STA 141C Final Report

3/22/2019

Evan Batzer, Damon Gaylor, Wyatt Robertson

Abstract

In this project, we aimed to provide a quantitative analysis of several topics related the current debate on illegal immigration and border security in the United States. Using the USASpending transactions database, in addition to several other publicly available datasets, we asked 4 key questions related to:

    1) Temporal patterns in border security spending.
    2) The intended purposes of border security spending.
    3) The location of border security spending.
    4) Border security and rates of illegal immigration.

Together, we find evidence for a general decrease in border security spending, accompanied by changing priorities in spending directives. As expected, this investment appeared highest at key border states, but appears to show little relationship with recent trends in illegal immigration.

Introduction

Concerns over illegal immigration have become a common theme in modern political discourse. Recent presidential and congressional campaigns have often made a point to highlight contrasting policy approaches to attract voters, as perspectives on the issue increasingly fall along party lines. This is perhaps best illustrated by the 2016 presidential election, where Donald Trump and Hillary Clinton clashed repeatedly over funding for border security, prompting claims to either "build the wall" or "abolish I.C.E". Conflict over the issue shows little sign of stopping -- the President's recent use of executive powers to bypass budget constraints prompted outcry from both Republican and Democratic legislators.

In addition to polarized public opinion[1], there appears to be little consensus on key facts underlying the current debate on illegal immigration and border security -- rates of illegal immigration, approaches to managing illegal immigration, where and how much money is spent on border security, etc. Without a source of shared information, it will be difficult to attain any sort of agreement between members of the public on the value of changing our national borders security policy.

USASpending.gov, a database of publicly available government transactions, presents a unique opportunity to generate a baseline set of information on border security using statistics provided directly from the agencies responsible for its enforcement. Using this data, in addition to other datasets provided by the federal government, we aim to answer a series of key questions:

1. Are trends in border security-related funding consistent over time? Do patterns of border funding follow presidential and congressional party affiliations?

2. How is border security funding spent? Have priorities (in terms of technology, services provided by the military, physical barriers, etc.) changed over time?

3. Where is border funding being spent? Is it concentrated along the Southwestern United States? Is more money spent on border security for the coast vs the southern and northern border?

4. Does border funding have a negative relationship with rates of immigration (legal and illegal)? Are these trends national or isolated to certain states?

In doing so, we hope to produce a comprehensive analysis that may be useful when evaluating the claims and policies put forth by our elected representatives.

Data Selection

What constitutes border spending?

The USAspending database contains a number of variables that may be useful in determining what constitutes border spending. Initially, we attempted to find funding related to border security through a keyword search, under the assumption that certain descriptors (including the transaction description, product or service description, and business category) would contain words such as "border" or "immigration". However, descriptions of individual transactions were quite sparse, so we chose to examine

	Variable Name	Description
General Description	award_id	Funding award ID number
	action_date	Date of transaction (numeric)
	fiscal_year	Fiscal year of transaction
	total_obligation	Total cost of award
	generated_pragmatic_obligation	Net cost of transaction (can be negative)
	transaction_description	Text description of transaction
Location	pop_country_name	Place of performance country name
	pop_state_code	Place of state code (numeric)
	pop_county_name	Place of performance county name
	pop_zip5	Place of performance zip code
Award Type	award_category	Type of award (grant/contract)
	naics_code	North American Industry Classification System code
	business_categories	Business category description
Agency Info	subtier_agency_id	Subtier agency numeric ID
	awarding_subtier_agency_name	Name of subtier ("level 2") agency awarding transaction
	funding_subtier_agency_name	Name of subtier ("level 2") agency funding transaction

Table 1: Description of variables queried from the USASpending dataset.

spending patterns of several key agencies as an alternative.

In a recent document released by the current Administration, President Trump highlighted contributions to 3 key agencies to increase border security[2]. These agencies included the Department of Immigration and Customs Enforcement (I.C.E.), the Department of Customs and Border Protection (C.B.P.) and the Executive Office of the President for Immigration Review (E.O.I.R). These agencies were all found to be represented within the "subtier agency" variable found within the USAspending transactions database, though we found no individual transactions that could be directly tied to the E.O.I.R.

Using the USASpending SQL database, we queried the transaction table ("universal_transaction_matview") and joined it with information on subtier agencies via the "subtier_agency" table. Then we filtered this joined dataset to those columns where either the funding or awarding agency matched the agency IDs of C.B.P., I.C.E. or E.O.I.R. This command, when run on the cluster, proved far more efficient than subsetting of the general "transaction.csv" file in bash.

In addition to columns related to subtier agencies, we also selected a number of other variables that we thought may be useful to our analysis (Table 1). The resulting dataset was roughly 100mb in size, which we then downloaded locally.

How to filter unique transactions?

We attempted to filter out transactions in this database by examining the uniqueness of each individual row across a number of columns. We considered a transaction distinct if it was unique relative to every other observation on at least one of the following variables: "award_id", "action_date", "total_obligation", "generated_pragmatic_obligation", and "transaction_description".

This filtering process led to a number of transactions with shared "award_id" and "total_obligation" values, yet different "action_date" and "generated_pragmatic_obligation". This led us to conclude that awards reflect total contracts or grants generated, though the actual expenditure of available funding is reflected in the pragmatic obligation.

How to determine the value of spending?

Given two separate variables that may be used to determine spending ("total_obligation" and "generated_pragmatic_obligation"), we calculated the reported annual spending of our two key agencies, C.B.P. and I.C.E.. This group-by computation was conducted for both types of obligation and compared to reported budgets for the two agencies. The result suggests that summation over "total_obligation" dramatically over-reports government spending, while summation over "generated_pragmatic_obligation" under-reports spending. Given our previous observations, we chose to focus on "generated_pragmatic_obligation" under the assumption

that this variable better captures the amount spent at a given point in time, rather than a contract that may be not be spent in full, or spent long after it has been awarded.

Prior to analysis, we created inflation-adjusted spending amounts using the Consumer Price Index provided by the Federal Reserve (averaged annually) to standardize dollar values. Additional comparison with

Other datasets:

We accessed several other datasets in this analysis:

- NAICS codes and descriptions [3]
- DHS Yearbook of Immigration Statistics [4]

Question 1 - Are trends in border security-related funding consistent over time? Do patterns of border funding follow presidential affiliations?

To characterize spending over time, we performed a group-by computation on each month by summing up the adjusted generated pragmatic obligation (which does not contain duplicates) to generate a list of sums. Plotting these sums below for each month from 2005 to present day, there appears to be a general increase in overall border funding.
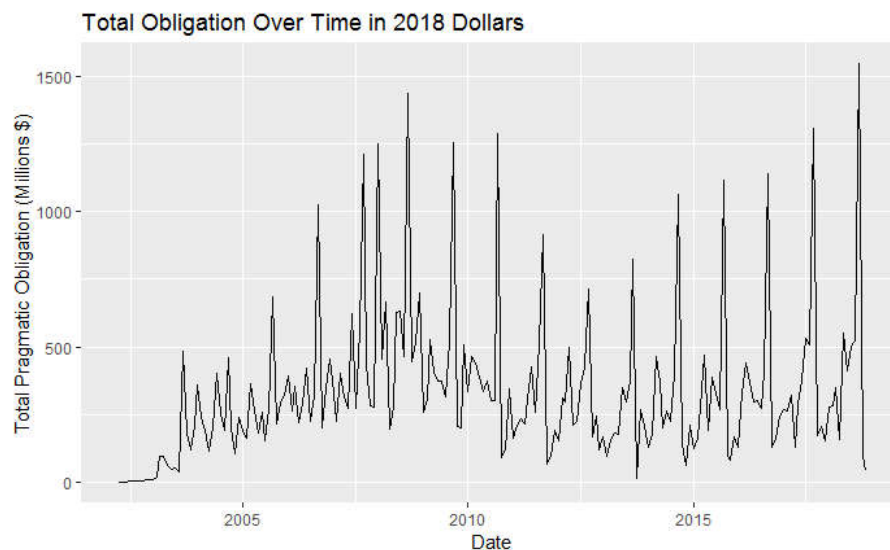
Figure 1: Generated pragmatic obligation over time, aggregated by month

There also appear to be a handful of spikes corresponding to the beginnings of each year, which may suggest that the majority of funds for each year are allocated early to put various programs into place. Additionally, funding before 2005 appears to be smaller than after 2005, which is consistent with limitations of the data listed in the data section. By grouping adjusted pragmatic obligation by presidential terms this difference becomes more apparent. Therefore,

the data provided in the USA Spending database before 2005 may be unreliable, which limits our results for analyses including this time period.
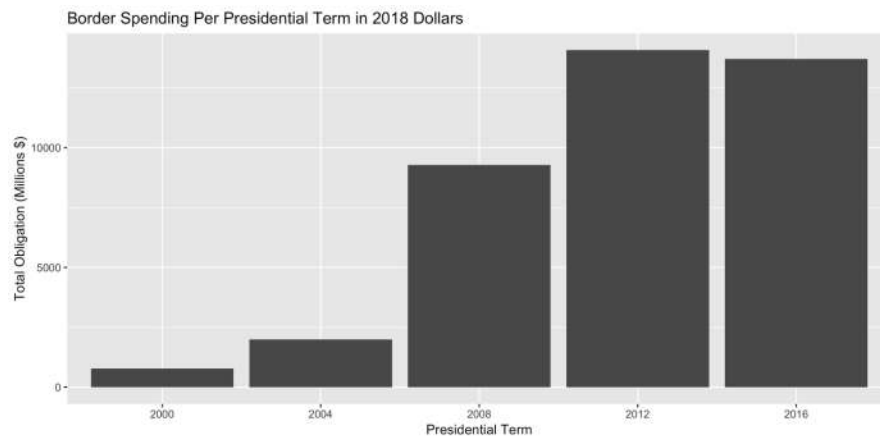
Figure 2: Generated pragmatic obligation over time, aggregated by presidential term.

Question 2 - How is border security funding spent? Have priorities (transaction objectives) changed over time?

To determine patterns in border funding priorities, we examined NAICS (North American Industrial Classification System) codes associated with each transaction in the last 10 fiscal years. NAICS codes present a standardized, hierarchical method of classifying transaction types, consisting of 2 digit "Sector" codes, followed by up to 4 additional digits used to add specificity. We began by filtering to transactions with NAICS codes present, excluding roughly 7% of our dataset.

Using a series of string selection operations, we broke down NAICS code character strings into new strings, 2 - 6 digits in length, that could then be joined to our NAICS term dictionary. Following the generation of a graph denoting nodes (agency funding sources and NAICS codes) and edges (aggregate pragmatic obligation), we created a Sankey Diagram to show "flow" of money to different purposes. We used this resulting figure to help characterize the general transaction types of both C.B.P. and I.C.E., as well as priority -- in this case, we selected the top three 2-digit NAICS codes in terms of aggregate pragmatic obligation.

Note: An interactive version of the Sankey Diagram is attached to this report as an html file.

Interestingly, we found that the vast majority of NAICS-labelled expenditures made by the C.B.P. and I.C.E. fell under "56: Administrative Support and Waste Management and Remediation Services" and "54: Professional, Scientific, and Technical Services".

# Border Spending Sankey Diagram

Figure 3: Sankey Diagram depicting award funding to different NAICS (North American Industry Classification System) transaction description categories. NAICS description categories are hierarchical, and movement of funds represent the subdivision of total funding pools into increasingly fine-grained categories. To reduce clutter, data is only shown for four hierarchical levels:

- Awarding agency: Departments of Customs and Border Protection, Immigration and Customs Enforcement
- Sector: 2-digit code
- Subsector: 3-digit code
- Industry Group: 4-digit code

In addition, sector codes displayed were limited to the three categories with the largest funding totals: Professional, Scientific, and Technical Services; Administrative and Support and Waste Management and Remediation Services; and Information. Together, these three categories comprised a substantial majority of the transactions with recorded NAICS codes.
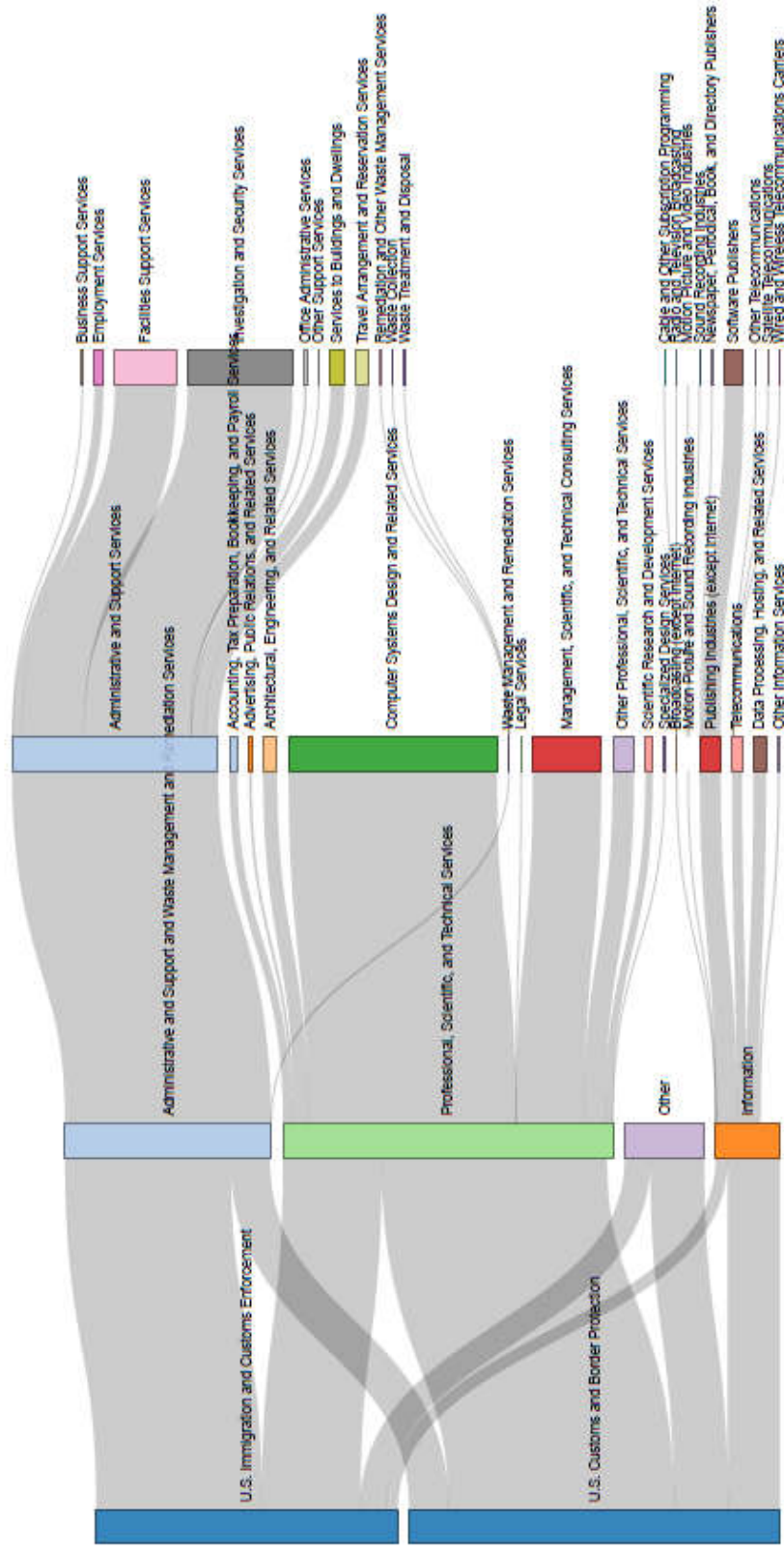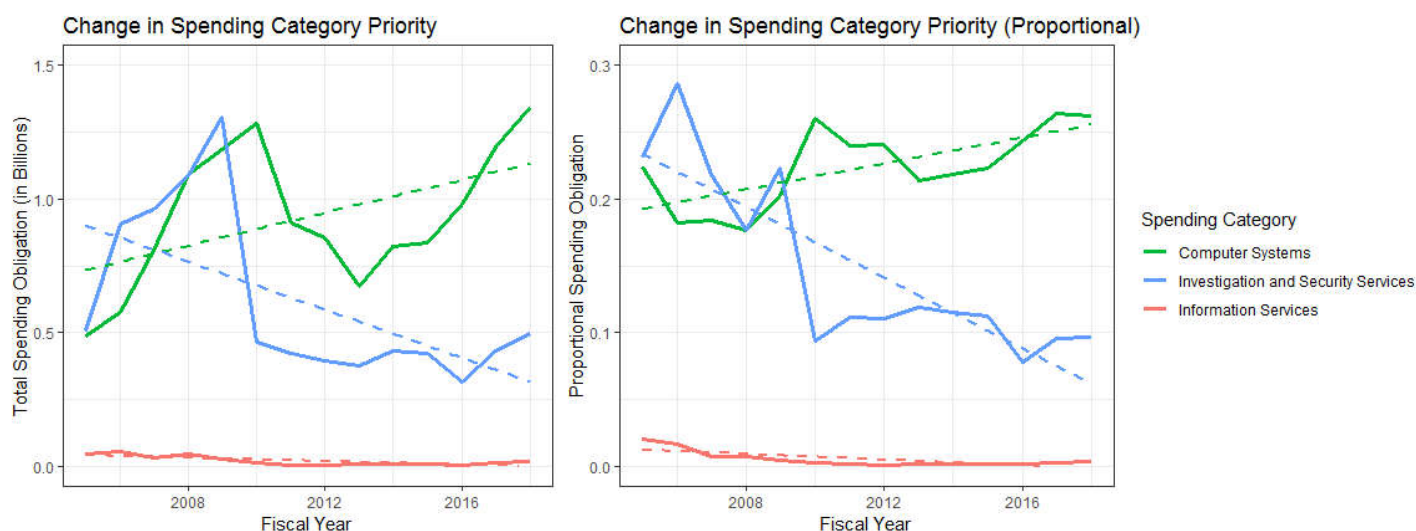
Figures 4,5: Generated pragmatic obligation over time, aggregated by fiscal year and by NAICS industry classification.

Over time, it appears that the relative priority of these 3 primary spending categories is changing. In particular, focusing on 3 key industry-level codes that appeared prominently in the above figure -- 5415: Computer Systems, 5616: Investigation and Security Services, 5191: Information Services -- suggests that both absolute and proportional spending in these three industries varies significantly over time. Investment in computer systems services appear to be increasing, while funds devoted to security services have remained flat.

This finding may reflect changing priorities in immigrations and customs enforcement, where computer-based approaches may be more effective than traditional "boots on the ground" efforts. It may also reflect general patterns of change in administrative procedure, where increasing digitization of many systems has led to increased investment in computer services. While a more fine-grained analysis may be able to tease apart small variations (e.g. border security agents vs. prison guards), these distinctions did not appear feasible with NAICS codes.

Question 3 - Where is border funding being spent? Is it concentrated along the Southwestern United States? Is more money spent on border security for the coast vs the southern and northern border?

To see where border funding is spent, we performed a group-by computation on states by summing adjusted generated pragmatic obligation in millions and plotted the results in a choropleth map.
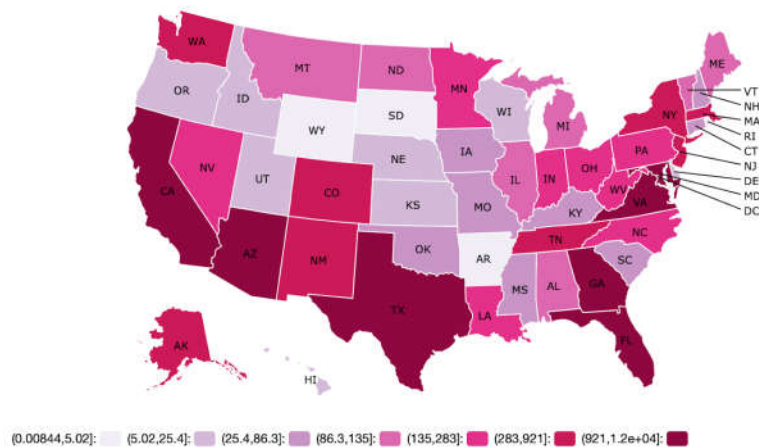
Figure 6: Choropleth map depicting variation in total pragmatic obligation, 2005 - 2019

From this plot we can see a heavy utilization of border funds along the states bordering Mexico with the notable exception of New Mexico, as well as heavy funding for coastal states with large ports. Additionally, there is a notable lack of funding for states in the northern midwest, suggesting that border security generally requires less resources on the USA-Canada border.

Grouping by fiscal quarters, we can also see which states receive more consistent streams of funding around the year. This allows us to see which states may require additional funding during certain seasons. Below is an animated map showing funding allocation in millions each quarter from 2003 to 2019.
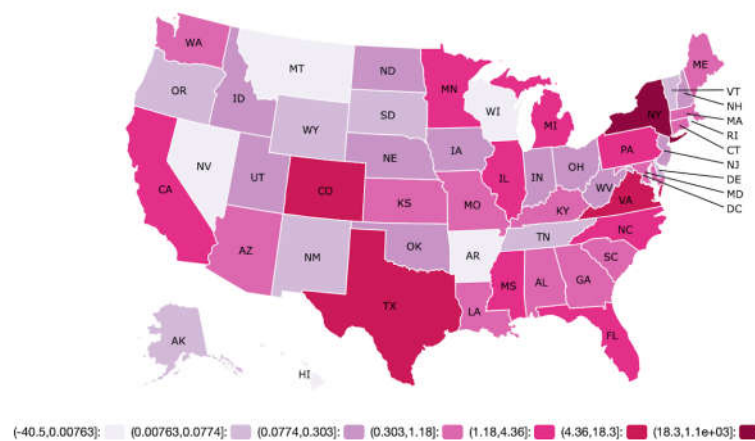
Figure 7: Choropleth map depicting variation in total pragmatic obligation, 2005 - 2019, animated by fiscal quarter.
Note: If embedded animation is not functional, a copy of this figure is attached as an html file.

Question 4 - Does border funding have a negative relationship with rates of immigration (legal and illegal)? Are these trends national or isolated to certain states?

To assess relationships between agency spending and the frequency of illegal immigration, we accessed immigration data provided by the Department of Homeland Security [4]. This data provides recorded frequencies of four key actions -- apprehension of immigrants, deportation, admission of lawful immigrants, or acceptance of refugees. Over time, this data suggests that the relative frequency of apprehensions and deportations, key actions often applied to illegal immigrants attempting to enter the country, has decreased in recent years.

Research has shown that the frequency of apprehensions may serve as an effective proxy to measure rates of illegal immigration [5]. Therefore, to estimate the relationship between spending by C.B.P., I.C.E. and rates of illegal immigration, we aggregated the total frequency of apprehensions in each year between 2005 - 2018.

The resulting figure shows a weak relationship between spending and apprehension frequency -- correlations between spending and both total immigration and immigration from Mexico were less than 0.25. It is difficult to determine a causal relationship between these two measures, as increased spending may lead to greater security presence and frequency of apprehension. However, given reduced emphasis on traditional security measures (Figures 4,5), it appears unlikely that recent spending increases may lead to greater apprehension frequencies in this fashion.
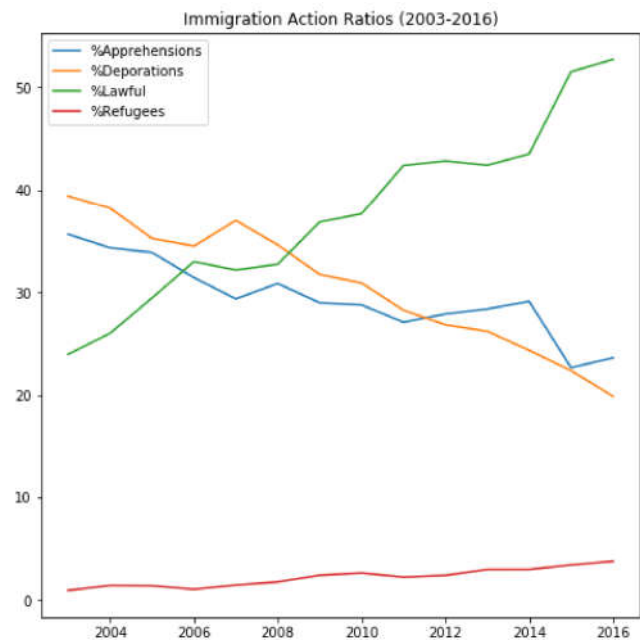

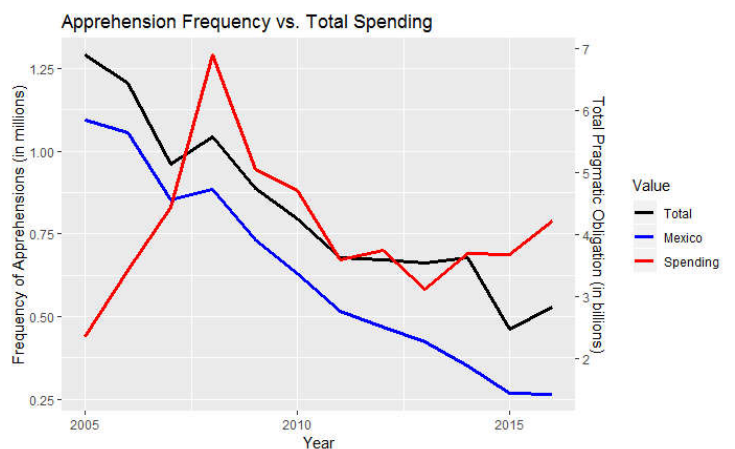**Figure 8:** Frequency of Department of Homeland Security reported immigration actions from 2003 - 2018

Figure 9: Frequency of reported immigration actions and aggregated pragmatic obligation from 2005 - 2018

Conclusions

Together, our report provides clarifying evidence on a number of key patterns in spending by the Department of Customs and Border Patrol and the Department of Immigration and Customs Enforcement. While border funding appears to be seeing a recent resurgence following a decline from 2008 - 2015, these funding increases appear to be accompanied by changing priorities in spending directives -- namely, greater investment into computing services that may reflect modernization of immigration enforcement practice. Additionally, this funding appears concentrated at hotspots of immigration activity, in states such as California, Texas, Florida, and Arizona. Despite increased funding in these states, however, we found no evidence of a recent change in the rate of illegal immigration.

Our project also highlights some of the challenges associated working with large datasets, namely the incompleteness of records that may make long-term comparisons difficult. We encourage continued development of open government databases as a public resources that may be used to encourage development of a more informed civic population.

References

[1] Gramlich, John. 2019. "How Americans see illegal immigration, the border wall and political compromise". Pew Research.
https://www.pewresearch.org/fact-tank/2019/01/16/how-americans-see-illegal-immigration-the-border-wall-and-political-compromise/

[2]
https://www.whitehouse.gov/wp-content/uploads/2018/02/FY19-Budget-Fact-Sheet_Border-Security.pdf

[3] https://www.census.gov/eos/www/naics/downloadables/downloadables.html

[4] https://www.dhs.gov/immigration-statistics/yearbook

[4] Espenshade, Thomas J. "Using INS Border Apprehension Data to Measure the Flow of Undocumented Migrants Crossing the U.S.-Mexico Frontier." *The International Migration Review*, vol. 29, no. 2, 1995, pp. 545–565. *JSTOR*, www.jstor.org/stable/2546793