

Nation of Immigrants

A Longitudinal Study of the Relationship of Immigration Laws with Changes in
Demographics and GDP in the United States

Jose Zuniga

May 01, 2018

Contents

Abstract	1
Introduction	1
Overview	1
Theory	2
Hypothesis	2
Question	2
Background	3
Definitions	3
Literature Review	3
Primary Data Sources	4
Data Characteristics	7
Legislation	7
Demographics	8
Immigration	9
GDP by Industry	10
Results	10
Correlation Analysis	10
Bidirectional Impact of Policy Lag	11
Legislation and Demographic Factors	11
Legislation and Economic Factors	12
Demographic and Economic Factors	13
Causality Analysis	14
Model Selection	15
Coefficient Analysis	15
Partial Least Squares (PLS)	15
Robust Linear Model (RLM)	16
Discussion	17
Correlations	17
Causation	18
Regression	18
Conclusion	19
Appendix	21
Tables and Figures	21
Code Listings	51

Abstract

Immigrants come to the United States looking to benefit from opportunities available in the U.S. labor market that they are unable to find in their country of origin. The current presidential administration is proposing immigration reforms focusing on merit-based admissions and enforcement of existing laws that it asserts will benefit the economy. Political opponents claim the proposed policies have racial undertones and are meant to reduce diversity in the United States. This paper conducts correlational, causality, and regression coefficient analyses to dissect and describe the underlying relationships between legislation on immigration, demographics, and the economy. Based on these empirics, the likely outcome of the proposed immigration reforms is then conjectured.

Introduction

Overview

Since the United States passed its first immigration law in 1790,¹ the U.S. Congress has passed over 200 other immigration laws and amendments an average of once every year.² Legislation has been exclusionary in some cases such as the Chinese Exclusion Act of 1882, and inclusionary on other occasions such as the Chinese Exclusion Repeal Act of 1943. The most recent immigration law passed was the Javier Vega, Jr. Memorial Act of November 2, 2017. The law named a Border Patrol checkpoint in Texas after fallen Border Patrol agent Javier Vega, Jr. According to the White House, the next round of immigration reforms will focus on enforcement, DACA legalization, sponsorship limitations, and visa restrictions.³ Even though the impact of immigration laws has varied over time, high-impact reforms like those proposed tend to bring racial and/or ethnic intent into question. The intent behind the proposed reforms is currently a controversial topic which cannot be quantified, but it is quite clear that a change in direction is definitely happening. On February 23, 2018 the United States Citizenship and Immigration Services removed the phrase “nation of Immigrants” from its mission statement.

Table 1: U.S. Citizenship and Immigration Services

Former Mission Statement	Current Mission Statement
USCIS secures America’s promise as a <i>nation of Immigrants</i> by providing accurate and useful information to our customers, granting immigration and citizenship benefits, promoting an awareness and understanding of citizenship, and ensuring the integrity of our immigration system. ⁴	U.S. Citizenship and Immigration Services administers the nation’s lawful immigration system, safeguarding its integrity and promise by efficiently and fairly adjudicating requests for immigration benefits while protecting Americans, securing the homeland, and honoring our values. ⁵

These changes are happening in the context of a growing number of movements in opposition to demographic changes in the United States. Several news outlets are conflating the proposed immigration policy changes with these factions. Therefore, the significance of the proposed reforms is not just rooted in the economic impact that the United States’ has on the global economy, but also on the shift of the country’s long-standing identity from a ‘Nation of Immigrations’ to something new. This research study will analyze the proposed immigration reforms in the context of the relationships between immigration laws, demographics, and the economy. There are empirical studies on these individual factors, but there are not many that look at the relationship between immigration laws, demographics, and the economy throughout history. Studies by the

¹<https://www.migrationpolicy.org/research/timeline-1790>

²<https://www.archives.gov/files/research/naturalization/420-major-immigration-laws.pdf>

³<https://www.whitehouse.gov/briefings-statements/white-house-framework-immigration-reform-border-security/>

⁴<https://web.archive.org/web/20180221163615/https://www.uscis.gov/aboutus>

⁵<https://www.uscis.gov/aboutus>

U.S. Census Bureau on historic demographics do not address immigration law or the economy. The Migration Policy Institute⁶ has reports on immigration flows based on data from the U.S. Department of Homeland Security, but this information is mostly demographic. A recent study was published in the Washington Post on the impact of the proposed immigration reforms on demographics and the economy, but it does not address anything prior to 2015. This research project looks at the relationship between immigration laws, demographics, and the economy at a very granular level.

Theory

In the current political climate, restrictions on immigration are rationalized as means to protect the average worker in the United States from the influx of low-cost labor. Kaufman and Hotchkiss (2006) find however, that the claim has not been supported by empirics.⁷ From an empirical and historical perspective, the proposed restrictions on immigration appear to be motivated by desires to reduce the rate of racial and ethnic diversity in United States. It is true that Immigrants come to the United States looking to benefit from opportunities available in the U.S. labor market that they are unable to find in their country of origin. It is also true that standard economic theories about competitive markets posit that immigration increases the supply of labor which in turn decreases the cost of labor (wages). Yet Kaufman and Hotchkiss note that the impact of immigration on wages has been found to be either nonexistent or small; even when the influx is large and sudden. Kaufman and Hotchkiss show how this contradiction has led to explanations such as those claiming that immigrants are not perfect substitutes for native workers due to language. They also describe researchers examining the same data with techniques that were “sufficiently refined and sophisticated to capture the true [negative] effect of immigration on wages.” Nowadays such repeated measures, if not controlled for, would constitute data dredging. Although the impact of immigration on wages is debatable, Kaufman and Hotchkiss are unambiguous about the positive correlation between immigration and economic growth—economic booms and immigration go hand-in-hand.

Hypothesis

The frequency of legislation regarding immigration fluctuates when matters related to immigration are under review. Therefore, the number of laws being passed on immigration is a useful sentiment indicator for how the United States feels about past and future immigration. The hypothesis of this analysis is that there exists a feedback loop between the number of immigration laws enacted and the demographic composition of the United States. The number of immigration laws enacted in the future is a function of demographic changes—and demographics change in response to the number of immigration laws enacted in the past. It is further hypothesized that the positive correlation that Immigrants share with various sectors of the U.S. economy holds regardless of the immigrant’s country of origin. Therefore, the proposed immigration restrictions based on country of origin are more likely to impact the diversity rather than the economy of the United States.

Question

The current presidential administration is proposing immigration reforms focusing on merit-based admissions and enforcement of existing laws that it asserts will benefit the economy. Political opponents claim the proposed policies have racial undertones and are meant to reduce the diversity of Immigrants. How have immigration reforms impacted the economy and the demographics of the United States historically and what is the likely outcome of the proposed immigration reforms?

⁶<https://www.migrationpolicy.org/>

⁷Kaufman, Bruce E., and Julie L. Hotchkiss. *The Economics of Labor Markets*. Thomson/South-Western, 2006.

Background

Definitions

The U.S. Census Bureau defines Natives as natural born citizens and children born abroad to citizen parent(s). Foreign-born are defined as permanent residents, naturalized citizens, authorized migrants, and unauthorized migrants. Other Race is a less well-defined category that has been found to capture different racial and ethnic categories which survey respondents feel are not reflected in the defined categories. Some examples are Hispanics before a category was added, individuals who identify race with country of origin, and multiracial individuals. The U.S. Department of Homeland Security refers to Foreign-born as aliens and subdivides their population into Immigrants, Nonimmigrants, and Inadmissible. Immigrants are foreigners permanently relocating to the U.S. and Nonimmigrants are foreigners relocating to the U.S. temporarily.⁸ Foreigners who are Inadmissible are those “not permitted by law to enter or remain in the United States.” A visa is an authorization to enter the United States. “A citizen of a foreign country who seeks to enter the United States generally must first obtain a U.S. visa.”

Literature Review

The U.S. Census Bureau has a demographic study examining the Foreign-born population by country from 1850 to 2010.⁹ Interesting revelations in this study are the steady decline in the Foreign-born population from 1910 to 1970 and then steady increase thereafter. In the last period examined, 2010, the Foreign-born population is still not as high as the period between 1860 and 1910. The study breaks down Natives and Foreign-born by State, Country or Origin, education level, preferred language, participation in the labor force, occupation, income, and age.

An article published by the U.S. National Archives in 2002 by Senior Historian Marian Smith takes an in depth look at demographics and immigration law. *Race, Nationality, and Reality* goes into great detail about how immigration policy is underlined by a long history of racism where Immigrants were excluded based on race and ethnicity. Immigration and nationality were legislated separately in the early years of the United States. Early Immigration laws focused on keeping out the poor and sick. Early Nationality laws focused on granting “free white persons” who immigrated from Europe citizenship while excluding African-Americans and untaxed Native-Americans. Federal citizenship was not even seen as that important until after the Civil War. So, when immigration and naturalization were placed under the administration of one agency, the coordination and uniform application of different laws designed to restrict race (as defined by color) and geographic origin became a problem. Attempts to mitigate the confusion by focusing only on ethnicity exacerbated the problems. Citizenship was being granted and denied based on color, origin, or ethnicity depending on the individual case. Several pieces of legislation tried, but failed, to codify the “common understanding” of who was entitled to citizenship. Exceptions either for and against citizenship became commonplace. The lack of clarity and guidance was resulting in courts flooded with appeals. Marian states that “there was no settled opinion on what constituted evidence of race [and] as long as nationality law contained racial requirements for naturalization, and immigration law excluded those ineligible to naturalize,” there was no way to stop the problem. Then, in 1952, immigration and naturalization laws were finally combined into a more cogent form with racial requirements completely removed.¹⁰

Another study published in the Washington Post last on February 26, 2018 looks at how the cuts to legal immigration proposed by the current presidential administration will impact demographics. The report states that the “immigration plan could keep whites in U.S. majority for up to five more years... [by] greatly slashing the number of Hispanic and black African Immigrants entering the United States.” This study looks at the proposed immigration policies in combination with population projections, fertility figures, and visa backlog information. It concludes by postulating that the proposed reforms will decrease the size of the

⁸https://help.cbp.gov/app/answers/detail/a_id/72

⁹https://www.census.gov/newsroom/pdf/cspan_fb_slides.pdf

¹⁰<https://www.archives.gov/publications/prologue/2002/summer/immigration-law-1.html>

population, which will therefore decrease the size of the workforce, which will therefore decrease the size of the economy.¹¹

Primary Data Sources

U.S. Census Bureau

The U.S. Census Bureau has wealth of information on their website¹² regarding the U.S. population going back as far as 1790.¹³ The data compiled on race and Hispanic origin consists of the *Historical Census Statistics on Population Totals by Race and by Hispanic Origin*¹⁴ which spans 1790 to 1990 on race and 1850 to 1990 on Hispanic origin. Hispanic origin is defined as Spanish mother tongue, culture, surname, or origin regardless of race. More recent observations of these data are available in the *Profile of General Demographic Characteristics* (DP-1) data tables from the 2000 Census¹⁵ and the 2010 Census.¹⁶ The most recent observations of these data, from 2011 to 2016, are found in the 1-year *American Community Survey Demographic and Housing Estimates* (DP5) data tables.

The data on Foreign-born population comes from tables in the *Historical Census Statistics on the Foreign-born Population of the United States*¹⁷ which spans 1850 to 2000. The *Nativity of the Population*¹⁸ table outlines the number of persons that are native and Foreign-born. The *Region of Birth of the Foreign-born Population*¹⁹ table breaks down the origin of the Foreign-born population into Europe, Asia, Africa, Oceania, Latin America, and Northern America. The details on which countries are sub-regions of these macro geographic regions is outlined on pages 46-47 of the United Nations 2002 Demographic Yearbook.²⁰ The most recent observations of these data, from 2011 to 2016, are found in the 1-year *Selected Characteristics of the Native and Foreign-born Populations* (S0501) data tables.

The dataset compiled from all these Census data is quite impressive, but it is not yet ready for use. Several nuances must be addressed. Although analyzing the cause of these discrepancies is not a subject of this paper, it is worth noting that -while not surprising given the history of the hegemony in the United States-the discrepancies are quite unsettling. Notes from the metadata explain that:

1. Although the Census includes data on the free and slave African-American population, “information on nativity was not collected for slaves.”
2. “The attempt to enumerate all American Indians started in 1890” with earlier estimates being inaccurate because “American Indians not taxed (i.e., living in tribal society) were not included in the enumeration.”
3. The 1940 Hispanic population figure only included the “White population of Spanish mother tongue.”
4. In 1940 and 1950, there is Foreign-born data missing “because data on the Foreign-born population by country of birth in census publications for these years are limited almost entirely to the White population.”

To address the discrepancy caused by (1), the number of slaves disembarking in the United States²¹ is used as a proxy for Foreign-born Africans entering the United States from 1790 to 1880. To address (2), the Native-American population figures are backfilled using the percent change²² of the Native-American population starting from the earliest period without discrepancies which is 1890. Native born and total population figures are also adjusted to reflect the revised Native-American population figures. To address (3), the value was removed since it was not representative of all Hispanics. Then, to complete the dataset,

¹¹<https://www.washingtonpost.com/news/wonk/wp/2018/02/06/immigration-plan>

¹²<https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml>

¹³<https://www.census.gov/prod/www/decennial.html>

¹⁴<https://www.census.gov/population/www/documentation/twps0076/twps0076.html>

¹⁵<https://www.census.gov/prod/cen2000/dp1/2kh00.pdf>

¹⁶<https://www.census.gov/history/pdf/orangenj-92017.pdf>

¹⁷<https://www.census.gov/population/www/documentation/twps0081/twps0081.html>

¹⁸<https://www.census.gov/population/www/documentation/twps0029/tab01.html>

¹⁹<https://www.census.gov/population/www/documentation/twps0029/tab02.html>

²⁰<http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm>

²¹<http://www.slavevoyages.org/assessment/estimates>

²²<https://nativestudy.wordpress.com/>

imputation is used. The missing data outlined in (4) for the Foreign-born in 1940 and 1950 is filled in using linear interpolation between 1930 and 1960. Linear interpolation is also used to fill in missing annual Census data. This is because the Census data obtained is decennial from 1790 to 2010 and then annual until 2016 while all the other non-Census data are annual.

$$y = y_1 + (x - x_1) \left(\frac{y_2 - y_1}{x_2 - x_1} \right)$$

Linear interpolation²³ makes no assumption on distribution. It draws a straight line between two known points and returns a value along that line for a given point in the domain. The remainder of the missing data is filled in using linear regression based on the natural log of the total population and the natural log of the variable being imputed.

$$\hat{y} = e^{(\beta_0 + \ln(x_i)\beta_1)}$$

The exceptions to the imputation via interpolation and regression are the statistics on Foreign-born persons from Europe and North America. Linear regression is not suited for these nonlinear data. Instead, imputation for other regions are subtracted from the total Foreign-born population and distributed according to weights based on the historical averages for the two regions. The resulting dataset is a complete dataset²⁴ from 1790 to 2016 on race, Hispanic origin, nativity, and country of origin for Foreign-born.

U.S. Department of State

Immigrant visas are issued to foreigners permanently relocating to the U.S. and Nonimmigrant visas are issued to foreigners relocating to the U.S. temporarily. Data regarding visas is available both from the U.S. Department of Homeland Security²⁵ and the U.S. Department of State.²⁶ The data²⁷ from the U.S. Department of State shows every Nonimmigrant visa type²⁸ by country from 1997 to 2017. This dataset has an extraordinary amount of detail, but only on Nonimmigrant visas issued by consular offices which is only a fraction of the Nonimmigrant visas. Therefore, these data are not suited for this analysis.

U.S. Department of Labor

Certain Immigrant and Nonimmigrant employment related visas require approval from the U.S. Department of Labor. Employers “must obtain foreign labor certification from the U.S. Department of Labor, prior to filing a petition” for these workers. The process of obtaining certification from the U.S. Department of Labor involves substantiating²⁹ that “there are not sufficient U.S. workers able, willing, qualified and available to accept the job opportunity in the area of intended employment and that employment of the foreign worker will not adversely affect the wages and working conditions of similarly employed U.S. workers.” The types of employment related visas subject to this certification process are:

- **E3:** Australian Free Trade Agreement principal’s family.
- **H1B:** Temporary workers in specialty occupations.
- **H1B1:** Chile and Singapore Free Trade Agreement aliens.
- **H2A/H2B:** Agricultural and Nonagricultural workers.
- **D1:** Crewman on a marine vessel or aircraft.

²³<http://mathworld.wolfram.com/Interpolation.html>

²⁴https://github.com/jzuniga123/SPS/blob/master/DATA%20698/US_Census_Data.csv

²⁵<https://www.uscis.gov/tools/reports-studies/immigration-forms-data>

²⁶<https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics.html>

²⁷<https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics.html>

²⁸<http://www.ustraveldocs.com/in/in-niv-typeall.asp#>

²⁹<https://www.foreignlaborcert.doleta.gov/perm.cfm>

The data³⁰ maintained by the U.S. Department of Labor on these employment related visas lists all information, including status, on every application from 2008 to 2017. There also exists an external source³¹ that has harvested data dating back to 2000. Although the U.S. Department of Labor data has an extraordinary amount of detail, it only encompasses a small number of the available employment related visas.³²

U.S. Department of Homeland Security

The U.S. Citizenship and Immigration Services division of the U.S. Department of Homeland Security has compiled the most complete list of immigration legislation available for the period between 1790 and 1996. The list is not currently available from the source³³ but it continues to exist in secondary sources such as the Immigration Daily³⁴ website and the book *Immigration, Assimilation, and Border Security* by Yoku Shaw-Taylor from 2012. The list is fairly exhaustive, but it is missing about eight pieces of legislation from a smaller list³⁵ maintained by the U.S. National Archives. Legislation from 1997 on can be found by querying a database³⁶ maintained by the U.S. Congress. The database contains legislation since the 93rd Congress in 1973. The legislation on immigration that has become law since the 105th Congress in 1997 from this database, when combined with the information from the U.S. Citizenship and Immigration Services and U.S. National Archives, produces complete list³⁷ of immigration legislation from 1790 to 2017.

The U.S. Citizenship and Immigration Services division also maintains data on immigration application forms by type (Naturalization, Employment, Family, Humanitarian, Other) and approval status from 2012 to 2017. Data on I-129 applications used for employment are available since 2009. Combining all years of available data from the U.S. Citizenship and Immigration Services results in an unremarkable dataset.³⁸ The most useful data³⁹ on immigration relevant to this research are the statistics produced by the U.S. Department of Homeland Security in their annual yearbooks. The data on Immigrant visas from their *Persons Obtaining Lawful Permanent Resident Status by Type and Major Class Of Admission*⁴⁰ tables spans 1986 to 2016. The data on *Nonimmigrant visas from their Nonimmigrant Admissions by Class of Admission*⁴¹ tables spans 1981 to 2016 with aggregate data on Immigrant visas since 1820. There are also data on *Aliens Apprehended*⁴² from 1925 to 2016, as well as on *Aliens Removed or Returned*⁴³ from 1892 To 2016. Apprehensions are foreigner arrests, Removals are compulsory based on a court order, Returns represent immigrants stopped (without a court order) from entering at the borders. From these data, a subset of the below Immigrant and Nonimmigrant employment visas has been carved out for additional analysis:

- **CW1/CW2:** CNMI-only transitional workers and family.
- **E1, E2, E2C:** Treaty traders or investors and family.
- **E3:** Australian Free Trade Agreement principal's family.
- **H1B:** Temporary workers in specialty occupations.
- **H1B1:** Chile and Singapore Free Trade Agreement aliens.
- **H1C:** Registered nurses participating in the Nursing Relief for Disadvantaged Areas.
- **H2A/H2B/H2R:** Agricultural/Nonagricultural/Returning workers.
- **H3/H4:** Trainees and Spouses and children of H1, H2, or H3.
- **I1:** Representatives of foreign information media and family.
- **L1/L2:** Intracompany transferees and family.
- **O1/O2/O3:** Workers with extraordinary ability, assistants, and family.

³⁰<https://www.foreignlaborcert.doleta.gov/performance/cfm>

³¹<http://www.flcdatacenter.com/>

³²<https://travel.state.gov/content/travel/en/us-visas/visa-information-resources/all-visa-categories.html>

³³<https://guides.law.fsu.edu/c.php?g=84898&p=546805>

³⁴<https://ilw.com/resources/>

³⁵<https://www.archives.gov/files/research/naturalization/420-major-immigration-laws.pdf>

³⁶<https://www.congress.gov/search>

³⁷https://github.com/jzuniga123/SPS/blob/master/DATA%20698/ImmigrationLaws_1790to2012.csv

³⁸https://github.com/jzuniga123/SPS/blob/master/DATA%20698/USCIS_2009to2017.csv

³⁹<https://www.dhs.gov/immigration-statistics/yearbook>

⁴⁰<https://www.dhs.gov/immigration-statistics/yearbook/2016/table6>

⁴¹<https://www.dhs.gov/immigration-statistics/yearbook/2016/table25>

⁴²<https://www.dhs.gov/immigration-statistics/yearbook/2016/table33>

⁴³<https://www.dhs.gov/immigration-statistics/yearbook/2016/table39>

- **P1/P2/P3/P4:** Internationally recognized athletes, entertainers, or artists, and family.
- **Q1/Q2/Q3:** Workers in international cultural exchange programs and family.
- **R1/R2:** Workers in religious occupations and family.
- **TN/TD:** NAFTA professional workers and family.

The full join of these Immigrant and Nonimmigrant visa data are made uniform across their combined time domain by imputing missing years. Nonimmigrant data are imputed for 1982-1984, 1986-1988, and 1997 using linear interpolation between years where the data are present: [[1981, 1985], [1985, 1989], [1996, 1998]]. The missing Immigrant data from 1981 to 1985 is filled in by first imputing total visas issued using the historic average proportion of Nonimmigrant visas relative to total visas, then by subtracting Nonimmigrant visas by the derived number of total visas. The resulting dataset⁴⁴ has complete cases on employment and total visas for both Immigrants and Nonimmigrants from 1981 to 2016. The dataset also includes Removals since 1892, Apprehensions since 1925, and Returns since 1927. No imputation was needed for these three enforcement actions since there are no missing data during the 1981 to 2016 period.

U.S. Bureau of Economic Analysis

The U.S. Bureau of Economic Analysis has data on *Gross Output by Industry*⁴⁵ from 1987 to 2016. The data are very granular and consists almost entirely of complete cases. The exceptions are from 1987 to 1996 in the subcategories of (1) Housing, (2) Other Real Estate, (3) General Government National Defense, (4) General Government Nondefense, and (5) Nursing and Residential Care Facilities. For the (1), (2), (3), and (4) subcategories, imputations are done using the average proportion of the child node under the parent node from the data that are present. This imputation is relatively easy since these subcategories are the only child nodes of Real Estate and General Government. The imputation for (5) is done using the average proportion of the sum of its sibling nodes which were Ambulatory Health Care Services and Hospitals. The two existing sibling nodes that the imputations are based on, are then adjusted downward to reflect the new sibling node under the Health Care parent. Failing to adjust them downward would inflate Health Care. The adjustment of the existing sibling nodes is based on the average proportion of each node relative to each other in the existing data. The data also require format changes as final step. The format they are provided in is great for display purposes, it is not amenable to analysis. The final compilation is a complete dataset⁴⁶ from 1987 to 2016 that breaks down GDP by sector: Agriculture, Mining, Utilities, Construction, Manufacturing, Wholesale, Retail, Transportation, Information, FIRE, Service, Education, Entertainment, Other, Federal, State and Local.

Data Characteristics

The dataset⁴⁷ constructed from the arduous data gathering and cleaning process described above consists of Immigration Laws and Census data from 1790 to 2016, employment visa data from 1981 to 2016, enforcement data from 1925 to 2016, and GDP by Industry from 1987 to 2016. Correlational research is being conducted on the dataset to describe relationships between immigration law, demographics, employment visas, and the economy. The significance of the relationships between variables is evaluated using formal statistical tests.

Legislation

Historic immigration law numbers have an interesting pattern. The period between 1790 and 1931 saw an average of 0.303 pieces legislation per decade. The relative frequency of legislation did increase around the time of Mexican-American War in 1846, but the number of laws passed continued to remain relatively low. Then in 1932, in the midst the Great Depression, the number and frequency immigration laws started to steadily increase. Since 1932 there have been an average of 2.059 immigration laws being passed every year.

⁴⁴https://github.com/jzuniga123/SPS/blob/master/DATA%20698/DHS_1820to2016.csv

⁴⁵<https://www.bea.gov/iTable/iTable.cfm?ReqID=51&step=1#reqid=51&step=2&isuri=1>

⁴⁶https://github.com/jzuniga123/SPS/blob/master/DATA%20698/GDPbyIndustry_1987to2016.csv

⁴⁷https://github.com/jzuniga123/SPS/blob/master/DATA%20698/DATA698_Project.csv

Looking only at the period since 1997 shows an average of 3.65 immigration laws passed annually. The last 21 years have seen legislation on immigration being enacted at a frequency over 10 times greater than what was witnessed during the first 141 years of Federal legislative action on immigration.

Notable periods of increased legislative activity that were outliers in their era are:

- **1798:** Three laws passed in response to the Quasi-War with France.
- **1882:** Two laws passed, one being the Chinese Exclusion Act.
- **1887:** Two laws passed restricting labor contracts and property rights.
- **1903:** Two laws passed, one restricting anarchists.
- **1932:** Four laws passed, three restricting labor contracts.
- **1940:** Four laws passed, three enhancing enforcement activities.
- **1950:** Four laws passed, three providing relief to orphans, military, and sheepherders.
- **1978:** Seven laws passed, two restrictive to agriculture and travel, five progressive: addressing bias in admissions, expulsion of Nazis, human smuggling, and adoption.
- **2000:** Thirteen laws passed during an election year, eleven prior to elections.
- **2002:** Ten laws passed in the wake of the September 11, 2001 terrorists attacks.
- **2008:** Six laws passed during an election year, all prior to elections.

Demographics

Race and Hispanic Origin (RAHO)

The purely numerical data from the decennial Census on race and Hispanic origin has two interesting features. As is usual with most social phenomena, the distributions appear mostly exponential. The two exceptions are the White race which can be described as linear since about 1870 and the Native-American race which has a positive parabolic shape with a minimum at the year 1900. For the purposes of aggregate analysis, the free and enslaved African-American population that the Census counts separately until 1860 is being counted together.

When the Census data on race and Hispanic origin are viewed as a percentage of total population however, some striking details emerge. Hispanics, Asians, Pacific Islanders, and Other races grow exponentially as a percentage of total population over time. Native-Americans have decreased exponentially as a percentage of total population over time, from 14.2% in 1790 to 0.7% in 2016. The African-American percentage of the population has decreased overall but in a cubic manner with a minimum of 9.5% in 1930 and two local maximas of 17.7% in 1810 and 11% in 1990. The White percentage of the population has a negative parabolic shape with a maximum of 89% in 1920 and an all time low of 61.7% in 2016. In aggregate, many of these individual characteristics seem less remarkable, but a demographic shift is clear.

Native and Foreign-born (NAFB)

The decennial numbers on the Native and Foreign-born population also follow what can be described as an exponential pattern. There was a decline in the Foreign-born population from 1940 to 1970, but by 1980 the number returns to its prior levels and continues on its prior trajectory. When aggregated, the fluctuation in the Foreign-born population is much less noticeable.

The proportion of Native and Foreign-born persons in the population are mirror images of each other. The median proportion of Foreign-born persons in the United States since 1790 has been 12.4%. In 1850 and 1970 the percentage of Foreign-born persons in the United States reached local minmas of 9.5% and 4.8%, respectively. The proportion of Foreign-born persons in the United States has peaks in 1870, 1890, and 1910 at 14.3%, 14.8%, and 14.7%; respectively. In 2016 the percentage of Foreign-born sits at 13.5%. When viewed in aggregate, the decline in the percentage of Foreign-born persons in the United States in 1970 appears to be significant relative to its proportion over time.

Foreign-born Country of Origin (FBCO)

The decennial numbers on country of origin for Foreign-born are a little more interesting. Persons from Africa, Asia, Oceania, and Latin America arrive at an exponential rate. `grangertest(ChickEgg[, 1], ChickEgg[, 2], order = 3)` Arrivals from Europe and North America however, show exponential growth until around 1930, and then steady declining thereafter. When aggregated, these data have a very interesting characteristic. There is a large hump in the number of Foreign-born Europeans that begins to deflate when race stopped being a factor in admissions around 1950.

The proportion of Foreign-born persons in the United States by country of origin shows exponential increases from Asia, Oceania, and Latin America just like the pure numbers. The large increases start around 1860 for Oceania, and 1970 for Latin America and Asia. Foreign-born persons from Africa follow a less natural pattern that coincides with slavery. The percentage of Foreign-born persons from Africa was high from 1790 to 1820, but relatively low thereafter. Looking at Foreign-born persons from Africa in the post-slavery era shows a more natural exponential pattern with a large increase starting in 1980. The proportion of Foreign-born persons from Europe remained above 80% until 1950 when a permanent and rapid decline started leading to its current low of 10.9% in 2016. The proportion of Foreign-born persons from North America has a very strange and unnatural pattern hovering around 8.9% until 1970 when a rapid decline started leading to its current low of 1.8% in 2016. The rapid decline of Europeans in 1950 and the rapid decline by North Americans starting 1970 are temporally proximate to the when race stopped being a factor in admissions around 1950. When aggregated, the fall from dominance by Foreign-born Europeans and the rise to dominance by Latin Americans and Asians over time is quite remarkable.

Immigration

Visas

The numeric data on Immigrant Visas since 1820 is best described as chaotic. It has wild swings up and down starting with an upward trend from 1820 to about 1910, followed by a drastic decline until about 1945, and then another trend upward with a similar trajectory until 2016. The graph also shows a huge spike around 1990. When aggregated, a clear upward trend is notable, but fluctuations seem relatively random.

Visa data since 1981 has been less chaotic. Employment related Immigrant visas over the last 30 years are not as noisy. The number of Immigrant visas related to employment held steady from 1981 to 1991, was chaotic from 1992 to 2007, and then steady ever since. There were large decreases around 1999 and 2004, and large increases around 1993, 2001, and 2006. Total immigrant visas follow a similar pattern with the exception of the huge spike around 1990 mentioned earlier. Nonimmigrant visas have a less interesting pattern. Employment related visas have a steadily increasing trend with only a slight amount of variability. All Nonimmigrant visas also have an increasing trend, but with less variability. Nonimmigrant visas have made up between 89% and 95% of all visas for the last 30 years.

Enforcement

Enforcement data has some interesting patterns. Removals remained at a constant rate until 1996 when they increased substantially. Apprehensions and Returns follow almost identical patterns. There was a large increase in enforcement in the ten-year period between 1945 and 1955 with its peak being reached in 1954. The increase started again in 1965 and has remained elevated ever since with peaks in 1986, 1996, and 2000. Around the time Removals started steadily increasing, Apprehensions and Returns started steadily decreasing. When aggregated, the three enforcement actions follow the pattern of the latter two enforcement actions which make up the bulk of the actions. Another characteristic that emerges in these data when aggregated is the recent decline in Apprehensions and Returns and the recent increase in Removals.

GDP by Industry

GDP by industry over the past 30 years has the same homoskedastic upward trend for nearly every industry. The exceptions are Agriculture Mining, Utilities and Construction which appear to be more susceptible to business cycles due to their heteroskedasticity. The homoscedastic upward trend dominates aggregated income with the exception of the period encompassing the Great Recession.

When examining the industries in aggregate as a proportion of the economy, their shares fluctuate slightly but remain relatively constant over the last 30 years. Proportionally, the bulk of GDP by order of share is maintained by the Manufacturing, FIRE (Finance, Insurance, and Real Estate), Service, and Education sectors. Over time there has been a slight decrease in the proportion of the economy captured by the Manufacturing sector and slight increases in the FIRE and Service sectors.

Results

Correlation Analysis

Correlation measures the relationship between two variables. The most commonly used measure of correlation is Pearson⁴⁸ correlation (r). Pearson correlation measures the relationship between normal linear homoskedastic variables. Measuring relationships between variables that are not normal, linear, or homoskedastic (inherently or through transformation) with the Pearson correlation formula produces misleading results. Spearman⁴⁹ correlation (ρ) and Kendall⁵⁰ correlation (τ) are non-parametric measures of correlation based on monotonic rank. Spearman correlation is more computationally efficient than Kendall correlation, but less robust. For this analysis, the accuracy of Kendall correlation is being prioritized over the speed of Spearman correlation.

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}} \quad (\text{Pearson})$$

$$\rho = \frac{\sum (x' - m_{x'})(y'_i - m_{y'})}{\sqrt{\sum (x' - m_{x'})^2 \sum (y' - m_{y'})^2}} \quad (\text{Spearman})$$

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (\text{Kendall})$$

The Census data obtained is decennial from 1790 to 2010 and then annual until 2016. All other non-Census data is annual and complete from 1987 until 2016. To conduct a proper analysis of the past 30 years, imputation is required on the Census data for missing periods. This is done with linear interpolation⁵¹ between periods that there are data points present. Linear interpolation makes no assumption on distribution. It draws a straight line between two known points and returns a value along that line for a given point in the domain.

$$y = y_1 + (x - x_1) \left(\frac{y_2 - y_1}{x_2 - x_1} \right)$$

⁴⁸<http://mathworld.wolfram.com/CorrelationCoefficient.html>

⁴⁹<http://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html>

⁵⁰<https://onlinecourses.science.psu.edu/stat509/node/158>

⁵¹<http://mathworld.wolfram.com/Interpolation.html>

Bidirectional Impact of Policy Lag

There is a lag between the enactment of a law and its full implementation. Changes in laws necessitate the adjustment of institutional processes which are mired by the rules and procedures guiding the bureaucracies implementing the laws. It is worth noting that bureaucracies have been maligned with the negative connotation behind red tape, but the bureaucratic adherence to rules is a feature—and not a bug—that hedges against chaos. The point is that there is no value judgment being made here. It is simply true that Immigration applications are processed on an ongoing basis while laws change and agencies readjust. Applications in progress must be managed around changing immigration laws. When looking at the relationship between legislation and other variables, it is therefore important to also look at the lagged relationships; and not just in one direction. When laws are reactive rather than proactive, there is also a lag between the motivation for a law and its enactment.

Legislation and Demographic Factors

First the Kendall correlation between the number of new, one- and two-year old, and one- and two-year out immigration laws and demographic factors is examined. When it comes to population based on Race and Hispanic Origin, legislation has positive correlations between $\tau = 0.49$ and $\tau = 0.56$ for Hispanic origin and nearly every race. The strongest correlations are with the number of laws passed one year in the future. The exception to this pattern is Native-American population which has a positive correlation that hovers around $\tau = 0.25$ which holds the strongest relationship with laws passed two years in the future. All of these correlations are significant at $\alpha = 0.05$. Increases in racial and ethnic subsections of the population are associated with future increases in the number of immigration laws. The strongest correlations are with Other Races then Asian and Pacific Islanders.

Table 2: Laws with Race and Hispanic Origin

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
White	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes
Black	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes
Native Am	0.225, Yes	0.236, Yes	0.255, Yes	0.272, Yes	0.282, Yes
Asian Pac	0.517, Yes	0.527, Yes	0.530, Yes	0.534, Yes	0.530, Yes
Other Race	0.513, Yes	0.529, Yes	0.543, Yes	0.560, Yes	0.555, Yes
Hispanic	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes

For the Native and Foreign-born population, the correlations with the number of immigration laws passed are also moderate and statistically significant at $\alpha = 0.05$. Natives have positive correlations that hover around $\tau = 0.5$. Foreign-born have positive correlations that hover around $\tau = 0.43$. The strongest correlations are with the number of laws passed one year in the future. Increases in the native and foreign-born subsections of the population are associated with future increases in the number of immigration laws.

Table 3: Laws with Native and Foreign-born

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
Native Born	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes
Foreign Born	0.424, Yes	0.433, Yes	0.434, Yes	0.434, Yes	0.425, Yes

The correlations with the number of immigration laws passed and Foreign-born Country of Origin are higher

for Asia, Oceania, and Latin America which have correlations between $\tau = 0.49$ and $\tau = 0.51$. The strongest correlations are with the number of laws passed one year in the future. The correlations for Africa are around $\tau = 0.38$ and exhibit the strongest relationship with laws passed two years in the future. The correlations of Europe and North America with immigration laws are between $\tau = 0.18$ and $\tau = 0.27$ with the lower correlations being with Europe. The strongest of these correlations are associated with the number of laws passed two years earlier. All of these correlations are significant at $\alpha = 0.05$. Increases in the foreign-born population from Asia, Oceania, Latin America, and Africa are associated with future increases in the number of immigration laws. The number of immigration laws passed is associated with future increases in the foreign-born population from Europe and North America.

Table 4: Laws with Foreign-born Country of Origin

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
Europe	0.221, Yes	0.218, Yes	0.206, Yes	0.195, Yes	0.182, Yes
Asia	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes
Africa	0.353, Yes	0.367, Yes	0.383, Yes	0.398, Yes	0.404, Yes
Oceania	0.495, Yes	0.506, Yes	0.507, Yes	0.507, Yes	0.501, Yes
Latin Am	0.493, Yes	0.504, Yes	0.504, Yes	0.505, Yes	0.498, Yes
North Am	0.267, Yes	0.267, Yes	0.258, Yes	0.251, Yes	0.237, Yes

Legislation and Economic Factors

The Kendall correlation between the number of new, one-year old, two-year old, and three-year old immigration laws and economic factors is examined next. The correlations are higher for employment related Visas than they are for non-employment related Visas. Although the correlations are larger, they are not very high. The correlations for employment related Visas are between $\tau = 0.16$ and $\tau = 0.315$, the largest correlations found two years after immigration laws are passed. Yet none of these correlations are significant at $\alpha = 0.05$.

Table 5: Laws with Immigration Visas

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
IV Emp	0.126, No	0.167, No	0.160, No	0.163, No	0.315, No
IV NonEmp	0.113, No	0.091, No	0.115, No	0.062, No	0.194, No
NIV Emp	0.115, No	0.150, No	0.189, No	0.237, No	0.305, No
NIV NonEmp	0.106, No	0.078, No	0.133, No	0.052, No	0.170, No

Enforcement actions have low correlations with the number of immigration laws passed. The correlations are fairly constant across lags ranging between $\tau = 0.19$ and $\tau = 0.24$. None of these correlations are significant at $\alpha = 0.05$.

Table 6: Laws with Immigration Enforcement Actions

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
Removed	0.167, No	0.191, No	0.209, No	0.223, No	0.214, No
Apprehended	0.199, No	0.184, No	0.207, No	0.213, No	0.238, No
Returned	0.220, No	0.195, No	0.215, No	0.205, No	0.215, No

The largest correlations with the number of immigration laws passed and GDP by Industry can be found after three years after immigration laws are passed. Just like with Visas however, although these correlations are larger, they are not very high. Excluding the highest and lowest correlations, these correlations range between $\tau = 0.25$ and $\tau = 0.39$ and none significant at $\alpha = 0.05$.

Table 7: Laws with the Economy

	Correlation, Significant				
	Past Laws 2yr	Past Laws 1yr	Current Laws	Future Laws 1yr	Future Laws 2yr
Agriculture	-0.067, No	0.018, No	0.086, No	0.160, No	0.199, No
Mining	-0.037, No	0.073, No	0.232, No	0.201, No	0.270, No
Utilities	0.017, No	0.108, No	0.116, No	0.191, No	0.285, No
Construction	-0.002, No	0.013, No	0.076, No	0.181, No	0.260, No
Manufacturing	-0.057, No	0.013, No	0.136, No	0.155, No	0.290, No
Wholesale	-0.082, No	-0.028, No	0.101, No	0.150, No	0.290, No
Retail	-0.077, No	-0.028, No	0.091, No	0.145, No	0.295, No
Transportation	-0.077, No	-0.028, No	0.121, No	0.165, No	0.280, No
Information	-0.087, No	-0.023, No	0.096, No	0.160, No	0.290, No
FIRE	-0.082, No	-0.013, No	0.096, No	0.160, No	0.270, No
Service	-0.087, No	-0.023, No	0.101, No	0.160, No	0.285, No
Education	-0.092, No	-0.023, No	0.091, No	0.170, No	0.290, No
Entertainment	-0.087, No	-0.023, No	0.101, No	0.160, No	0.285, No
Other Industries	-0.082, No	-0.023, No	0.096, No	0.165, No	0.285, No
Govt Federal	-0.027, No	0.013, No	0.131, No	0.186, No	0.305, No
Govt StateLoc	-0.092, No	-0.023, No	0.091, No	0.170, No	0.290, No

Demographic and Economic Factors

As previously mentioned, Kendall correlations between the economy and the number of immigration laws passed are statistically insignificant. This is not true for Kendall correlations between industries in the economy and many other factors. Correlations between industries and Race and Hispanic Origin are high and statistically significant at $\alpha = 0.05$ for every demographic segment. Income for every industry increases with increases in racial and ethnic subsections of the population.

The same is true for the Native and Foreign-born population and every industry. Correlations are high and statistically significant at $\alpha = 0.05$. Income for every industry increases with increases in native and foreign-born subsections of the population.

Country of origin does not show the same overarching relationship with the economy however. At an $\alpha = 0.05$ Asia, Africa, Oceania, and Latin America have high and statistically significant correlations with every industry, but Europe and North America have all statistically insignificant correlations. Income for every industry increases with increases in immigration from Asia, Africa, Oceania, and Latin America.

There is also a dichotomy between Immigrant and Nonimmigrant Visas and their relationship with the economy. At an $\alpha = 0.05$, employment-related Nonimmigrant Visas have statistically significant correlations with every industry, but employment related Immigrant Visas only have statistically significant correlations with Utilities. Non-employment related Nonimmigrant and Immigrant Visas do not have statistically significant correlations with any industry. Income for every industry increases with increases in employment-related Nonimmigrant Visas. Income for Utilities increases with increases in all employment-related Immigrant Visas. This relationship between employment-related Immigrant Visas and Utilities could also be a statistical anomaly. Within the context of statistical tests, at least one in twenty false positives are expected at a 95% confidence level.

Enforcement actions also have mixed results. Apprehensions and Returns have statistically insignificant

correlations with every industry in the economy, but Removals have statistically significant positive correlations with every industry in the economy at $\alpha = 0.05$. Income for Utilities for every industry increases with increases in Removals. As stated earlier, Removals are compulsory actions based on a court order and Returns represent immigrants stopped (without a court order) from entering at the borders.

The largest statistically significant correlations are found between the two years after immigration laws are passed and demographics factors. The difference between two-year lagged legislation and new legislation is not very large however, $\tau \pm 0.015$ to be exact.

Causality Analysis

$$Y_t = (a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p}) + (b_1 X_{t-1} + \cdots + b_p X_{t-p}) + u_t \quad (\text{Equation 1})$$

$$X_t = (c_0 + c_1 X_{t-1} + \cdots + c_p X_{t-p}) + (d_1 Y_{t-1} + \cdots + d_p Y_{t-p}) + v_t \quad (\text{Equation 2})$$

Granger Causality tests are useful for evaluating time series pairs for causality such that “ X is said to Granger-cause Y if Y can be better predicted using the histories of both X and Y than it can by using the history of Y alone.”⁵² Essentially, the test checks whether one variable is useful in predicting the other taking lagging of both variables into consideration. In terms of the above two equations, if in the first equation $H_A : \exists b_i \neq 0$ holds true, then X is a Granger-cause of Y . If in the second equation $H_A : \exists d_i \neq 0$ holds true, then Y is a Granger-cause of X . The test has limitations in multivariate cases and does not rule out confounding variables, but it remains one of the most useful tools for analyzing causality in model building. To conduct the test effectively, the maximum order of integration d for the time series group must first be determined.

The `auto.arima()` function in **R** chooses ARIMA models automatically using a variation of the Hyndman and Khandakar algorithm which combines unit root tests, minimization of the AICc, and MLE to obtain an ARIMA model. The results suggest that the maximum order of integration is $d = 2$. Using this maximum order of integration, tests are then run to check for the existence of Granger Causality.

Table 8: Legislation-Demographic Causality Test

	Granger Causality Statistic, Significant	
	Number of Laws (Antecedent)	Number of Laws (Consequence)
White	1.61, No	12.58, Yes
Black	0.83, No	13.51, Yes
Native Am	2.68, No	8.12, Yes
Asian Pac	2.85, No	5.86, Yes
Other Race	1.28, No	8.47, Yes
Hispanic	4.17, Yes	12.46, Yes
Native Born	3.94, Yes	13.36, Yes
Foreign Born	0.86, No	9.21, Yes
Europe	1.10, No	1.22, No
Asia	0.33, No	6.47, Yes
Africa	5.02, Yes	3.58, Yes
Oceania	1.78, No	9.98, Yes
Latin Am	5.81, Yes	11.49, Yes
North Am	2.97, No	2.27, No

The test results indicate that changes in most demographics exhibit the characteristics necessary for being a Granger cause of the number of immigration laws are passed. The exceptions are foreign-born of European

⁵²<http://davegiles.blogspot.co.uk/2011/04/testing-for-granger-causality.html>

and North American origin. Granger Causality does not imply true causality, but the results do indicate that changes in the demographic variables displaying Granger Causality are useful predictors for forecasting future changes in number of immigration laws enacted. The bidirectional causality relationship with some variables (Hispanic, Native Born, Africa, and Latin America) suggests a feedback loop which implies that changes in the number of immigration laws enacted is also a useful predictor for forecasting future changes in those four demographic elements.

Model Selection

Based on the strength of the one-year correlations and bidirectional Granger causality for some of the correlated factors, a few models will be constructed and tested. First, a model that attempts to explain the number of laws that will be passed one year in the future based on changes to multiple demographic factors will be constructed. Then, four smaller models that attempt to explain changes to the Hispanic, Native born, foreign-born from Africa, and foreign-born from Latin America population segments based on the number of laws passed one year prior will be constructed. Given the oddities in the distributions of some of these series', multiple regression model types are tuned and examined to see which model type is most suited for these data.

- Linear Regression Models
 - Robust Linear Regression (RLM)
 - Principal Component Regression (PCR)
 - Partial Least Squares (PLS)
 - Elastic Net Regression (ENET)
- Nonlinear Regression Models
 - Artificial Neural Networks (ANN)
 - Multivariate Adaptive Regression Splines (MARS)
 - Support Vector Machines (SVM)
 - K -Nearest Neighbors (KNN)
- Tree-Based Regression Models
 - Classification and Regression Tree (CART)
 - Random Forest (RF)
 - Stochastic Gradient Boosting (SGB)
 - Rule-Based Cubist (CUBE)

All the models perform almost equivalently on these data. As a matter of fact, the optimal RMSE, R^2 , and MAE training set resampling performance metrics are associated with different models depending on the random seed. The need to validate training set resampling results—which are highly optimistic as a result of the repeated sampling—with a test set is highlighted by cases such as this where the resampling performance metrics are nearly identical. In this run the respectively optimal linear, nonlinear, and tree-based models based on the RMSE of the training set resampling are the Partial Least Squares (PLS), K -Nearest Neighbors (KNN), and Stochastic Gradient Boosting (SGB) models. These training set resampling rankings hold for the PLS and KNN models after running the test set validation. Since the model performance is essentially equivalent and the goal of this modeling process is explanation of relationships between variables rather than prediction, the non-covariance based KNN model which cannot be cleanly summarized will not be used.⁵³ The focus will be on covariance based linear models.

Coefficient Analysis

Partial Least Squares (PLS)

The first model attempts to explain the number of laws that will be passed one year in the future based on changes to multiple demographic factors. There exists a very high level of multicollinearity in these

⁵³Kuhn, Max and Kjell Johnson, Applied Predictive Modeling, Springer Science+Business Media, 2013

demographic factors however. In fact, attempting to remove highly correlated predictors removes nine of the twelve variables leaving only foreign-born from Africa, Native-Americans, and Whites. These were the three variables with very unusual distributions due to slavery, genocide, and admissions being based on race until around 1950. Using a PLS model—which is essentially supervised Principal Component Analysis—is a regression technique that works well when a high level of multicollinearity is present. Applying the PLS model with two components (determined by the tuning process) to Box-Cox transformations of these data and examining the residuals for normality with a Shapiro-Wilks test suggests that the PLS model provides an appropriate fit for these data at $\alpha = 0.05$. Examining legislation as a consequence where the cause is demographic factor changes and the effect is the number of laws enacted one year later, the PLS model suggests that the variation in the number of immigration laws enacted is explained in varying amounts by changes in each of the demographic factors. It is important to note that the PLS coefficients reflect the amount of change explained by the demographic factor and not the direction of the change. The direction of change was however, noted in the Correlation Analysis where all were found to have a positive relationship. Therefore, it can be surmised that when there are increases in these specific demographic elements, the number of immigration laws enacted one year later increases. All of these jackknife variance estimates are statistically significant at $\alpha = 0.05$.

Table 9: Number of Laws (Consequence)

	Ordered Partial Least Squares (PLS) Coefficients			
	Estimate	Std. Error	t value	Pr(> t)
Other Race	0.1794285	0.0598980	2.995566	0.0150643
Native Am	0.1733696	0.0530547	3.267750	0.0097176
Africa	0.1642788	0.0325645	5.044722	0.0006953
Asian Pac	0.1376798	0.0442626	3.110526	0.0125082
Black	0.0836869	0.0104169	8.033746	0.0000214
Native Born	0.0757040	0.0084934	8.913244	0.0000092
White	0.0737364	0.0086038	8.570215	0.0000127
Foreign Born	0.0626821	0.0106390	5.891742	0.0002314
Hispanic	0.0555109	0.0092094	6.027655	0.0001958
Oceania	0.0480144	0.0083247	5.767740	0.0002702
Latin Am	0.0434573	0.0071833	6.049785	0.0001905
Asia	0.0404559	0.0070373	5.748799	0.0002767

Robust Linear Model (RLM)

The remaining four models attempt to explain changes to four demographic factors based on the number of laws that were passed one year prior. Addressing multicollinearity is unnecessary for these small models with one X variable. There are irregularities in some of the distributions however which suggest the need for a robust method, hence a Robust Linear Model (RLM). Applying the RLM model with bisquare weights and an intercept (determined by the tuning process) to Box-Cox transformations of these data and examining the residuals for normality with a Shapiro-Wilks test suggests that the RLM model provides an appropriate fit for most of these data at $\alpha = 0.05$. The exception is the model for foreign-born from Africa which has a Shapiro-Wilks p -value of 0.06 and an unusual distribution due to the period of slavery. Post-slavery the distribution of the transformed Africa variable is more linear and therefore more suited for an RLM model. Examining legislation as an antecedent where the effect is demographic factor changes and the cause is the number of laws enacted one year prior, the RLM suggests that as the number of immigration laws enacted increases, specific demographic elements change such as the number of Native Born \uparrow , Hispanics \uparrow , and foreign-born from specific countries of origin (Africa \uparrow , Latin America \uparrow). These relationships are statistically significant at $\alpha = 0.05$.

Table 10: Ordered Number of Laws (Antecedent)

	Ordered Robust Linear Model (RLM) Coefficients			
	Estimate	Std. Error	t value	Pr(> t)
Native Born	0.4700338	0.0498572	9.427596	0
Hispanic	0.4345160	0.0459628	9.453639	0
Latin Am	0.3201317	0.0369501	8.663889	0
Africa	0.3142223	0.0481317	6.528390	0

Discussion

Correlations

Kendall correlations between the number of future and past immigration laws enacted and the demographic factors examined are statistically significant at $\alpha = 0.05$ while the correlations between the number of future and past immigration laws passed and all economic factors examined are not. The Correlation Analysis between Legislation and Demographic Factors shows statistically significant positive correlations between the number of future and past immigration laws passed and demographic factors. With a just few exceptions, nearly all of these correlations are around $\tau = 0.5$. The relationship between the number of future and past immigration laws and Native-Americans hovers around $\tau = 0.25$. This lower correlation makes sense intuitively since Native-Americans are not generally associated with legislation on immigration. The relationships between the number of future and past immigration laws and foreign-born from Europe and North America are all between $\tau = 0.18$ and $\tau = 0.27$. This is likely related to European and North American arrivals having an unnatural distribution that was inflated until race stopped being a factor in admissions around 1950.

Although not necessary for this analysis, in keeping with the meticulousness of the correlational analysis, every combinatoric possibility for correlations was examined. In doing this, it was found that the correlations between demographic factors and various industries in the economy contradict several stereotypes. The Correlation Analysis show that every race, Hispanic, Natives, and Foreign-born have a strong and statistically significant positive correlations of $\tau = 0.6$ or higher with every industry in the economy. This finding was refreshing, but it was also a little interesting to see when examining the individual countries of origin of Foreign-born, Europe and North America did not have statistically significant correlations with any industry. This, like the low correlations those regions had with legislation, is likely due to the unnatural distribution that European and North American arrivals have due to admissions being based on race until around 1950. It was surprising to see in the Correlation Analysis that employment-related Nonimmigrant (temporary) Visas had strong statistically significant correlations of $\tau = 0.72$ or higher with every sector of the economy while nearly all employment-related Immigrant (permanent) Visas had statistically insignificant correlations. The exception to these findings of statistically insignificant correlation for Immigrant Visas is the $\tau = 0.63$ relationship between Immigrant Visas and Utilities. This may however, just be a statically anomaly since one in twenty false positives are likely at a 95% confidence level. This finding if true however, would not be counterintuitive as it is easy to see how permanent relocations can be associated with increases in Utilities. It is strange however that this demographic is has no statistically significant relationship with other sectors in the economy. In the final section of the Correlation Analysis, every industry in the economy is found to have a strong and statistically significant positive correlations of $\tau = 0.67$ or higher with Removals. Yet no other enforcement action was found to have statistically significant correlations with the economy. At first glance this is strange but it makes sense when looking at the definitions. Apprehensions and Returns are swift actions that involve a limited number of parties. Court ordered Removals are prolonged processes that involve several institutions such as government agencies, private prison systems, and law firms.

Causation

It cannot be stressed enough that Granger Causality tests, like every other statistical test, does not prove or disprove any hypothesis. The results simply provide statistical support for or against the argument made by the hypothesis. It can be further stated that Granger Causality does not hypothesize true causality, but rather the existence predictive ability such that if X can help predict Y , then X is a Grange cause of Y . These tests use lagged time series observations to see if prior observations of X are useful in predicting future observations of Y , and then taste the same for the converse relationship. Again, this test has limitations in multivariate cases and does not rule out confounding variables, but it remains one of the most useful tools for analyzing causality in model building. The first step to conducting this lagged time series test appropriately involves determining the ARIMA order of differencing for each time series variable and then using the maximum order in the Granger Causality test. In this case the maximum order of differencing was found to be $d = 2$.

The Granger Causality test returns some unexpected results. The bidirectional relationship between the number of future and past immigration laws enacted and demographic factors does exist, but in a very disparate way. Legislation on immigration is a consequence of many demographic changes but an antecedent for change to only a few of those demographic factors. Specifically, changes in every demographic factor except Foreign-born from Europe and North America are found to be a Granger cause for change in number of immigration laws enacted in the future. Yet the laws that are enacted are found to be Granger causes for future change in only four demographic factors: Hispanic, Native-born, and foreign born from Africa and Latin America. Hence, legislation that is enacted in response to various demographic changes impacts those demographic elements disproportionately. It could be argued that this bidirectional relationship does not exist and that these lagged results are independent of each other, but it is highly unlikely that laws are enacted by the U.S. Congress without regard to both past events and future outcomes.

Regression

The findings from both the correlation and causality analysis support the existence of a relationship between demographics and legislation backward and forward. Capturing this bidirectional relationship calls for the creation of multiple models to capture the affects of the feedback loop. The first model looks at the number of laws on immigration enacted as a consequence of demographic changes. This is a multivariate model where the independent variables are the demographic factors and the dependent variable is the number of immigration laws enacted one year later. The next set of models look at the number of laws on immigration enacted as an antecedent for change to the individual demographic factors. These are small univariate models where the dependent variable is a given demographic factor and the independent variable is the number of immigration laws enacted one year prior. The data have some peculiarities which complicate the modeling process however.

Although social phenomena tend to follow an exponential distribution, some of these data are not exponential. The impact of African slavery, Native American genocide, and race being used as a factor for admission into the country until 1950 distort many of the distributions. There was therefore a need to tune and test several candidate models on transformations these data to see which provided the most appropriate fit for the data. The model selection process evaluated four linear models, four nonlinear models, and four tree-based models. Interestingly enough, all of the models performed almost equivalently on these data. The average performance metrics of the training set resampling were $RMSE = 1.11$, $R^2 = 0.41$, and $MAE = 0.74$. These metrics are highly optimistic because they are based on repeated sampling, but even after validating the results with the test set there was very little variation in model performance. The average performance metrics of the test set validation were $RMSE = 1.68$, $R^2 = 0.33$, and $MAE = 0.95$. This homogeneity in performance is beneficial in that it negates the need to balance the tradeoff between model complexity and interpretability. Therefore, given the benefit of model indifference and the need to analyze coefficients rather than predict outcomes, linear models were chosen.

Another issue with these data was the high level of multicollinearity. As mentioned in the modeling process, if all the highly correlated variables were removed from the model, only three of the twelve variables would

be left. Those variables are foreign-born from Africa, Native-Americans, and Whites which are the variables with the distorted distributions mentioned at the beginning of the previous paragraph. To address the multicollinearity a Partial Least Square (PLS) regression model is used. PLS regression is an enhanced Principal Component Analysis (PCA) regression model that takes the response variable into account when constructing the model components. The variance in the number of immigration laws passed one in the future is explained in a statistically significant way by changes in every demographic factor that was found to be a Granger cause for the change. The significance of the variance estimates was derived using jackknife resampling. Although PLS coefficients reflect variance and direction, the Correlational Analysis showed that the relationship was positive. Meaning that as each of these demographic factors increase, so do the number of laws the following year. The variables that explain most of the variance are the demographic segments of Other Race, Native American, Asian Pacific, and foreign-born from Africa. Other Race seems rather odd until the definition is examined. Other Race is a catchall that has meant different things at different times; from Hispanic when there was no such category, multiracial, and those who do not feel that the given categories describe them. It can be said that other race is the most diverse category of all. It also pulls back the curtain on the social construct of race and ethnicity, but that is not the topic of this analysis. The results for Native Americans and foreign-born from Africa can may or may not be accurate given their population distributions being distorted due to slavery and genocide. The two demographic segments and the number immigration laws enacted all have semi-parabolic distributions. In the Literature Review it was noted that early immigration and naturalization laws focused on granting free White men citizenship and that race was a factor in admission until around 1950. Many of the early laws on immigration could have very well resulted from changes the number of Native Americans and foreign-born from Africa. Changes in the number of Asian and Pacific Islanders impacting the number of immigration laws passed the next year is not surprising. Many immigration laws focused entirely on those demographic segments such as the Alien and Sedition Acts of 1798 and the Chinese Exclusion Act of 1882, with laws from the former being exercised again during World War II to facilitate the implementation of Japanese Internment Camps.

The small models had fewer modelling issues but were just as interesting. The model used for these data was a Robust Linear Model (RLM) which was chosen for its ability to handle outliers and irregularities effectively. The results of the model show that the number of laws enacted one year prior has a statistically significant impact on the four demographic segments that were found to have a correlational and causal relationship with legged legislation. These are Native Born, Hispanics, and foreign-born from Latin America and Africa. The largest relationship was the positive one between the number of Native born in the population and immigration laws past one year prior. There is nothing extraordinary about this relationship, when the number of immigration laws enacted increases, the number of Native born increases one year later. The rest of the finding are not so clear cut however. When the number of immigration laws enacted increases the number of Hispanics and foreign-born from Latin America and Africa increases. There is something about this which does not seem quite right, but this phenomena has been studied and explained before. Princeton Professor Douglass Massey has found that immigration is a circular flow where immigrants come to work then return home.⁵⁴ When there are no restrictions, immigrants come and go as they please. When immigration laws place restrictions on entry, outflow decreases. Immigrants who would normally come and go freely now risk the ability to return if they leave, therefore they chose to stay. Hence the increase in foreign-born from Africa and Latin America, and by extension Hispanics.

Conclusion

The current presidential administration is proposing immigration reforms that it claims will improve the economy. Legislation on immigration has a temporal relationship with demographics in the United States. Changes in demographics impact the number of laws on immigration enacted one year later; and the number of laws enacted one year prior impact demographics. This bidirectional relationship is significant in terms of correlation, causation, and regression. Legislation on immigration does not have a statistically significant relationship with the economy. Throughout history the economy has improved regardless of the demographic

⁵⁴<https://www.princeton.edu/news/2016/04/20/tighter-enforcement-along-us-mexico-border-backfired-researchers-find>

makeup of immigrants. Politicians claiming things will improve if a change is made, knowing that improvement is likely regardless of the change, are exercising political savvy in advancing their interests. Much can be said about the ethics behind such self-serving behavior,⁵⁵ but only a reference to further reading on the topic will be provided in this analysis. Although intent cannot be inferred, this research does show that it is more likely that opponents of the current presidential administration are correct in stating that the proposed reforms will not improve the economy and impact the overall diversity of the country. They are not correct however in the alleged impact on diversity that those changes will have. Yes the number of Native born will increase, but due to interruptions in natural immigration outflows, there will also be increases in foreign-born from Africa and Latin America, and by extension Hispanics.

⁵⁵<http://www.hbs.edu/faculty/Publication%20Files/self-serving+justifications>

Appendix

Tables and Figures

Table 11: Legislation on Immigration (1790-2016)

Decade	1700	1800	1900	2000
Laws During the 00s	NA	1	5	45
Laws During the 10s	NA	1	4	16
Laws During the 20s	NA	1	8	NA
Laws During the 30s	NA	0	9	NA
Laws During the 40s	NA	1	20	NA
Laws During the 50s	NA	1	17	NA
Laws During the 60s	NA	2	12	NA
Laws During the 70s	NA	2	20	NA
Laws During the 80s	NA	7	15	NA
Laws During the 90s	5	4	22	NA

Legislation on Immigration (1790–2016)

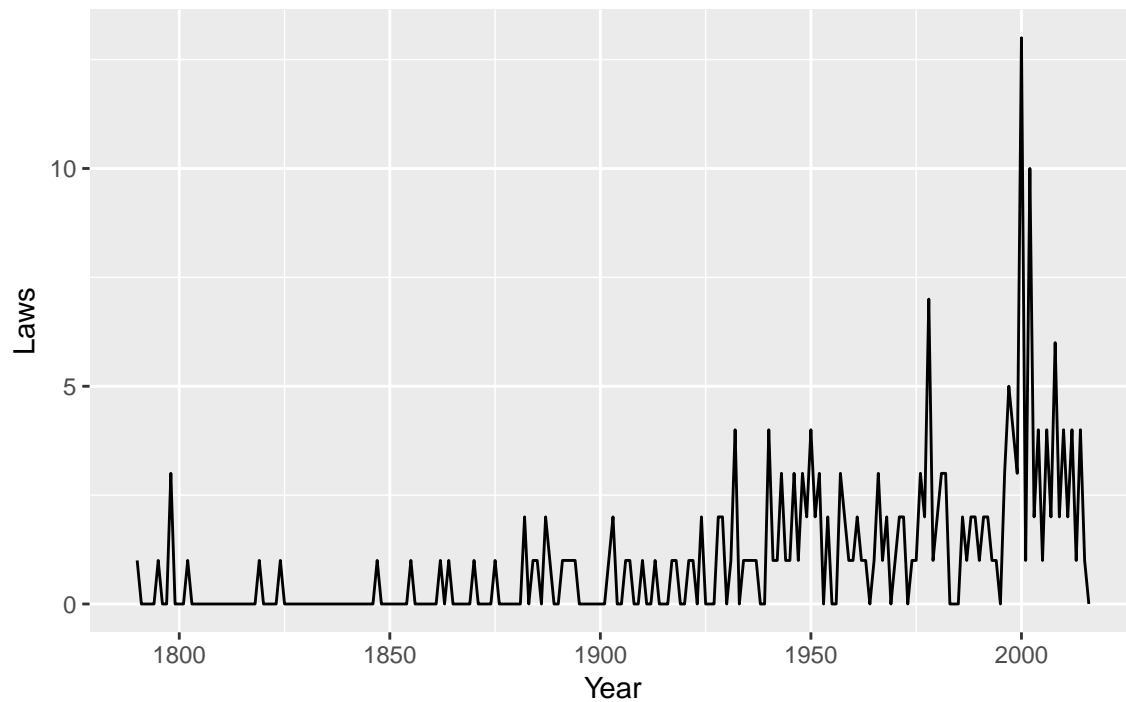
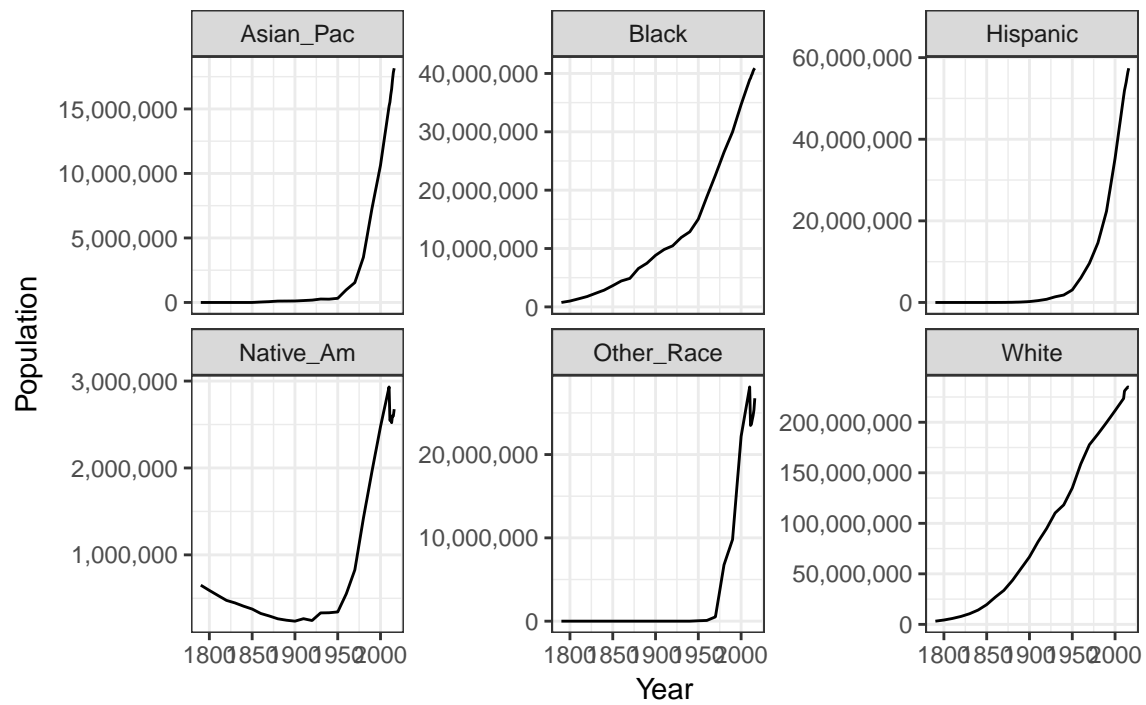


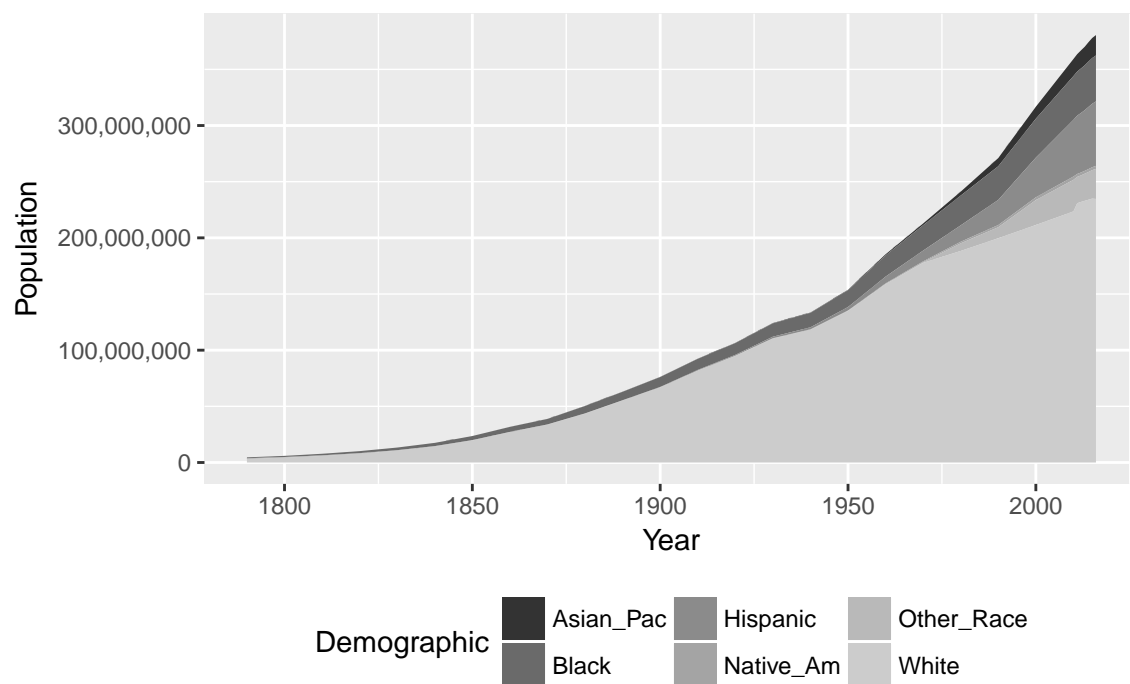
Table 12: Race and Hispanic Origin (1790-2016)

	White	Black Free	Black Slave	Native Am	Asian Pac	Other Race	Hispanic
2016	234644039	40893369	0	2676399	18152921	26760787	57398719
2015	234940100	40695277	0	2597249	17828723	25357472	56496122
2014	233963128	40379066	0	2601714	17244114	24669034	55279452
2013	232969901	39919371	0	2521131	16537870	24180566	53986412
2012	231992377	39623138	0	2563505	16098728	23636292	52961017
2011	230838975	39189528	0	2547006	15526436	23489974	51939916
2010	223553265	38929319	0	2932248	15214265	28116441	50477594
2000	211460626	34658190	0	2475956	10641833	22185301	35305818
1990	199686070	29986060	0	1959234	7273662	9804847	22354059
1980	188371622	26495025	0	1420400	3500439	6758319	14608673
1970	177748975	22580289	0	827255	1538721	516686	9589216
1960	158831732	18871831	0	551669	980337	87606	6020852
1950	134942028	15042286	0	343410	321033	48604	3068684
1940	118214870	12865518	0	333969	254918	0	1818764
1930	110286740	11891143	0	332397	264766	0	1386967
1920	94820915	10463131	0	244437	182137	0	776611
1910	81731957	9827763	0	265683	146863	0	452788
1900	66809196	8833994	0	237196	114189	0	216138
1890	55101258	7488676	0	248253	109527	0	104164
1880	43402970	6580793	0	264532	105613	0	43855
1870	33589377	4880009	0	296885	63254	0	16018
1860	26922537	488070	3953760	325802	34933	0	7320
1850	19553068	434495	3204313	376681	0	0	2314
1840	14189705	386293	2487355	409641	0	0	726
1830	10532060	319599	2009043	446881	0	0	253
1820	7866797	233634	1538022	476294	0	0	87
1810	5862073	186446	1191362	533271	0	0	31
1800	4306446	108435	893602	591042	0	0	11
1790	3172006	59527	697681	651041	0	0	4

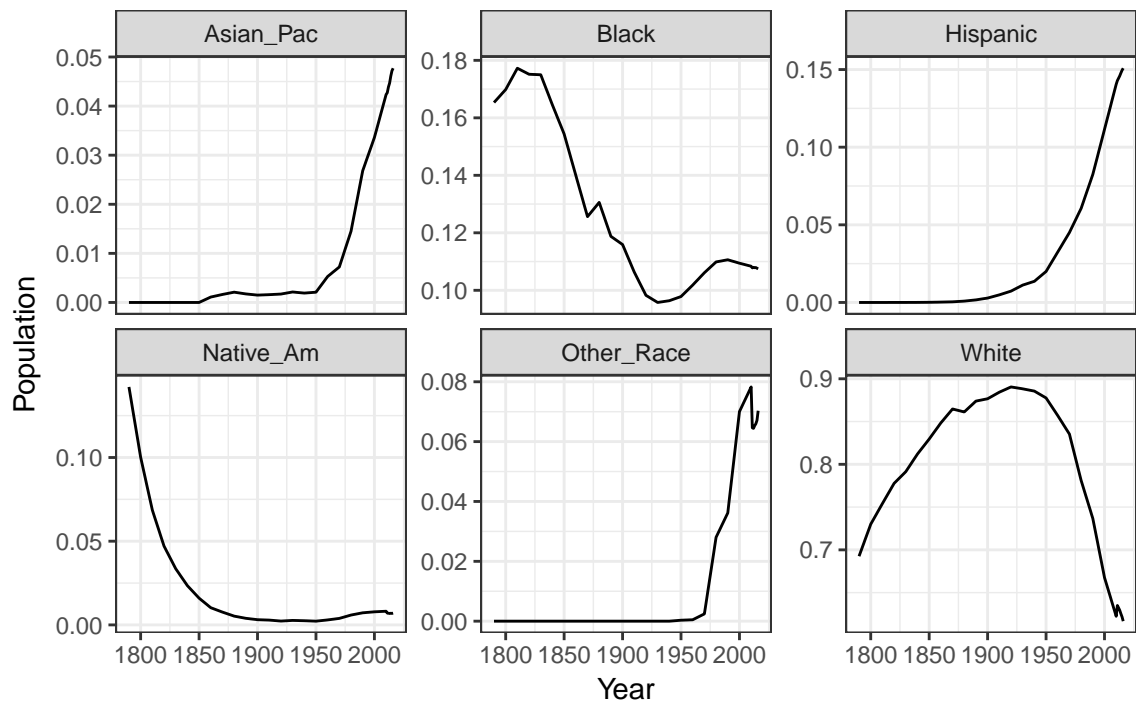
Race and Hispanic Origin (1790–2016)



Race and Hispanic Origin (1790–2016)



Race and Hispanic Origin Proportions (1790–2016)



Race and Hispanic Origin Proportions (1790–2016)

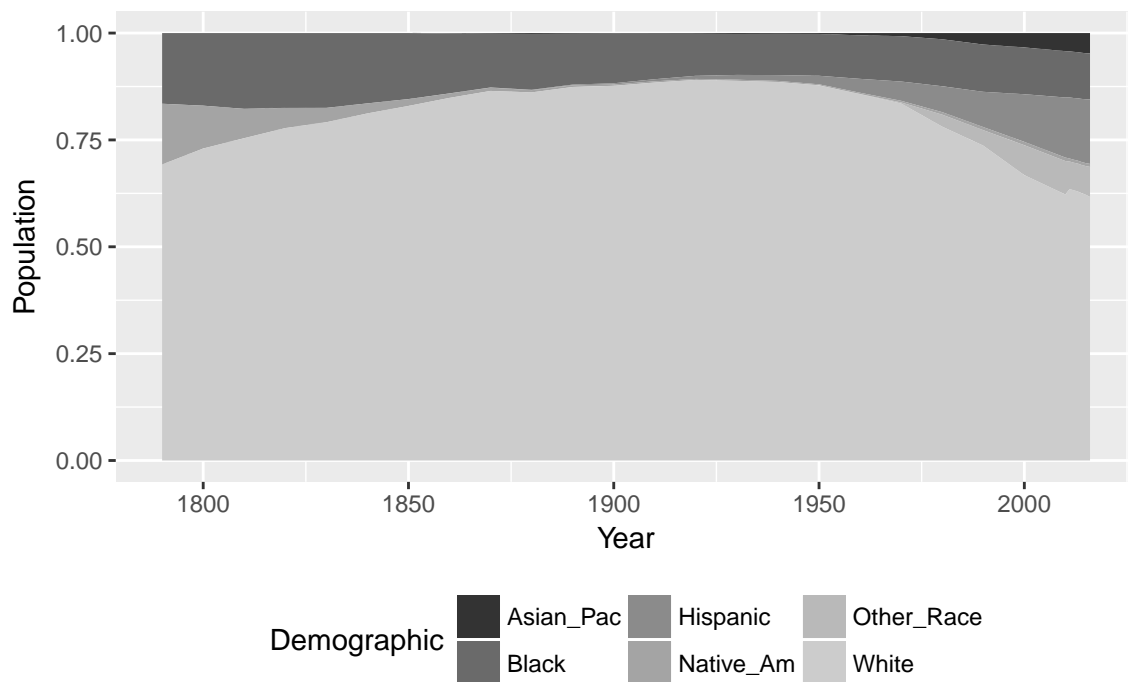
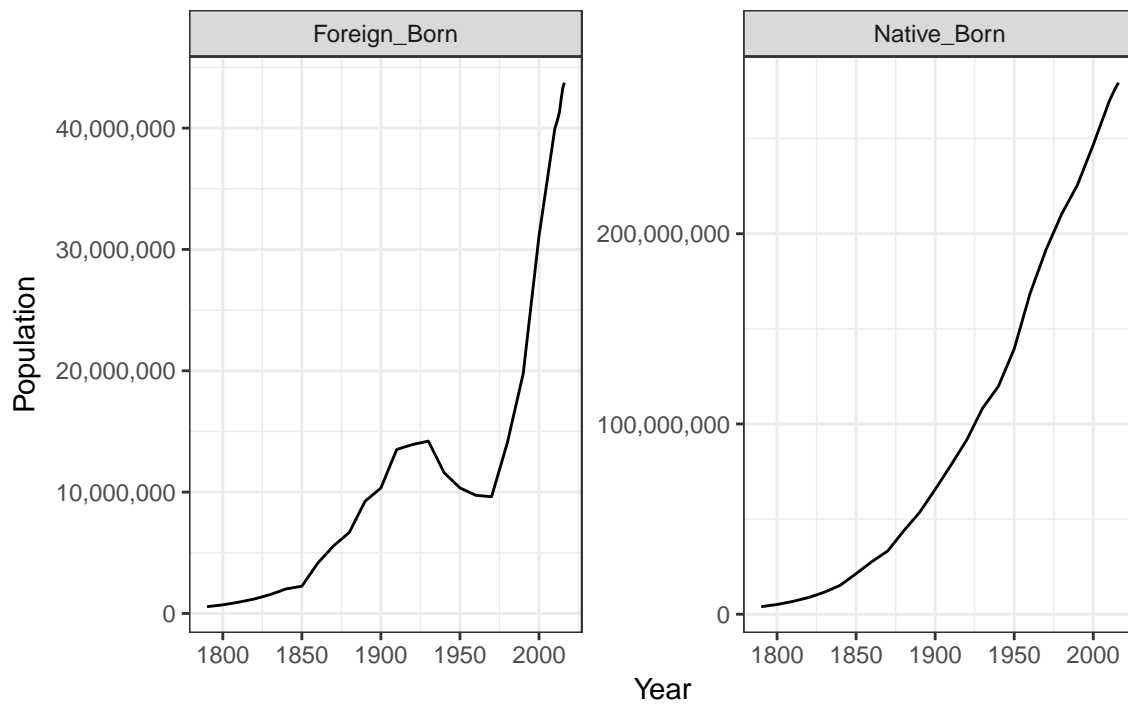


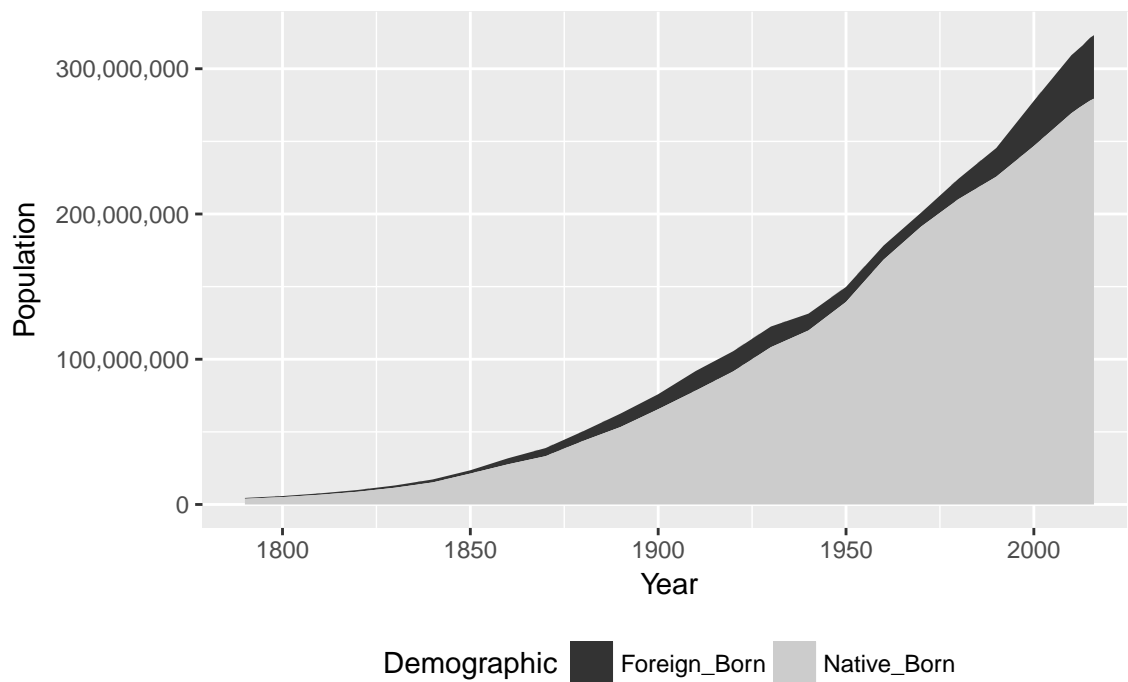
Table 13: Native, Foreign-born, and Origin (1790-2016)

	Native Born	Foreign Born	Europe	Asia	Africa	Oceania	Latin Am
2016	279388170	43739345	4767589	13471718	2143228	262436	22307066
2015	278128449	43290372	4805231	13246854	2077938	259742	22121380
2014	276465262	42391794	4747881	12759930	1950023	254351	21874166
2013	274780773	41348066	4796376	12197679	1819315	206740	21459646
2012	273089382	40824658	4817310	11920800	1714636	244948	21310471
2011	271214059	40377860	4885721	11548068	1655492	201889	21238754
2010	269393835	39955854	4834658	11267551	1598234	199779	21216558
2000	246786466	31107889	4915557	8226254	881300	168046	16086974
1990	225695826	19767316	4350403	4979037	363819	104145	8407837
1980	210322697	14079906	5149572	2539777	199723	77577	4372487
1970	191329489	9619302	5740891	824887	80143	41258	1803970
1960	168525645	9738091	7256311	490996	35355	34730	908309
1950	139442390	10347395	8765544	419219	29679	28934	869486
1940	119795254	11594896	10274777	347442	24002	23139	830663
1930	108304188	14204149	11784010	275665	18326	17343	791840
1920	91659045	13920692	11916048	237950	16126	14626	588843
1910	78381104	13515886	11810115	191484	3992	11450	279514
1900	65583225	10341276	8881548	120248	2538	8820	137458
1890	53362371	9249547	8030347	113383	2207	9353	107307
1880	43713577	6679943	5751823	107630	2204	6859	90073
1870	33250537	5567229	4941049	64565	2657	4028	57871
1860	27655574	4138697	3807062	36796	526	2140	38315
1850	21312894	2244602	2031867	1135	551	588	20773
1840	15219062	2021834	1785806	626	3	263	831
1830	11570196	1556632	1374269	253	1334	141	325
1820	8778327	1196148	940373	102	132090	76	126
1810	6734427	928923	791554	42	33380	41	51
1800	5101895	712801	621461	17	9743	22	20
1790	3954395	559007	439469	7	61851	12	8

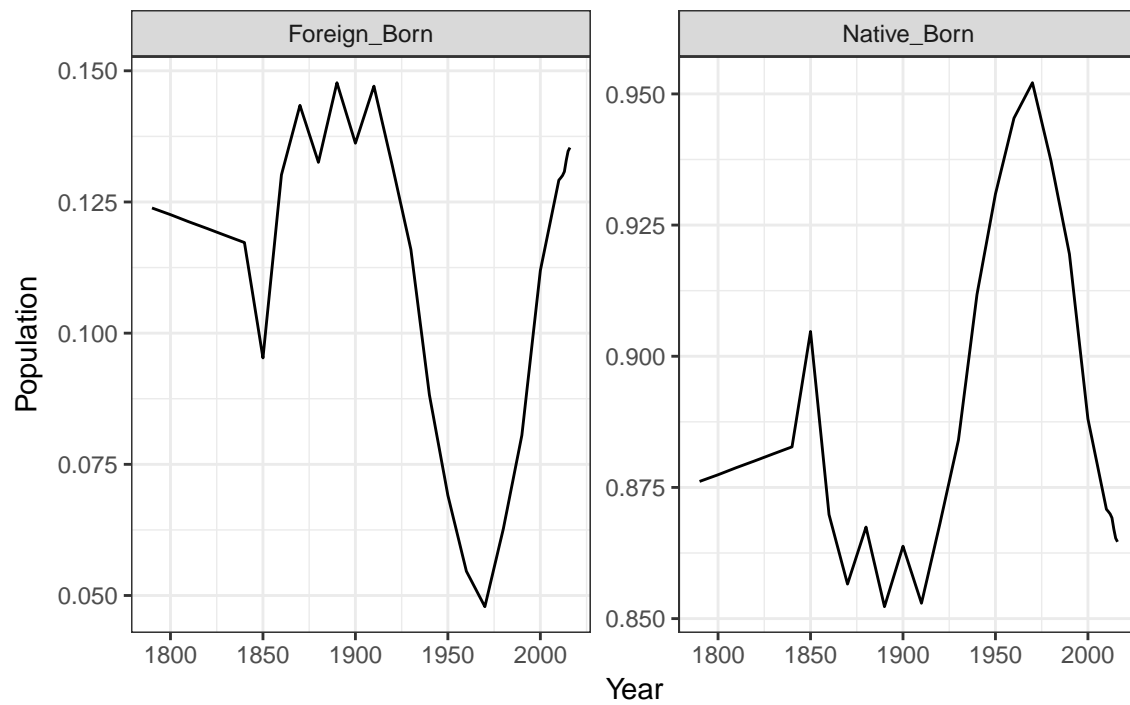
Native and Foreign-born (1790–2016)



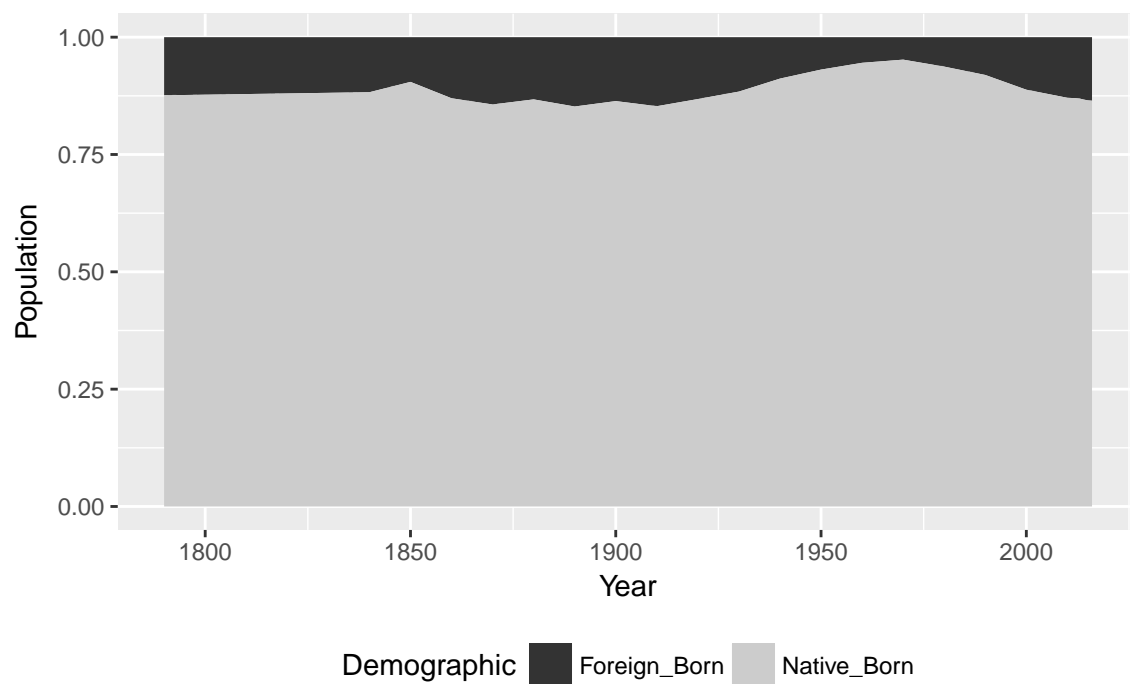
Native and Foreign-born (1790–2016)



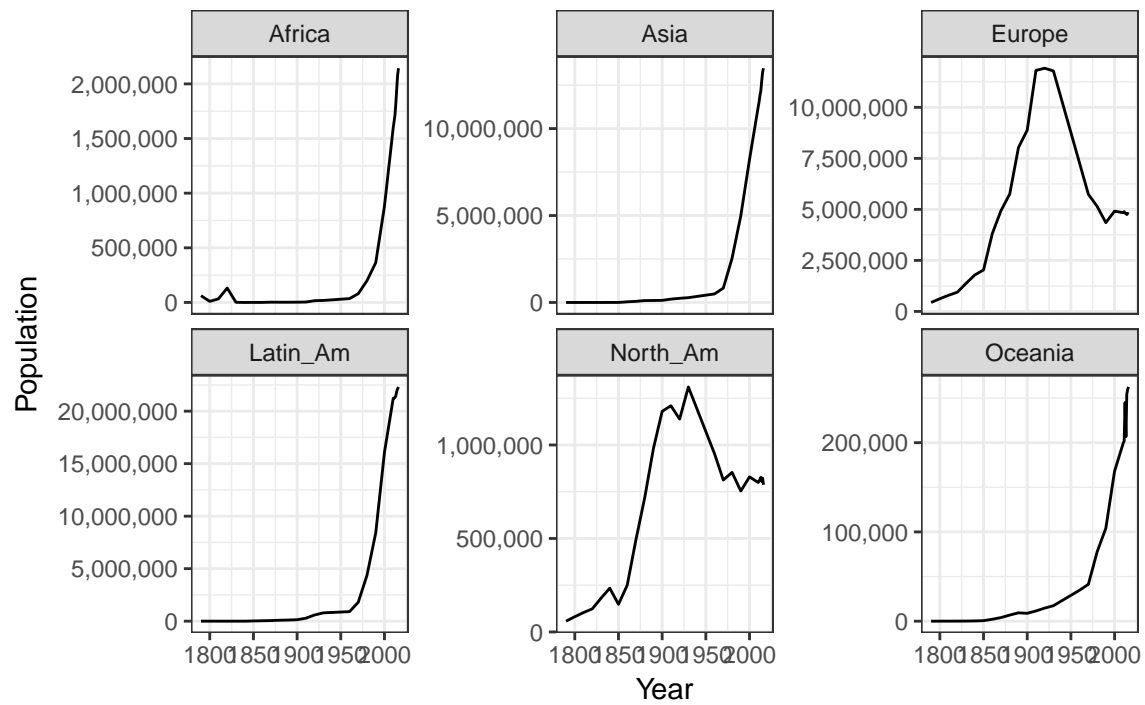
Native and Foreign-born Proportions (1790–2016)



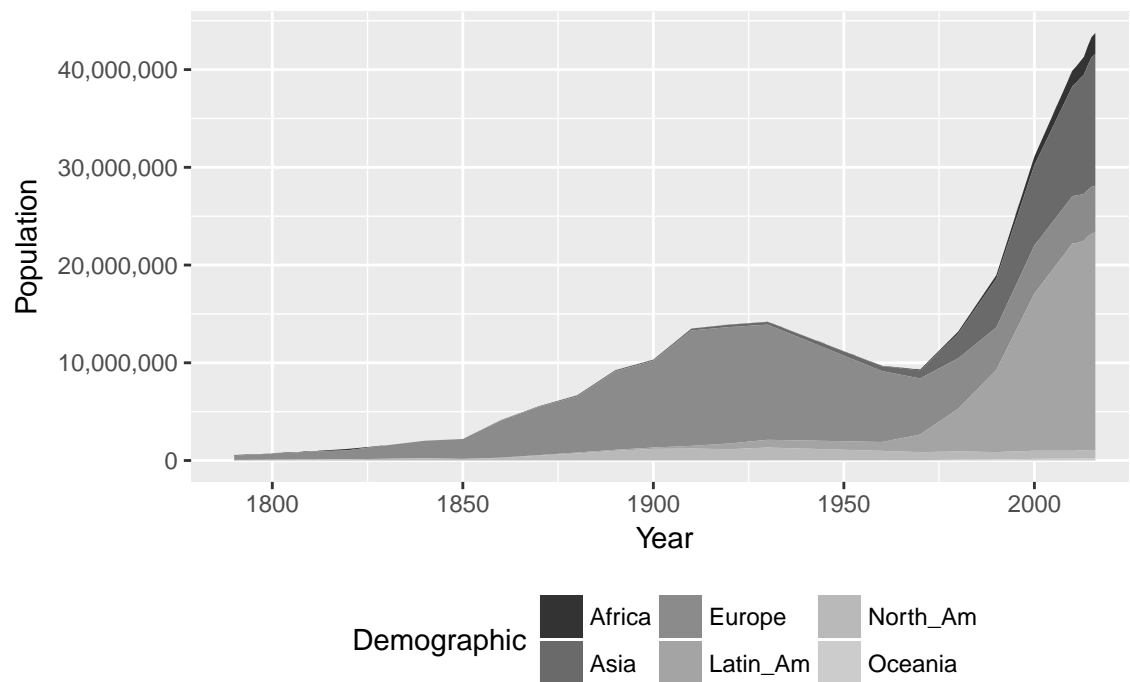
Native and Foreign-born Proportions (1790–2016)



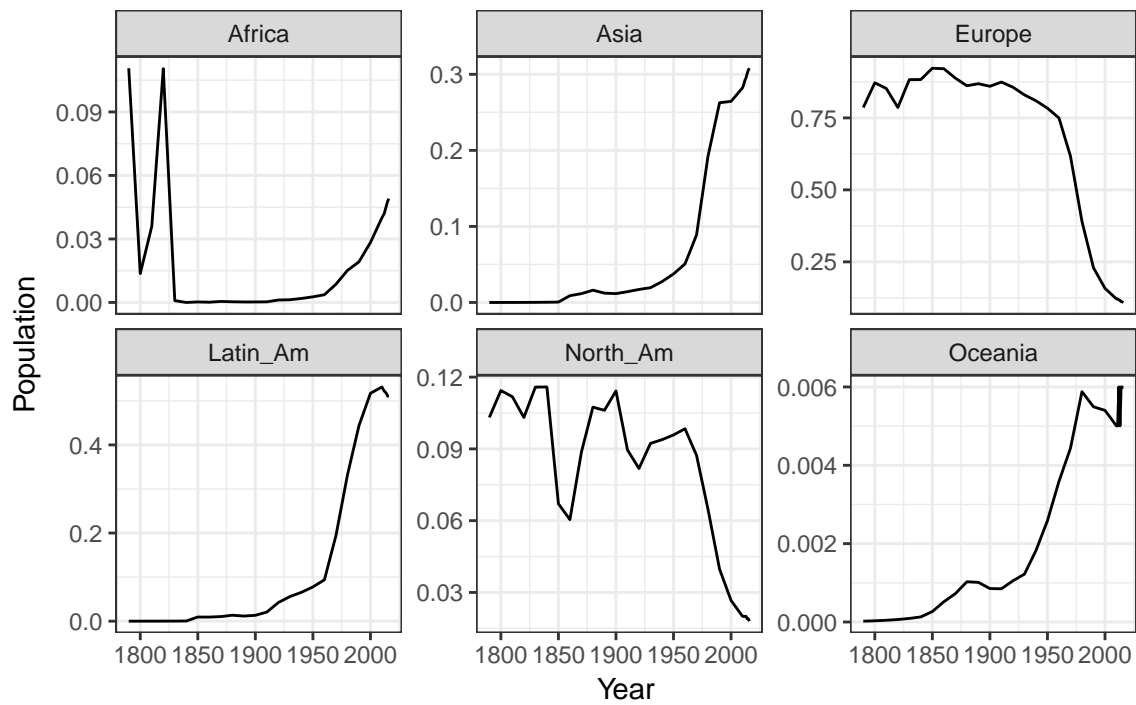
Foreign-born Country of Origin (1790–2016)



Foreign-born Country of Origin (1790–2016)



Foreign-born Country of Origin Proportions (1790–2016)



Foreign-born Country of Origin Proportions (1790–2016)

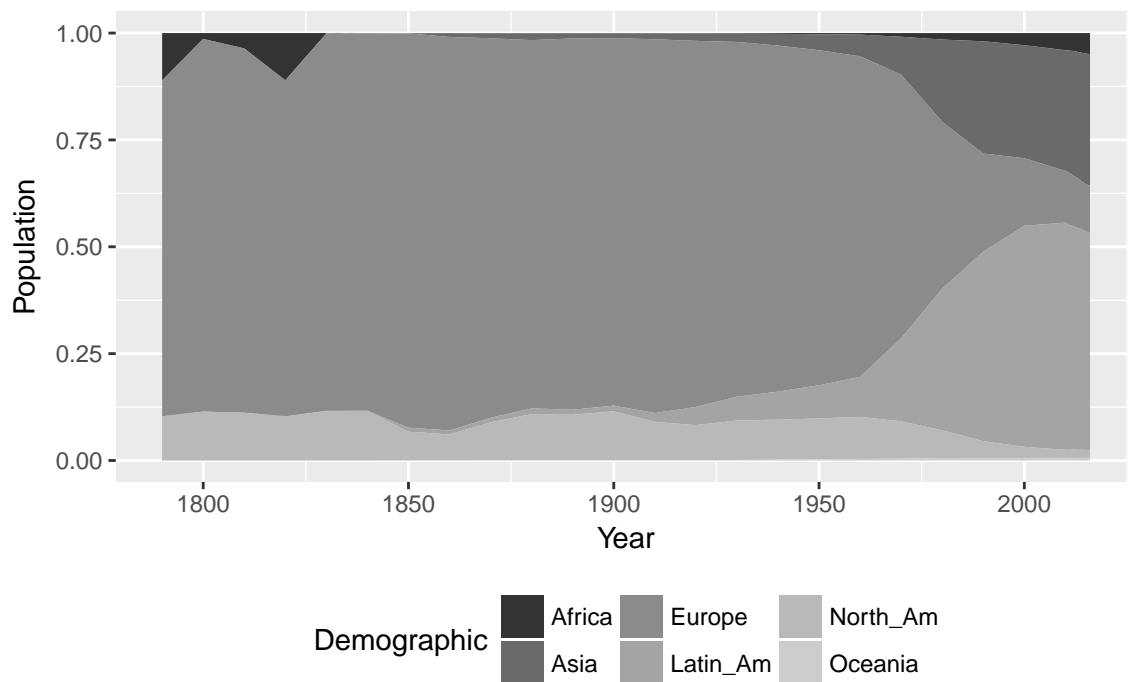
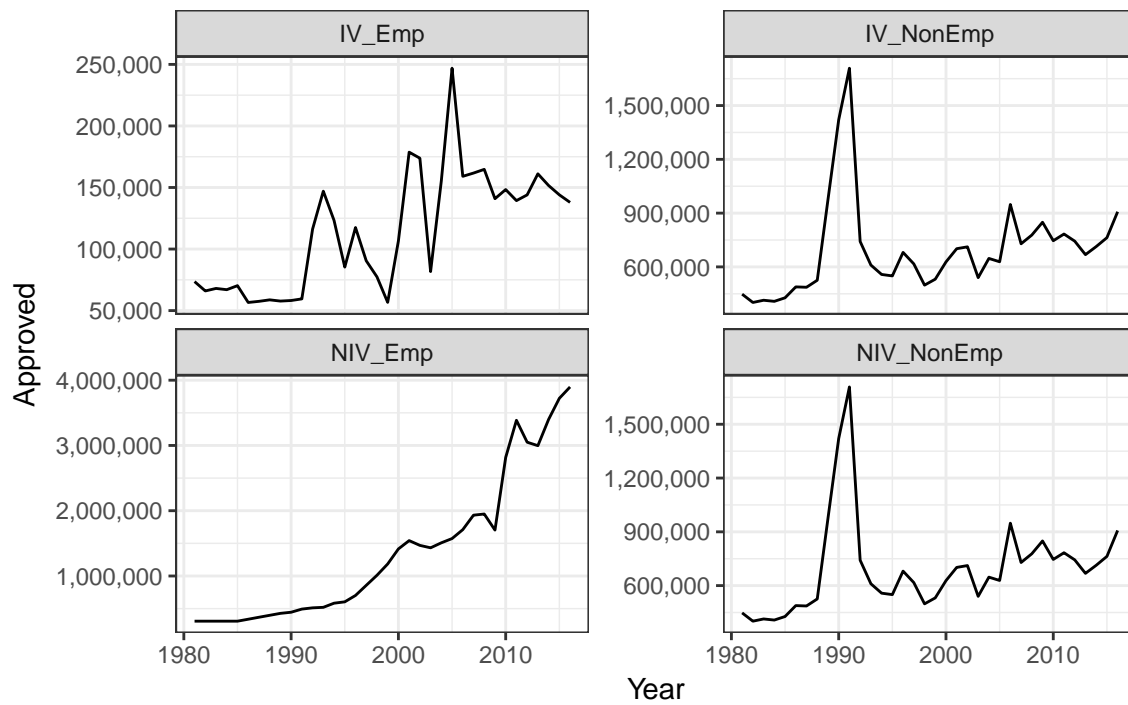


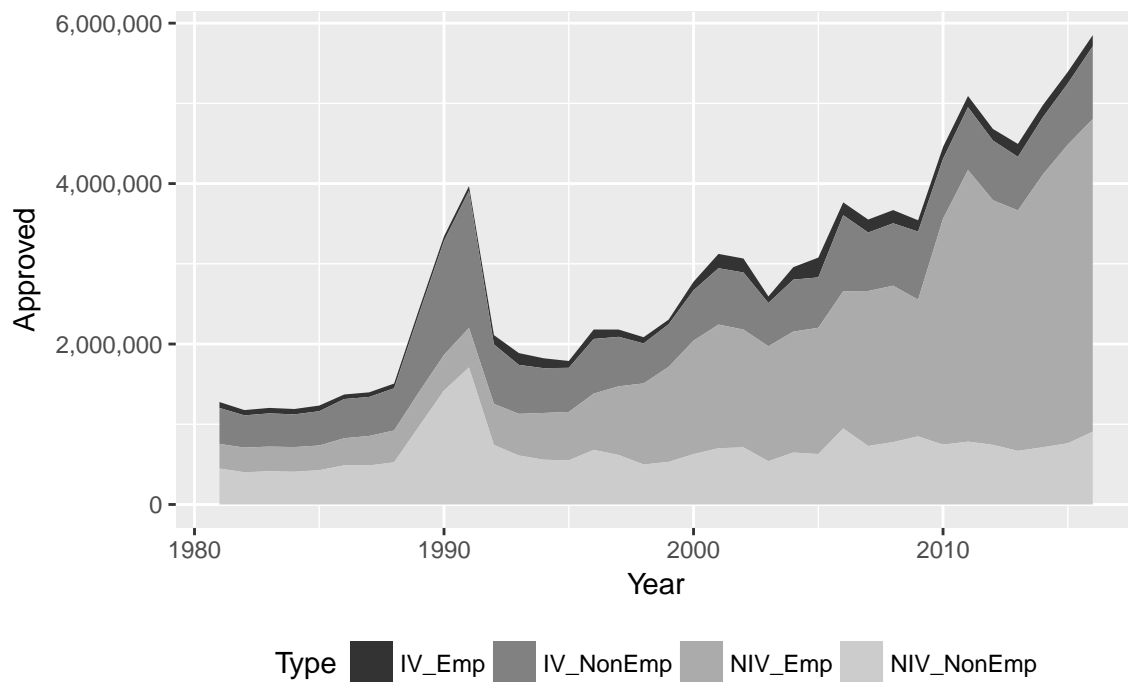
Table 14: Visas and Enforcement (1981-2016)

	IV Emp	IV All	NIV Emp	NIV All	Removed	Apprehended	Returned
2016	137893	1183505	3896674	76786751	340056	530250	106167
2015	144047	1051031	3722543	76638236	326962	462388	129429
2014	151596	1016518	3398961	74930606	405589	679996	163223
2013	161110	990553	2996743	61052260	433034	662483	178663
2012	143998	1031631	3049419	53887286	415900	671327	230333
2011	139339	1062040	3385775	53082286	385778	678606	322073
2010	148343	1042625	2816485	46471569	381593	796587	474166
2009	140903	1130818	1703697	36231554	391283	889212	582584
2008	164741	1107126	1949695	39381928	359795	1043759	811263
2007	161733	1052415	1932075	37149651	319382	960673	891390
2006	159081	1266264	1709953	33667328	280974	1206408	1043381
2005	246878	1122373	1573631	32003435	246431	1291065	1096920
2004	155330	957883	1508340	30781330	240665	1264232	1166576
2003	81727	703542	1431717	27849443	211098	1046422	945294
2002	173814	1059356	1468760	27907139	165168	1062270	1012116
2001	178702	1058902	1541902	32824088	189026	1387486	1349371
2000	106642	841002	1414453	33660320	188467	1814729	1675876
1999	56678	644787	1184072	31446054	183114	1714035	1574863
1998	77413	653206	1010756	30174627	174813	1679439	1570127
1997	90490	797847	856188	27508565	114432	1536520	1440684
1996	117499	915900	701619	24842503	69680	1649986	1573428
1995	85336	720461	603988	22640540	50924	1394554	1313764
1994	123291	804416	583023	22118706	45674	1094719	1029107
1993	147012	904292	519591	21566404	42542	1327261	1243410
1992	116198	973977	511939	20910880	43671	1258481	1105829
1991	59525	1827167	492577	18920045	33189	1197875	1061105
1990	58192	1536483	444504	17574055	30039	1169939	1022533
1989	57741	1090924	427350	16144576	34427	954243	830890
1988	58727	643025	397419	14493402	25829	1008145	911790
1987	57519	601516	367488	12842228	24336	1190488	1091203
1986	56617	601708	337556	11191054	24592	1767400	1586320
1985	70251	568149	307625	9539880	23105	1348749	1041296
1984	66995	541811	307625	9539880	18696	1246981	909833
1983	68014	550052	307625	9539880	19211	1251357	931600
1982	65982	533624	307625	9539880	15216	970246	812572
1981	73573	595014	307625	9539880	17379	975780	823875

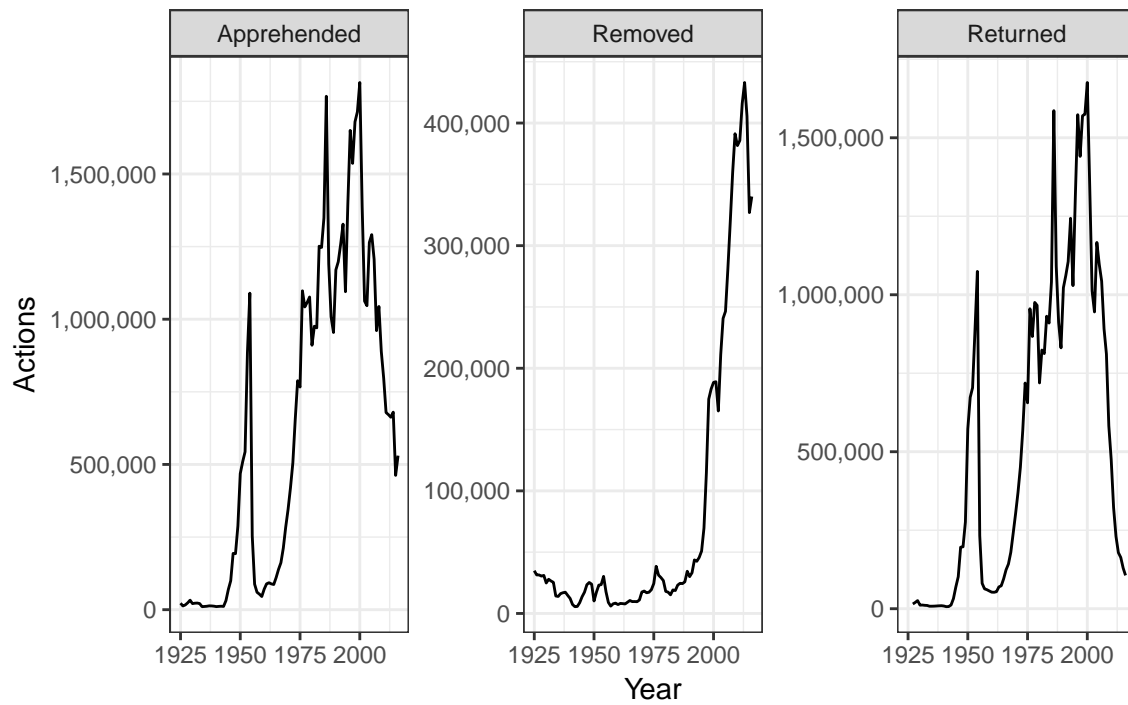
Immigrant and Nonimmigrant Visas by Type (1981–2016)



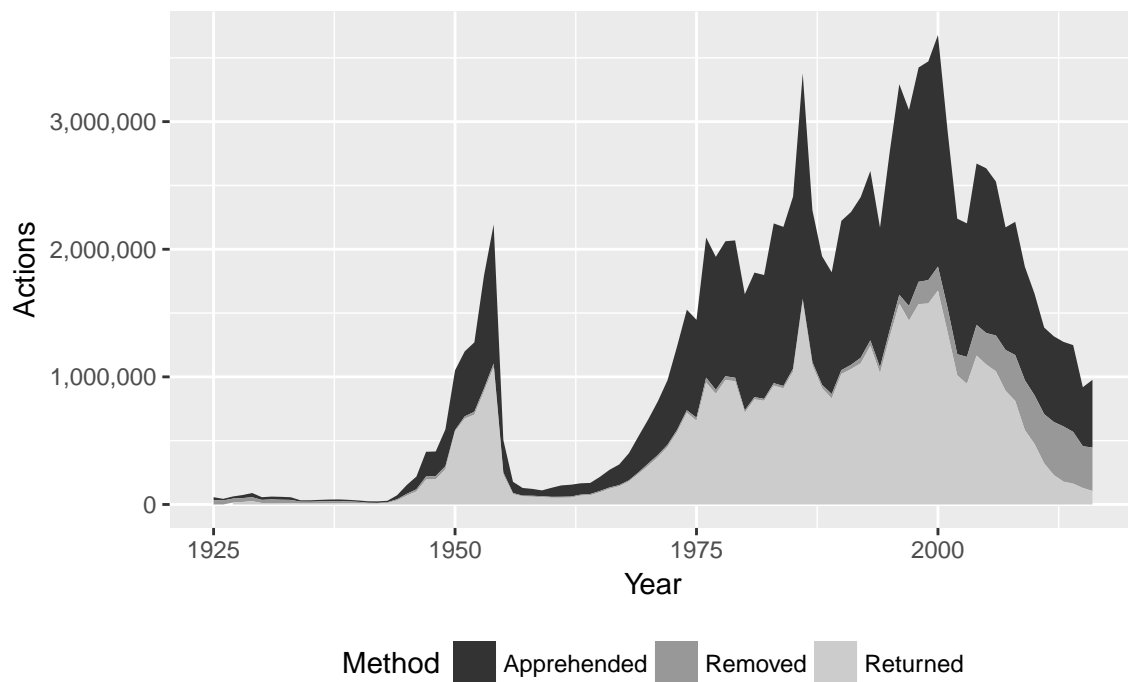
Immigrant and Nonimmigrant Visas by Type (1981–2016)



Enforcement by Method (1981–2016)



Enforcement by Method (1981–2016)



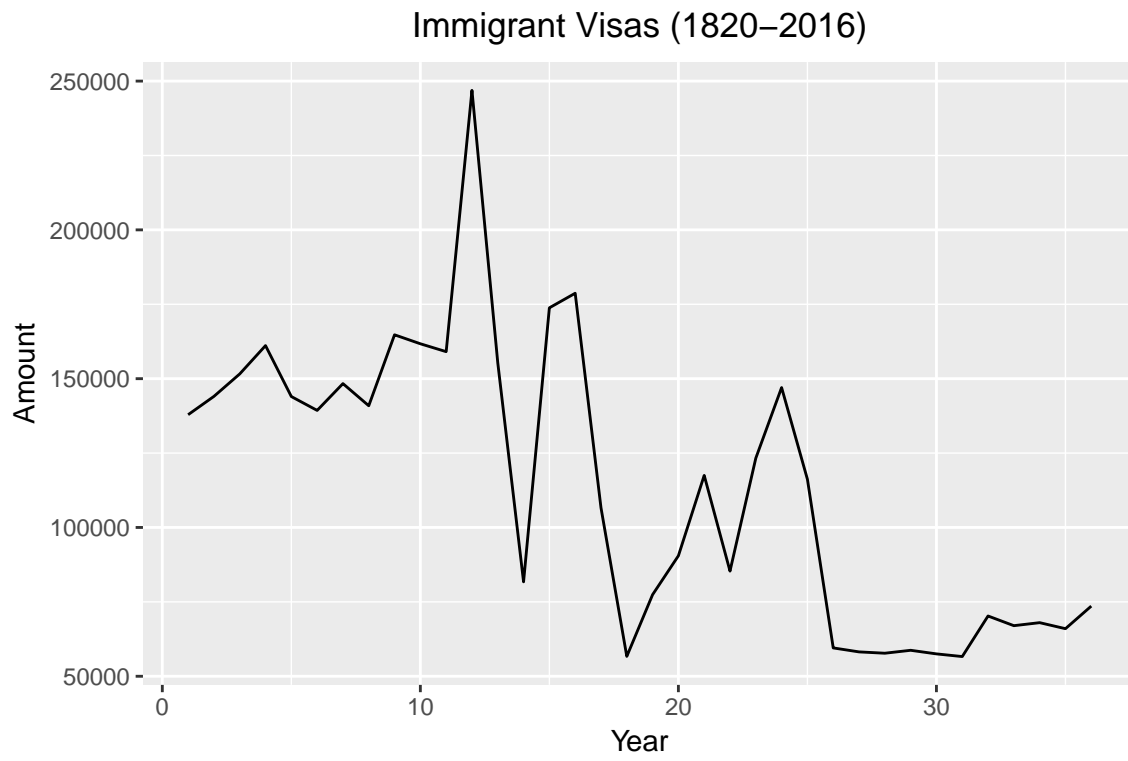


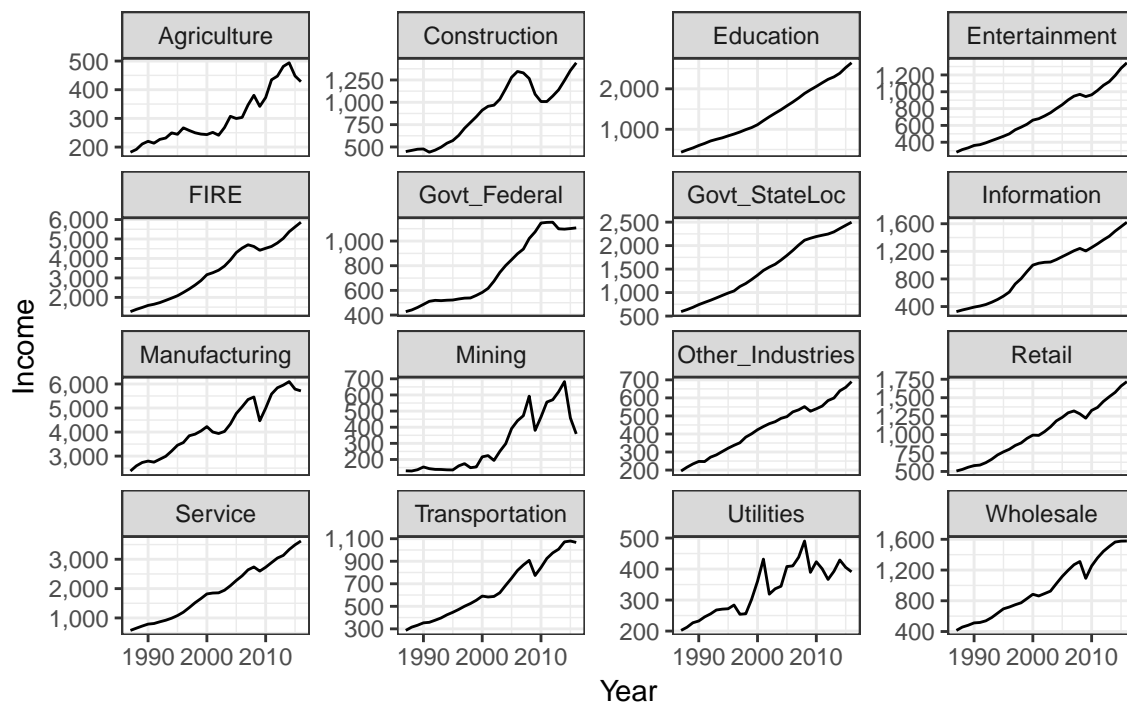
Table 15: GDP by Sector Part 1 (1987-2016)

	AGRI	MINE	UTIL	CONST	MFG	WHOLE	RETAIL	TRANS
2016	428.1	358.6	391.1	1442.0	5712.6	1577.8	1718.0	1067.0
2015	447.9	457.7	405.4	1355.5	5785.0	1576.3	1657.7	1079.6
2014	493.8	682.7	429.2	1244.5	6098.4	1564.8	1572.4	1072.0
2013	481.2	619.5	392.6	1139.7	5953.3	1506.8	1512.5	1007.1
2012	447.3	570.2	366.5	1069.9	5841.6	1443.2	1449.2	974.4
2011	434.7	556.1	400.8	1008.6	5581.9	1357.8	1367.6	924.9
2010	372.8	463.3	423.5	1008.3	4992.5	1253.1	1327.5	844.7
2009	341.9	380.4	389.3	1090.4	4469.3	1092.0	1222.5	773.5
2008	380.5	592.4	490.4	1266.0	5457.8	1311.4	1280.7	907.9
2007	346.9	471.4	438.2	1329.4	5354.4	1269.5	1320.7	866.9
2006	303.5	439.9	410.1	1345.0	5052.8	1196.5	1295.1	818.0
2005	299.6	391.8	408.2	1280.1	4770.9	1116.8	1233.7	749.9
2004	307.2	297.8	344.5	1152.5	4336.0	1023.0	1185.5	686.4
2003	268.2	250.5	336.3	1035.1	4025.8	925.8	1104.6	620.7
2002	241.3	194.0	319.0	967.9	3939.9	893.9	1040.4	587.5
2001	251.2	224.2	432.0	955.2	4001.4	862.6	989.4	582.5
2000	243.5	216.3	360.7	911.8	4225.3	884.3	992.8	590.9
1999	245.2	153.9	302.6	838.4	4045.6	828.0	947.8	553.6
1998	249.7	149.1	255.8	774.5	3913.1	773.5	886.5	524.5
1997	257.8	174.2	254.2	709.6	3846.0	748.5	852.2	500.6
1996	266.9	160.3	283.9	629.4	3563.9	718.5	801.0	472.3
1995	244.5	135.2	272.0	571.7	3452.1	695.2	766.0	446.8
1994	249.5	135.9	270.7	542.0	3209.9	638.2	724.2	424.4
1993	231.5	138.1	267.9	497.7	2991.7	581.7	665.7	396.7
1992	226.8	138.5	255.3	464.7	2870.6	539.2	619.4	376.7
1991	213.5	143.1	245.6	441.4	2748.5	517.6	586.8	358.0
1990	220.2	153.6	232.5	477.6	2792.7	512.0	579.3	353.2
1989	210.7	136.5	226.7	474.6	2734.2	481.4	557.4	331.8
1988	191.8	128.6	212.3	461.6	2592.4	457.4	527.5	315.1
1987	182.1	129.7	202.1	448.7	2379.8	414.3	506.3	286.1

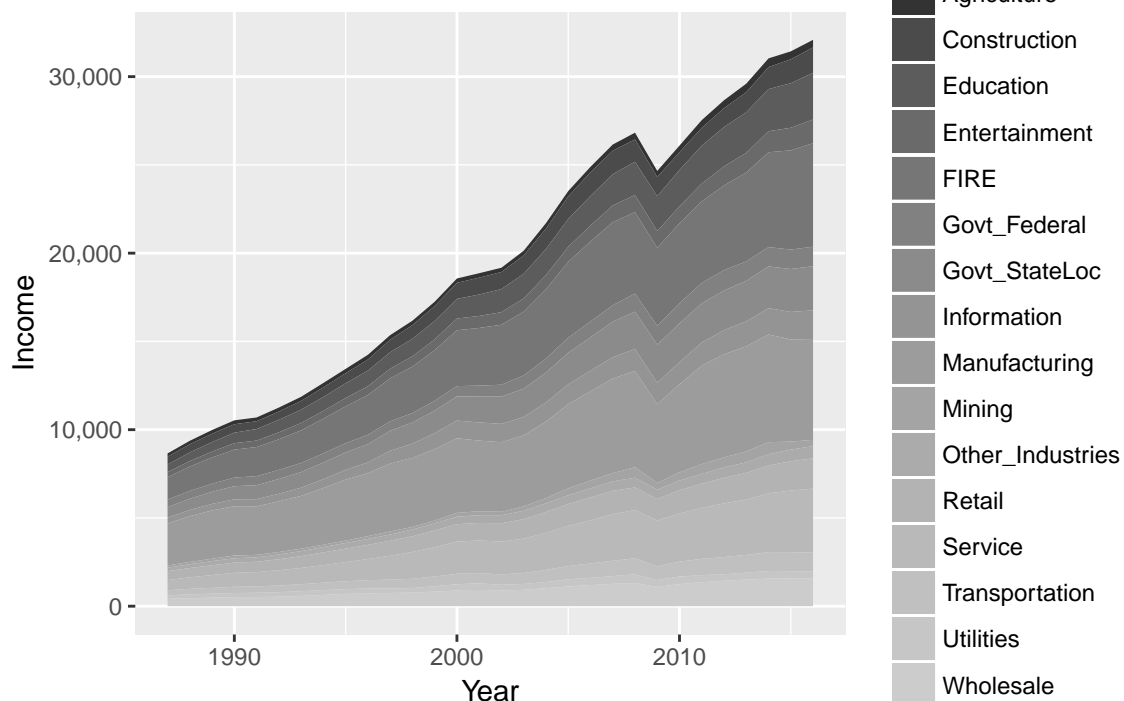
Table 16: GDP by Sector Part 2 (1987-2016)

	INFO	FIRE	SRVC	EDU	ENT	OTHR	FED	STLOC
2016	1622.3	5862.9	3621.7	2643.2	1344.1	690.2	1106.8	2498.5
2015	1557.4	5620.6	3492.4	2525.0	1279.3	658.4	1101.2	2432.1
2014	1493.1	5374.3	3331.1	2385.7	1193.0	639.0	1096.4	2363.6
2013	1418.8	5031.9	3131.4	2295.4	1120.5	599.1	1099.0	2292.5
2012	1366.7	4797.3	3041.2	2234.8	1076.2	585.3	1152.7	2246.5
2011	1308.8	4618.7	2885.9	2148.6	1015.2	554.7	1151.3	2220.6
2010	1255.9	4522.5	2727.5	2058.8	964.0	539.6	1146.1	2193.4
2009	1205.4	4420.7	2596.3	1975.2	943.1	526.0	1074.2	2157.1
2008	1242.7	4614.8	2738.0	1885.8	969.6	551.6	1021.0	2114.9
2007	1208.6	4702.8	2639.2	1774.7	948.0	533.8	934.6	2012.1
2006	1165.0	4533.8	2436.4	1673.5	902.4	522.3	896.6	1897.0
2005	1123.2	4302.0	2277.0	1580.0	846.1	496.2	848.5	1790.7
2004	1080.2	3921.2	2099.1	1483.3	800.1	486.6	801.7	1692.3
2003	1045.2	3611.7	1945.4	1397.9	750.7	468.1	745.6	1603.5
2002	1041.2	3396.6	1856.7	1308.7	712.9	457.0	674.8	1543.2
2001	1027.9	3266.4	1849.2	1213.0	679.9	441.1	616.3	1470.8
2000	1000.9	3165.5	1818.1	1113.8	662.8	423.6	583.6	1370.6
1999	906.3	2865.7	1661.8	1043.1	614.0	401.1	559.3	1278.4
1998	804.2	2634.8	1520.2	991.4	579.1	382.7	538.7	1193.5
1997	727.6	2433.2	1353.3	932.8	548.1	351.3	537.1	1128.9
1996	610.6	2252.7	1196.5	882.4	499.9	337.2	530.4	1034.1
1995	551.2	2084.9	1081.6	837.6	471.2	321.2	522.0	990.2
1994	502.6	1963.0	991.2	788.9	444.8	302.7	520.3	938.8
1993	462.9	1845.7	922.5	752.0	419.3	284.1	516.8	886.6
1992	429.5	1732.0	869.0	711.2	393.1	271.3	519.1	836.3
1991	407.4	1647.9	810.6	650.9	371.1	247.1	511.3	790.4
1990	392.1	1586.7	790.7	599.5	361.7	247.9	486.3	743.0
1989	369.8	1481.6	723.7	538.7	333.5	233.8	461.7	685.3
1988	350.4	1384.4	651.3	489.5	313.0	215.8	441.2	637.0
1987	326.8	1269.5	575.6	433.8	281.7	194.8	428.0	594.8

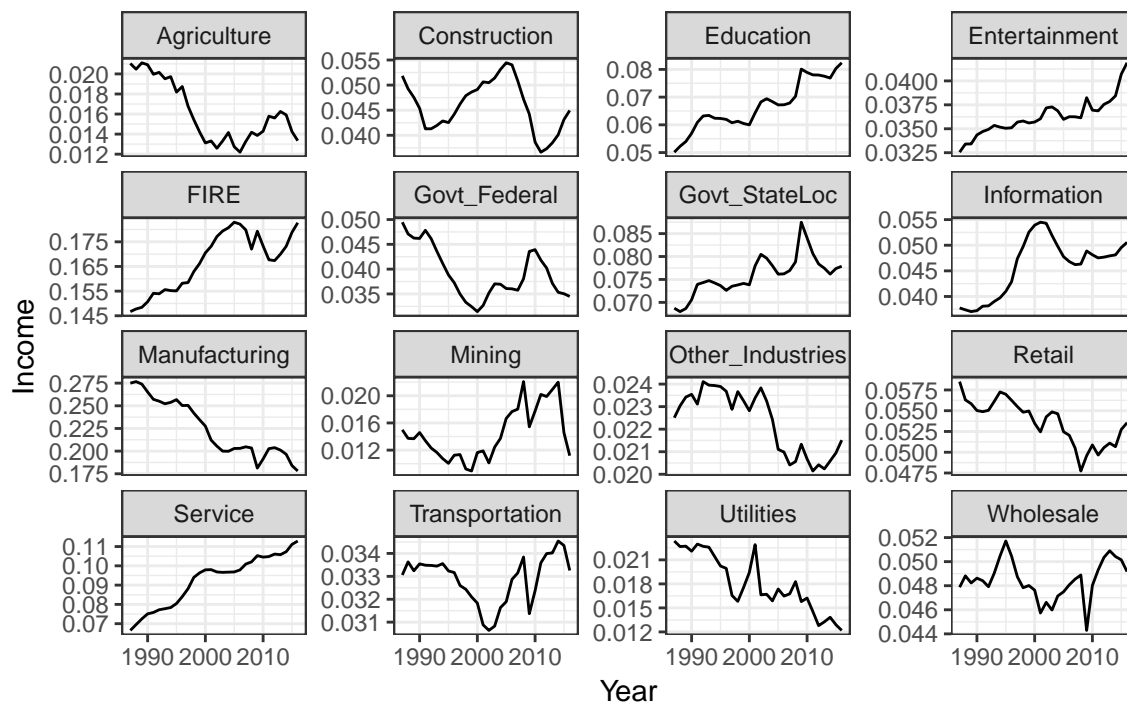
GDP by Sector (1987–2016)



GDP by Sector (1987–2016)



GDP by Sector Proportions (1987–2016)



GDP by Sector Proportions (1987–2016)

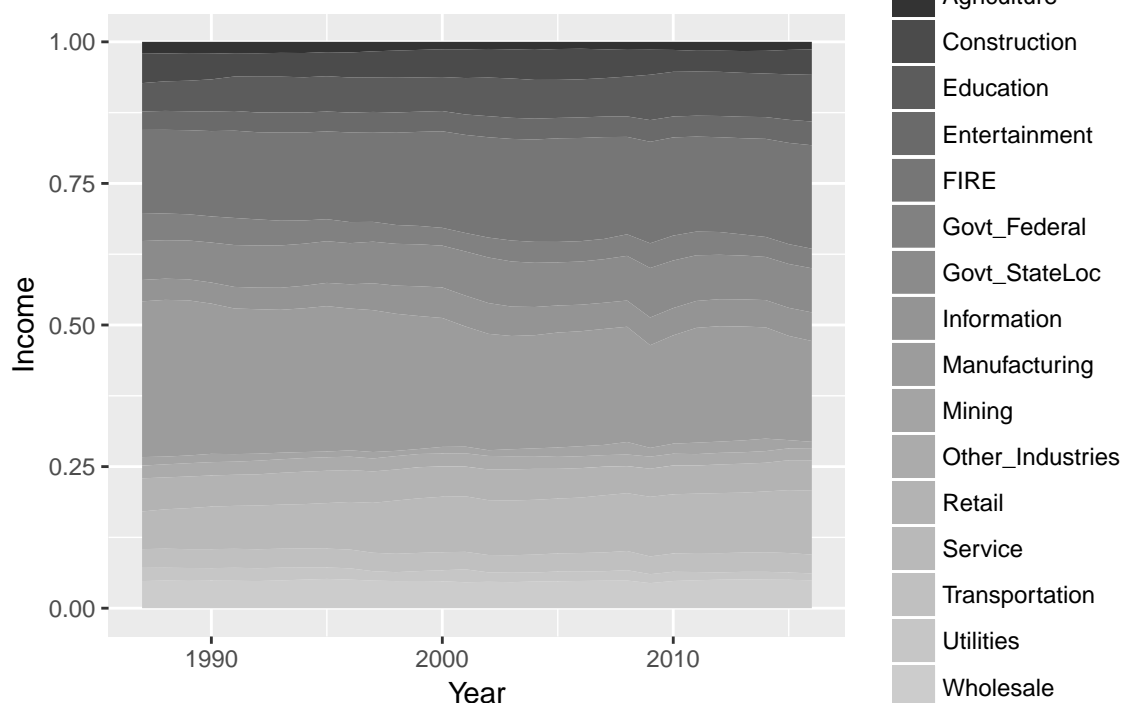


Table 17: Economy with Race and Hispanic Origin

	Correlation, Significant					
	White	Black	Native Am	Asian Pac	Other Race	Hispanic
Agriculture	0.834, Yes	0.830, Yes	0.595, Yes	0.830, Yes	0.660, Yes	0.830, Yes
Mining	0.770, Yes	0.766, Yes	0.632, Yes	0.766, Yes	0.669, Yes	0.766, Yes
Utilities	0.701, Yes	0.697, Yes	0.738, Yes	0.697, Yes	0.747, Yes	0.697, Yes
Construction	0.793, Yes	0.798, Yes	0.756, Yes	0.798, Yes	0.830, Yes	0.798, Yes
Manufacturing	0.903, Yes	0.899, Yes	0.664, Yes	0.899, Yes	0.729, Yes	0.899, Yes
Wholesale	0.963, Yes	0.968, Yes	0.724, Yes	0.968, Yes	0.798, Yes	0.968, Yes
Retail	0.963, Yes	0.968, Yes	0.724, Yes	0.968, Yes	0.798, Yes	0.968, Yes
Transportation	0.963, Yes	0.959, Yes	0.715, Yes	0.959, Yes	0.789, Yes	0.959, Yes
Information	0.986, Yes	0.991, Yes	0.747, Yes	0.991, Yes	0.821, Yes	0.991, Yes
FIRE	0.959, Yes	0.963, Yes	0.729, Yes	0.963, Yes	0.802, Yes	0.963, Yes
Service	0.982, Yes	0.986, Yes	0.743, Yes	0.986, Yes	0.816, Yes	0.986, Yes
Education	0.995, Yes	1.000, Yes	0.756, Yes	1.000, Yes	0.830, Yes	1.000, Yes
Entertainment	0.982, Yes	0.986, Yes	0.743, Yes	0.986, Yes	0.816, Yes	0.986, Yes
Other Industries	0.977, Yes	0.982, Yes	0.738, Yes	0.982, Yes	0.811, Yes	0.982, Yes
Govt Federal	0.931, Yes	0.936, Yes	0.747, Yes	0.936, Yes	0.802, Yes	0.936, Yes
Govt StateLoc	0.995, Yes	1.000, Yes	0.756, Yes	1.000, Yes	0.830, Yes	1.000, Yes

Table 18: Economy with Native and Foreign-born

	Correlation, Significant	
	Native Born	Foreign Born
Agriculture	0.83, Yes	0.83, Yes
Mining	0.77, Yes	0.77, Yes
Utilities	0.70, Yes	0.70, Yes
Construction	0.80, Yes	0.80, Yes
Manufacturing	0.90, Yes	0.90, Yes
Wholesale	0.97, Yes	0.97, Yes
Retail	0.97, Yes	0.97, Yes
Transportation	0.96, Yes	0.96, Yes
Information	0.99, Yes	0.99, Yes
FIRE	0.96, Yes	0.96, Yes
Service	0.99, Yes	0.99, Yes
Education	1.00, Yes	1.00, Yes
Entertainment	0.99, Yes	0.99, Yes
Other Industries	0.98, Yes	0.98, Yes
Govt Federal	0.94, Yes	0.94, Yes
Govt StateLoc	1.00, Yes	1.00, Yes

Table 19: Economy with Foreign-born Country of Origin

	Correlation, Significant					
	Europe	Asia	Africa	Oceania	Latin Am	North Am
Agriculture	0.24, No	0.83, Yes	0.83, Yes	0.83, Yes	0.83, Yes	0.30, No
Mining	0.30, No	0.77, Yes	0.77, Yes	0.76, Yes	0.77, Yes	0.31, No
Utilities	0.46, No	0.70, Yes	0.70, Yes	0.69, Yes	0.70, Yes	0.36, No
Construction	0.43, No	0.80, Yes	0.80, Yes	0.79, Yes	0.80, Yes	0.41, No
Manufacturing	0.34, No	0.90, Yes	0.90, Yes	0.89, Yes	0.90, Yes	0.41, No
Wholesale	0.34, No	0.97, Yes	0.97, Yes	0.96, Yes	0.97, Yes	0.38, No
Retail	0.34, No	0.97, Yes	0.97, Yes	0.96, Yes	0.97, Yes	0.38, No
Transportation	0.34, No	0.96, Yes	0.96, Yes	0.95, Yes	0.96, Yes	0.39, No
Information	0.31, No	0.99, Yes	0.99, Yes	0.99, Yes	0.99, Yes	0.36, No
FIRE	0.33, No	0.96, Yes	0.96, Yes	0.96, Yes	0.96, Yes	0.39, No
Service	0.32, No	0.99, Yes	0.99, Yes	0.98, Yes	0.99, Yes	0.37, No
Education	0.31, No	1.00, Yes	1.00, Yes	1.00, Yes	1.00, Yes	0.35, No
Entertainment	0.32, No	0.99, Yes	0.99, Yes	0.98, Yes	0.99, Yes	0.37, No
Other Industries	0.31, No	0.98, Yes	0.98, Yes	0.98, Yes	0.98, Yes	0.36, No
Govt Federal	0.36, No	0.94, Yes	0.94, Yes	0.94, Yes	0.94, Yes	0.34, No
Govt StateLoc	0.31, No	1.00, Yes	1.00, Yes	1.00, Yes	1.00, Yes	0.35, No

Table 20: Economy with Immigration Visas

	Correlation, Significant			
	IV Emp	IV NonEmp	NIV Emp	NIV NonEmp
Agriculture	0.43, No	0.20, No	0.81, Yes	0.20, No
Mining	0.48, No	0.31, No	0.76, Yes	0.29, No
Utilities	0.63, Yes	0.30, No	0.72, Yes	0.24, No
Construction	0.48, No	0.30, No	0.78, Yes	0.25, No
Manufacturing	0.45, No	0.23, No	0.89, Yes	0.21, No
Wholesale	0.43, No	0.28, No	0.95, Yes	0.26, No
Retail	0.42, No	0.28, No	0.95, Yes	0.26, No
Transportation	0.43, No	0.27, No	0.94, Yes	0.26, No
Information	0.42, No	0.28, No	0.96, Yes	0.27, No
FIRE	0.44, No	0.28, No	0.94, Yes	0.27, No
Service	0.43, No	0.29, No	0.96, Yes	0.27, No
Education	0.41, No	0.29, No	0.95, Yes	0.28, No
Entertainment	0.43, No	0.29, No	0.96, Yes	0.27, No
Other Industries	0.42, No	0.28, No	0.95, Yes	0.27, No
Govt Federal	0.39, No	0.30, No	0.91, Yes	0.29, No
Govt StateLoc	0.41, No	0.29, No	0.95, Yes	0.28, No

Table 21: Economy with Immigration Enforcement

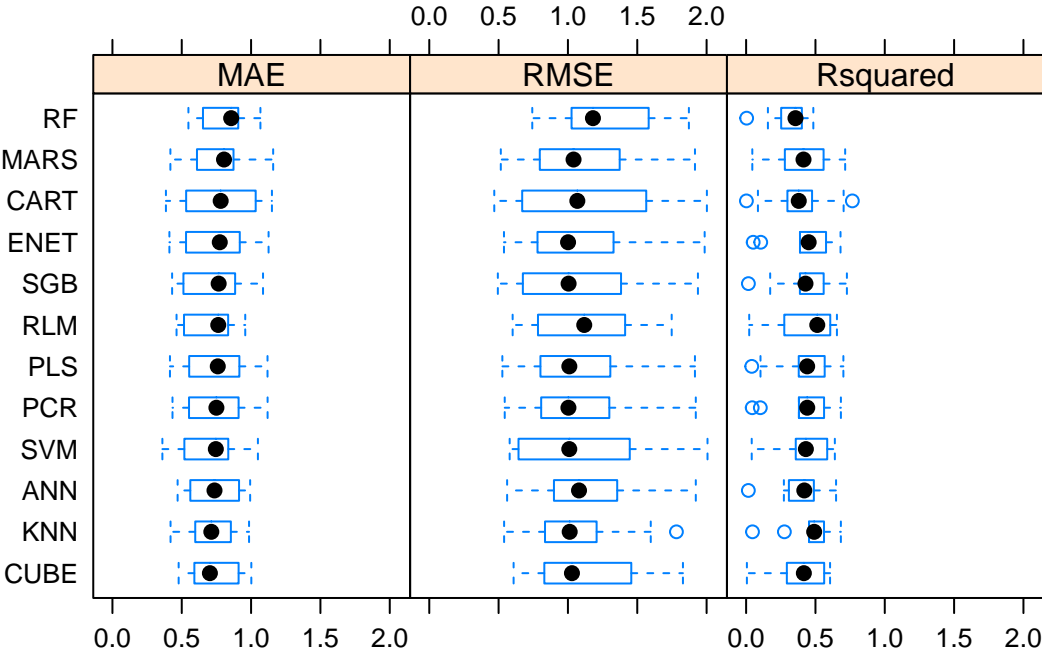
	Correlation, Significant		
	Removed	Apprehended	Returned
Agriculture	0.81, Yes	-0.39, No	-0.41, No
Mining	0.80, Yes	-0.35, No	-0.39, No
Utilities	0.67, Yes	-0.19, No	-0.22, No
Construction	0.71, Yes	-0.27, No	-0.31, No
Manufacturing	0.87, Yes	-0.33, No	-0.37, No
Wholesale	0.85, Yes	-0.38, No	-0.43, No
Retail	0.85, Yes	-0.39, No	-0.43, No
Transportation	0.86, Yes	-0.38, No	-0.42, No
Information	0.88, Yes	-0.40, No	-0.45, No
FIRE	0.85, Yes	-0.38, No	-0.43, No
Service	0.87, Yes	-0.40, No	-0.45, No
Education	0.89, Yes	-0.41, No	-0.46, No
Entertainment	0.87, Yes	-0.40, No	-0.45, No
Other Industries	0.87, Yes	-0.40, No	-0.45, No
Govt Federal	0.90, Yes	-0.38, No	-0.41, No
Govt StateLoc	0.89, Yes	-0.41, No	-0.46, No

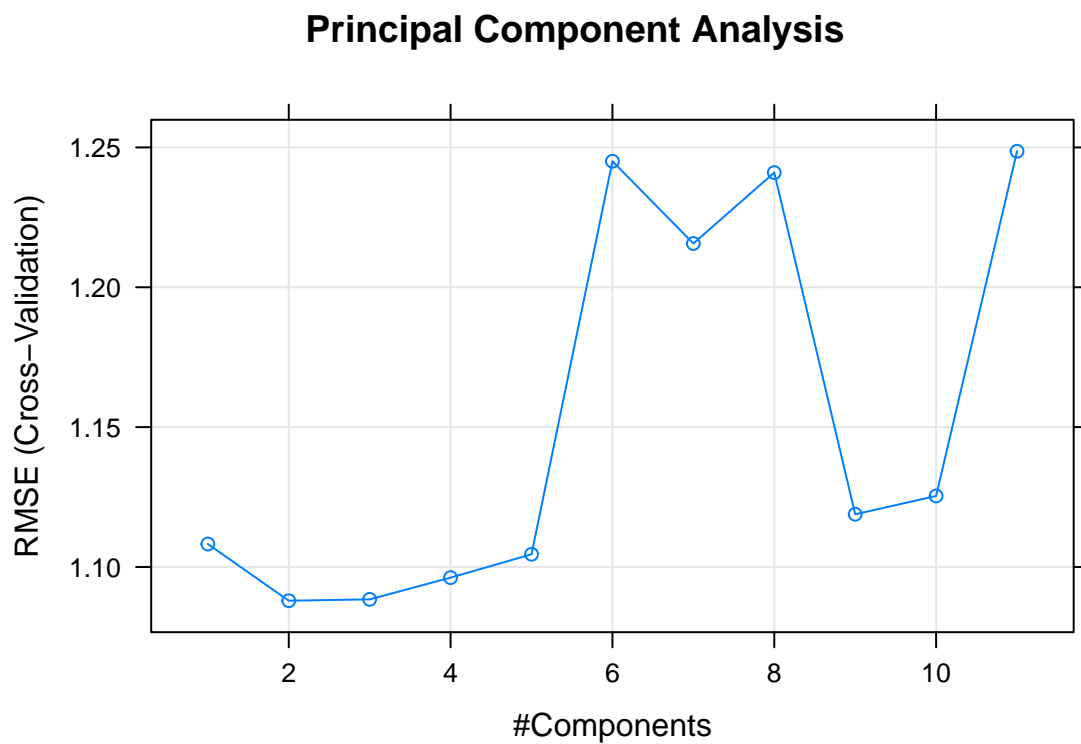
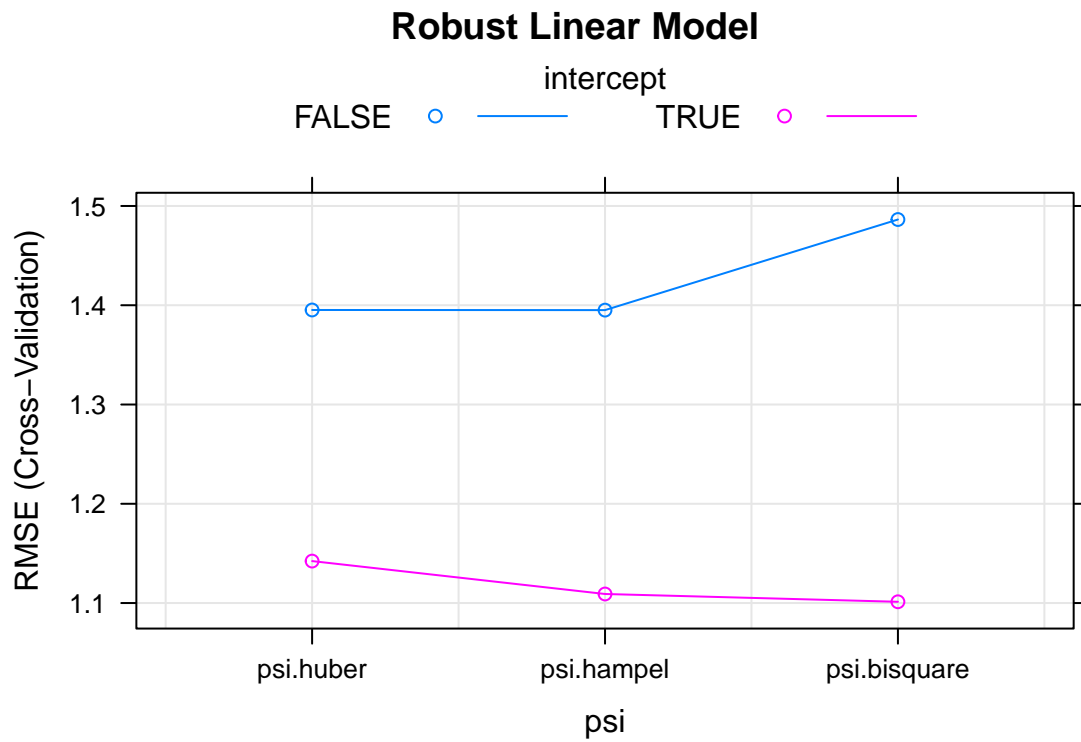
Table 22: Time Series ARIMA Modeling

	Stationary	Box-Cox Transform	p	d	q
Laws	Yes	0.274	5	0	0
White	No	0.374	0	2	1
Black	No	0.22	1	1	0
Native Am	Yes	-0.366	5	0	0
Asian Pac	No	0.255	1	2	2
Other Race	No	0.255	1	2	2
Hispanic	Yes	0.081	0	0	0
Native Born	No	0.302	1	2	1
Foreign Born	No	-0.004	0	2	0
Europe	No	0.217	2	2	3
Asia	No	0.12	1	1	1
Africa	No	0.44	1	2	2
Oceania	No	-0.019	4	2	1
Latin Am	No	0.113	1	1	1
North Am	No	0.54	2	2	2

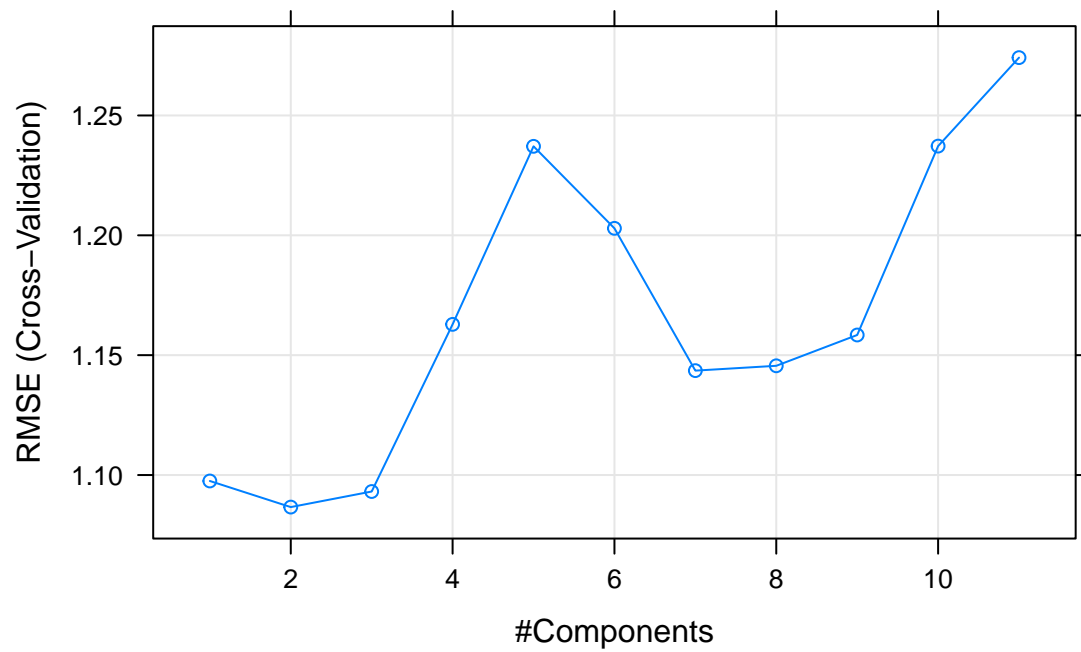
Table 23: Performance Metrics

	Training Set Resampling			Test Set Validation		
	RMSE	Rsquared	MAE	RMSE	Rsquared	MAE
RLM	1.101253	0.4215269	0.7128411	1.717429	0.3141630	0.9216199
PCR	1.087937	0.4378742	0.7331757	1.673304	0.3429144	0.9411131
PLS	1.086624	0.4257641	0.7380014	1.672647	0.3426580	0.9397111
ENET	1.090662	0.4184976	0.7339712	1.697809	0.3291550	0.9539869
ANN	1.134243	0.3915584	0.7437596	1.743432	0.2716640	0.9865738
MARS	1.116175	0.3960809	0.7643166	1.683975	0.3233568	0.9352765
SVM	1.104148	0.4127897	0.7148267	1.743417	0.2898321	0.8877538
KNN	1.057420	0.4609521	0.7081444	1.633050	0.3684154	0.9598214
CART	1.127081	0.3890166	0.7731064	1.826424	0.1877542	1.0078656
RF	1.242846	0.3149483	0.8172722	1.489037	0.4894702	1.0142196
SGB	1.071784	0.4307816	0.7269328	1.699391	0.3222708	0.9427416
CUBE	1.116010	0.3955091	0.7271977	1.603782	0.3949694	0.9653687

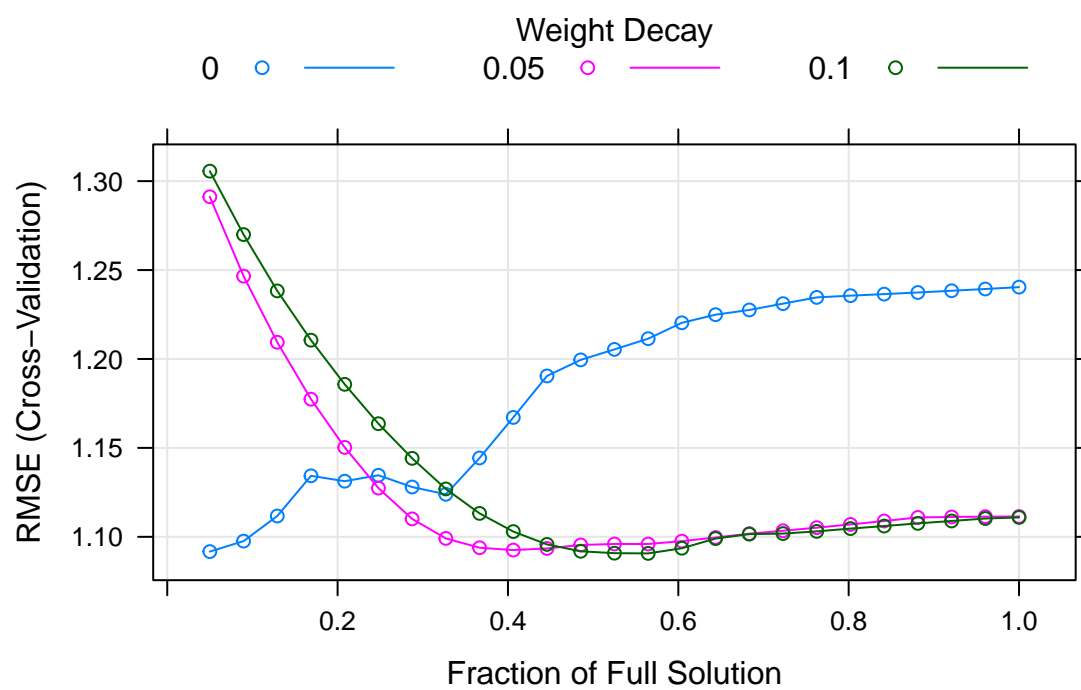


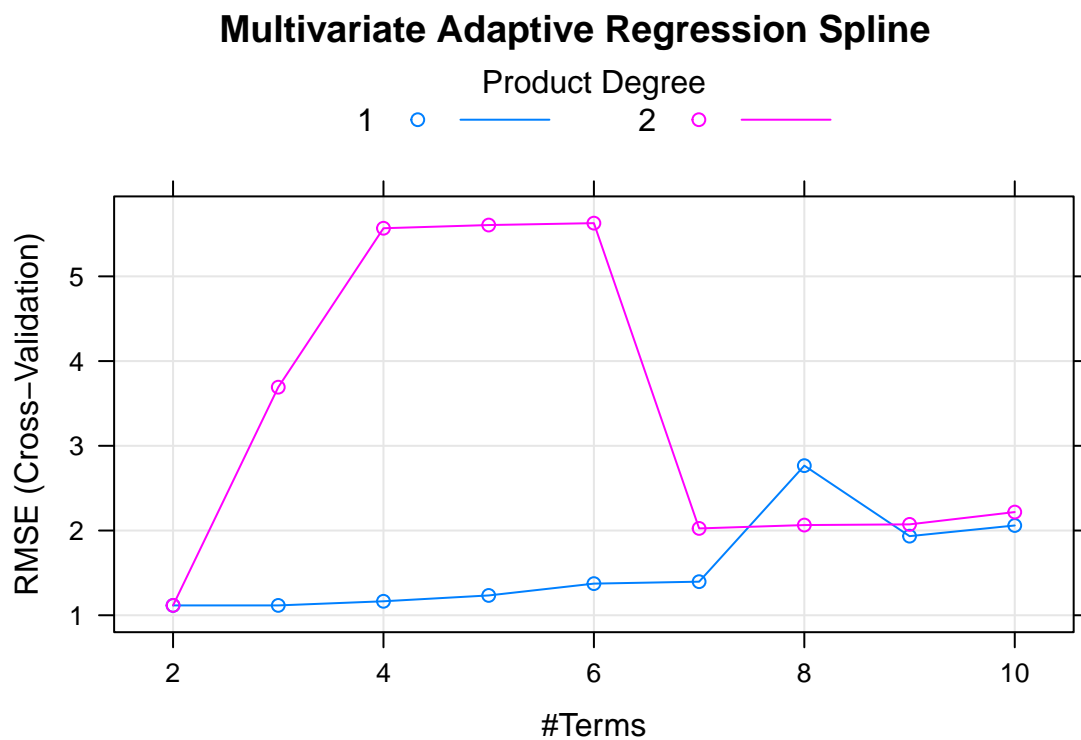
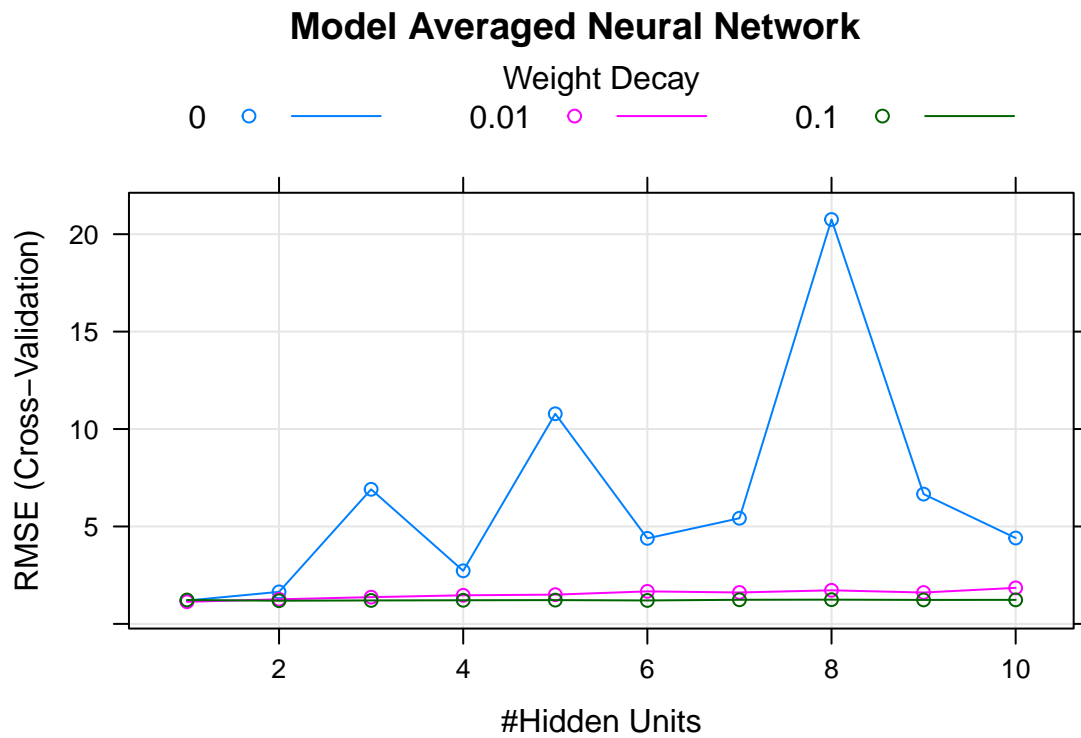


Partial Least Squares

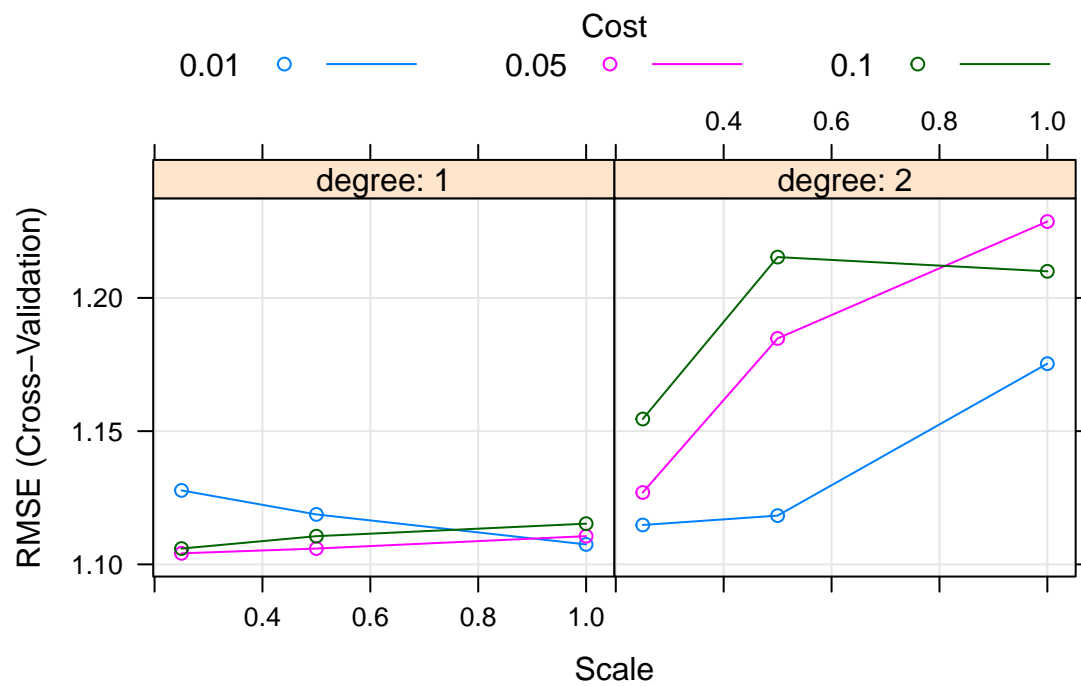


Elasticnet

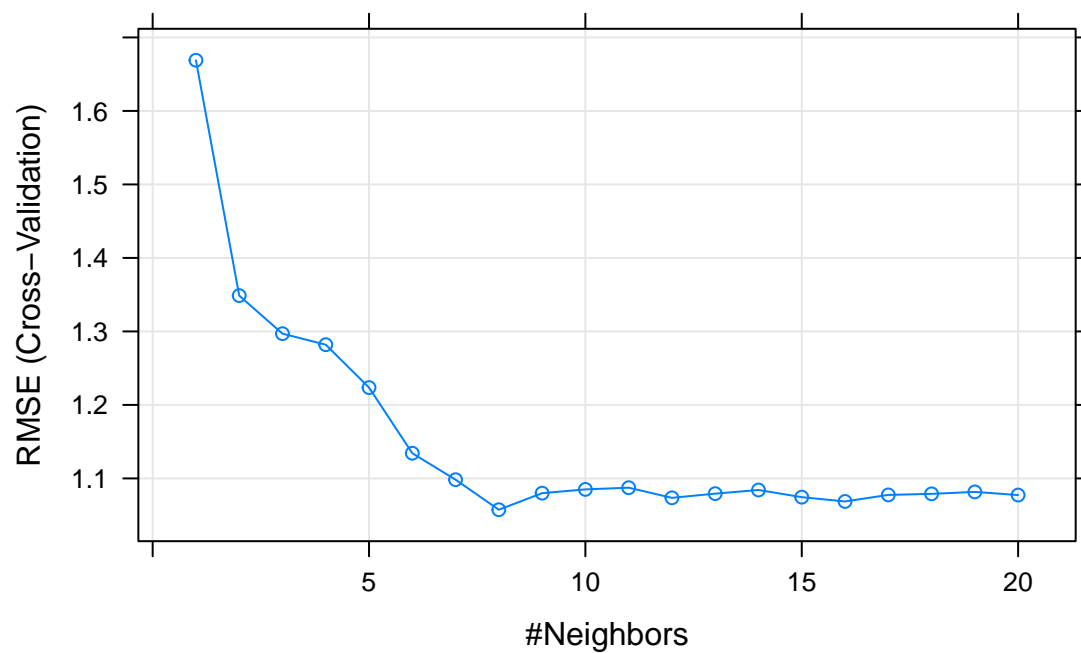


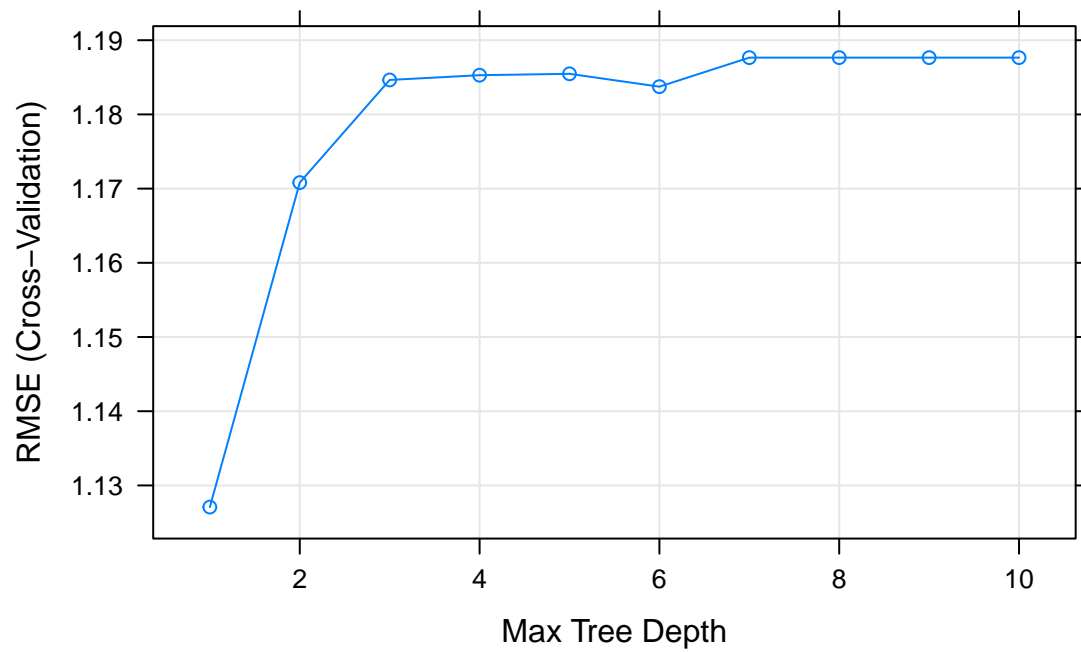
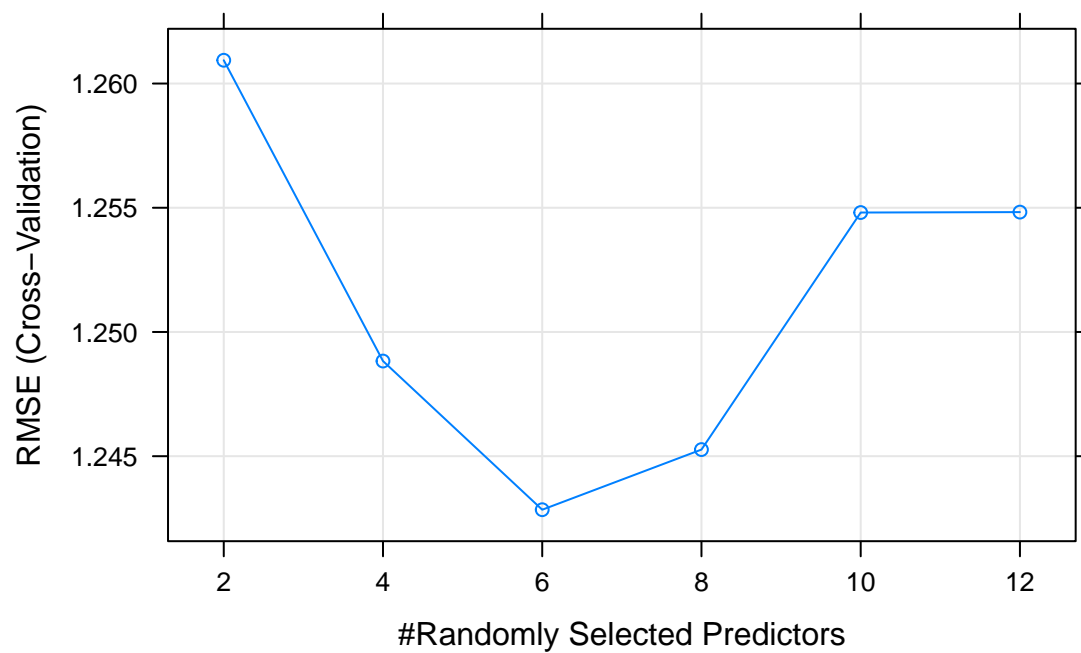


Support Vector Machines with Polynomial Kernel

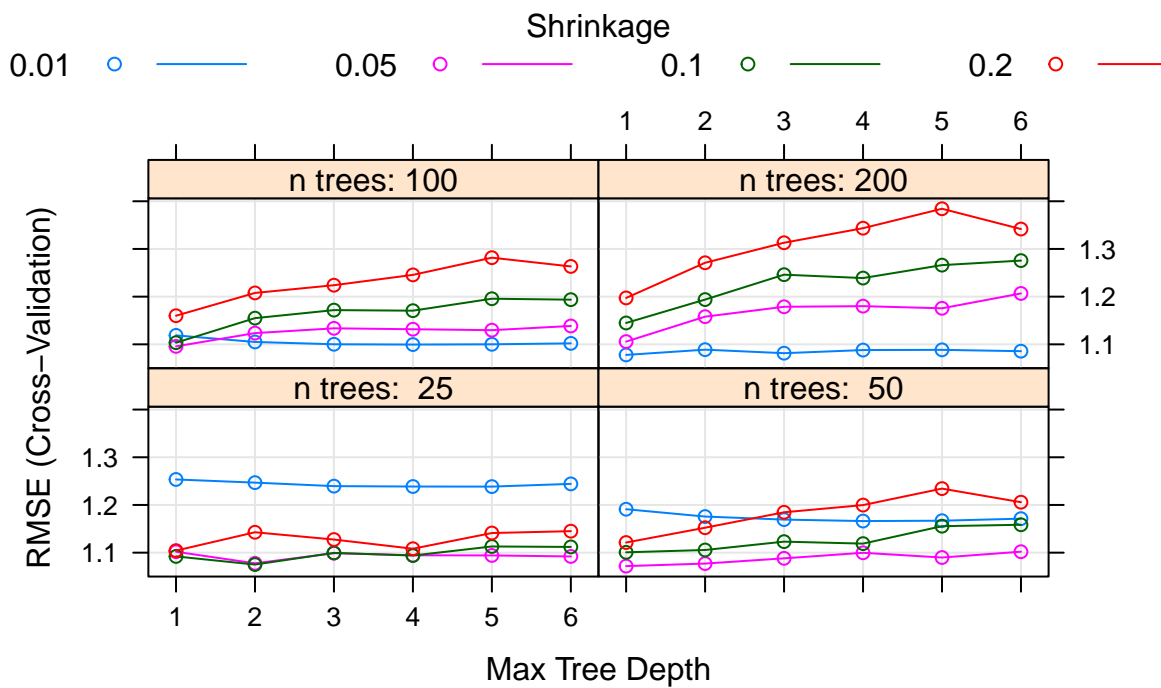


k-Nearest Neighbors

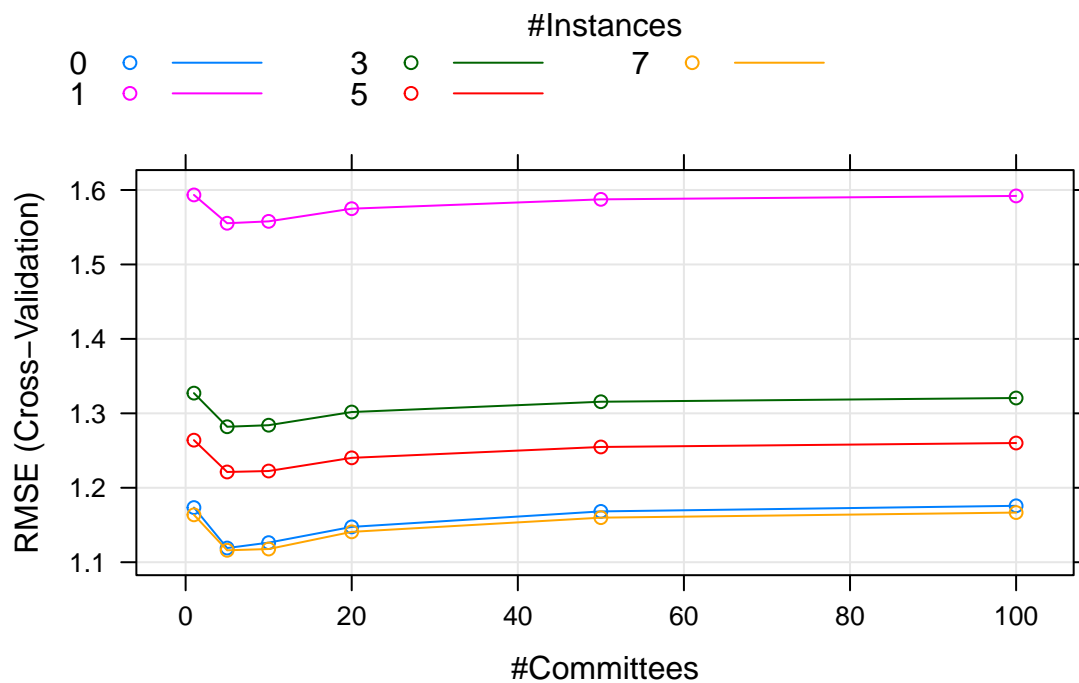


CART**Random Forest**

Stochastic Gradient Boosting



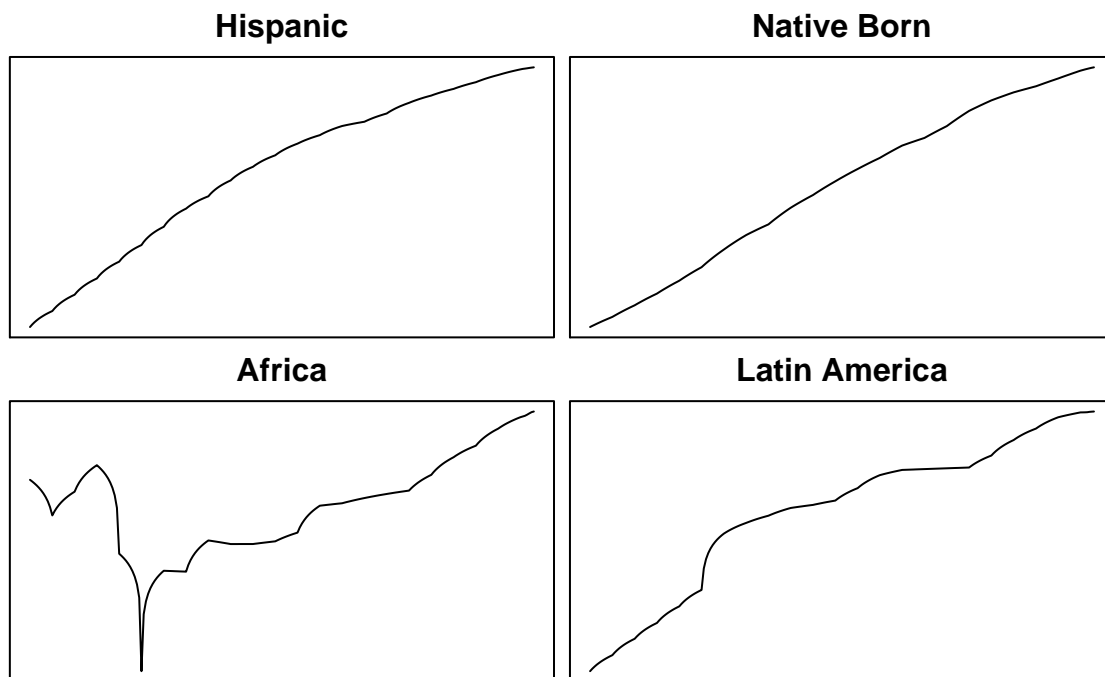
Cubist



Partial Least Squares Regression

```
## Data:      X dimension: 227 12
## Y dimension: 227 1
## Fit method: kernelppls
## Number of components considered: 2
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##      (Intercept)  1 comps  2 comps
## CV           1.601   1.337   1.320
## adjCV        1.601   1.335   1.319
##
## TRAINING: % variance explained
##           1 comps  2 comps
## X           75.80   93.59
## LawsFuture   32.07   34.69
##
##
## Shapiro-Wilk normality test
##
## data: fit1$residuals
## W = 0.76207, p-value < 2.2e-16
```

Box-Cox Transformations

Box-Cox Transformations

Robust Linear Regressions

```
##
## Call:
## lmRob(formula = Hispanic ~ LawsPast, data = LawsCause, control = ctrl)
##
```

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.08282 -0.67745  0.01269  0.58536  1.66576
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.33150    0.08138  -4.073 6.43e-05 ***
## LawsPast     0.43452    0.04596   9.454 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8366 on 225 degrees of freedom
## Multiple R-Squared: 0.2381
##
## Test for Bias:
##              statistic    p-value
## M-estimate      21.927 1.732e-05
## LS-estimate      7.122 2.841e-02
##
## Shapiro-Wilk normality test
##
## data:  fit2$residuals
## W = 0.96259, p-value = 1.132e-05
##
## Call:
## lmRob(formula = Native_Born ~ LawsPast, data = LawsCause, control = ctrl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3563 -0.6769 -0.0247  0.6093  1.8930
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.41201    0.07348  -5.607 6e-08 ***
## LawsPast     0.47003    0.04986   9.428 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9191 on 225 degrees of freedom
## Multiple R-Squared: 0.282
##
## Test for Bias:
##              statistic    p-value
## M-estimate      1.040 0.5945
## LS-estimate     -8.895 1.0000
##
## Shapiro-Wilk normality test
##
## data:  fit3$residuals
## W = 0.96916, p-value = 7.518e-05
##
## Call:

```

```
## lmRob(formula = Africa ~ LawsPast, data = LawsCause, control = ctrl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2357 -0.6354  0.0730  0.6202  2.2801
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.29173    0.07553  -3.862 0.000147 ***
## LawsPast      0.31422    0.04813   6.528 4.36e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9608 on 225 degrees of freedom
## Multiple R-Squared: 0.1994
##
## Test for Bias:
##              statistic p-value
## M-estimate      1.532  0.4649
## LS-estimate    -10.745  1.0000
##
## Shapiro-Wilk normality test
##
## data:  fit4$residuals
## W = 0.98855, p-value = 0.06734
##
## Call:
## lmRob(formula = Latin_Am ~ LawsPast, data = LawsCause, control = ctrl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.900863 -0.646202  0.009532  0.337137  1.424240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06467    0.06345  -1.019   0.309
## LawsPast      0.32013    0.03695   8.664 9.01e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6345 on 225 degrees of freedom
## Multiple R-Squared: 0.1498
##
## Test for Bias:
##              statistic p-value
## M-estimate     19.94 4.680e-05
## LS-estimate     32.35 9.442e-08
##
## Shapiro-Wilk normality test
##
## data:  fit5$residuals
## W = 0.92388, p-value = 2.025e-09
```

Code Listings

```

library(lmtest)
library(tidyr)
library(dplyr)
library(fma)
library(zoo)
library(pls)
library(caret)
library(knitr)
library(psych)
library(scales)
library(robust)
library(ggplot2)
library(forecast)
library(kableExtra)
knitr::opts_chunk$set(echo = T, warning=F, message=F, fig.align='center',
  fig.width=ifelse(knitr::is_html_output(),7,6), fig.height=ifelse(knitr::is_html_output(),5,4))
if (knitr::is_html_output(excludes = "markdown")) {
  options(knitr.table.format = "html") # HTML
} else if (knitr::opts_knit$get('rmarkdown.pandoc.to') == 'docx') {
  options(knitr.table.format = "markdown") # WORD
} else if (knitr::is_latex_output()) {
  options(knitr.table.format = "latex") # PDF
}
github <- "https://raw.githubusercontent.com/jzuniga123"
file <- "/SPS/master/DATA%20698/DATA698_Project.csv"
data698 <- read.csv(paste0(github, file), header=1)
data698 <- data698[order(data698$Year, decreasing=T), ]
na.omit(data.frame(data698[, 3:17], row.names = data698$Year))
temp_df <- data698[, 1:2] %>%
  mutate(Century = floor(data698$Year / 100) * 100) %>%
  mutate(Decade = floor((data698$Year %>% 100) / 10) * 10) %>%
  select(-Year) %>%
  group_by(Century, Decade) %>%
  summarise(Laws = sum(Laws)) %>%
  spread(Century, Laws) %>%
  mutate(Decade = ifelse(Decade==0,
    paste0("Laws During the 0", Decade, "s"),
    paste0("Laws During the ", Decade, "s")))
temp_df
na.omit(data.frame(data698[, 19:25], row.names = data698$Year))
na.omit(data.frame(data698[, 27:42], row.names = data698$Year))
github <- "https://raw.githubusercontent.com/jzuniga123"
file <- "/SPS/master/DATA%20698/DATA698_Project.csv"
data698 <- read.csv(paste0(github, file), header=1)
data698 <- data698[order(data698$Year), ]
data698 <- ts(data698[, -1], frequency=1,
  start=min(data698[, 1]), end=max(data698[, 1]))
Laws <- data698[, 1]
ggplot() + ggtitle("Legislation on Immigration (1790-2016)") +
  geom_line(aes(y = Laws, x = time(Laws)), stat="identity") +
  xlab("Year") + theme(plot.title = element_text(hjust = 0.5))

```

```

Census <- data698[, 2:17]
RAHO <- Census[, 2:8]
RAHO[, 2] <- RAHO[, 2] + RAHO[, 3]
colnames(RAHO)[2] <- gsub("_.*", "", colnames(RAHO)[2])
RAHO <- RAHO[, -3]
numb_df <- data.frame(na.approx(RAHO), Year=time(RAHO))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Race and Hispanic Origin (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Race and Hispanic Origin (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(RAHO)
Proportions <- RAHO / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Race and Hispanic Origin Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Race and Hispanic Origin Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
round(na.omit(data.frame(as.data.frame(Proportions), row.names=time(Proportions))), 6)
NAFB <- Census[, 9:10]
numb_df <- data.frame(na.approx(NAFB), Year=time(NAFB))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Native and Foreign-born (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Native and Foreign-born (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(NAFB)
Proportions <- NAFB / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +

```

```

  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Native and Foreign-born Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Native and Foreign-born Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
round(na.omit(data.frame(as.data.frame(Proportions), row.names=time(Proportions))), 6)
FBCO <- Census[, 11:16]
numb_df <- data.frame(na.approx(FBCO), Year=time(FBCO))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Foreign-born Country of Origin (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Foreign-born Country of Origin (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(FBCO)
Proportions <- FBCO / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Foreign-born Country of Origin Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Foreign-born Country of Origin Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
round(na.omit(data.frame(as.data.frame(Proportions), row.names=time(Proportions))), 6)
Visas <- data698[, 18:21]
Visas[, 2] <- Visas[, 2] - Visas[, 1]
Visas[, 4] <- Visas[, 2] - Visas[, 1]
colnames(Visas) <- c("IV_Emp", "IV_NonEmp", "NIV_Emp", "NIV_NonEmp")
IV_All <- na.omit(data698[, 19])
numb_df <- data.frame(IV_All, Year=time(IV_All))
ggplot() + ggtitle("Immigrant Visas (1820-2016)") + xlab("Year") +
  geom_line(aes(y = IV_All, x = time(IV_All)), stat="identity") +
  ylab("Amount") + theme(plot.title = element_text(hjust = 0.5))
Visa_Types <- window(Visas, start = 1981)
Visa_Types[,2] <- Visa_Types[,2] - Visa_Types[,1]
colnames(Visa_Types)[c(2,4)] <- c("IV_NonEmp", "NIV_NonEmp")

```

```

numb_df <- data.frame(Visa_Types, Year=time(Visa_Types))
plot_df <- gather(numb_df, "Type", "Approved", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Approved)) + geom_line() +
  facet_wrap(~ Type, scales="free_y") + theme_bw() +
  ggtitle("Immigrant and Nonimmigrant Visas by Type (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Approved, fill= Type)) + geom_area() +
  ggtitle("Immigrant and Nonimmigrant Visas by Type (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Enforce <- data698[, 22:24]
Enforce_Window <- window(Enforce, start = 1925)
numb_df <- data.frame(Enforce_Window, Year=time(Enforce_Window))
plot_df <- gather(numb_df, "Method", "Actions", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Actions)) + geom_line() +
  facet_wrap(~ Method, scales="free_y") + theme_bw() +
  ggtitle("Enforcement by Method (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Actions, fill= Method)) + geom_area() +
  ggtitle("Enforcement by Method (1981-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Economy <- data698[, 26:41]
Economy_Window <- window(Economy, start = 1987)
numb_df <- data.frame(Economy_Window, Year=time(Economy_Window))
plot_df <- gather(numb_df, "Sector", "Income", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Income)) + geom_line() +
  facet_wrap(~ Sector, scales="free_y") + theme_bw() +
  ggtitle("GDP by Sector (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Income, fill= Sector)) + geom_area() +
  ggtitle("GDP by Sector (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8)
GDP <- rowSums(Economy_Window)
Proportions <- Economy_Window / GDP
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Sector", "Income", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Income)) + geom_line() +
  facet_wrap(~ Sector, scales="free_y") + theme_bw() +
  ggtitle("GDP by Sector Proportions (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Income, fill= Sector)) + geom_area() +
  ggtitle("GDP by Sector Proportions (1987-2016)") +

```

```

theme(plot.title = element_text(hjust = 0.5)) +
scale_y_continuous(labels = scales::comma) +
scale_fill_grey(start = 0.2, end = 0.8)
round(na.omit(data.frame(as.data.frame(Proportions), row.names=time(Proportions))), 6)
Legislation <- na.approx(window(cbind(
  Past_Laws_2yr=stats::lag(Laws, k=2),
  Past_Laws_1yr=stats::lag(Laws, k=1),
  Current_Laws=Laws,
  Future_Laws_1yr=stats::lag(Laws, k=-1),
  Future_Laws_2yr=stats::lag(Laws, k=-2)),
  start=1790, end=2016), rule = 2)
r <- corr.test(na.approx(RAHO), Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(na.approx(RAHO), Legislation, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with Race and Hispanic Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(na.approx(NAFB), Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(na.approx(NAFB), Legislation, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with Native and Foreign-born") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(na.approx(FBCO), Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(na.approx(FBCO), Legislation, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with Foreign-born Country of Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Visas, Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(Visas, Legislation, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with Immigration Visas") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Enforce, Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(Enforce, Legislation, use="complete.obs", method="kendall")$p < 0.05

```



```

table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with Immigration Enforcement Actions") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Economy, Legislation, use="complete.obs", method="kendall")$r
p <- corr.test(Economy, Legislation, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Laws with the Economy") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- round(corr.test(Economy, na.approx(RAHO), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(RAHO), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Race and Hispanic Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- round(corr.test(Economy, na.approx(NAFB), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(NAFB), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Native and Foreign-born") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- round(corr.test(Economy, na.approx(FBCO), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(FBCO), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Foreign-born Country of Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Economy, Visas, use="complete.obs", method="kendall")$r
p <- corr.test(Economy, Visas, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", "),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))

```

```

rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
      align="c", caption = "Economy with Immigration Visas") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Economy, Enforce, use="complete.obs", method="kendall")$r
p <- corr.test(Economy, Enforce, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=" ", ),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
      align="c", caption = "Economy with Immigration Enforcement") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
Correlated <- data.frame(Laws, RAHO, NAFFB, FBCO)
Arima_Orders <- data.frame(matrix(NA, nrow = ncol(Correlated), ncol = 5))
for (i in 1:ncol(Correlated)){
  ts_var <- na.approx(Correlated[, i])
  adf <- tseries::adf.test(ts_var)$p.value < 0.05
  Arima_Orders[i, 1] <- ifelse(adf, "Yes", "No")
  l <- BoxCox.lambda(ts_var)
  Arima_Orders[i, 2] <- format(round(l, 3), nsmall = 2)
  aa <- auto.arima(ts_var, stepwise=F, approximation=F, d=ifelse(adf, 0, NA), lambda=1)
  Arima_Orders[i, 3] <- aa$arima[1]
  Arima_Orders[i, 4] <- aa$arima[6]
  Arima_Orders[i, 5] <- aa$arima[2]
}
colnames(Arima_Orders) <- c("Stationary", "Box-Cox Transform", "p", "d", "q")
rownames(Arima_Orders) <- gsub('_', ' ', colnames(Correlated))
kable(Arima_Orders, longtable = T, booktabs = T, row.names=T, linesep = "",
      align="c", caption = "Time Series ARIMA Modeling") %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
Causality <- data.frame(matrix(NA, nrow = ncol(Correlated[, -1]), ncol = 2))
for (i in 2:ncol(Correlated)){
  for (j in 1:2) {
    cause <- as.data.frame(na.approx(Correlated[, ifelse(j==1, 1, i)]))
    effect <- as.data.frame(na.approx(Correlated[, ifelse(j==1, i, 1)]))
    d <- max(Arima_Orders[, "d"])
    gtest <- grangertest(cause, effect, order = d)
    f <- format(round(gtest$F[2], 2), nsmall = 2)
    p <- gtest$`Pr(>F)`[2] < 0.05
    Causality[i-1, j] <- paste(f, ifelse(p, "Yes", "No"), sep=" ", )
  }
}
colnames(Causality) <- c("Number of Laws (Antecedent)",
  "Number of Laws (Consequence)")
rownames(Causality) <- gsub('_', ' ', colnames(Correlated[, -1]))
kable(Causality, longtable = T, booktabs = T, row.names=T, linesep = "",
      align="c", caption = "Legislation-Demographic Causality Test") %>%
  add_header_above(c(" ", "Granger Causality Statistic, Significant" = ncol(Causality))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
model_vars <- na.approx(Correlated[, c(1, 3:14)])

```

```

set.seed(698)
rows_train <- createDataPartition(model_vars[,1], p=0.75, list=F)
X_train <- model_vars[rows_train, -1]
X_test <- model_vars[-rows_train, -1]
Y_train <- model_vars[rows_train, 1]
Y_test <- model_vars[-rows_train, 1]
set.seed(698)
ctrl <- trainControl(method = "cv", number = 10)
set.seed(698)
tune01 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "rlm", trControl = ctrl)
plot(tune01, main=tune01$modelInfo$label)
set.seed(698)
tune02 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "pcr", trControl = ctrl, tuneLength = 25)
plot(tune02, main=tune02$modelInfo$label)
set.seed(698)
tune03 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "pls", trControl = ctrl, tuneLength = 25)
plot(tune03, main=tune03$modelInfo$label)
set.seed(698)
tg <- expand.grid(lambda = c(0, 0.05, .1), fraction = seq(0.05, 1, length = 25))
tune04 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "enet", trControl = ctrl, tuneGrid = tg)
plot(tune04, main=tune04$modelInfo$label)
set.seed(698)
tg <- expand.grid(.decay = c(0, 0.01, .1), .size = c(1:10), .bag = F)
tune05 <- train(x = lag(X_train, k=1), y = Y_train,
  method = "avNNet", tuneGrid = tg, trControl = ctrl, linout = T,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  trace = F, MaxNWts = 10 * (ncol(X_train) + 1) + 10 + 1, maxit = 500)
plot(tune05, main=tune05$modelInfo$label)
set.seed(698)
tg <- expand.grid(degree = c(1:2), nprune = c(2:10))
tune06 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "earth", tuneGrid = tg, trControl = ctrl)
plot(tune06, main=tune06$modelInfo$label)
set.seed(698)
tg <- expand.grid(C=c(0.01,0.05,0.1), degree=c(1,2), scale=c(0.25,0.5,1))
tune07 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "svmPoly", tuneGrid = tg, trControl = ctrl)
plot(tune07, main=tune07$modelInfo$label)
set.seed(698)
tg <- data.frame(.k = 1:20)
tune08 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox", "center", "scale", "knnImpute"),
  method = "knn", tuneGrid = tg, trControl = trainControl(method = "cv"))

```

```

plot(tune08, main=tune08$modelInfo$label)
set.seed(698)
tg <- expand.grid(maxdepth= seq(1,10,by=1))
tune09 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox","center","scale","knnImpute"),
  method = "rpart2", tuneGrid = tg, trControl = ctrl)
plot(tune09, main=tune09$modelInfo$label)
set.seed(698)
P <- ncol(X_train)
tg <- expand.grid(mtry=seq(2, P, by = floor(P/5)))
tune10 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox","center","scale","knnImpute"),
  method = "rf", tuneGrid = tg, trControl = ctrl)
plot(tune10, main=tune10$modelInfo$label)
set.seed(698)
tg <- expand.grid(interaction.depth=seq(1,6,by=1), n.trees=c(25,50,100,200),
  shrinkage=c(0.01,0.05,0.1,0.2), n.minobsinnode=10)
tune11 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox","center","scale","knnImpute"),
  method = "gbm", tuneGrid = tg, trControl = ctrl, verbose=F)
plot(tune11, main=tune11$modelInfo$label)
set.seed(698)
tg <- expand.grid(committees = c(1,5,10,20,50,100), neighbors = c(0,1,3,5,7))
tune12 <- train(x = lag(X_train, k=1), y = Y_train,
  preProcess = c("BoxCox","center","scale","knnImpute"),
  method = "cubist", tuneGrid = tg, trControl = ctrl)
plot(tune12, main=tune12$modelInfo$label)
metrics <- function(tune) {
  RMSE = min(tune$results$RMSE)
  Rsquared = max(tune$results$Rsquared)
  MAE = min(tune$results$MAE)
  return(cbind(RMSE, Rsquared, MAE)) }
resampling <- data.frame(rbind(metrics(tune01), metrics(tune02),
  metrics(tune03), metrics(tune04), metrics(tune05), metrics(tune06),
  metrics(tune07), metrics(tune08), metrics(tune09), metrics(tune10),
  metrics(tune11), metrics(tune12)), row.names = c("RLM","PCR","PLS",
  "ENET", "ANN", "MARS", "SVM", "KNN", "CART", "RF", "SGB", "CUBE"))
validation <- data.frame(row.names = c("RLM","PCR","PLS", "ENET",
  "ANN", "MARS", "SVM", "KNN", "CART", "RF", "SGB", "CUBE"), rbind(
  postResample(pred = predict(tune01, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune02, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune03, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune04, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune05, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune06, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune07, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune08, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune09, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune10, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune11, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune12, newdata = X_test), obs = Y_test)))
kable(cbind(resampling, validation), longtable = T, booktabs = T, row.names=T, linesep = "",
  caption = "Performance Metrics") %>%

```

```

  add_header_above(c("", "Training Set Resampling" = 3, "Test Set Validation" = 3)) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
fits <- list(RLM=tune01, PCR=tune02, PLS=tune03, ENET=tune04, ANN=tune05, MARS=tune06,
  SVM=tune07, KNN=tune08, CART=tune09, RF=tune10, SGB=tune11, CUBE=tune12)
bwplot(resamples(fits))
LagFwd <- data.frame(stats::lag(na.approx(data698[,1]), k=-1))
LagBwd <- data.frame(stats::lag(na.approx(data698[,1]), k=1))
X <- data.frame(na.approx(data698[,c(3:11,13:16)]))
X[, 2] <- X[, 2] + X[, 3]
colnames(X)[2] <- gsub("_.*", "", colnames(X)[2])
X <- X[, -3]
prepro <- preprocess(X, method=c("BoxCox", "center", "scale", "knnImpute"))
demographics <- predict(prepro, X)
LawsEffect <- data.frame(LagFwd, demographics)
colnames(LawsEffect) <- c("LawsFuture", colnames(demographics))
LawsCause <- data.frame(LagBwd, demographics)
colnames(LawsCause) <- c("LawsPast", colnames(demographics))
fit1 <- plsr(LawsFuture ~ ., data = LawsEffect, ncomp = 2, validation = "CV", jackknife = T)
shapiro.test(fit1$residuals)
jack <- jack.test(fit1, ncomp = fit1$ncomp, use.mean = T)
model <- data.frame(jack[c(1:3,5)])
model <- model[order(model[,4]>0.05, -abs(model[,1])),]
colnames(model) <- c("Estimate", "Std. Error", "t value", "Pr(>|t|)")
rownames(model) <- gsub('_', ' ', rownames(model))
kable(model, longtable = T, booktabs = T, row.names=T, linesep = "",
  caption = "Number of Laws (Consequence)") %>%
  add_header_above(c("", "Ordered Partial Least Squares (PLS) Coefficients" = ncol(model))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
ctrl <- lmRob.control(weight=c("Bisquare", "Bisquare"))
fit2 <- lmRob(Hispanic ~ LawsPast, data = LawsCause, control=ctrl)
fit3 <- lmRob(Native_Born ~ LawsPast, data = LawsCause, control=ctrl)
fit4 <- lmRob(Africa ~ LawsPast, data = LawsCause, control=ctrl)
fit5 <- lmRob(Latin_Am ~ LawsPast, data = LawsCause, control=ctrl)
par(mfrow=c(2, 2), mar = c(0, 0.5, 2, 0), oma = c(0.5, 0.5, 2, 0.5))
plot(LawsCause$Hispanic, type="l", xaxt = "n", yaxt = "n", main="Hispanic")
plot(LawsCause$Native_Born, type="l", xaxt = "n", yaxt = "n", main="Native Born")
plot(LawsCause$Africa, type="l", xaxt = "n", yaxt = "n", main="Africa")
plot(LawsCause$Latin_Am, type="l", xaxt = "n", yaxt = "n", main="Latin America")
title("Box-Cox Transformations", outer=TRUE)
shapiro.test(fit2$residuals)
shapiro.test(fit3$residuals)
shapiro.test(fit4$residuals)
shapiro.test(fit5$residuals)
models <- rbind(Hispanic=summary(fit2)$coefficients[-1,],
  Native_Born=summary(fit3)$coefficients[-1,],
  Africa=summary(fit4)$coefficients[-1,],
  Latin_Am=summary(fit5)$coefficients[-1,])
models <- models[order(models[,1], decreasing = T), ]
rownames(models) <- gsub('_', ' ', rownames(models))
kable(models, longtable = T, booktabs = T, row.names=T, linesep = "",
  caption = "Ordered Number of Laws (Antecedent)") %>%
  add_header_above(c("", "Ordered Robust Linear Model (RLM) Coefficients" = ncol(models))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))

```



```

github <- "https://raw.githubusercontent.com/jzuniga123"
file <- "/SPS/master/DATA%20698/DATA698_Project.csv"
data698 <- read.csv(paste0(github, file), header=1)
data698 <- data698[order(data698$Year, decreasing=T), ]
kable(temp_df, booktabs = T, linesep = "",
      caption = "Legislation on Immigration (1790-2016)" %>%
      kable_styling(latex_options = c("striped", "hold_position"))
ggplot() + ggtitle("Legislation on Immigration (1790-2016)") +
  geom_line(aes(y = Laws, x = time(Laws)), stat="identity") +
  xlab("Year") + theme(plot.title = element_text(hjust = 0.5))
table_df1 <- na.omit(data.frame(data698[, 4:10], row.names = data698$Year))
colnames(table_df1) <- gsub('_', ' ', colnames(table_df1))
kable(table_df1, longtable = T, booktabs = T, row.names=T, linesep = "",
      caption = "Race and Hispanic Origin (1790-2016)" %>%
      kable_styling(latex_options = c("striped", "hold_position", "repeat_header")))
numb_df <- data.frame(na.approx(RAHO), Year=time(RAHO))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Race and Hispanic Origin (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Race and Hispanic Origin (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(RAHO)
Proportions <- RAHO / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Race and Hispanic Origin Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Race and Hispanic Origin Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
table_df2 <- na.omit(data.frame(data698[, 11:17], row.names = data698$Year))
colnames(table_df2) <- gsub('_', ' ', colnames(table_df2))
kable(table_df2, longtable = T, booktabs = T, row.names=T, linesep = "",
      caption = "Native, Foreign-born, and Origin (1790-2016)" %>%
      kable_styling(latex_options = c("striped", "hold_position", "repeat_header")))
numb_df <- data.frame(na.approx(NAFB), Year=time(NAFB))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Native and Foreign-born (1790-2016)") +

```

```

  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Native and Foreign-born (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(NAFB)
Proportions <- NAFB / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Native and Foreign-born Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Native and Foreign-born Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
numb_df <- data.frame(na.approx(FBCO), Year=time(FBCO))
plot_df <- gather(numb_df, "Demographic", "Population", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Foreign-born Country of Origin (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Foreign-born Country of Origin (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Population <- rowSums(FBCO)
Proportions <- FBCO / Population
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Demographic", "Population", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Population)) + geom_line() +
  facet_wrap(~ Demographic, scales="free_y") + theme_bw() +
  ggtitle("Foreign-born Country of Origin Proportions (1790-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Population, fill= Demographic)) + geom_area() +
  ggtitle("Foreign-born Country of Origin Proportions (1790-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
table_df <- na.omit(data.frame(data698[, 19:25], row.names = data698$Year))
colnames(table_df) <- gsub('_', ' ', colnames(table_df))

```

```

kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
      caption = "Visas and Enforcement (1981-2016)") %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
Visa_Types <- window(Visas, start = 1981)
Visa_Types[,2] <- Visa_Types[,2] - Visa_Types[,1]
colnames(Visa_Types)[c(2,4)] <- c("IV_NonEmp", "NIV_NonEmp")
numb_df <- data.frame(Visa_Types, Year=time(Visa_Types))
plot_df <- gather(numb_df, "Type", "Approved", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Approved)) + geom_line() +
  facet_wrap(~ Type, scales="free_y") + theme_bw() +
  ggtitle("Immigrant and Nonimmigrant Visas by Type (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Approved, fill= Type)) + geom_area() +
  ggtitle("Immigrant and Nonimmigrant Visas by Type (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
Enforce_Window <- window(Enforce, start = 1925)
numb_df <- data.frame(Enforce_Window, Year=time(Enforce_Window))
plot_df <- gather(numb_df, "Method", "Actions", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Actions)) + geom_line() +
  facet_wrap(~ Method, scales="free_y") + theme_bw() +
  ggtitle("Enforcement by Method (1981-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Actions, fill= Method)) + geom_area() +
  ggtitle("Enforcement by Method (1981-2016)") + xlab("Year") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8) +
  theme(legend.position="bottom")
IV_All <- na.omit(data698[, 19])
numb_df <- data.frame(IV_All, Year=time(IV_All))
ggplot() + ggtitle("Immigrant Visas (1820-2016)") + xlab("Year") +
  geom_line(aes(y = IV_All, x = time(IV_All)), stat="identity") +
  ylab("Amount") + theme(plot.title = element_text(hjust = 0.5))
table_df1 <- na.omit(data.frame(data698[, 27:34], row.names = data698$Year))
colnames(table_df1) <- c("AGRI", "MINE", "UTIL", "CONST", "MFG", "WHOLE", "RETAIL", "TRANS")
kable(table_df1, longtable = T, booktabs = T, row.names=T, linesep = "",
      caption = "GDP by Sector Part 1 (1987-2016)") %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
table_df2 <- na.omit(data.frame(data698[, 35:42], row.names = data698$Year))
colnames(table_df2) <- c("INFO", "FIRE", "SRVC", "EDU", "ENT", "OTHR", "FED", "STLOC")
kable(table_df2, longtable = T, booktabs = T, row.names=T, linesep = "",
      caption = "GDP by Sector Part 2 (1987-2016)") %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
Economy_Window <- window(Economy, start = 1987)
numb_df <- data.frame(Economy_Window, Year=time(Economy_Window))
plot_df <- gather(numb_df, "Sector", "Income", 1:(ncol(numb_df)-1))
ggplot(plot_df, aes(x=Year, y=Income)) + geom_line() +
  facet_wrap(~ Sector, scales="free_y") + theme_bw() +

```



```

ggtitle("GDP by Sector (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Income, fill= Sector)) + geom_area() +
  ggtitle("GDP by Sector (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8)
GDP <- rowSums(Economy_Window)
Proportions <- Economy_Window / GDP
prop_df <- data.frame(na.approx(Proportions), Year=time(Proportions))
plot_df <- gather(prop_df, "Sector", "Income", 1:(ncol(prop_df)-1))
ggplot(plot_df, aes(x=Year, y=Income)) + geom_line() +
  facet_wrap(~ Sector, scales="free_y") + theme_bw() +
  ggtitle("GDP by Sector Proportions (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
ggplot(plot_df, aes(x=Year, y=Income, fill= Sector)) + geom_area() +
  ggtitle("GDP by Sector Proportions (1987-2016)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_grey(start = 0.2, end = 0.8)
r <- round(corr.test(Economy, na.approx(RAHO), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(RAHO), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 3), nsmall = 3), ifelse(p, "Yes", "No"), sep=", ",
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Race and Hispanic Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- round(corr.test(Economy, na.approx(NAFB), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(NAFB), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", ",
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Native and Foreign-born") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- round(corr.test(Economy, na.approx(FBCO), use="complete.obs", method="kendall")$r, 5)
p <- corr.test(Economy, na.approx(FBCO), use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", ",
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Foreign-born Country of Origin") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Economy, Visas, use="complete.obs", method="kendall")$r

```

```

p <- corr.test(Economy, Visas, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", ", ),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Immigration Visas") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
r <- corr.test(Economy, Enforce, use="complete.obs", method="kendall")$r
p <- corr.test(Economy, Enforce, use="complete.obs", method="kendall")$p < 0.05
table_df <- matrix(paste(format(round(r, 2), nsmall = 2), ifelse(p, "Yes", "No"), sep=", ", ),
  nrow=nrow(r), dimnames=dimnames(r) )
colnames(table_df) <- gsub('_', ' ', colnames(table_df))
rownames(table_df) <- gsub('_', ' ', rownames(table_df))
kable(table_df, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Economy with Immigration Enforcement") %>%
  add_header_above(c(" ", "Correlation, Significant" = ncol(table_df))) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
kable(Arima_Orders, longtable = T, booktabs = T, row.names=T, linesep = "",
  align="c", caption = "Time Series ARIMA Modeling") %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
metrics <- function(tune) {
  RMSE = min(tune$results$RMSE)
  Rsquared = max(tune$results$Rsquared)
  MAE = min(tune$results$MAE)
  return(cbind(RMSE, Rsquared, MAE)) }
resampling <- data.frame(rbind(metrics(tune01), metrics(tune02),
  metrics(tune03), metrics(tune04), metrics(tune05), metrics(tune06),
  metrics(tune07), metrics(tune08), metrics(tune09), metrics(tune10),
  metrics(tune11), metrics(tune12)), row.names = c("RLM", "PCR", "PLS",
  "ENET", "ANN", "MARS", "SVM", "KNN", "CART", "RF", "SGB", "CUBE"))
validation <- data.frame(row.names = c("RLM", "PCR", "PLS", "ENET",
  "ANN", "MARS", "SVM", "KNN", "CART", "RF", "SGB", "CUBE"), rbind(
  postResample(pred = predict(tune01, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune02, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune03, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune04, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune05, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune06, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune07, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune08, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune09, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune10, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune11, newdata = X_test), obs = Y_test),
  postResample(pred = predict(tune12, newdata = X_test), obs = Y_test)))
kable(cbind(resampling, validation), longtable = T, booktabs = T, row.names=T, linesep = "",
  caption = "Performance Metrics") %>%
  add_header_above(c(" ", "Training Set Resampling" = 3, "Test Set Validation" = 3)) %>%
  kable_styling(latex_options = c("striped", "hold_position", "repeat_header"))
fits <- list(RLM=tune01, PCR=tune02, PLS=tune03, ENET=tune04, ANN=tune05, MARS=tune06,
  SVM=tune07, KNN=tune08, CART=tune09, RF=tune10, SGB=tune11, CUBE=tune12)
bwplot(resamples(fits))

```

```
plot(tune01, main=tune01$modelInfo$label)
plot(tune02, main=tune02$modelInfo$label)
plot(tune03, main=tune03$modelInfo$label)
plot(tune04, main=tune04$modelInfo$label)
plot(tune05, main=tune05$modelInfo$label)
plot(tune06, main=tune06$modelInfo$label)
plot(tune07, main=tune07$modelInfo$label)
plot(tune08, main=tune08$modelInfo$label)
plot(tune09, main=tune09$modelInfo$label)
plot(tune10, main=tune10$modelInfo$label)
plot(tune11, main=tune11$modelInfo$label)
plot(tune12, main=tune12$modelInfo$label)
summary(fit1)
shapiro.test(fit1$residuals)
par(mfrow=c(2, 2), mar = c(0, 0.5, 2, 0), oma = c(0.5, 0.5, 2, 0.5))
plot(LawsCause$Hispanic, type="l", xaxt = "n", yaxt = "n", main="Hispanic")
plot(LawsCause$Native_Born, type="l", xaxt = "n", yaxt = "n", main="Native Born")
plot(LawsCause$Africa, type="l", xaxt = "n", yaxt = "n", main="Africa")
plot(LawsCause$Latin_Am, type="l", xaxt = "n", yaxt = "n", main="Latin America")
title("Box-Cox Transformations", outer=TRUE)
summary(fit2)
shapiro.test(fit2$residuals)
summary(fit3)
shapiro.test(fit3$residuals)
summary(fit4)
shapiro.test(fit4$residuals)
summary(fit5)
shapiro.test(fit5$residuals)
```