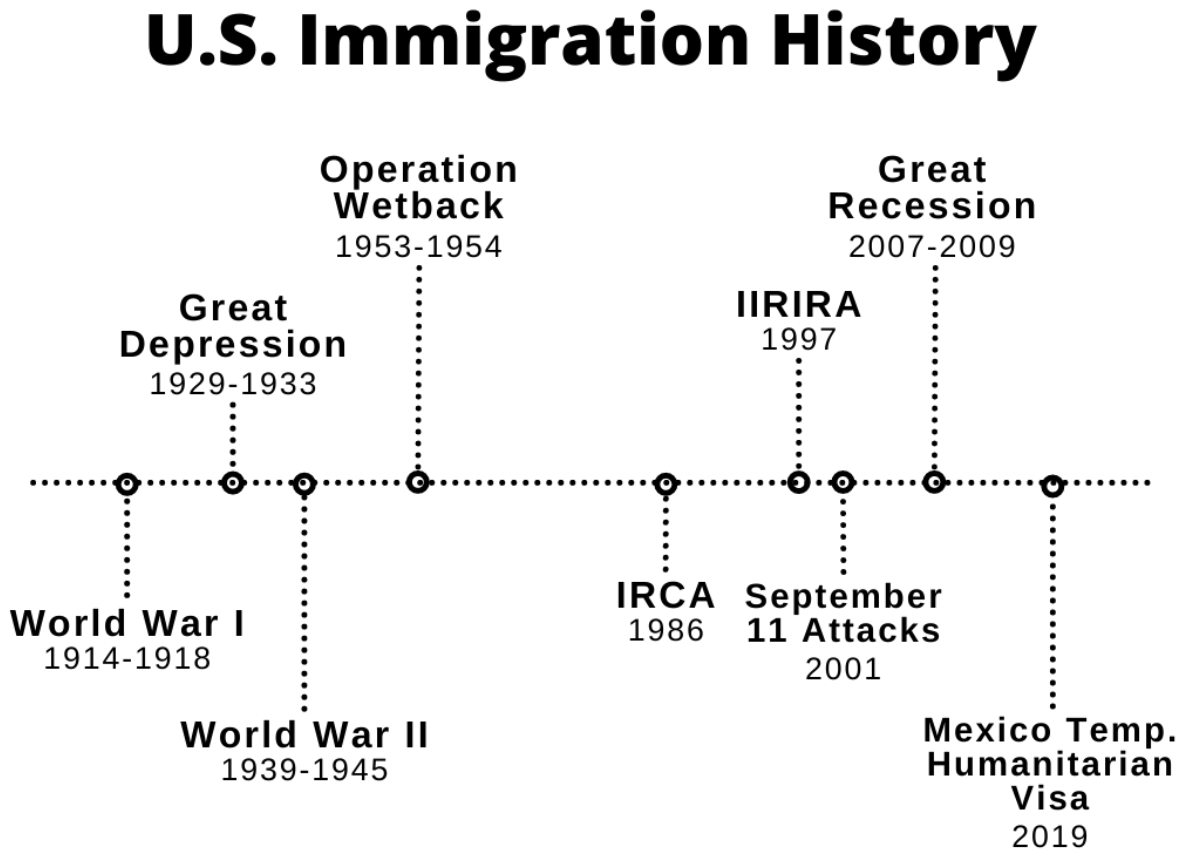


Figure 2.3. Intervention events timeline.

2.3.1 Political Events

2.3.1.1 World War I & World War II

The outbreak of World War I (WWI) was the start of an important period in international migration (Rystad, 1992). Since there were mass mobilizations of individuals due to the war, the introduction of systematic immigration controls, such as passports, were established to prevent spies from entering the country (Rystad, 1992). As anti-immigrant sentiment grew after WWI, the entry of certain types of immigrants were barred (Donato & Amuedo-Dorantes, 2020). The Asiatic Barred Zone Act of 1917, for instance, excluded the Chinese and immigrants from the Middle East, Southeast Asia, and India from U.S. entry (Donato & Amuedo-Dorantes, 2020). This legislation even imposed a literacy test for all incoming immigrants (Donato & Amuedo-Dorantes, 2020). After the end of WWI, these immigration controls remained (Rystad, 1992).

Exclusionary immigration policies largely continued in World War II (WWII) with even more complexity (Donato & Amuedo-Dorantes, 2020). The Alien Registration Act of 1941 required registration information and fingerprints of all incoming immigrants (Donato & Amuedo-Dorantes, 2020). It prevented citizenship for Korean and Japanese women and all non-Chinese Southeast Asians (Donato & Amuedo-Dorantes, 2020). However, it allowed naturalization for Filipinos, Indians, and Chinese wives of U.S. citizens (Donato & Amuedo-Dorantes, 2020). The Bracero Program also began around this time due to agricultural labor shortages which opened the door for Mexican migrants to work (refer to Section 2.3.1.2) (Donato & Amuedo-Dorantes, 2020).

2.3.1.2 Bracero Program & Operation Wetback

In the last half of the 20th century, Mexican immigrants were framed as a "threat", but certain policies prevented them from migrating and yielded unfavorable outcomes (Massey, 2020). One example is the Bracero Program that began in 1942 (Massey, 2020). The Bracero Program was an agreement between the U.S. and Mexico that issued temporary work to Mexican migrants (Massey, 2020). During 1942, World War II ensued and labor shortages in agriculture rose (Massey, 2020). For that reason, the program began with a small number of migrant workers. However, the next decade showed a push and pull between the need for migrant workers and appeasing the public's anti-immigration sentiment. A cutback on the program was always followed with an increase. By 1953 a militarization effort, called Operation Wetback, was implemented. It removed over two million Mexican immigrants from the U.S. (Massey, 2020). In some cases, even U.S.-born Mexican Americans were mistakenly deported as well (Massey, 2020).

After Operation Wetback ended in 1954, a steady decline in undocumented immigration continued until 1964 where the Bracero Program finally terminated (Massey, 2020). A Border Patrol official predicted that the end to the Bracero Program would lead to an increased number in undocumented immigrants crossing the U.S. border. This official's prediction came true especially after the U.S. Congress pushed the Immigration and Nationality Act of 1965 (Massey, 2020). Fewer opportunities for these migrant workers meant a greater urgency to move to the U.S. and permanently settle. Eventually this influx of migration steadied and cycled back down.

Figure 2.4 highlights the history of Mexican immigration to the U.S. (Massey, 2020). The chart provides a rough view of the temporal changes in border enforcement (Massey, 2020). For instance, the number of undocumented migrants from 1950 to 1953 indicate the political climax of Operation Wetback. However, the large spike may only suggest the increase in border enforcement. The entry for undocumented immigrants is a proxy from dividing the number of border apprehensions by the number of U.S. Border Patrol officers (Massey, 2020). The number of temporary workers around the mid-1940s to 1965 show the length of the Bracero Program until its end. After 1965, there were very few temporary workers and undocumented migrants. However, the 1970s begin to show a steady increase in the number of undocumented migrants to the U.S. again.

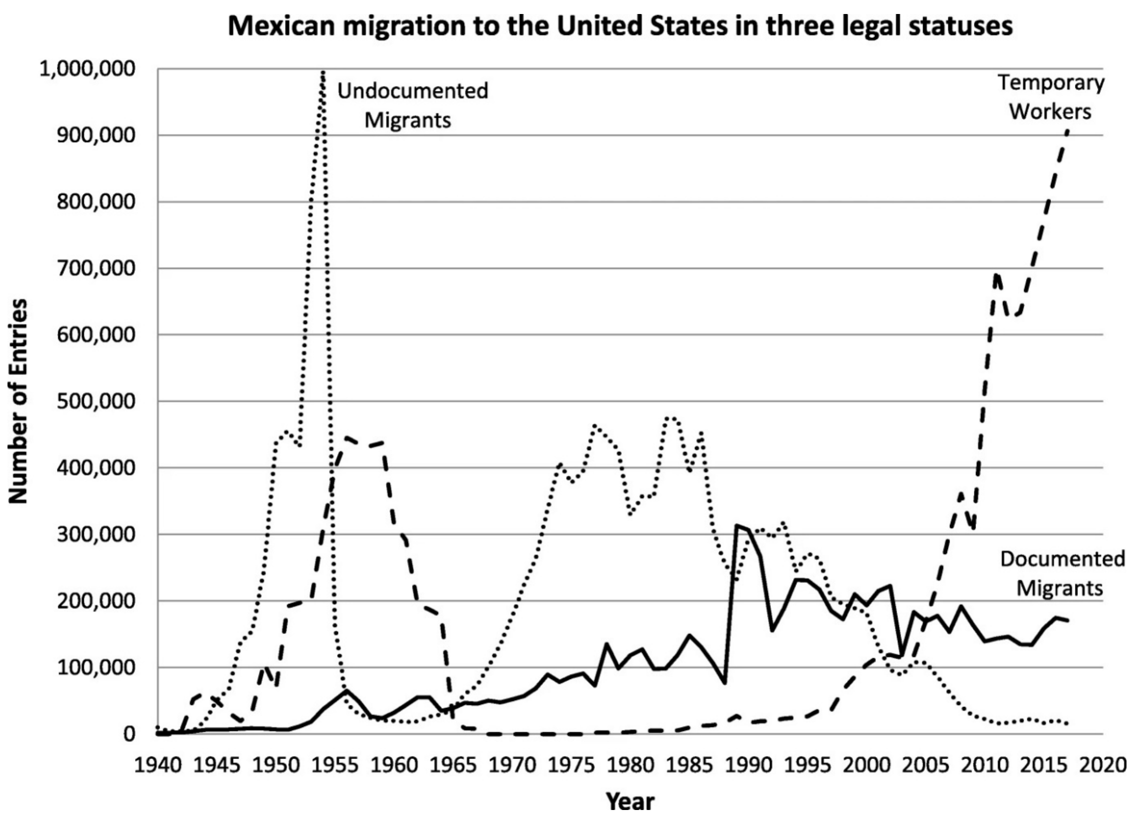


Figure 2.4. Mexican migration.

2.3.1.3 Immigration Reform and Control Act of 1986 (IRCA)

In response to the increased number of undocumented immigrants in the U.S. after the Bracero Program was dismantled in 1964, the Immigration Reform and Control Act (IRCA) was passed in 1986 (Phillips & Massey, 1999). The goal of IRCA was to control undocumented immigration in three ways (Phillips & Massey, 1999). First, it eliminated the attraction of U.S. jobs by imposing sanctions on employers who knowingly hire undocumented migrants (Phillips & Massey, 1999). Second, it discouraged people from entering the U.S. illegally by increasing funding to the U.S. Border Patrol which would thus increase border enforcement (Phillips & Massey, 1999). Lastly, it authorized amnesty for undocumented migrants who have already resided in the U.S. since 1982 (Phillips & Massey, 1999). IRCA also specialized legalizing undocumented farm workers who worked at least 90 days during the past year preceding May 1986 (Phillips & Massey, 1999). By 1990, there were two legalization programs that granted documentation to more than three million people: Legally Authorized Workers (LAWs) and Special Agricultural Workers (SAWs) (Phillips & Massey, 1999).

In Philips and Massey's (1999) paper, they observed the labor market’s consequence of IRCA using a rich cross section of data that relates the wages of workers with a variety of characteristics. The authors not only wanted to confirm if undocumented workers had a significant wage decrease after IRCA was passed, but also to locate the reason for this change. Overall, the results showed IRCA had no effect on lowering wages (Phillips & Massey, 1999). Before IRCA, wages were determined by human capital (e.g., education, English literacy, years of experience in the U.S.). However, after IRCA was passed, social capital (e.g., networking and relationships in the U.S.) became the main determinant for wages (Phillips & Massey, 1999).

2.3.1.4 Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (IIRIRA)

Abrego et al. (2017) critically reviews immigrant criminalization described in previous literature and traces back specific laws that link undocumented immigrants with criminality. One of these laws is the 1996 Illegal Immigration Reform and Immigration

Responsibility Act (IIRIRA). The IIRIRA reclassified undocumented immigrants as deportable and/or admissible and curtailed immigrants' rights for due process. It set forth provisions like the 287(g) program that enlisted local law enforcement agencies to participate in the enforcement of immigration law (Abrego et al., 2017). Thus, these agencies can make immigration-related arrests and take immigrants into custody (Abrego et al., 2017).

2.3.1.5 SEPTEMBER 11 ATTACKS

Another pivotal moment in immigration history is the September 11 terrorist attacks in 2001. An Islamic extremist group called al Qaeda hijacked four planes where two flown into the World Trade Center in New York City, one hit the Pentagon, and the last one crashed just outside Washington, D.C. (History.com Editors, 2021). Almost 3,000 people were killed and many more were injured (History.com Editors, 2021). This incident remains to be the deadliest terrorist attack in human history which led to the War on Terror and the enactment of the USA Patriot Act soon after (Panetta, 2005). The USA Patriot Act expanded the power of the Immigration and Naturalization Service (INS) to detain suspected terrorist (Panetta, 2005). The act also handed power over to the attorney general to detain aliens without charge or through the legal process (Panetta, 2005). Next, Congress passed the Homeland Security Act in 2002 to disband the INS and created three new federal agencies under the Department of Homeland Security (DHS) (U.S. Citizenship and Immigration Services, 2019):

- Customs and Border Protection (CBP): Monitors trade and people coming in and out of the U.S.
- Immigration and Customs Enforcement (ICE): Enforces immigration policies and combats transnational crime.
- U.S. Citizenship and Immigration Services (USCIS): Oversees the lawful immigration and naturalization process.

2.3.1.6 MEXICO TEMPORARY HUMANITARIAN VISA

Around late 2018 to 2019, caravans of migrants from Central America transmigrated through Mexico with the intention of crossing the U.S.-Mexico border (Marchand, 2020). By January 2019, over 15,000 Mexico humanitarian visas were handed out at the southern border (Figure 2.5) (Marchand, 2020). Marchand (2020) mentions the reason why these visas were handed out was to keep track of the inflow and outflow of migrants. The humanitarian visas gave visa holders access to the country for one year and allowed them to work (usually with wages less than Mexican workers) (Marchand, 2020). It gave the migrants an alternative to remain in Mexico although crossing the border to the U.S was still the better opportunity. In June 2019, the sheer number of migrants reaching the U.S.-Mexico border could not be ignored (Marchand, 2020). President Donald Trump threatened to increase tariffs on Mexican exports if they did not take drastic measures in preventing migrants from reaching the U.S. border (Marchand, 2020). With negotiations between Mexican and American diplomats already ongoing at that time, the Mexican government agreed to better protect its border (Marchand, 2020). As a result, militarization increased at the southern Mexican border and the number of humanitarian visas issuances began to decrease.

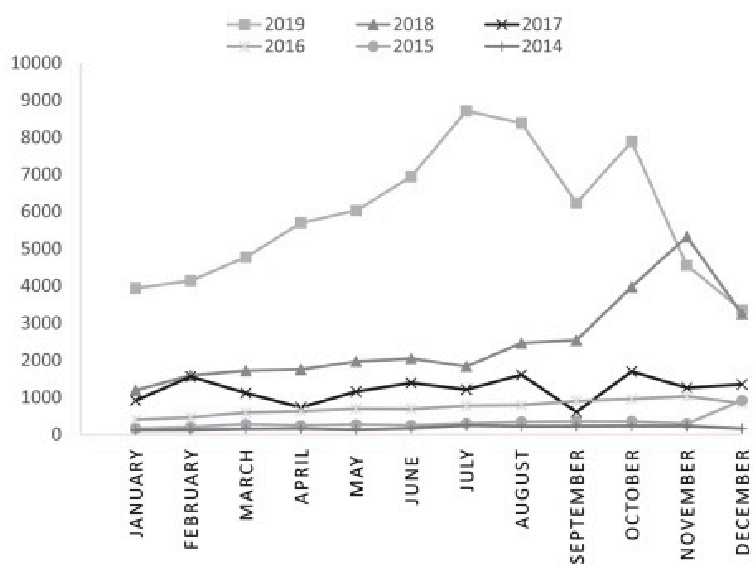


Figure 1. Monthly evolution of asylum claimants in Mexico (2014–2019).
Source: Comisión Mexicana de Ayuda a Refugiados (2020) *Estadísticas de solicitantes de la condición de refugiado en México* (Blog). March. Available from https://www.gob.mx/cms/uploads/attachment/file/544676/CIERRE_DE_MARZO_2020__1-abril-2020_-2__1_.pdf. Unidad de Política Migratoria y Comisión Mexicana de Ayuda a Refugiados (2014–2018). *Boletín Estadístico de Solicitantes de Refugio en México.* Mexico, D.F.: Secretaría de Gobernación. http://portales.segob.gob.mx/es/PoliticaMigratoria/Refugio.

Figure 2.5. Issuance of Mexico temporary humanitarian visas (2014-2019).

2.3.2 Economic Events

Global events like economic recessions have been one of the main drivers in immigration policy making. For instance, the Great Depression indirectly set forth the merging of the Bureau of Immigration and the Bureau of Naturalization into the Immigration and Naturalization Services (INS) with the Department of Labor (Baxter & Nowrasteh, 2021). This allowed the Secretary of Labor William N. Doak to deport illegal immigrants to free up jobs for natives (Baxter & Nowrasteh, 2021). More than one million Mexicans and U.S.-born Mexican Americans were deported as a result. However, the policy goals were not met, and the return efforts increased unemployment rates (Baxter & Nowrasteh, 2021). Understanding why certain immigration policies are created due to economic events can prevent unexpected outcomes.

Another significant economic recession is the Great Recession from 2007 to 2009 (Kalleberg & Wachter, 2017). It was the largest and longest economic downturn since the Great Depression (Kalleberg & Wachter, 2017). Often the two events are compared due to its prolonged effects on the economy. However, the Great Recession did not have as big of a global impact as the Great Depression. One of the main suspects for the Great Recession was the bursting of the housing bubble paired with the collapse of the subprime mortgage market (Kalleberg & Wachter, 2017). Other reasons for the recession include (Kalleberg & Wachter, 2017):

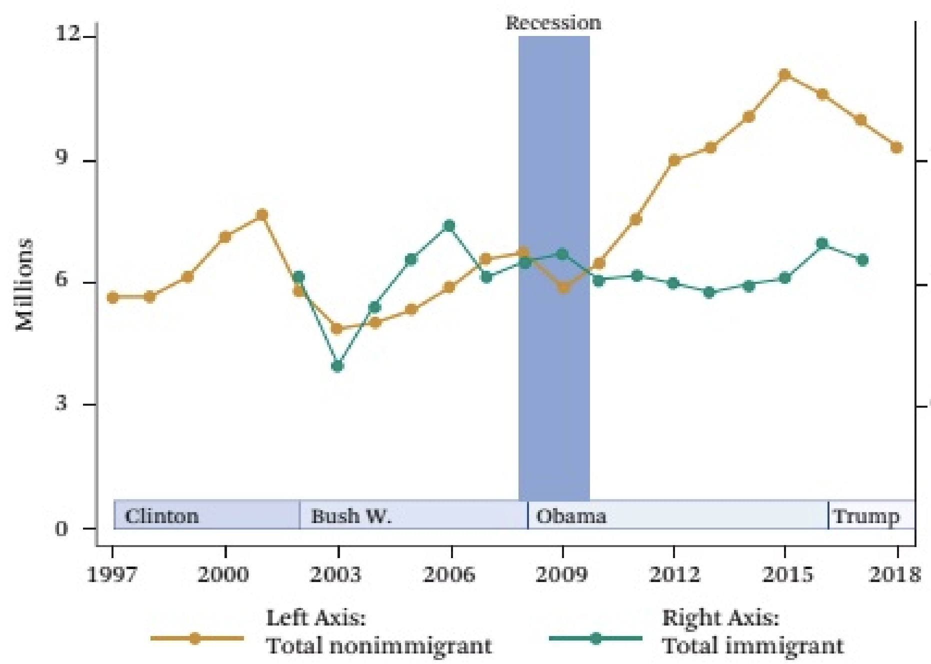
- Mismatch of the labor market
- Low incentives to work
- Uncertainty of economic events and public policy

2.3.3 Previous Studies

Past studies have investigated political and economic effects on U.S. immigration. In some cases, authors only focus on certain major migration eras. For instance, in Abramitzky and Boustan's (2017) essay, one period of interest is the Age of Mass Migration (1850-1913) where there was unrestricted migration to the U.S. from Europe. Another recent period of interest was a more constrained migration from Asia and Latin America (1965-present). In their study, they confirmed that migration selection changes over time. The expanding income gap between the U.S. and certain sending countries provided a monetary exchange for immigrants with valuable skills (Abramitzky & Boustan, 2017). The evolving selection process may also explain the changing immigration policy and the rising migration costs. Abramitzky and Boustan (2017) also looked over the economic effects of immigration where evidence shows that it reduces the wages of some natives, but overall, it has a net positive effect on the U.S. economy.

Unlike Abramitzky and Boustan (2017), Donato and Amuedo-Dorantes (2020) delves into five different periods that covers more than a century of immigration history (Donato & Amuedo-Dorantes, 2020). Each period focused not only immigration reform that was passed by Congress but also broad executive actions from presidents. They used visa trends to map shifts in governance and influence by highlighting the presidential terms and adding a bar on 2009 to signify the Great Recession (Figure 2.6) (Donato & Amuedo-Dorantes, 2020). These indicators may provide an understanding on what factors contribute to a shift in visa issuances.

Figure 1. Nonimmigrant and Immigrant Visas, 1997–2018



Source: U.S. Department of Homeland Security 2018; U.S. Department of State 2019.

Figure 2.6. Nonimmigrant and immigrant Visa trends.

Donato and Amuedo-Dorantes (2020) compared not only nonimmigrant (temporary) and immigrant (LPR) visa distribution but also between different temporary nonimmigrant employment visas. Figure 2.7 shows a plot where high-skilled workers are preferred over low-skilled workers since they have more issuances throughout the years (Donato & Amuedo-Dorantes, 2020). In Beine et al. (2016), the authors also compared high-skilled and low-skilled immigrants by using preliminary data from International Migration Law and Policy Analysis (IMPALA) that measures a variety of immigration regulations (Beine et al., 2016). The IMPALA database answers questions for high-skilled and low-skilled entry tracks where their coded stringency scores are calculated based on a set of selected and relevant questions. The results showed more restrictions for low-skilled migrants, variation in stringency scores between countries, and contrasting temporal trends between high-skilled and low-skilled immigrant workers. Observing these differences between high-skilled and low-skilled immigrants reveals regulatory complexity and the status of the U.S. economy from its labor market needs.

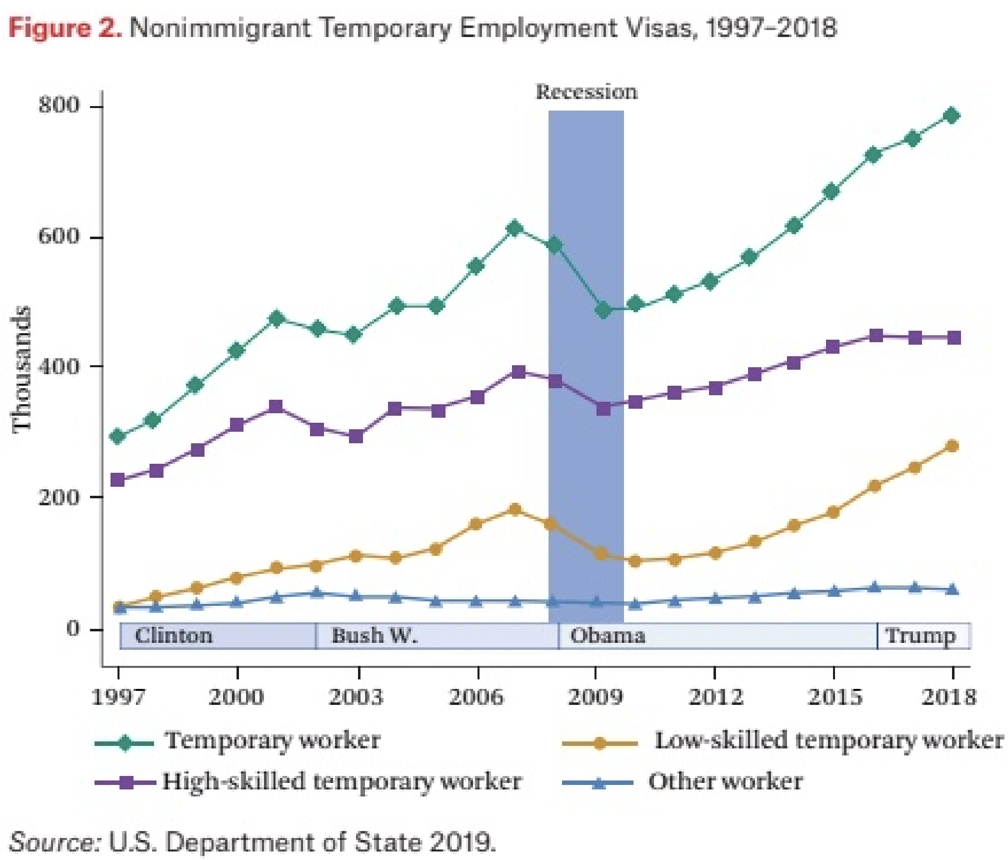


Figure 2.7. Nonimmigrant temporary employment Visa trends.

2.3.4 Mismatch in Policy-Making

The goal of immigration policies is "to manage the volume, origin, direction, and composition of immigration flows" (Department of Economic and Social Affairs, 2013). However, these policies through time have shown to have moral consequences and complexity. Czaika and de Haas (2013) outlined the conceptual framework of migration policy effects and effectiveness. Migration policy flows in four levels: public policy discourse, migration policies on paper, policy implementation, and migration outcomes. Between these levels there are gaps that can explain a policy's failure: discursive gap, implementation gap, and efficacy gap (Czaika & de Haas, 2013). Discursive gap is the discrepancy between public discourse and actual migration policy (Czaika & de Haas, 2013). This gap should not automatically be the reason for immigration policy failure since it is not necessarily concrete policy formation when it is only being discussed. However, the other two gaps can be analyzed further for policy failure. Implementation gap is the discrepancy between policies on paper and their actual implementation, and efficacy gap reflects the degree to which the implemented immigration policies have on migration flows (Czaika & de Haas, 2013). These policy levels and gaps can be seen in Figure 2.8 (Meyers, 2000).

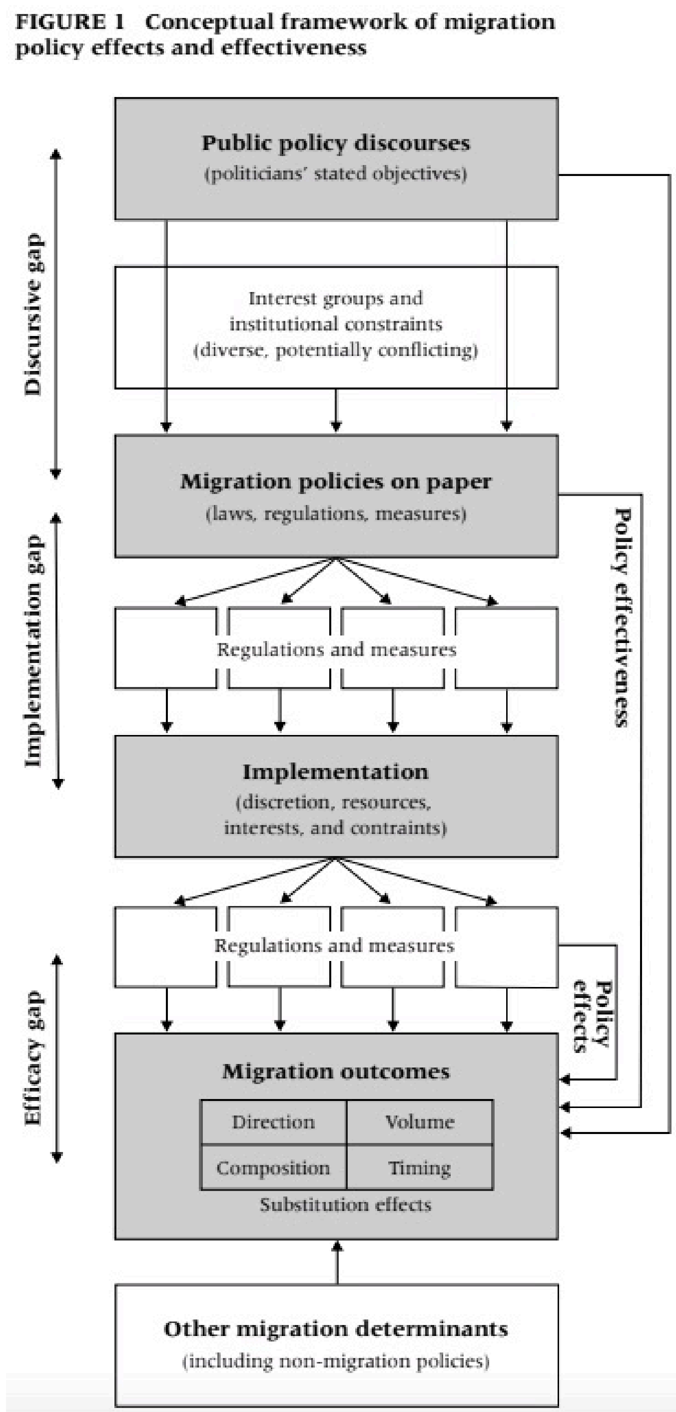


Figure 2.8. Identifying policy gaps.

Other than understanding the gaps in policy making, comparing and contrasting different immigration policy approaches can help identify current mismatches. These hurdles can be avoided and provide a more efficient process in immigration policy making. Meyers (2000) describes six different approaches: Marxism, realism, liberalism, the "national identity" approach, domestic politics, and internationalism. In this study, a domestic politics and globalization approach is further analyzed.

- The domestic politics approach is where a state has a neutral arena for social interests (Meyers, 2000). The critique on this approach is that it lacks general theory and only focuses on one country at a time (Meyers, 2000). It also does not consider refugee policies (Meyers, 2000).
- Globalization is subset under liberalism that challenges stability and territoriality as well as its capacity to control economic and welfare policies (Meyers, 2000). The issue with this approach is that it only depends on these outside forces while not considering the internal politics (Meyers, 2000).

CHAPTER 3

DATA

Listed below are the datasets used for intervention event analysis and for time series clustering. Note that data series is based on the U.S. federal fiscal year. This means the fiscal year starts on October 1 of the previous year and runs through to September 30 of the selected year. For instance, the 2022 fiscal year begins on October 1, 2021, to September 30, 2022.

3.1 INTERVENTION EVENT ANALYSIS

3.1.1 Monthly Apprehensions

The total monthly apprehensions come from the *U.S. Border Patrol Monthly Encounters Report* and ranges from the 2000 to 2020 fiscal year (October 1999 to September 2020) (U.S. Border Patrol, 2021). The report also contains the monthly apprehensions for each sector and border region. Apprehension is defined by the DHS as “the arrest of a removable alien” (Office of Immigration Statistics, 2018).

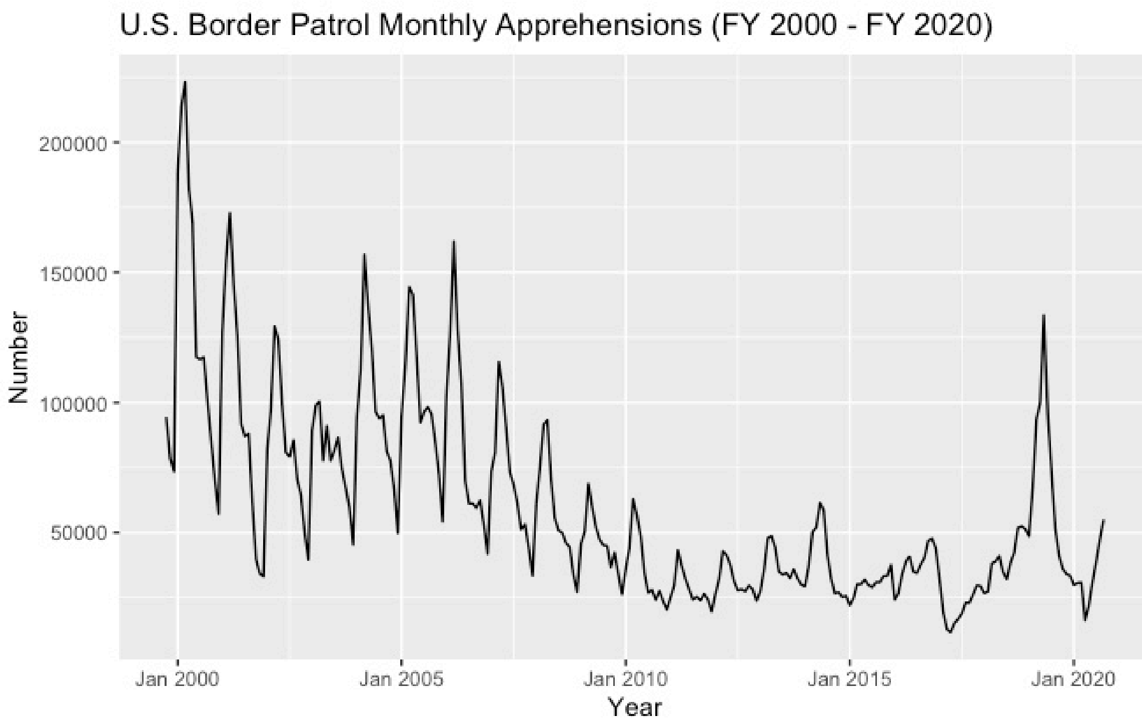


Figure 3.1. Monthly apprehension time series plot.

3.1.2 Annual Apprehensions

The annual apprehension count comes from the *Yearbook of Immigration Statistics 2020* and spans from the 1925 to 2020 fiscal year (Department of Homeland Security, 2022). The *Yearbook of Immigration Statistics* is a collection of variables such as the number of those granted asylum or refugee status to the number of undocumented immigrants that were removed or returned. An individual is counted again every time they are apprehended within the same fiscal year (Office of Immigration Statistics, 2018).

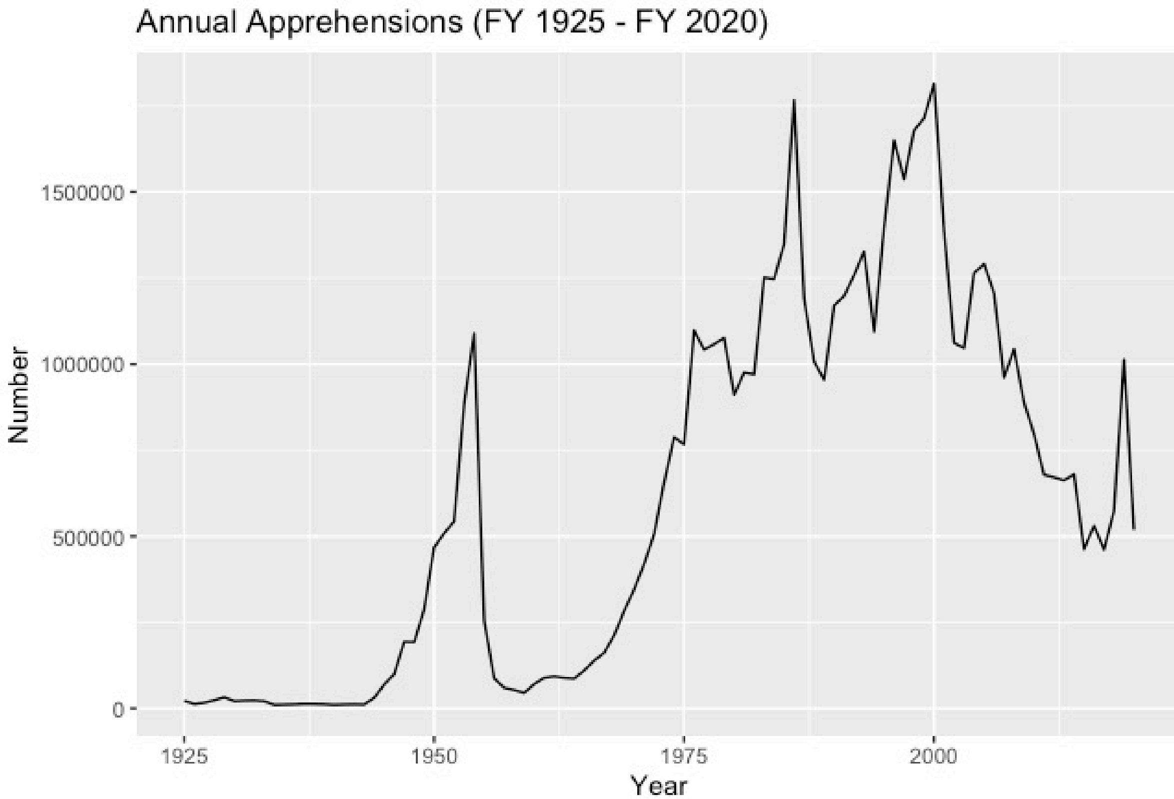


Figure 3.2. Annual apprehension time series plot.

3.1.3 Annual Lawful Permanent Resident (LPR) Status

The number of people that obtained lawful permanent resident (LPR) status for each fiscal year is taken from the *Yearbook of Immigration Statistics 2020* (Department of Homeland Security, 2022). The scope of the data ranges from 1820 to 2020. LPR, also known as "green card holders", are non-citizens authorized to live in the U.S. LPRs do not have the full rights of a U.S. citizen, like voting, but can work their way to become a naturalized citizen.

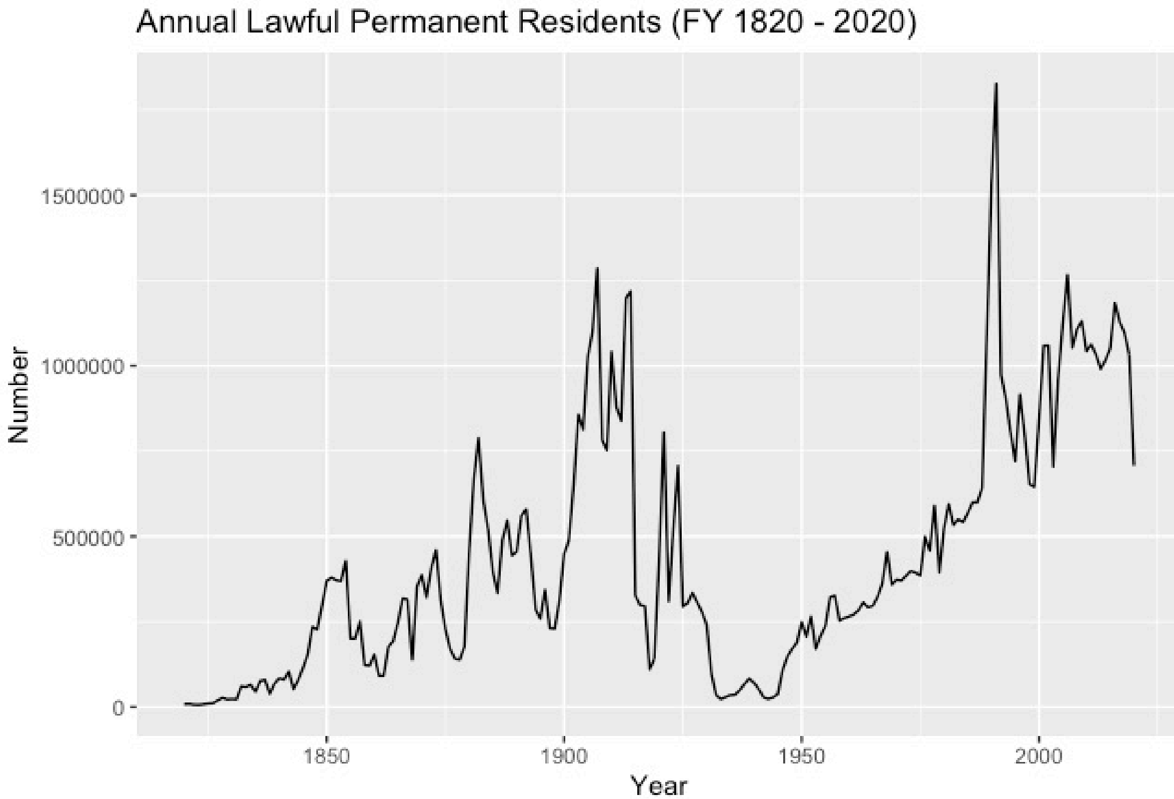


Figure 3.3. Annual lawful permanent resident status time series plot.

3.2 Time Series Clustering

The regions listed under these datasets are Africa, Asia, Europe, North America, Oceania, and South America. As for country of birth portion, the datasets have almost all the same countries. Note, the data also contain former countries (e.g., Czechoslovakia) labeled as "former" at the end of the name (e.g., Czechoslovakia_former).

3.2.1 Annual Apprehension by Region and Country of Birth

Just like the annual apprehension count, this dataset also comes from the *Yearbook of Immigration Statistics 2020*, but it ranges only from 2011 to 2020 fiscal year and is filtered by region and country of birth (Department of Homeland Security, 2022). The sum of all the region apprehensions and the sum of all the country apprehensions for each fiscal year should equal to the total apprehension count found in the annual apprehension dataset.

3.2.2 Annual Lawful Permanent Resident Status by Region and Country of Birth

This dataset also originates from the *Yearbook of Immigration Statistics 2020* (Department of Homeland Security, 2022). The LPRs is divided by region and country of birth and ranges from the 2011 to 2020 fiscal year (Department of Homeland Security, 2022). The sum of all the region LPRs and the sum of all the country LPRs for each fiscal year should equate to the total number of LPRs found in the annual LPR dataset.

3.2.3 Annual Naturalization by Region and Country of Birth

The annual naturalization dataset has the same structure as the annual LPR and annual apprehension datasets. It originates from the *Yearbook of Immigration Statistics 2020* and is divided up into regions and countries ranging from the 2011 to 2020 fiscal year (Department of Homeland Security, 2022). The naturalization of an immigrant is where they take an oath of allegiance to the U.S. and become citizens. Once immigrants are naturalized, they now have the same rights as citizens that were born in the U.S.

3.2.4 Annual Nonimmigrant Admission by Region and Country of Birth

A nonimmigrant is “an alien who seeks temporary entry to the United States for a specific purpose” (Office of Immigration Statistics, 2018). Temporary visas can range from student visas to various types of temporary work visas. The annual nonimmigrant dataset is an aggregate of all these issued temporary visas. Similar to the other time series clustering datasets, its source is the *Yearbook of Immigration Statistics 2020* (Department of Homeland Security, 2022). The dataset ranges from 2011 to 2020 and is divided by region and country of birth (Department of Homeland Security, 2022).

CHAPTER 4

METHODS

Preprocessing, modeling, and visualizations are performed on R (https://www.r-project.org/). The R code and data to perform the time series analysis is available on GitHub (https://github.com/ohkaaaaay/thesis-tsa).

4.1 INTERVENTION EVENT ANALYSIS

The R packages used for intervention event analysis are listed below:

- **dplyr**: Used for data manipulation.
- **forecast**: Contains time series functions imported from the stats package (e.g., autocorrelation function) and the BoxCox.lambda function to determine if a logarithm transformation is needed.
- **ggplot2**: Used for creating data visualizations.
- **lmtest**: Package was loaded for the coeftest function. It is used to determine what coefficients are significant in models.
- **TSA**: Provides the arimax function that can create ARIMA models by incorporating transfer functions.
- **tseries**: Used for the Augmented Dickey-Fuller Test function.
- **tsoutliers**: Detects outliers in a time series.
- **zoo**: Contains the yearmon function for indexing monthly data.

4.1.1 Data Preprocessing

The datasets from the *Yearbook of Immigration Statistics 2020* did not require much preprocessing since the CSV files were converted from Excel files provided for each table. Some data manipulation was performed on Excel by clearing the formatting and organizing the data sequentially under one column. Originally the data was divided into two columns, which would not load correctly in R. As for the *U.S. Border Patrol Monthly Encounters Report*, the PDF was scraped using the pdftools package. Further unnecessary columns and rows were removed using the tidyverse package since the scraping did not split the table evenly. Only the monthly total was selected, leaving out the sector and border region rows.

4.1.2 Exploratory Data Analysis

Before performing any intervention event analysis, the following descriptive statistics were calculated for each dataset:

- Mean
- Standard Deviation
- Maximum
- Minimum

Time series plots were then created to observe any trends and abnormalities (Shumway & Stoffer, 2017). If the dataset is found to be non-stationary, it will need to be transformed by applying the logarithm and then the difference (Shumway & Stoffer, 2017). The logarithm transformation shown below suppresses large fluctuations (Shumway & Stoffer, 2017):

$$y_t = \log x_t. \tag{4.1}$$

Differencing ensures the removal of any general trends (Shumway & Stoffer, 2017). Additional differencing may be needed based on the data series. Another method to determine non-stationarity is the Augmented Dickey-Fuller unit root test. If the null hypothesis (H_0) is not rejected, the series has at least one unit root indicating it is non-stationary. However, if the null hypothesis is rejected, the alternative hypothesis (H_a) determines the series has no unit root and is stationary. If the p-value is less than 0.05, H_0 is rejected, and the data is stationary. However, if the p-value is greater than 0.05, H_0 is not rejected, and the data will need to be differenced. As for data with seasonal trends, the dataset will need to be seasonally differenced meaning it will be differenced by its period. To better understand the seasonal trends in a dataset, a boxplot can be made for further analysis.

4.1.3 Models

Once the datasets are transformed, the autoregressive integrated moving average (ARIMA) model order can be determined. This procedure is selected over the Student's *t*-test because the observations before and after the event of interest do not behave independently (Box & Tiao, 1975). Since the data is a time series, observations are serially dependent on its past values (Box & Tiao, 1975). Here's the following steps to find the best fitting ARIMA model:

1. Plot the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to determine the autoregressive (AR) order "p" and the moving average (MA) order "q" (Shumway & Stoffer, 2017).
2. Create possible ARIMA models with a few selected preliminary values of p and q (Shumway & Stoffer, 2017). Then use Akaike information criterion (AIC) as a metric to determine the best fit (Shumway & Stoffer, 2017). The best model is one with the smallest AIC value (Shumway & Stoffer, 2017).
3. Perform model diagnostics to guarantee the model is the best fit.

For datasets with seasonality, steps 1 and 2 would need to be repeated twice to determine the seasonal order and the general order before moving on to step 3.

4.1.3.1 Autoregressive and Moving Average Models

Autoregressive (AR) models are a selected time point that can be explained by its past values in the following form where x_t is stationary, w_t is white noise, and ϕ_1, ϕ_2,...,ϕ_p are constants ($\phi_p \neq 0$) (Shumway & Stoffer, 2017):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t. \tag{4.2}$$

It can also be written with the backshift operator "B", shown in (4.3), or even simplified further in (4.4) (Shumway & Stoffer, 2017):

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) x_t = w_t, \tag{4.3}$$

$$\phi(B) x_t = w_t. \tag{4.4}$$

For MA models, it determines a selected time point based on past error values in the following form where w_t is white noise and θ_1, θ_2,...,θ_q ($\theta_q \neq 0$) are parameters (Shumway & Stoffer, 2017):

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_p w_{t-q}. \quad (4.5)$$

Unlike AR models, MA models is stationary for any values with the θ parameter (Shumway & Stoffer, 2017). The backshift operator "B" in (4.6) can be simplified further (4.7) (Shumway & Stoffer, 2017):

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \quad (4.6)$$

$$\phi(B) x_t = w_t. \quad (4.7)$$

4.1.3.2 Autocorrelation and Partial Autocorrelation Functions

An autocorrelation function (ACF) is essentially a correlogram that measures the direct and indirect correlation between a present value to a past value (Shumway & Stoffer, 2017). While partial autocorrelation function (PACF) is a conditional correlation between a present value and a past value (Shumway & Stoffer, 2017). This means that the direct correlation between the present value and past value is considered while the correlation for the other lags in between the present and selected past value are not (Shumway & Stoffer, 2017).

Table 4.1 Behavior of ACF and PACF for ARMA Models

Model	**ACF**	**PACF**
AR(p)	Tails off	Cuts off at lag p
MA(q)	Cuts off at lag q	Tails off

To determine the p and q orders, the ACF and PACF plots will need to be observed together. Table 4.1 describes the behavior of ACF and PACF for a particular order. To get a better idea on how to determine the model order based on the ACF and PACF plots, the following models in (4.8) and (4.9) are simulated as an example:

$$\text{AR(1): } x_t = 0.7x_{t-1} + w_t, \tag{4.8}$$

$$\text{MA(1): } x_t = w_t + 0.7w_{t-1}. \tag{4.9}$$

When plotting the time series along with their ACF and PACF plots (Figure 4.1), the AR(1) model is easily determined since the PACF plot cuts off at lag one while the ACF plot tails off. For the MA(1) plot (Figure 4.2), the ACF plot cuts off at lag one while the PACF plot tails off.

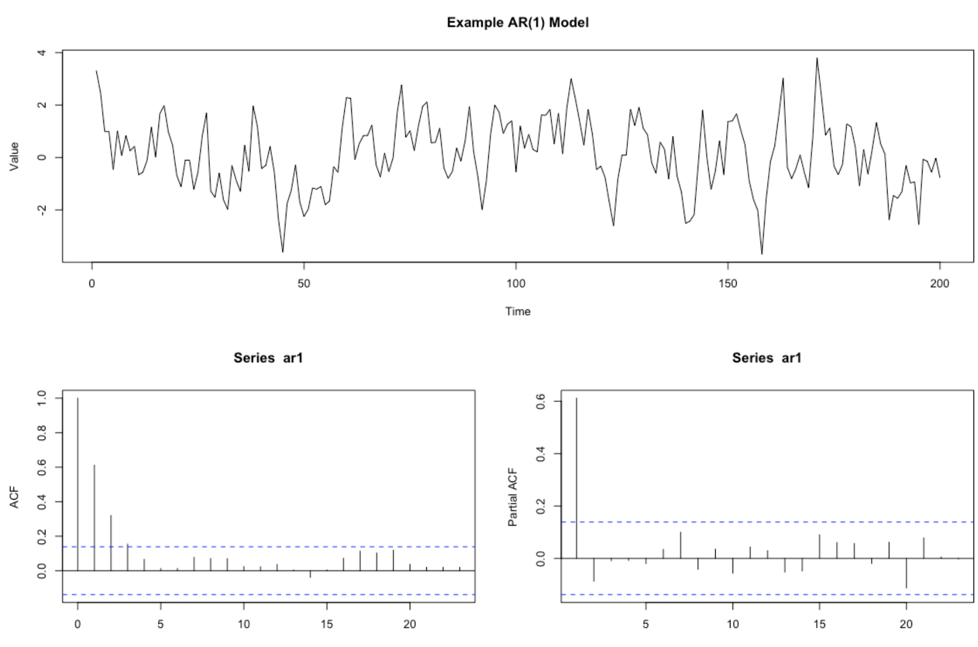


Figure 4.1. Example of AR(1) time series with its ACF and PACF plots.

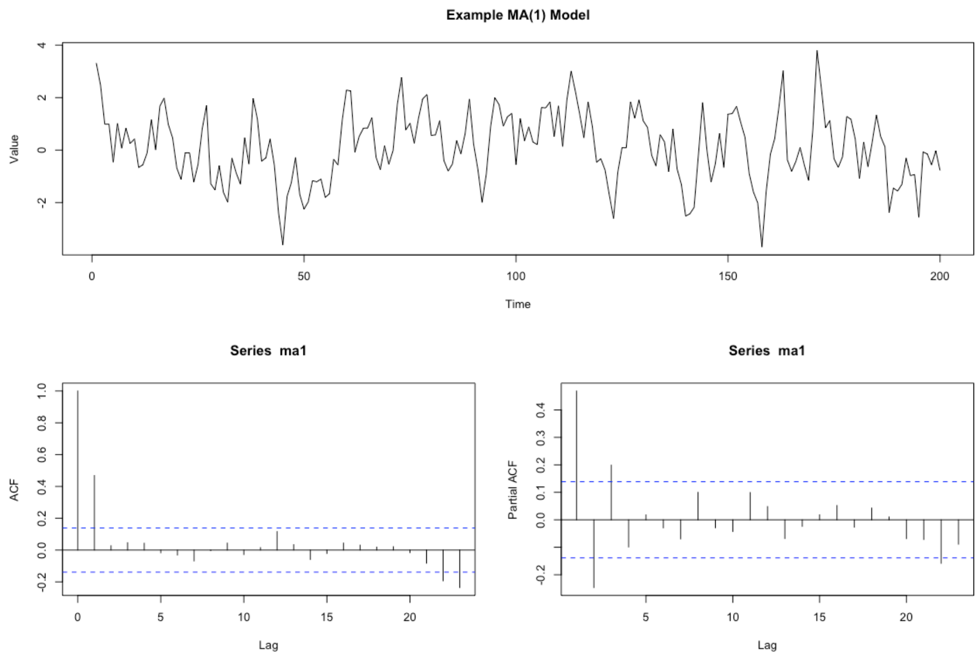


Figure 4.2. Example of MA(1) time series with its ACF and PACF plots.

4.1.3.3 Model Diagnostics

Before using an ARIMA model for further analysis, it needs to go through a few diagnostic tests.

1. **Residual Analysis**: The residuals (error) are analyzed to determine if they behave like white noise. A preliminary analysis is done by observing the ACF and PACF plots of the residuals. If the residuals appear to be stationary, then they are white noise. To guarantee the residuals are white noise, a Box-Ljung test is performed with the following null and alternative hypotheses listed below where ρ_i is the autocorrelations for each lag (Shumway & Stoffer, 2017):
    a. H_0: $\rho_1 = \rho_2 = \cdots = \rho_i = 0$
    b. H_a: At least one of the correlations is not zero.

If the p-value is less than 0.05, the null hypothesis is rejected. However, the residuals need to be white noise so p-value should be greater than 0.05.

2. **Check for Overspecification**: Remove coefficients from an ARIMA model that is not significant (Shumway & Stoffer, 2017). If this model has a smaller AIC, use this adjusted model instead.
3. **Check for Causality (Stationarity)**: Each root from the AR model should be outside the unit circle. Thus, the values should not be close to one (Shumway & Stoffer, 2017). If so, the dataset needs to be differenced further.

4.1.4 Step and Pulse Functions

Once the ARIMA passes the diagnostic tests, specific events explained in Section 2.3 can be analyzed. Table 4.2 summarizes the interventions events inspected.

Table 4.2 Intervention Events

Events	Time Point	Type
World War I	1915-1918	Political
Great Depression	1931-1938	Economic
World War II	1940-1945	Political
Operation Wetback	1953-1954	Political
IRCA	1986[1], 1989-1991[2]	Political
IIRIRA	1997	Political
September 11 Attacks	2001	Political
Great Recession	Dec 2007 – Jun 2009	Economic
Mexico Temp. Humanitarian Visa	Jan 2019 – Apr 2019	Political

[1]Time point analyzed on the annual apprehension dataset.
[2]Time point analyzed on the annual LPR dataset.

Next, the possible input (I_t) for each intervention event is listed below:

- **Step Indicator**: This is a permanent impact that occurs at time T and still in effect afterwards (Box & Tiao, 1975).

$$S_t^T \begin{cases} 0, & t < T; \\ 1, & t \geq T. \end{cases} \tag{4.10}$$

- **Pulse Indicator**: This is a temporary impact that occurs only at time T (Box & Tiao, 1975).

$$P_t^T \begin{cases} 0, & t \neq T; \\ 1, & t = T. \end{cases} \tag{4.11}$$

Based on the input, its effect can be measured in (4.12) (Box & Tiao, 1975):

$$E_t = \frac{\omega(B)}{\delta(B)} I_t, \tag{4.12}$$

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_i B_s, \tag{4.13}$$

$$\delta(B) = 1 - \delta_0 - \delta_1 B - \delta_2 B^2 - \cdots - \omega_i B_r. \tag{4.14}$$

The ω are the coefficients for the MA component and the δ are the coefficients for the AR component. From the input and the ARIMA model, the following responses can be estimated using maximum likelihood (Box & Tiao, 1975):

- Permanent
- Temporary
- Permanent & Temporary
- None

The best response model depends on finding smallest AIC value. Once the model is determined, the response can be plotted, as shown in Figure 4.3, and the percent change of the impact can be calculated from ω using (4.15) to better understand the event (Box & Tiao, 1975):

$$\%\ \text{Change} = (e^{\omega} - 1) \times 100. \tag{4.15}$$

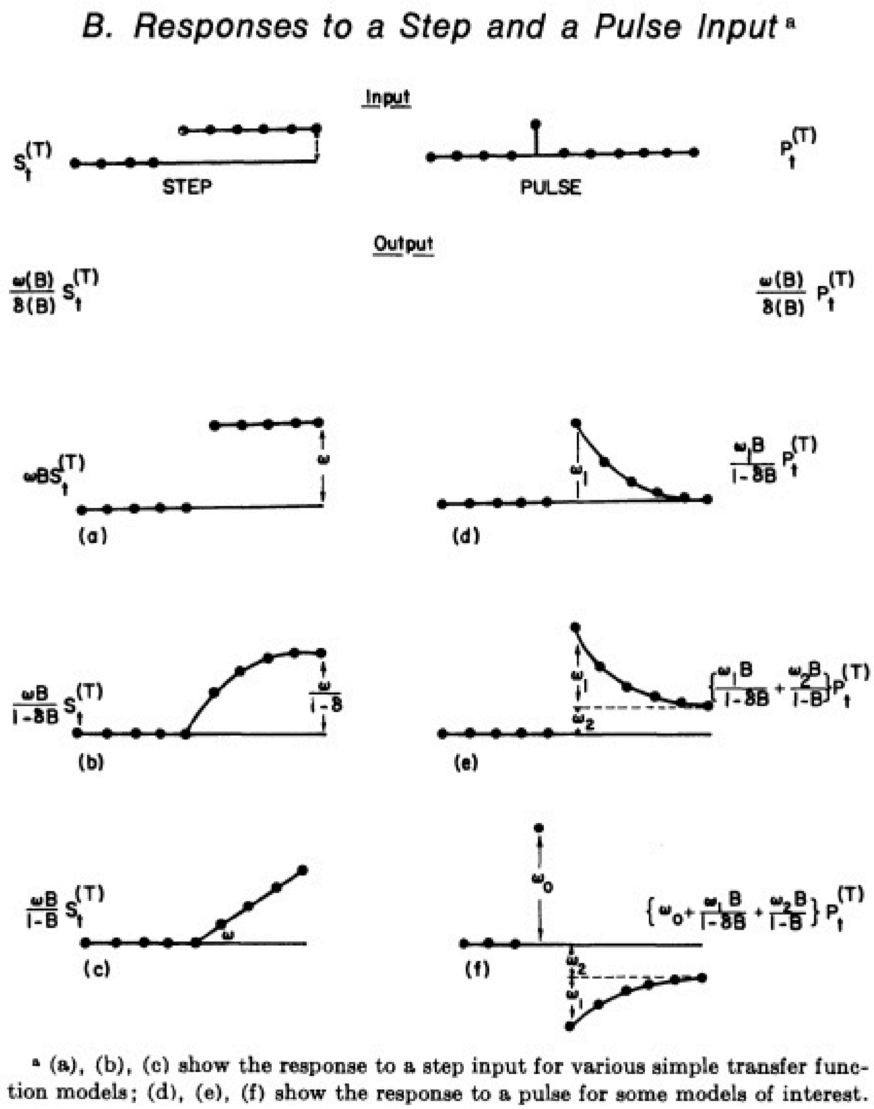


Figure 4.3. Responses to step and pulse input.

4.2 Time Series Clustering

The R packages used for intervention event analysis is listed below:

- **clue**: Measure the dissimilarity between different clusters using the cl_dissimilarity function.
- **cluster**: Contains various clustering functions such as the divisive clustering analysis (diana) function.
- **cowplot**: Arranges multiple plots made from the ggplot2 package into a grid.
- **dendextend**: Offers a set of functions for visualizing dendrograms.
- **dtwclust**: Used for building time series clusters.

- **ggplot2**: Used for creating data visualizations.
- **tictoc**: Nested timing functions in R to measure the computation length of clustering functions used.

4.2.1 Data Preprocessing

The datasets involved in time series clustering required more preprocessing. The tables in Excel were transposed and converted as nested lists so the datasets remain as integers. If they were not transposed and converted to a nested list, the values would change to characters. Next, the following non-numerical values were replaced with the zero integer:

- "-": Represent zero.
- "D": Data withheld to limit disclosure. These values are assumed zero.
- "NA": These values are unavailable and assumed zero.

4.2.2 Distance Measure

Before clustering, the distance measure must be determined first. It is arguable that the choice of distance measure is even more important that the clustering algorithm (Paparrizos & Gravano, 2015). Time series clustering uses classic clustering methods but replaces the default distance measure with one suitable for time series where similarities and invariances are determined (Paparrizos & Gravano, 2015). One popular distance measure that can be used outside of time series clustering is Euclidean distance (ED). In (4.16), two time series, $\vec{x} = (x_1, \dots, x_m)$ and $\vec{y} = (y_1, \dots, y_m)$, of length m is compared (Paparrizos & Gravano, 2015):

$$ED(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{4.16}$$

However, this distance was not used in the report since Paparrizos & Gravano (2015) shows that ED underperforms. The distance measures used for time series clustering that outperforms all the other measures are dynamic time warping (DTW) and shape-based distance (SBD) (Paparrizos & Gravano, 2015).

DTW is an extension of ED and creates a non-linear, local alignment called the warping path where $W = \{w_1, w_2, \dots, w_k\}$ on an m-by-m matrix M (Paparrizos & Gravano, 2015). With $k \geq m$, a set of matrix elements defines the mapping between $\vec{x}$ and $\vec{y}$ under the following conditions in (4.17) (Paparrizos & Gravano, 2015):

$$DTW(\vec{x}, \vec{y}) = min\sqrt{\sum_{i=1}^{k} w_i} \tag{4.17}$$

SBD manages distortions in amplitude and phase and uses cross-correlation (CC) to measure the similarity between two time series, $\vec{x} = (x_1, \ldots, x_m)$ and $\vec{y} = (y_1, \ldots, y_m)$, in (4.18) (Paparrizos & Gravano, 2015).

$$SBD(\vec{x}, \vec{y}) = 1 - \max_{m} \frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}} \tag{4.18}$$

4.2.3 Time Series Prototyping

This process involves summarizing the most important characteristics of time series from the same cluster (Sardá-Espinosa, 2019). Time series prototypes can also be referred to as the time series average or the time series centroid (Sardá-Espinosa, 2019). There are many prototyping methods that can be used when clustering, but in this case, the prototype function is selected based on the distance measure used (Sardá-Espinosa, 2019). This means that for SBD, shape extraction was selected as the centroid parameter in the tsclust function. As for DTW, DTW barycenter averaging (DBA) was used as the prototype method.

Shape extraction applies the k-Shape algorithm to compute an average sequence by minimizing the sum of squared distances to all the other time series sequences (Paparrizos & Gravano, 2015). The input series needs to be z-normalized first since the output series must be normalized as well (Sardá-Espinosa, 2019). DBA is similar to shape extraction except that the DTW alignment between each time series in a cluster and centroid is measured instead (Sardá-Espinosa, 2019).

4.2.4 Clustering

The clustering method used for analysis is hierarchical clustering. This method was selected over partitional clustering because a specific number of clusters does not need to be specified (Sardá-Espinosa, 2019). The focus of hierarchical clustering is to observe and measure the linkages between inter-groups visualized on a dendrogram (Sardá-Espinosa, 2019). A dendrogram is a binary tree where the height of each node indicates the inter-group dissimilarity (Sardá-Espinosa, 2019). In hierarchical clustering, there are two methods: agglomerative and divisive. For agglomerative hierarchical clustering, the average linkage criteria are compared with the divisive hierarchical clustering.

The different cluster methods can be compared by using the cl_dissimilarity function that measures the minimal Euclidean membership distance. The maximum distance between two cluster methods would be the square root of the number of time series that need to be clustered. A distance measure close to this value would deem the clusters different. If the distance is zero, the clusters are the same.

4.2.5 Exploratory Data Analysis

The datasets used for time series clustering contain many countries to analyze which would be hard to interpret the hundreds of time series plots and descriptive statistics. Thus, exploratory data analysis cannot easily be accomplished with this type of data. After applying the parameters to the tclust function for different cluster types (Table 4.3), the prototype and cluster dissimilarity measure can determine the number of clusters to split at. Given this is an unsupervised learning method, cluster analysis can be subjective. The best number of clusters to split at is when the prototypes produce distinct shapes. If the prototypes begin to appear similar in shape, then the cluster should not be used for final analysis.

Table 4.3 Time Series Clustering Parameters

Type	Splitting Method	Pre-Processing	Distance	Centroid
Hierarchical	Agglomerate (Average)	z-score	SBD	Shape Extraction
Hierarchical	Divisive	z-score	SBD	Shape Extraction
Hierarchical	Agglomerate (Average)	z-score	DTW	DBA
Hierarchical	Divisive	z-score	DTW	DBA

CHAPTER 5

RESULTS

Current results listed in this chapter are summarized on the Shiny Apps website built from R Shiny (https://ekayfabio.shinyapps.io/US_TSA/).

5.1 Intervention Event Analysis

5.1.1 Descriptive Statistics

Table 5.1 lists the mean, standard deviation, maximum, and minimum for each dataset used for intervention event analysis. The descriptive statistics can be better visualized in the histogram plots in Figure 5.1. The monthly apprehension and annual LPR datasets show the distribution to be skewed to the right. As for the annual apprehension histogram, the distribution appears somewhat uniformed.

Table 5.1 Intervention Event Dataset Descriptive Statistics

Dataset	Mean	Std. Deviation	Max	Min
Monthly Apprehension	60,447	38,625.3	223,305	11,677
Annual Apprehension	613,878	541,115	1,814,729	10,319
Annual LPR	431,331	363,892.9	1,826,595	6,354

To further analyze the seasonality in the monthly apprehension dataset, the boxplot in Figure 5.2 shows the descriptive statistics for each month. There appear to be more apprehensions from March through May versus the rest of the year. This seasonality may be due to a variety of factors such as mild weather in the spring for undocumented immigrants to cross the border, U.S. Border Patrol's employment and hiring trends, or vacations taken by the Border Patrol agents around the summer and the holidays.

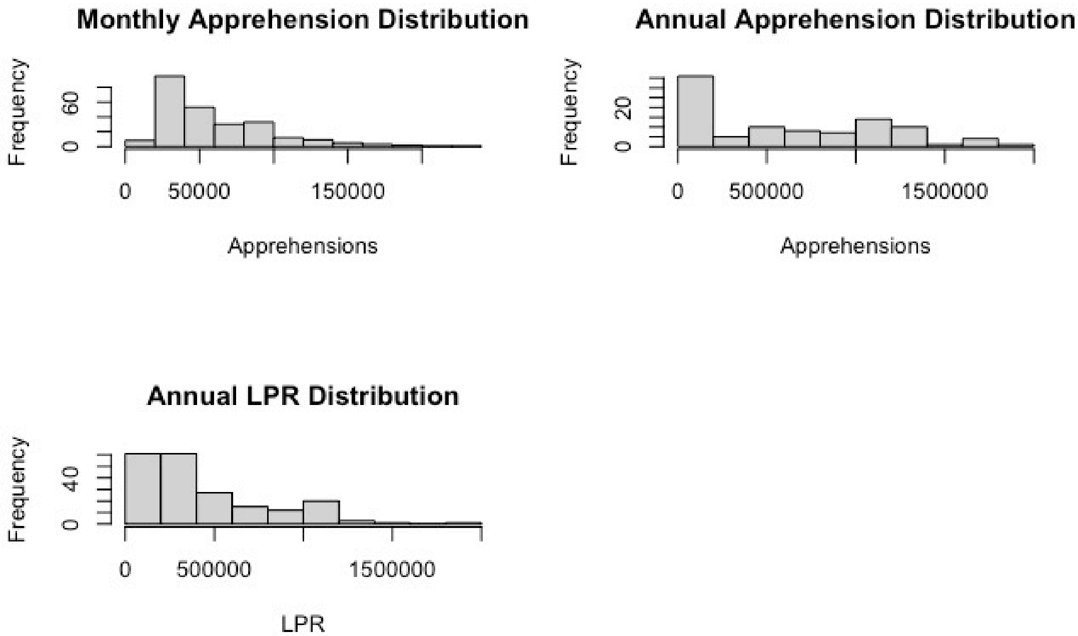


Figure 5.1. Histogram plots of the apprehension and LPR datasets.

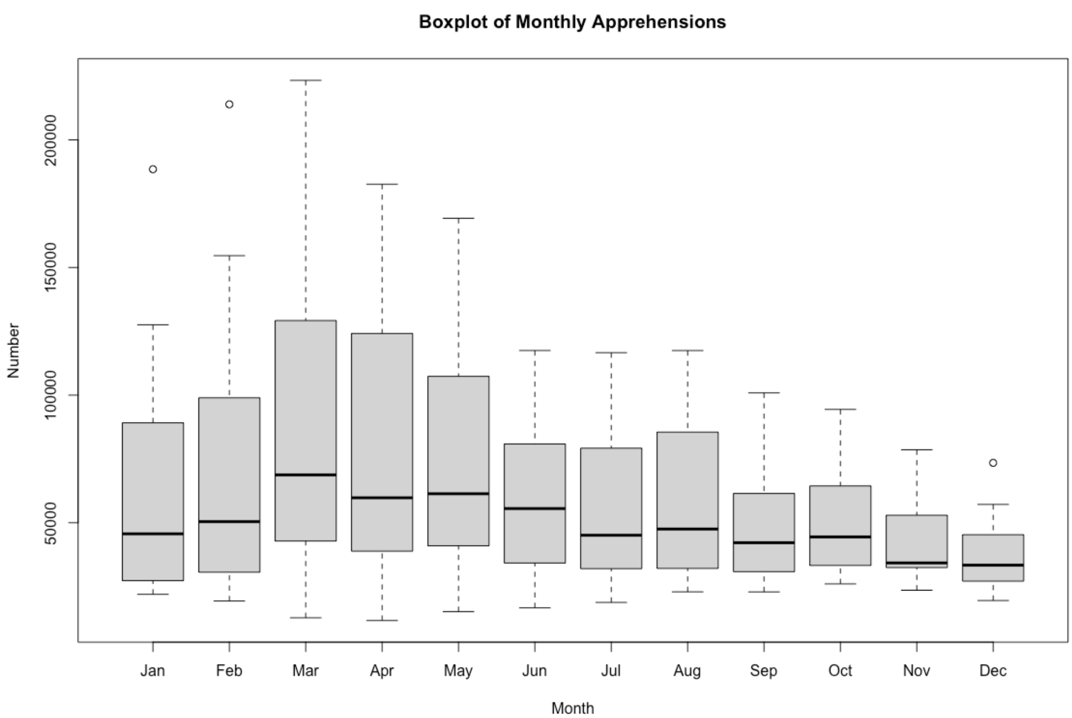


Figure 5.2. Boxplot analysis of the monthly apprehension dataset.

5.1.2 Time Series Plots

The time series plot in Figure 3.1 appears to be non-stationary with a seasonality period of twelve months or one year. Seasonality is also evident in Figure 5.2 where there is an increase in apprehensions around March through May. Figure 5.3 shows the same time series plot from Figure 3.1 but with the intervention events denoted. Just from the time series plot, the Great Recession does not appear to have an impact on the number of monthly apprehensions. The most prominent outlier on the time series plot is around when the temporary humanitarian visas were issued in Mexico.

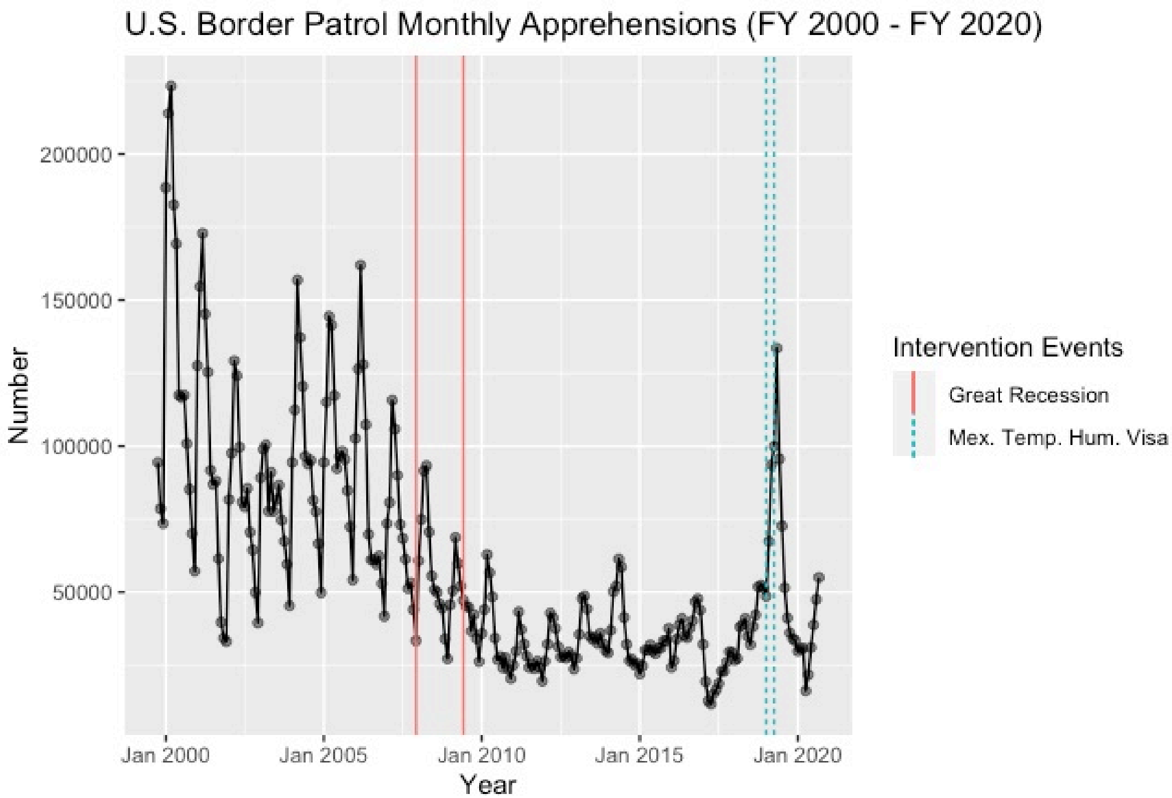


Figure 5.3. Monthly apprehensions time series plot with intervention events.

Figure 3.2 also shows non-stationarity with outliers present. However, there is no seasonality in the annual apprehension dataset. With the intervention events denoted in Figure 5.4, Operation Wetback appears to have greater impact versus IRCA and IIRIRA.

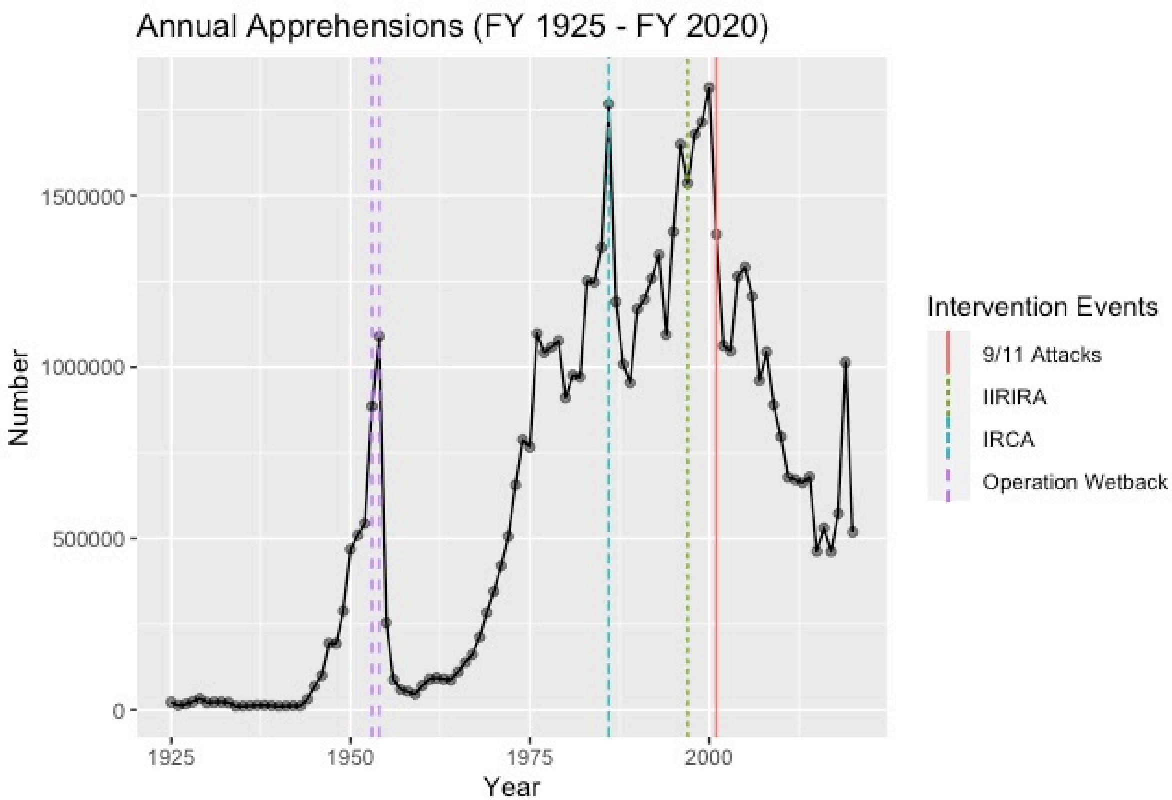


Figure 5.4. Annual apprehensions time series plot with intervention events.

Similar to the annual apprehension dataset, the time series plot for annual LPR in Figure 3.3 has no seasonality and is non-stationary. Outliers are also present in the time series plot. Some of these outliers may be due to the intervention events highlighted in Figure 5.5. IRCA especially appears to be a prominent intervention event with its temporary spike.

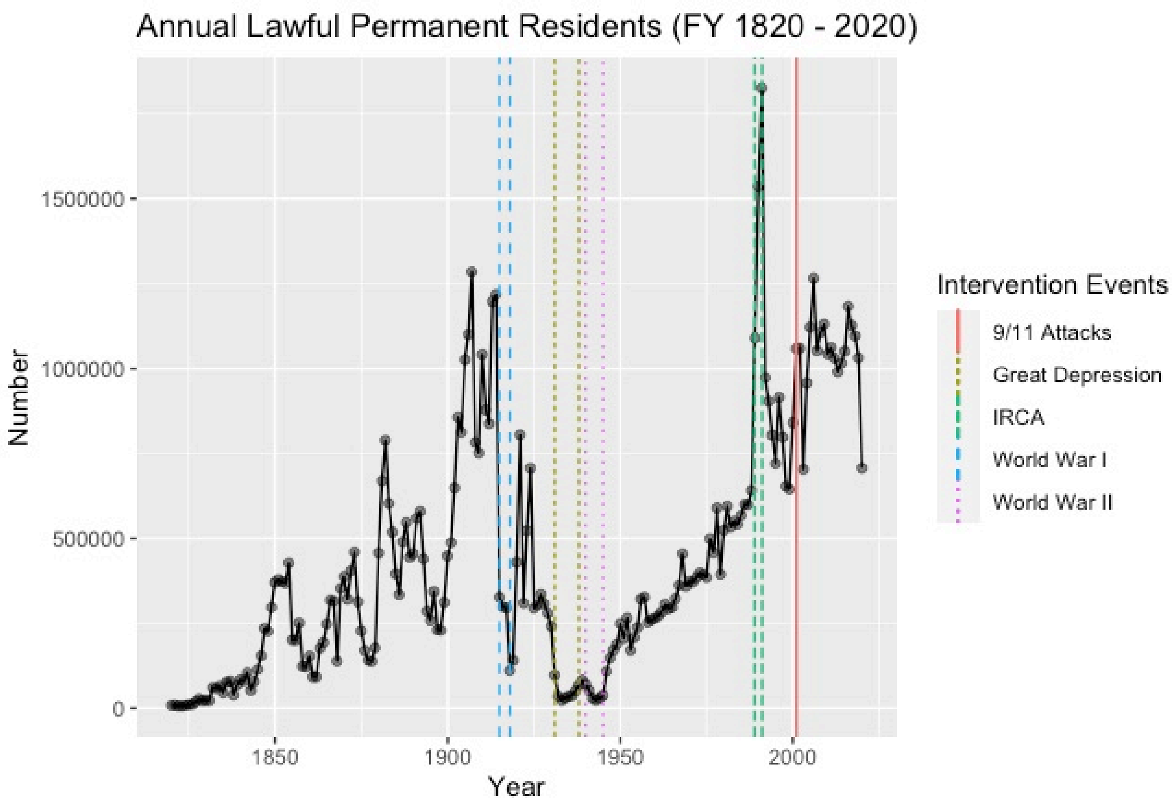


Figure 5.5. Annual LPR time series plot with intervention events.

5.1.3 Intervention Event Impact

The total intervention event analysis can be summarized in Table 5.2. Majority of the events appeared to have a temporary negative impact. Depending on the range of the datasets, not all events were analyzed.

Table 5.2 Intervention Event Impact Analysis

Events	Monthly Apprehension	Annual Apprehension	Annual LPR
World War I	-	-	Temporary (Negative)
Great Depression	-	-	Temporary (Negative)
World War II	-	-	Temporary (Negative)
Operation Wetback	-	Temporary (Negative)	-
IRCA	-	None	Temporary (Positive)
IIRIRA	-	None	-
September 11 Attacks	-	None	None
Great Recession	None	-	-
Mexico Temp. Humanitarian Visa	Temporary (Negative)	-	-

Note: The dashed line (-) indicates that the event was not analyzed for that dataset.

Based on Table 5.3, the seasonal ARIMA model selected for the monthly apprehension dataset is ARIMA(1,1,0) x $(0,1,3)_{12}$. This means the regular order is AR(1) and the seasonal order is SMA(3). All the coefficients were significant and thus not removed from the model. Out of the two events analyzed for this dataset, only the issuance of temporary humanitarian visas in Mexico around early 2019 had a temporary negative impact. This impact can be visualized in Figure 5.6 as a simple pulse function. The negative impact value is also the coefficient term for ω in Table 5.3.

Based on the impact plot in Figure 5.6, the decrease after the highest point of the outlier corresponds with the negative impact result. Around this time caravans of migrants from Central America would arrive at the U.S. border and wait to get processed for refugee status. Some may grow impatient and may make the decision to cross the border illegally. Due to the large volume of temporary humanitarian visas, the Mexican government was pressured from the U.S. to control the flow of Central American migrants from entering Mexico. This event indirectly led to an increase of monthly apprehensions when the temporary humanitarian visas were in place, and then a sudden decrease once the visas were cut.

Table 5.3 Monthly Apprehension Estimated ARIMA Model Coefficients

Events	**AR(1)**	**SMA(1)**	**SMA(2)**	**SMA(3)**	ω	**% Change**
Great Recession	0.392*** (0.061)	-0.572*** (0.073)	0.001 (0.096)	0.155 (0.084)	-	-
Mexico Temp. Humanitarian Visa	0.396*** (0.061)	-0.568*** (0.073)	0.002 (0.095)	0.148 (0.085)	-0.052 (0.085)	-5.07%

*p<0.05; **p<0.01; ***p<0.001

Note: The dashed line (-) indicates no coefficient for that event.

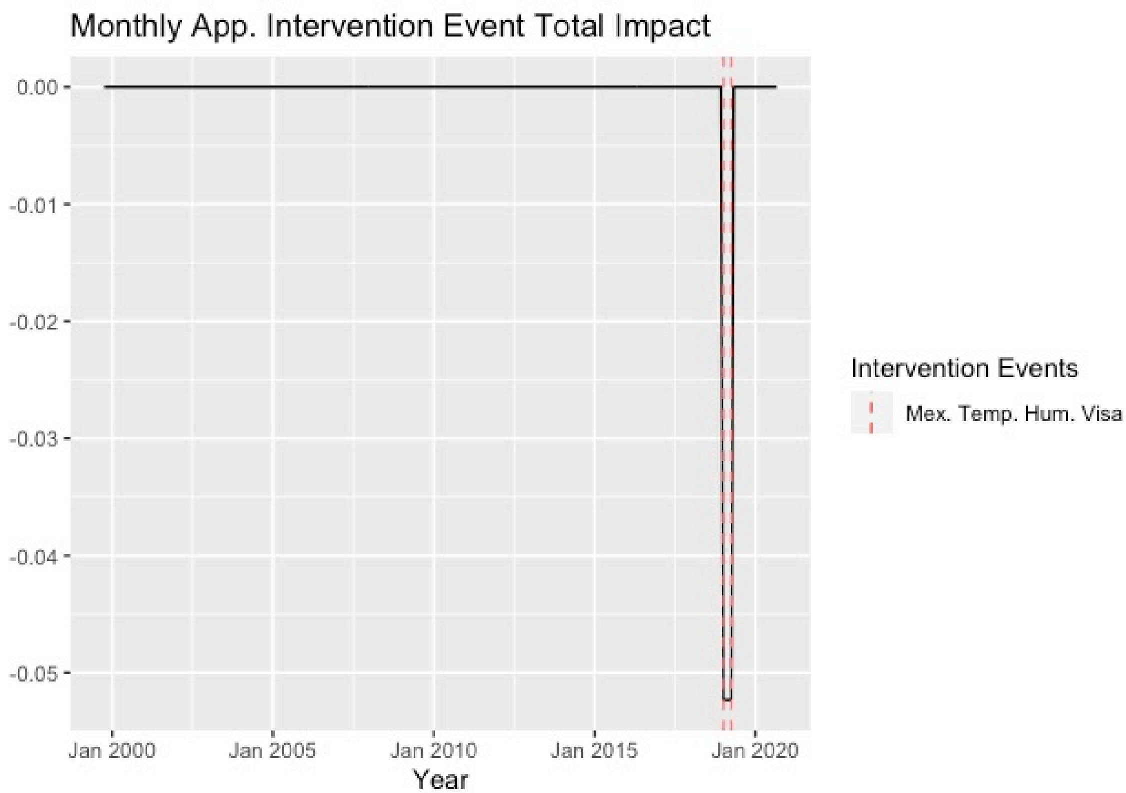


Figure 5.6. Monthly apprehension event impact.

For the annual apprehension dataset, the ARIMA model is ARIMA(11,0,0) with the majority of the AR coefficients removed, and leaving only the AR(1) and AR(11) terms. Based on Table 5.4 and Figure 5.7, only Operation Wetback had a significant negative pulse response. This is evident based on the coefficient terms for δ and ω. ω is the main contributor for the negative impact due to its step function component. As for δ, its output is the gradual increase curve back to zero. IRCA, IIRIRA, and the September 11 attacks have the same AR(1) and AR(11) coefficients to the general ARIMA model since they have no significant impact.

After Operation Wetback, the decrease in the number of annual apprehensions is reflected in the negative impact results from the intervention event analysis. With the buildup of the Bracero Program and the increased anti-immigrant sentiment, Operation Wetback was initiated as an effort to apprehend and remove undocumented Mexican immigrants. This directly affected the annual apprehension count around this time. After Operation Wetback terminated, the annual apprehension count decreased but began to gradually increase after the

1960s. IRCA and IIRIRA in the 1980s and 1990s had small spikes around this time, but it resulted in no significant impact. This may be due to the policies only affecting how undocumented immigrants are taken care of once they are in the U.S. versus policies related to border enforcement that have more of an impact on apprehensions. As for the September 11 Attacks, there was no significant impact. Even though it is a significant event in history, it may not have affected apprehension specifically.

Table 5.4 Annual Apprehension Estimated ARIMA Model Coefficients

Events	AR(1)	AR(11)	δ	ω	% Change
Operation Wetback	0.238* (0.101)	-0.245* (0.125)	0.551*** (0.124)	-0.534** (0.181)	-41.37%
IRCA	0.255** (0.096)	-0.336*** (0.098)	-	-	-
IIRIRA	0.255** (0.096)	-0.336*** (0.098)	-	-	-
September 11 Attacks	0.255** (0.096)	-0.336*** (0.098)	-	-	-

*p<0.05; **p<0.01; ***p<0.001
Note: The dashed line (-) indicates no coefficient for that event.

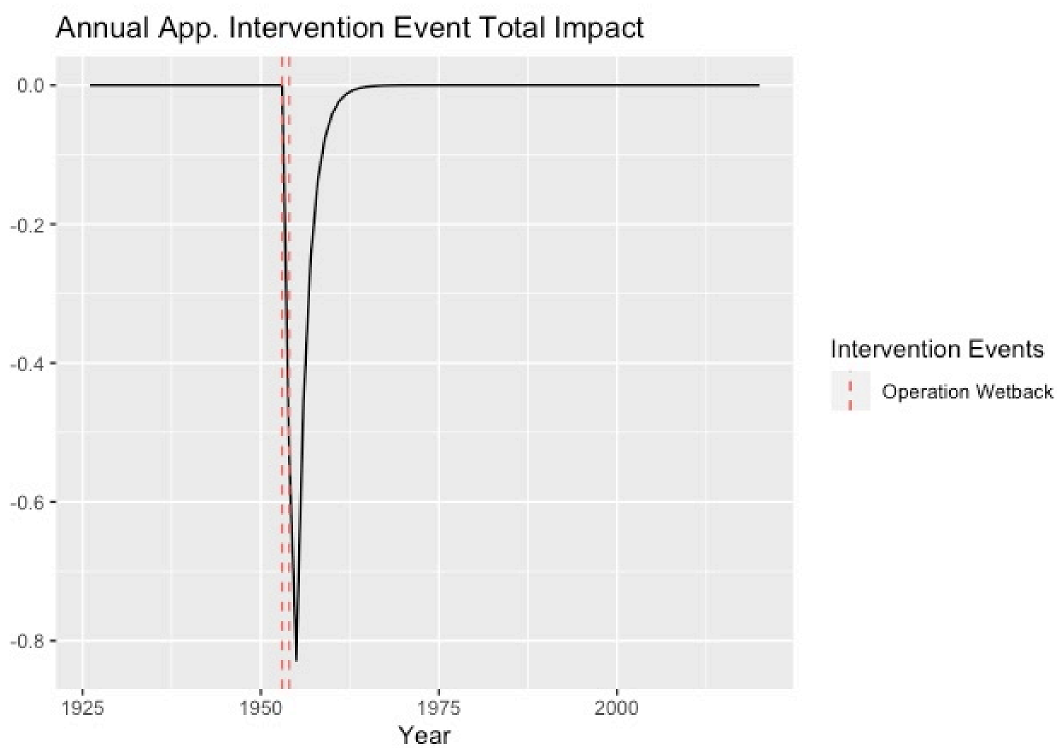


Figure 5.7. Annual apprehension event impact.

The ARIMA model for the annual LPR dataset is ARIMA(5,1,0) with all the AR coefficients removed other than AR(5). Unlike the other datasets, majority of the events analyzed had a significant impact which can be seen all together in Figure 5.8 and in Table 5.5. The negative pulse response is seen in the Great Depression, World War I, and World War II as these events were pivotal moments in history where accepting immigrants to permanently reside in the U.S. was out of the question. In the Great Depression, the unemployment rates were high, and the government put their efforts on taking care of their own citizens with their financial hardships. As for World War I and II, protecting citizens meant blocking immigrants from entering. September 11 Attacks appeared to have no effect on LPRs. To explore if September 11 Attacks are significant, other immigration time series data, like the asylum and refugee count, may need to be pulled for intervention analysis. This can provide an answer on how the event impacted U.S. immigration.

The only event with a positive pulse impact is IRCA. IRCA not only instilled employment sanctions to discourage illegal immigration but allowed undocumented immigrants already residing since 1982 to become permanent residents. With legalization programs, like the Legally Authorized Workers (LAWs) and Special Agricultural Workers (SAWs), underway many undocumented immigrants were granted LPR status around 1990 and 1991.

Table 5.5 Annual LPR Estimated ARIMA Model Coefficients

Events	**AR(5)**	ω	**% Change**
World War I	-0.234*** (0.069)	-0.656** (0.232)	-48.11%
Great Depression	-0.267*** (0.068)	-0.601** (0.229)	-45.17%
World War II	-0.240*** (0.069)	-0.494* (0.233)	-38.98%
IRCA	-0.259*** (0.068)	0.581* (0.229)	78.78%
September 11 Attacks	-0.252*** (0.068)	-	-

*p<0.05; **p<0.01; ***p<0.001
Note: The dashed line (-) indicates no coefficient for that event.

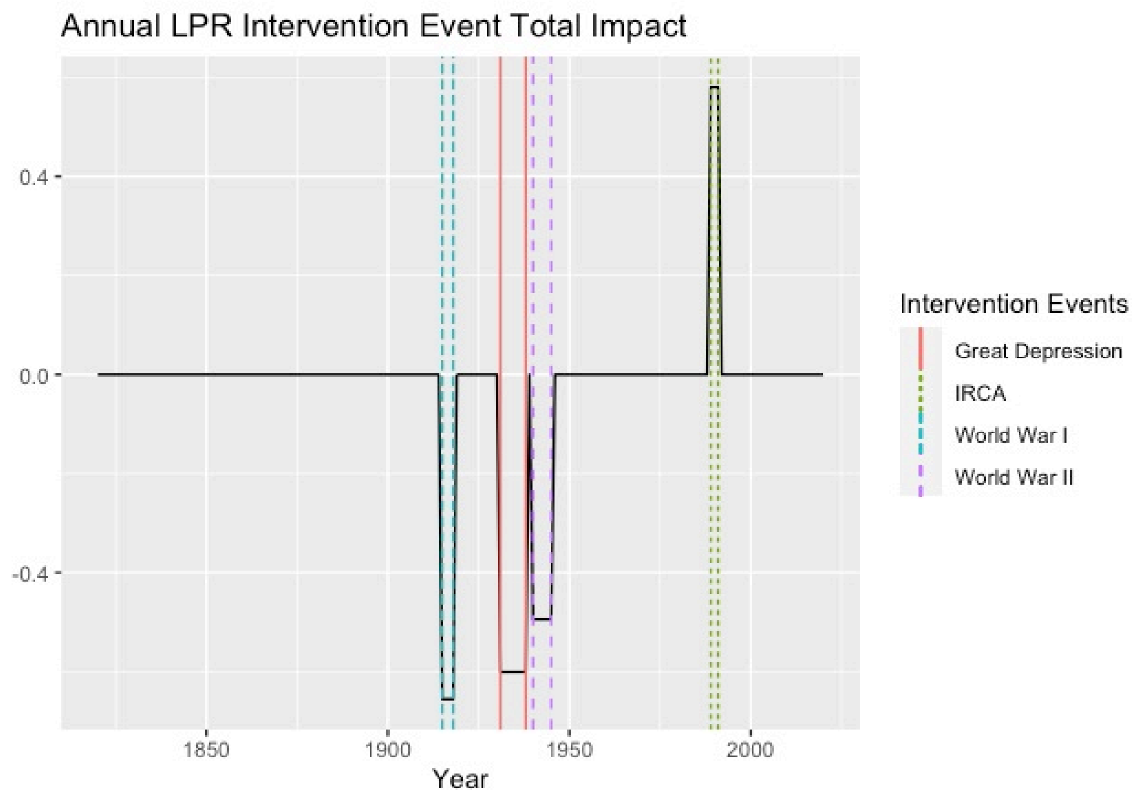


Figure 5.8. Annual LPR event impact.

5.2 TIME SERIES CLUSTERING

The cluster assignment tables for each dataset and method can be viewed in the Shiny Apps website from the Analysis menu under "Cluster Assignment" (https://ekayfabio.shinyapps.io/US_TSA/). Since three or more clusters resulted in similar prototypes with complex interpretations, two clusters were only for analyzing sending countries.

5.2.1 Dendrograms & Cluster Assignment

Figures 5.9-5.12 shows dendrograms for each dataset under different clustering methods. The different color branches (black and grey) indicate the cluster assignment for each sending country. For the apprehension dataset, the SBD agglomerative average dendrogram has an unbalanced number of sending countries for each cluster. Only two out of 159 sending countries are assigned to cluster two while the remainder falls under cluster one. The LPR and naturalization dataset has a similar disproportion in cluster assignment for the SBD agglomerate average method. The LPR dataset only has 16 sending countries out of 202 for cluster two and the naturalization dataset only has eight sending countries out of 199 for cluster two. As for the nonimmigrant dataset, both SBD and DTW agglomerate average methods also has a large disproportion in cluster assignment. Out of the 194 sending countries, the SBD dendrogram identifies six sending countries for cluster two and the DTW dendrogram identifies four sending countries for cluster two.

Overall, the dendrograms that appeared to show the best method with an even cluster assignment distribution is a divisive hierarchical cluster using DTW as a distance measure. A map of the cluster assignments using this method can be seen in Figures 5.13-5.16 where cluster one is labeled dark grey and cluster two is labeled light grey. To note, the apprehension and LPR dataset shows U.S. assigned a cluster. This may be due to immigrants who were born in the U.S. but not recognized as LPR or naturalized citizens. We can assume that sending countries under the same clusters have similar conditions that led to the trends in apprehension, LPR, naturalization, and nonimmigration status from 2011 to 2020. However, more research will need to be performed to determine what these conditions are. For instance, countries of different economic standings (developed, developing, and least developed) and geographical considerations can be studied further.

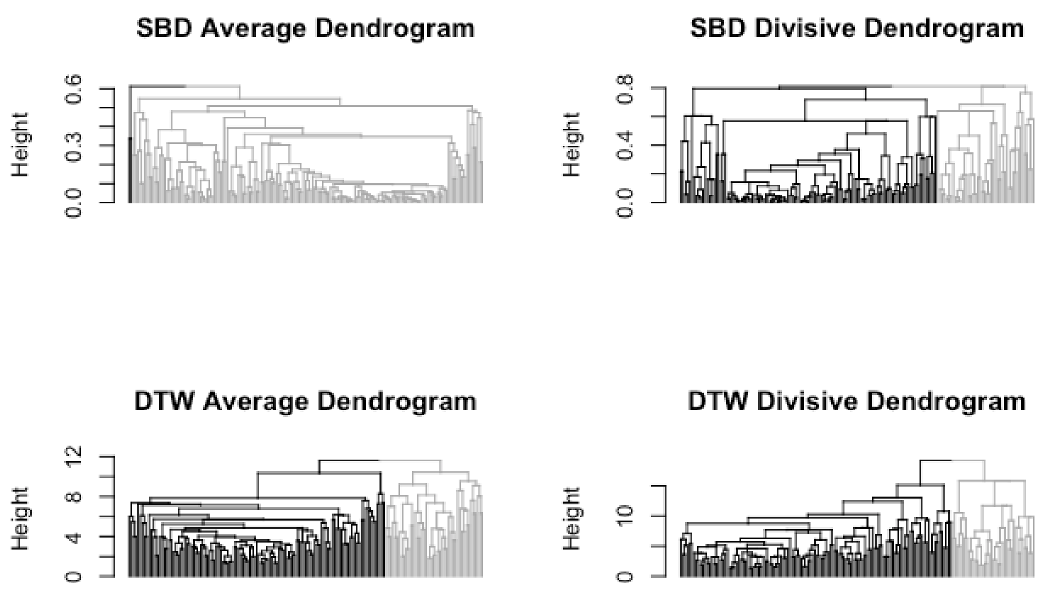


Figure 5.9. Apprehension dendrograms.

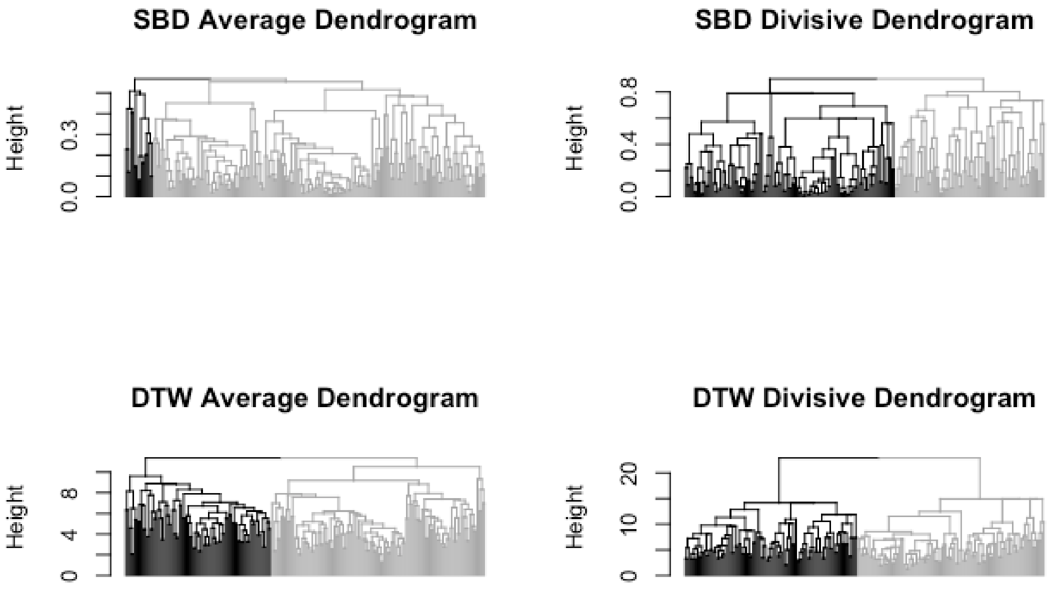


Figure 5.10. LPR dendrograms.

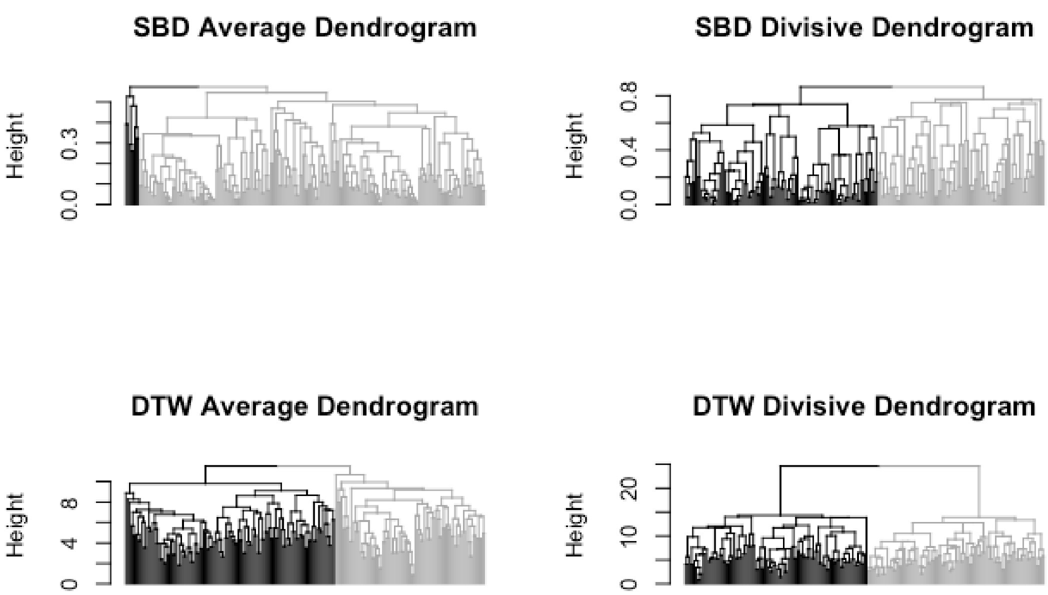


Figure 5.11. Naturalization dendrograms.

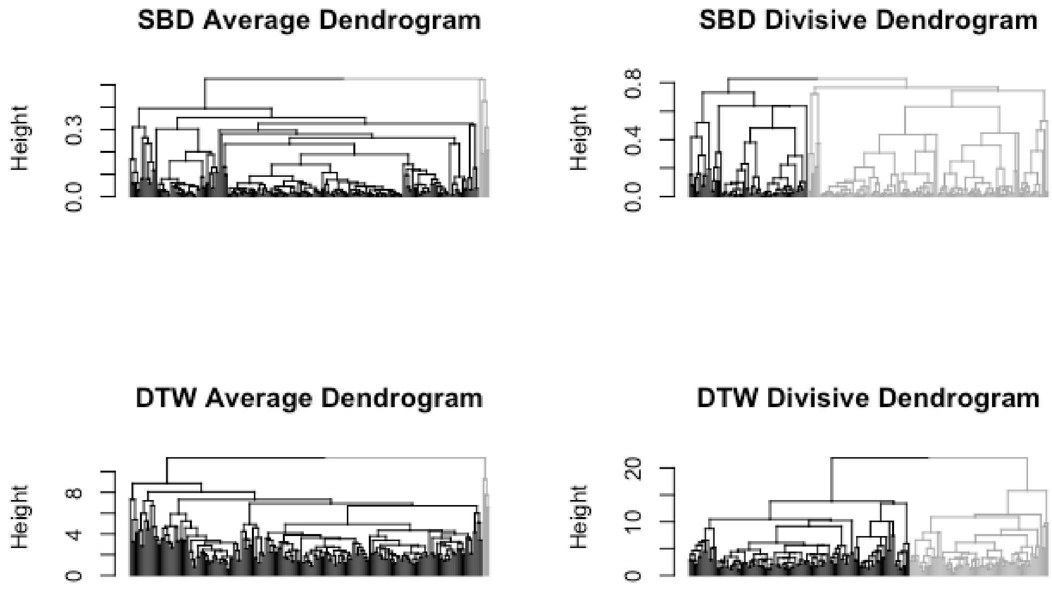


Figure 5.12. Nonimmigrant dendrograms.

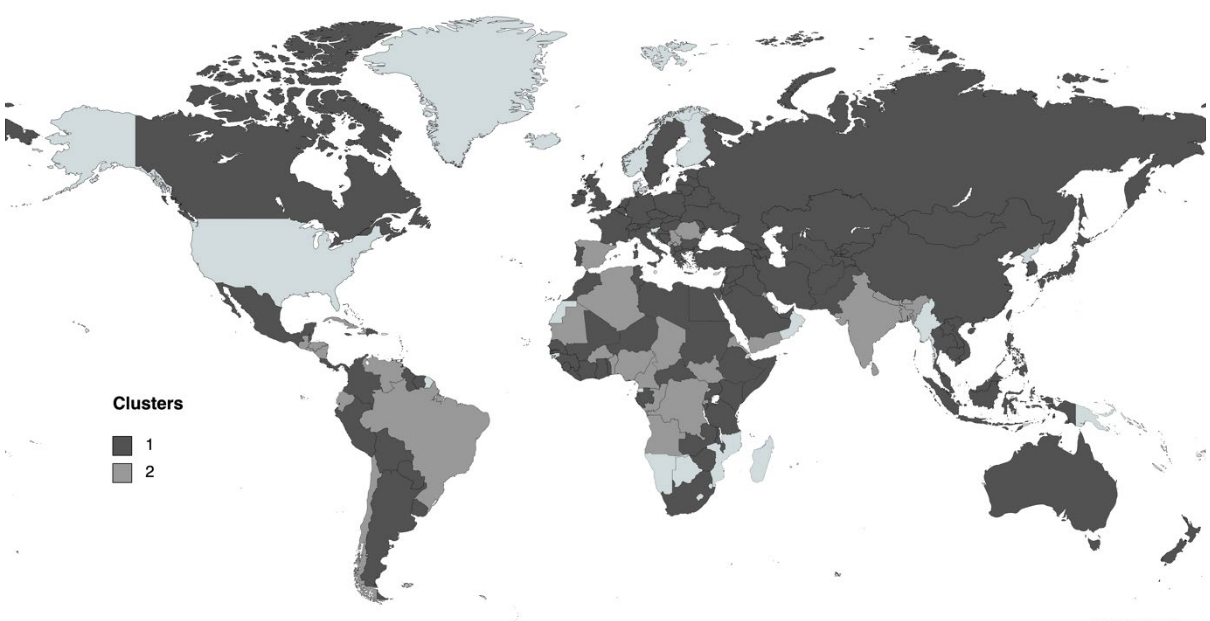


Figure 5.13. Apprehension DTW divisive clustering assignment map.

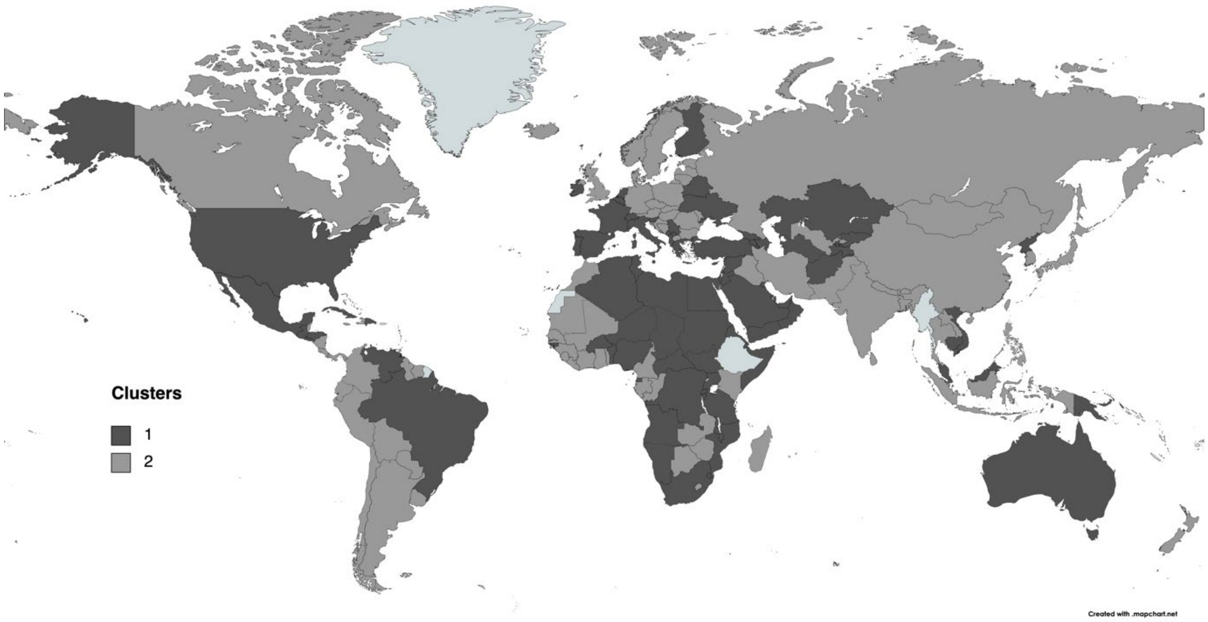


Figure 5.14. LPR DTW divisive clustering assignment map.

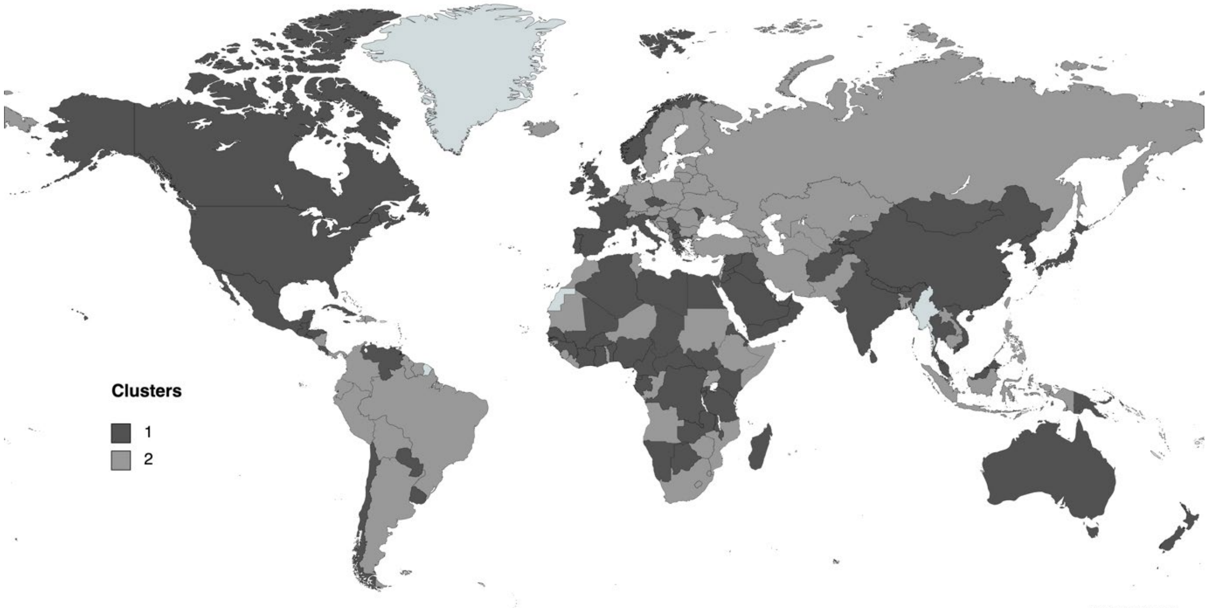


Figure 5.15. Naturalization DTW divisive clustering assignment map.

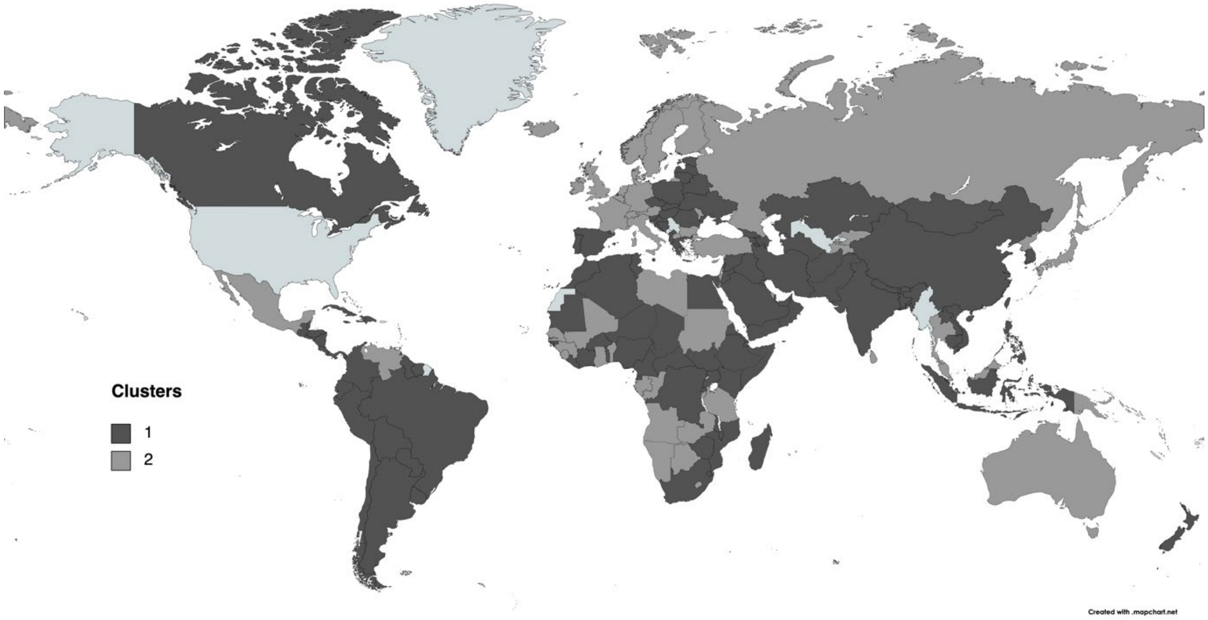


Figure 5.16. Nonimmigrant DTW divisive clustering assignment map.

5.2.2 Time Series Prototypes

Prototyping captures important trends of a group of sending countries with similar shape. The clustering algorithms and centroid methods resulted in similar prototypes as seen in Figures 5.17-5.20. However, there was a bigger distinction between SBD and DTW prototypes compared to average and divisive prototypes. This is because SBD prototypes used shape extraction for creating prototypes and DTW used DBA.

The apprehension dataset showed cluster one with a gradual decrease in apprehensions. As for cluster two, apprehensions increased until near the end where there is a dip. The SBD average prototype differed the most from the other prototypes for cluster two. The prototype for this clustering method demonstrated a zigzag pattern. For the LPR dataset, cluster one gradually increased and cluster two was steady until it dips down near the end. The SBD average prototype also differed from the other prototypes for cluster two. The LPR started near the bottom and moved up until it plateaus around the middle. Eventually the LPR dips back down near the end. For the remaining datasets, the naturalization and nonimmigrant prototypes were all similar in shape and direction for each cluster except cluster two for the SBD average prototypes. One cluster demonstrated an increase trend while the other would go in the opposite direction. The SBD average prototype for the naturalization dataset had a big dip by the midway point. As for the nonimmigrant dataset, the divisive prototypes appeared closest in shape. The average prototypes for SBD and DTW were still visibly different from each other regardless of the same negative trend.

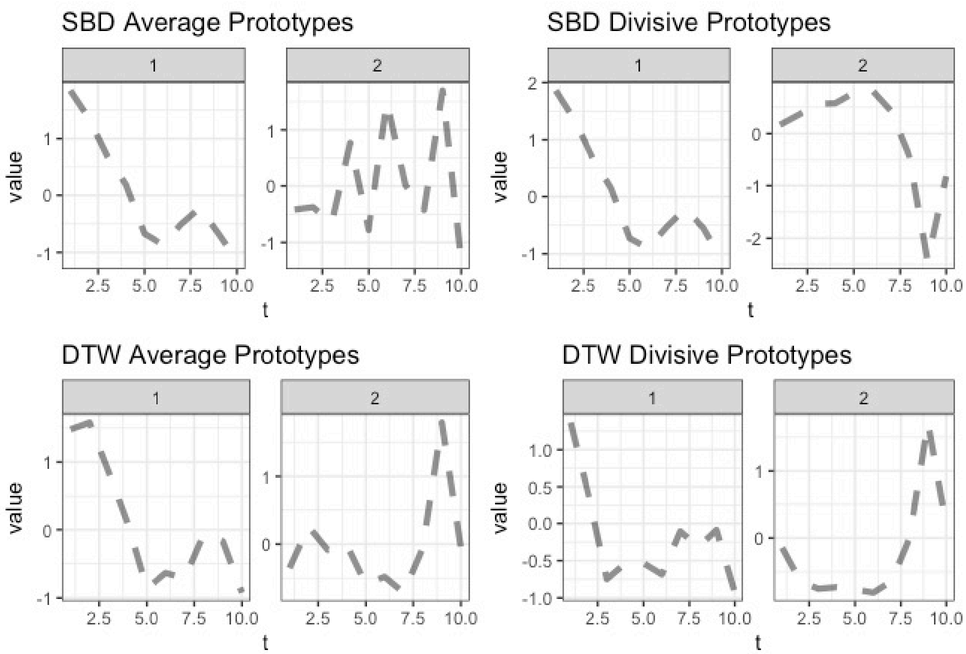


Figure 5.17. Apprehension cluster prototypes.

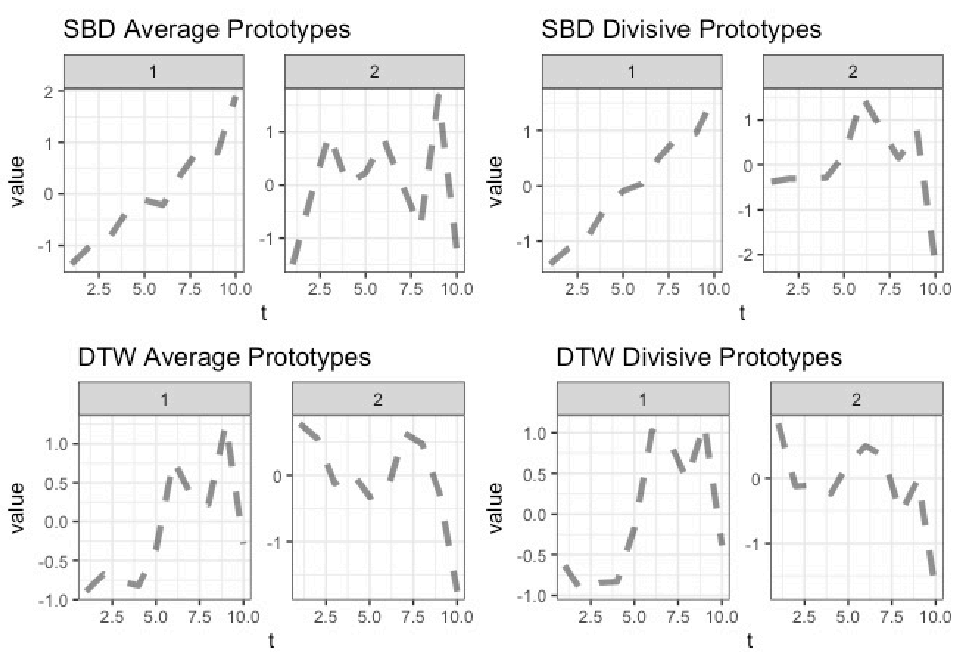


Figure 5.18. LPR cluster prototypes.

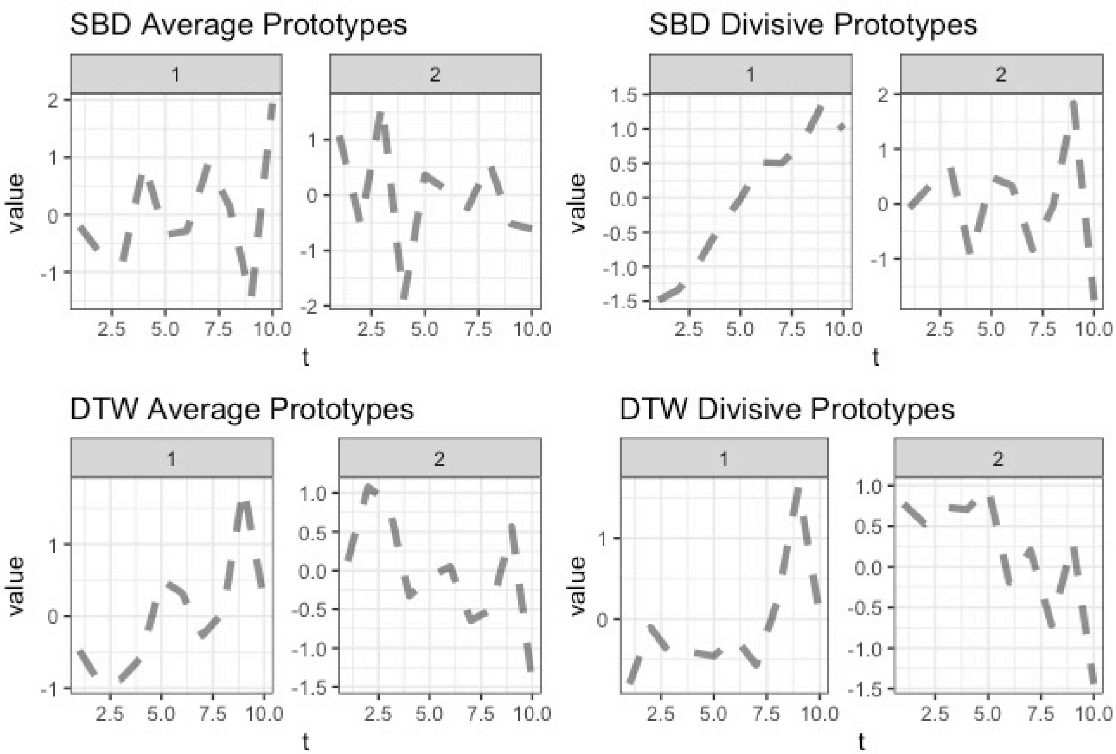


Figure 5.19. Naturalization cluster prototypes.

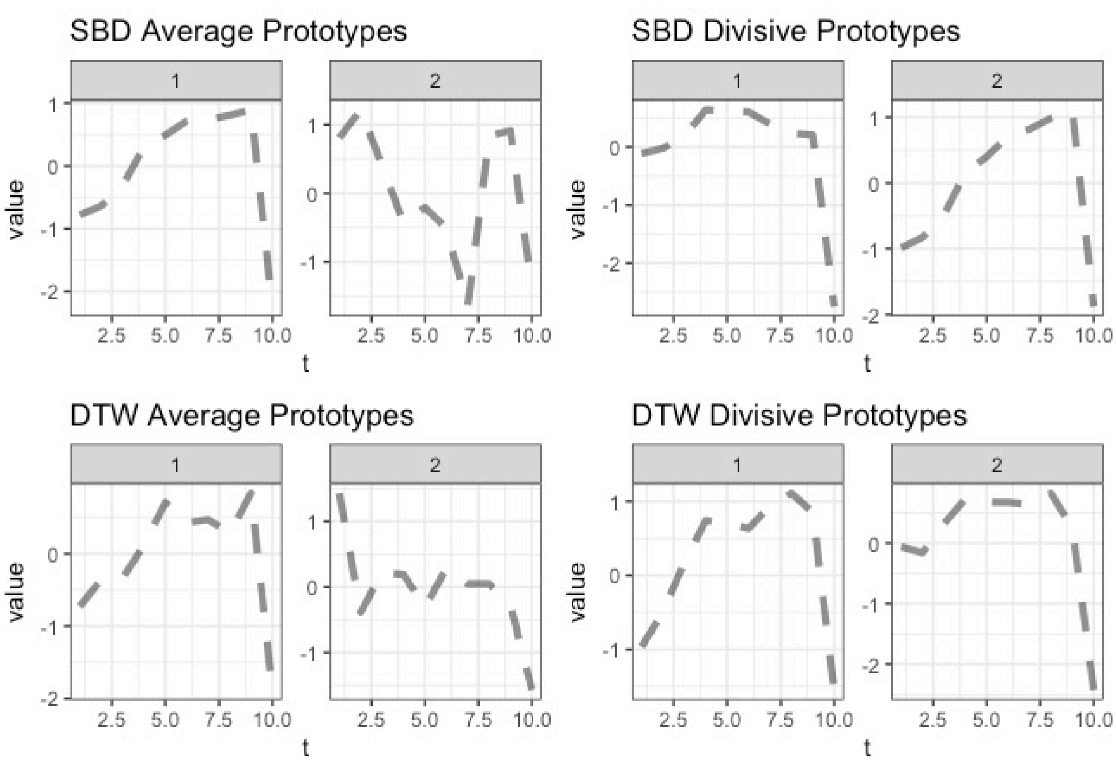


Figure 5.20. Nonimmigrant cluster prototypes.

5.2.3 Cluster Comparison

Other than noting the observable differences from the prototype plots above, using Euclidean distance can be useful in measuring cluster method dissimilarity and determine the best clustering model to use for this type of dataset. With there being 159 sending countries in the apprehension dataset, the cluster comparison with a value close to the square root of 159, which is 12.6, would have maximum dissimilarity. If there was a cluster dissimilarity value close to zero, then the clusters would be almost the same. In Table 5.6, the SBD average cluster had a higher dissimilarity value versus the other cluster comparisons.

Table 5.7 and 5.8 show greater dissimilarity between SBD and DTW cluster methods. For LPR datasets, the maximum dissimilarity for 202 sending countries is 14.2. This dissimilarity is exemplified in the Euclidean distances ranging from 13 to 14 when SBD and DTW clusters are compared. This is also evident in the naturalization dataset. In Table 5.9, the nonimmigrant dataset shows that average and divisive cluster methods have more weight for cluster dissimilarity. For instance, DTW average and SBD average clusters have a dissimilarity value of 2.8 which is closer to zero. Another example is SBD divisive and DTW divisive clusters with a dissimilarity value of 7.9. The other cluster comparisons still ranged around 11 to 12 which is closer to the maximum dissimilarity value of 13.9.

Based on the results it would be difficult to determine which clustering method is best especially with these datasets where there are no target variables to test on. However, the dissimilarity measurements provide a benchmark for these clustering methods. Even though SBD has a fast computation time, it provided a disproportionate clusters assignment. DTW, while it took longer to compute, still maintained an even split of sending countries to each cluster especially using a divisive hierarchical clustering method.

Table 5.6 Apprehension Cluster Dissimilarity Measurement

Clusters	sbd_average	sbd_diana	dtw_average
sbd_diana	9.165151	-	-
dtw_average	9.165151	2.828427	-
dtw_diana	8.602325	3.741657	4.242641

Note: Dissimilarity was measured using minimal Euclidean membership distance.

Table 5.7 LPR Cluster Dissimilarity Measurement

Clusters	sbd_average	sbd_diana	dtw_average
sbd_diana	11.832160	-	-
dtw_average	12.165525	14.142136	-
dtw_diana	13.490738	12.884099	7.071068

Note: Dissimilarity was measured using minimal Euclidean membership distance.

Table 5.8 Naturalization Cluster Dissimilarity Measurement

Clusters	sbd_average	sbd_diana	dtw_average
sbd_diana	12.961481	-	-
dtw_average	13.038405	13.416408	-
dtw_diana	13.856406	12.806248	7.745967

Note: Dissimilarity was measured using minimal Euclidean membership distance.

Table 5.9 Nonimmigrant Cluster Dissimilarity Measurement

Clusters	sbd_average	sbd_diana	dtw_average
sbd_diana	11.313708	-	-
dtw_average	2.828427	11.135529	-
dtw_diana	12.083046	7.874008	11.916375

Note: Dissimilarity was measured using minimal Euclidean membership distance.

5.2.4 Case Study: Mexico & Northern Triangle

In this case study, time series clustering results from Mexico and Northern Triangle countries (Guatemala, Honduras, and El Salvador) were investigated further. Based on the time series plots for each dataset in Figure 5.21, Mexico appears to be very different from the Northern Triangle countries. While their time series are similar, Mexico has more apprehensions, LPR, naturalizations, and nonimmigrant visa issuances. This difference in height can be deceiving in assuming that Mexico and the Northern Triangle countries fall under different clusters. Since the data is normalized using z-score, the shape of a time series is the determining factor when clustering.

The gap between Mexico and the Northern Triangle countries can be due to a variety of factors such as the accessibility to the U.S. Mexico brings in more traffic since it borders the country. However, the large inflow of Mexicans crossing the border seeking employment is slowly being replaced by the Northern Triangle countries seeking protection (Massey, 2020). In Figure 5.21, the number of Mexicans apprehended in 2019 converges with the apprehensions from Guatemala and Honduras. As for the LPR, naturalization, and nonimmigrant dataset, this increase inflow from the Northern Triangle countries is not reflected on their time series plots. The reason for this may be due to an annual quota of nonimmigrant visa issuances per country which then limit the number of LPRs and naturalizations.

Regarding cluster assignment of Mexico and the Northern Triangle countries, a divisive hierarchical clustering method with DTW as the distance measure is selected for analysis and summarized in Table 5.10. These cluster assignments can also be visualized on world maps in Figures 27-30. The results from the apprehension dataset show Mexico as a different cluster from the Northern Triangle countries. Since Mexico borders the U.S., there are more opportunities for Mexicans to cross the border. This may be the reason for the different apprehension trends between the two groups. The LPR and naturalization results show both Mexico and the Northern Triangle countries all under the same cluster. This indicates there is little variation between Mexico and the Northern Triangle countries regarding their LPR status and naturalization in the U.S. With the nonimmigrant dataset, Mexico, Honduras, and Guatemala are assigned to the same cluster while El Salvador is assigned to another. Since cluster one and two have similar cluster prototypes in Figure 5.20, it is hard to determine why there is a mismatch between El Salvador and the rest of the countries without further analysis.

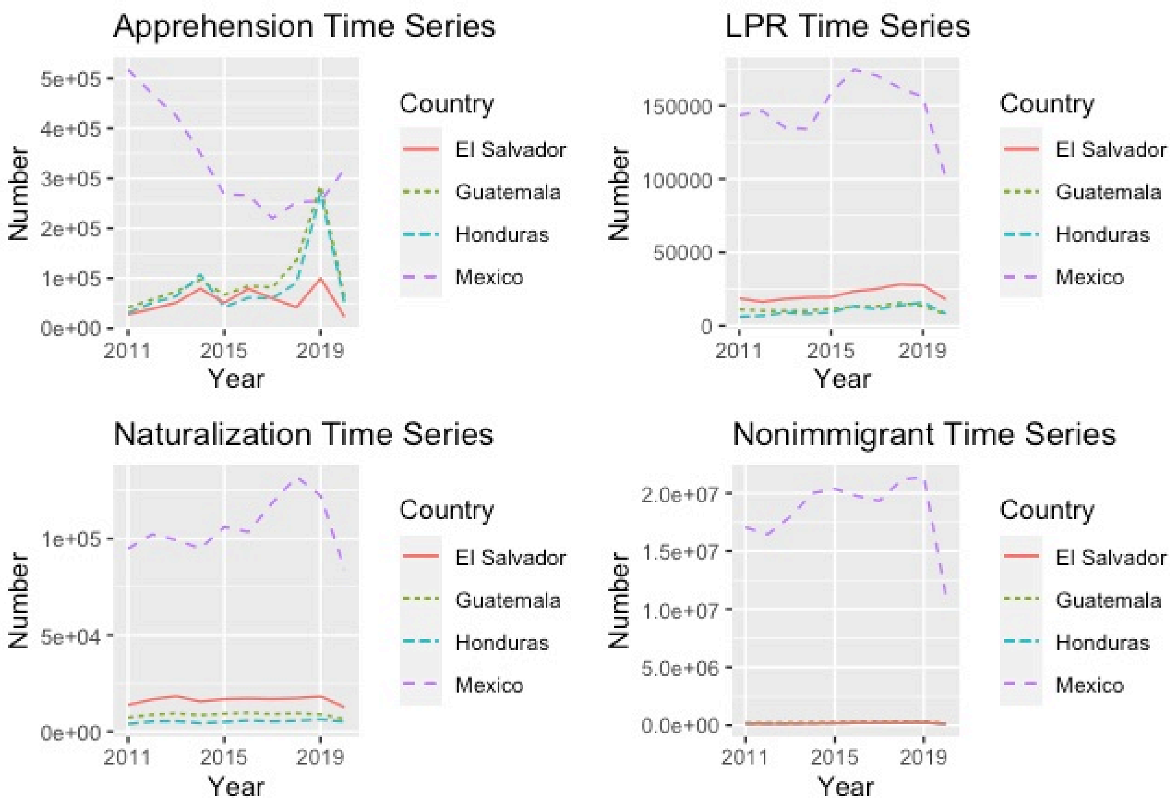


Figure 5.21. Case study time series plot.

Table 5.10 Case Study DTW Divisive Cluster Assignment

Dataset	Mexico	Guatemala	Honduras	El Salvador
Apprehension	1	2	2	2
LPR	1	1	1	1
Naturalization	1	1	1	1
Nonimmigrant	2	2	2	1

5.3 Discussion

Using intervention analysis and time series clustering accounts for all the factors that influence U.S. immigration. Intervention analysis directly investigates political and economic events that affect the opening and closing of U.S. borders, and time series clustering covers sending country's political and economic events that affect emigration trends.

Unauthorized immigration was directly affected by immigration policies but was not sensitive to changes in the economic environment. Specifically, the apprehension count dropped by 41 percent after Operation Wetback was implemented in 1953-1954 and by 5 percent after Mexico increased humanitarian visa issuances in 2019. On the other hand, legal immigration is related to both economic and political factors. The number of LPRs reduced by 48 percent after World War I, by 38 percent after World War II, and by 45 percent after the Great Depression. For IRCA, the number of LPRs jumped up almost 80 percent after it passed in 1986. It is surprising the September 11 Attacks had no significant impact on either apprehension or LPR datasets since it was such a pivotal moment in recent U.S. history. It may be that other policies have branched off from this event instead.

As seen from Mexico's humanitarian visa issued event, the immigration policies of sending countries can also play a critical role in the inflow of immigration to the U.S. Time series clustering looks further into the immigration behavior of more than two hundred sending countries. For interpretation purposes, the number of clusters created is set to two. The prototype results demonstrated one cluster with a monotone trend and the other cluster with a zigzag pattern. For example, the apprehension dataset shows undocumented immigrants from countries in cluster one declined while those in cluster two fluctuated. The LPR dataset exhibited the sending countries in cluster one with steady growth while the countries in cluster two fluctuated and gradually decreased. However, there are no obvious trends from the naturalization and nonimmigrant data.

A case study was also conducted to take a closer look at four countries that immigrate to the U.S. often. Based on Section 2.2, the macro-factors from Mexico and the Northern Triangle countries can be easily identified. The time series plots of the Northern Triangle countries showed an increase in apprehensions that eventually merged with Mexico's apprehension count. This points at a continuing political macro-factor the Northern Triangles countries are experiencing. Many individuals from these countries are choosing to emigrate due to the violence and political instability. Exploring the cluster that contains the Northern Triangle countries can help identify other possible sending countries with similar political macro-factors. These sending countries are mostly South American and African countries that may be going through a similar political crisis.

Analyzing major events using intervention analysis and time series clustering can help world leaders prepare for the future. The results can provide insight into past policy mismatches and a resolution on how to improve immigration flow and procedures.

CHAPTER 6

CONCLUSION

This study accomplished identifying political and economic factors that have impacted U.S. immigration through intervention analysis. Political events, such as Operation Wetback and IRCA, have mainly supported the domestic politics approach for policymaking. Many of these events reflected the views of the public. The policies that passed were driven by societal interest and complaints brought to the attention of politicians. Economic events primarily supported the globalization approach. The Great Depression is an example of an economic downturn that impacted more than just the U.S. market but the global market. Fewer immigrants were restricted from entering the U.S. because the government wanted to ensure enough of their own citizens would be supported first with jobs and security. From monitoring trends and outliers from these political and economic events, U.S. immigration flow can be efficiently monitored and managed.

Intervention analysis also fulfilled the goal in determining the length of these political and economic events on U.S. immigration. An impact plot for each event demonstrates either a permanent (step) or temporary (pulse) response. All of the significant events had a simple temporary impact with one exhibiting a gradual decay.

The emigration trends from sending countries were also determined with an unsupervised learning algorithm. A case study that focused on Mexico and Northern Triangle countries was used to analyze the effectiveness of time series clustering and opens the door for further analysis on additional sending countries using this method. The accuracy in time series clustering could be improved if there was more data. Currently, all the datasets range annually from 2011 to 2020 fiscal year, leaving only ten data points for each sending country. Time series clustering can still be performed, but the accuracy in prototyping and clustering assignment may come into question. This study has not only provided in-depth analysis on political and economic events in time series data, but also used time series clustering to understand sending countries' emigration history and behavior.

REFERENCES

Abramitzky, R., & Boustan, L. (2017). Immigration in American economic history. *Journal of Economic Literature*, *55*(4), 1311–1345. https://doi.org/10.1257/jel.20151189

Abrego, L., Coleman, M., Martínez, D. E., Menjívar, C., & Slack, J. (2017). Making immigrants into criminals: Legal processes of criminalization in the post-IIRIRA era. *Journal on Migration and Human Security*, *5*(3), 694–715. https://doi.org/10.1177/233150241700500308

Baxter, A. M., & Nowrasteh, A. (2021). *A brief history of U.S. immigration policy from the colonial period to the present day*. CATO Institute. https://www.cato.org/policy-analysis/brief-history-us-immigration-policy-colonial-period-present-day

Beine, M., Boucher, A., Burgoon, B., Crock, M., Gest, J., Hiscox, M., McGovern, P., Rapoport, H., Schaper, J., & Thielemann, E. (2016). Comparing immigration policies: An overview from the IMPALA database. *International Migration Review*, *50*(4), 827–863. https://doi.org/10.1111/imre.12169

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, *70*(349), 70–79. https://doi.org/10.1080/01621459.1975.10480264

Card, D. (2005). Is the new immigration really so bad? *The Economic Journal*, *115*(507), F300–F323. https://doi.org/10.1111/j.1468-0297.2005.01037.x

Castelli, F. (2018). Drivers of migration: Why do people move? *Journal of Travel Medicine*, *25*(1). https://doi.org/10.1093/jtm/tay040

Czaika, M., & de Haas, H. (2013). The effectiveness of immigration policies. *Population and Development Review*, *39*(3), 487–508. https://doi.org/10.1111/j.1728-4457.2013.00613.x

Department of Economic and Social Affairs. (2013). *International Migration Policies*. United Nations.

Department of Homeland Security. (2022). *Yearbook of immigration statistics 2020* [Dataset]. https://www.dhs.gov/immigration-statistics/yearbook/2020

Donato, K. M., & Amuedo-Dorantes, C. (2020). The legal landscape of U.S. immigration: An introduction. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *6*(3), 1-16. https://doi.org/10.7758/rsf.2020.6.3.01

History.com Editors. (2021). *September 11 Attacks*. History. https://www.history.com/topics/21st-century/9-11-attacks

Kalleberg, A. L., & Wachter, T. M. V. (2017). The U.S. labor market during and after the great recession: Continuities and transformations. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *3*(3), 1-19. https://doi.org/10.7758/rsf.2017.3.3.01

Marchand, M. H. (2020). The caravanas de migrantes making their way north: Problematising the biopolitics of mobilities in Mexico. *Third World Quarterly*, *42*(1), 141–161. https://doi.org/10.1080/01436597.2020.1824579

Massey, D. S. (2020). Immigration policy mismatches and counterproductive outcomes: Unauthorized migration to the U.S. in two eras. *Comparative Migration Studies*, *8*(1). https://doi.org/10.1186/s40878-020-00181-6

Meyers, E. (2000). Theories of international immigration policy: A comparative analysis. *International Migration Review*, *34*(4), 1245–1282. https://doi.org/10.1177/019791830003400407

Office of Immigration Statistics. *Definition of Terms*. (2018). Department of Homeland Security. https://www.dhs.gov/immigration-statistics/data-standards-and-definitions/definition-terms

National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Committee on Population, & Panel on the Integration of Immigrants into American Society. (2016). *The integration of immigrants into American society* (Illustrated ed.) (M. G. Pineau, & M. C. Waters, Ed.). National Academies Press.

Panetta, T. (2005). Human rights and the war on terror: The USA Patriot Act. *Human Rights & Human Welfare*, *5*(1), 163–182. https://digitalcommons.du.edu/hrhw/vol5/iss1/51

Paparrizos, J., & Gravano, L. (2015). k-Shape: Efficient and accurate clustering of time series. *SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870. https://doi.org/10.1145/2723372.2737793

Phillips, J. A., & Massey, D. S. (1999). The new labor market: Immigrants and wages after IRCA. *Demography*, *36*(2), 233–246. https://doi.org/10.2307/2648111

Rystad, G. (1992). Immigration history and the future of international migration. *International Migration Review*, *26*(4), 1168–1199. https://doi.org/10.1177/019791839202600405

Sardá-Espinosa, A. (2019). Time-series clustering in R using the dtwclust package. *The R Journal*, *11*(1), 22–43. https://doi.org/10.32614/rj-2019-023

Sequeira, S., Nunn, N., & Qian, N. (2019). Immigrants and the making of America. *The Review of Economic Studies*, *87*(1), 382–419. https://doi.org/10.1093/restud/rdz003

Sherman, A., Trisi, D., Stone, C., Gonzales, S., & Parrott, S. (2019). *Immigrants contribute greatly to U.S. economy, despite administration's "Public Charge" rule rationale*. Center on Budget and Policy Priorities. https://www.cbpp.org/research/poverty-and-inequality/immigrants-contribute-greatly-to-us-economy-despite-administrations

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

U.S. Border Patrol. (2021). *U.S. Border Patrol Monthly Encounters (FY 2000 - FY 2020)* [Dataset]. U.S. Customs and Border Protection. https://www.cbp.gov/document/stats/us-border-patrol-monthly-encounters-fy-2000-fy-2020?_ga=2.39847586.1406906253.1629969446-988257266.16293

U.S. Bureau of Labor Statistics, & U.S. Department of Labor. *Foreign-born workers made up 17.4 percent of labor force in 2019.* (2020). The Economics Daily. https://www.bls.gov/opub/ted/2020/foreign-born-workers-made-up-17-point-4-percent-of-labor-force-in-2019.htm

U.S. Citizenship and Immigration Services. *Post-9/11*. (2019). U.S. Citizenship and Immigration Services. https://www.uscis.gov/about-us/our-history/post-911