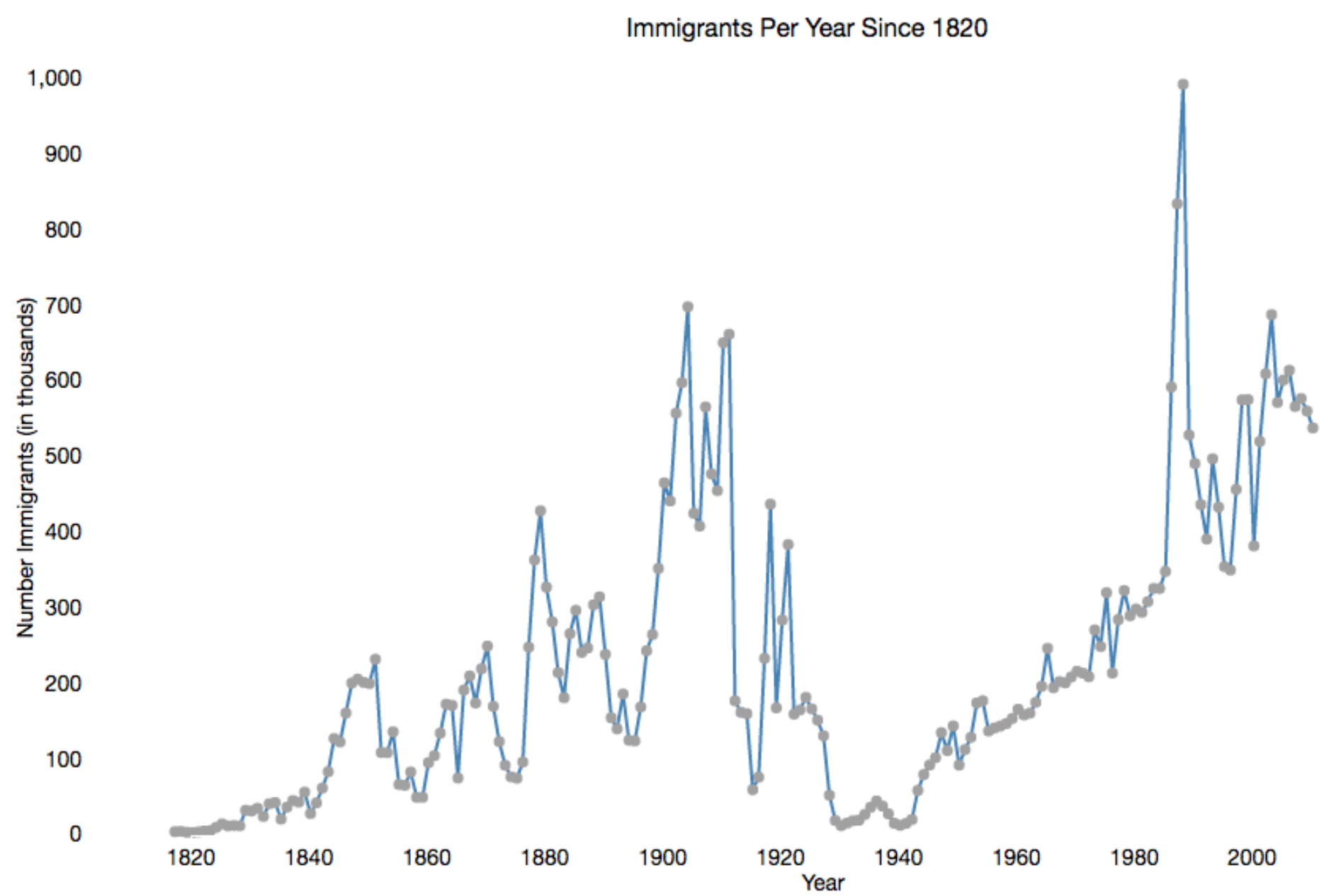


Immigration in the United States:

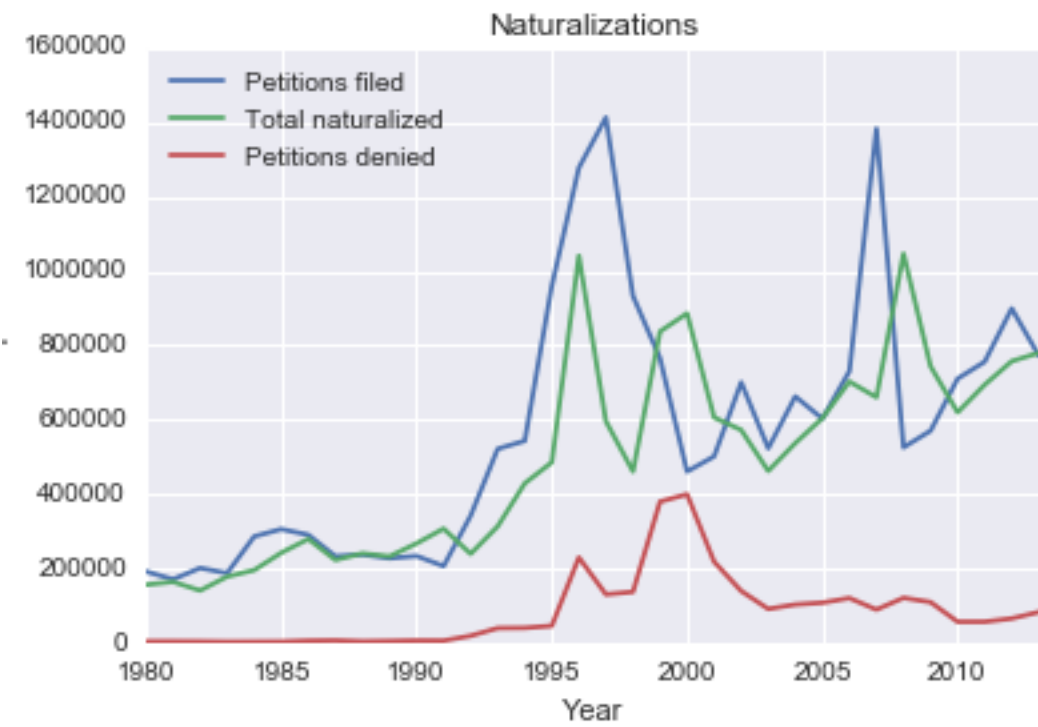
This election cycle has once again cast a limelight upon the issue of immigration in the United States, referencing both legal and undocumented immigrants, and even refugees fleeing war-torn countries. It’s amazing that we take the rhetoric that some politicians display as the norm nowadays, since 30 years ago, this entire space looked radically different. Immigration rates themselves have been quite polarized in recent years, varying by over 150 million in the past thirty years alone. We wanted to investigate exactly when and how certain parts of the United States’ attitude towards immigration became so polarized and hostile. To do this, we chose multiple routes to investigate, and performed various methods of analysis in order to determine what the most significant years were. We also attempted to recreate the story of their importance and context.



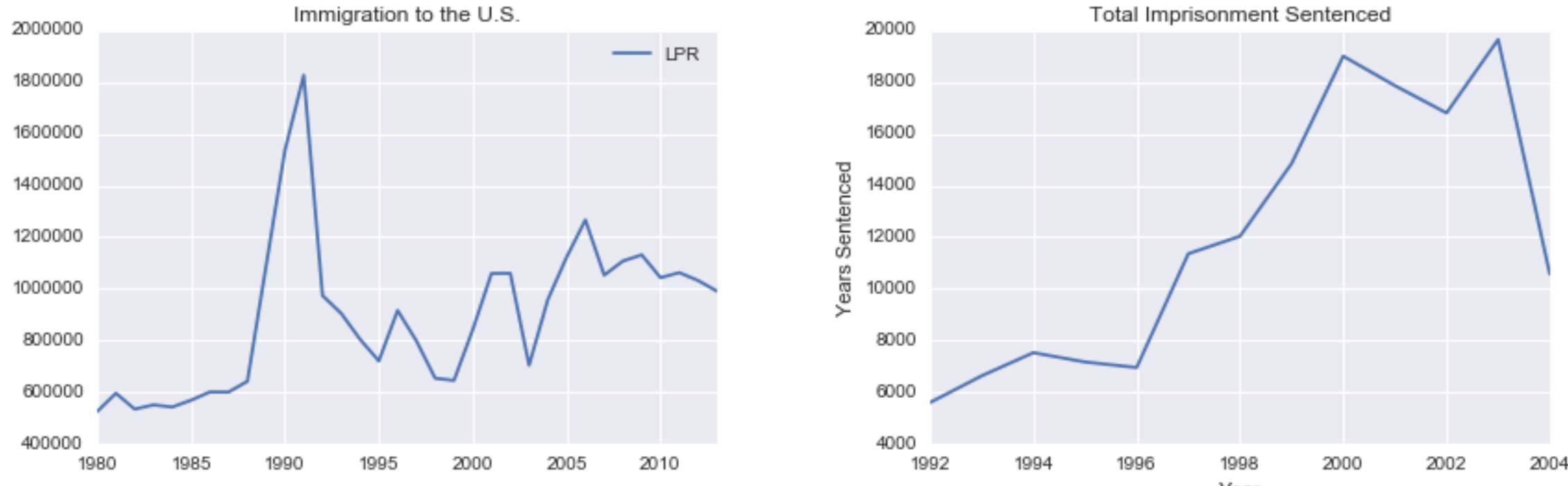
Homeland Security

By acquiring data from the Department of Homeland Security, we hoped to gain a view into how the executive side of government views immigration, and reacts to immigrants and foreigners. Among the numerous statistics we gathered were naturalization, border patrol arrests, prosecutions, deportations, imprisonment, and foreigners determined inadmissible. Focusing on the years between 1980 up until the published data (2013), we discovered two trends.

One was an increase in naturalization petitions and immigration peaking in the early 1990’s and into the late 90’s. We believe this to be a result of the fall of the Berlin Wall, which allowed many more immigrants to seek better lives in the United States.



The other was a sharp rise in actions taken by the border patrol, and resulting prosecutions and punishments, beginning in 1996. We believe that this ties into the relative conclusion of the first half of the Yugoslav war, and that it continues to rise because of the revival of conflict in the area, which the Clinton administration eventually responded to with force.S



We found a very interesting correlation between the peak years for many of these statistics and the peak years of congressional bills regarding immigration. It would not be surprising to assume that the increase in border regulation was caused by newly introduced legislation. However, it could very easily be the case that the DHS begins performing more actions, and as a result of their actions, the people and congressional representative feel that there is more of a need to protect and isolate the United States from its neighbors.

The Story

Not only did we manage to discover important trends and peaks in each of our datasets, we also found correlations between domains, which resulted in unexpected findings.

For our generation the terrorist attacks on September 11, 2001, are the defining moment of foreign danger, and as such, we often tend to associate this event with the development of the United States’ more hostile attitude towards immigrants. While it most definitely had an effect on many immigrants lives in the United States, and on many regulations involving travel, it does not appear to have had as large an effect on the

The fall of the Berlin Wall, and the Yugoslav War seem to have contributed much more to creating the polarized trends we see in the United States’ attitude towards immigrants. While our text analyses could not explicitly state many of our suspicions, we believe that much of our clustering work, combined with our research, lends credence to our theories surrounding the evolution of the United States’ attitude towards the foreign in the 1990’s.



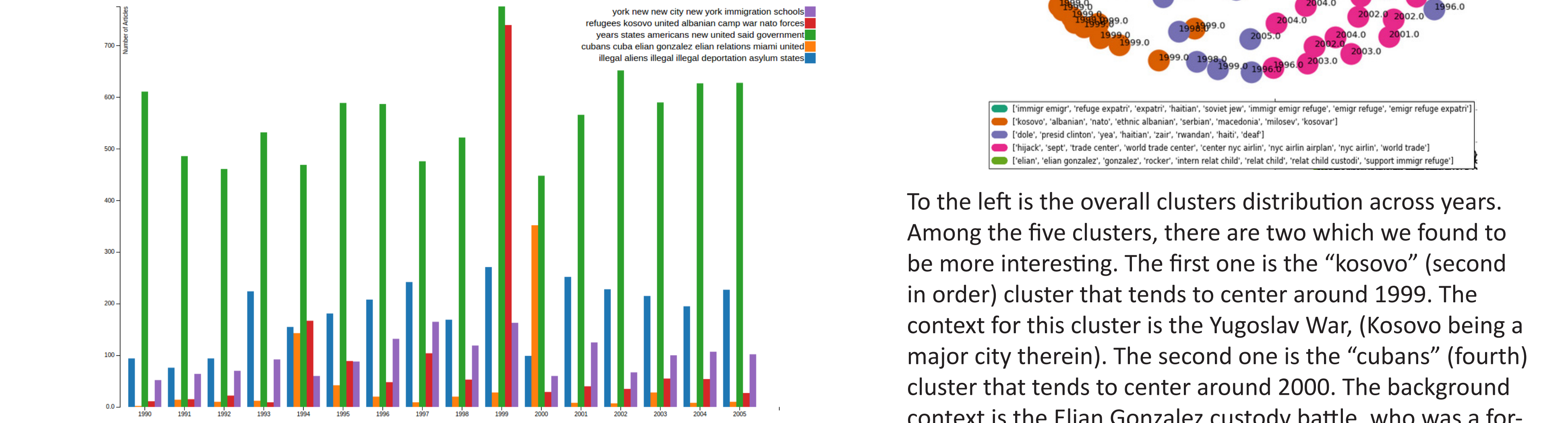
A Changing Perspective

Filippos Lymperopolous, Nur Shlapobersky, Zhecan Wang

New York Times Articles

Using on the New York Times API, we make a script to stream all the articles that mention the keyword “Immigration” from the years 1850 to 2005. We collected around 30,000 articles in total. However, in order to better compare this data set with other bills and budget data in time series, we focused on the articles in the range of 1900 to 2005 (around 15,000).

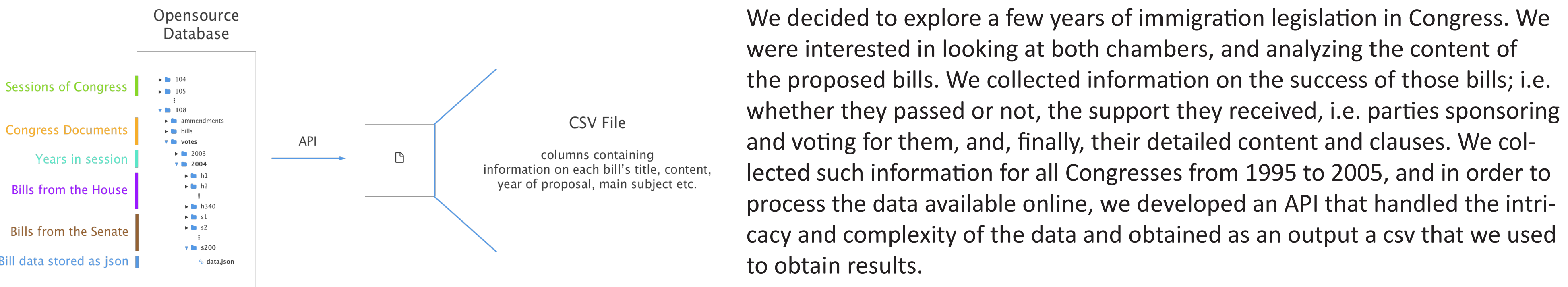
We used TF-IDF to analyze our articles and extract a feature matrix from our data set. In this matrix, each row represents either a group of articles or an article, and each column a feature word. After feeding this matrix into a K-means clustering algorithm, we end up with 5 clusters. On the right, each circle represents around 150 articles. Each circle is a feature vector that has been displayed using multi-dimensional scaling. As you may observe, the circles in a cluster tend to be closer together in spatial distance. Also, the year of circles in a cluster also tend to be closer in chronological distance. The clusters’ keywords are very distinct from each other. Overall, it is a reasonable clustering.



To the left is the overall clusters distribution across years. Among the five clusters, there are two which we found to be more interesting. The first one is the “kosovo” (second in order) cluster that tends to center around 1999. The context for this cluster is the Yugoslav War, (Kosovo being a major city therein). The second one is the “cubans” (fourth) cluster that tends to center around 2000. The background context is the Elian Gonzalez custody battle, who was a former Cuban castaway. The other clusters tend to be more general comparatively.

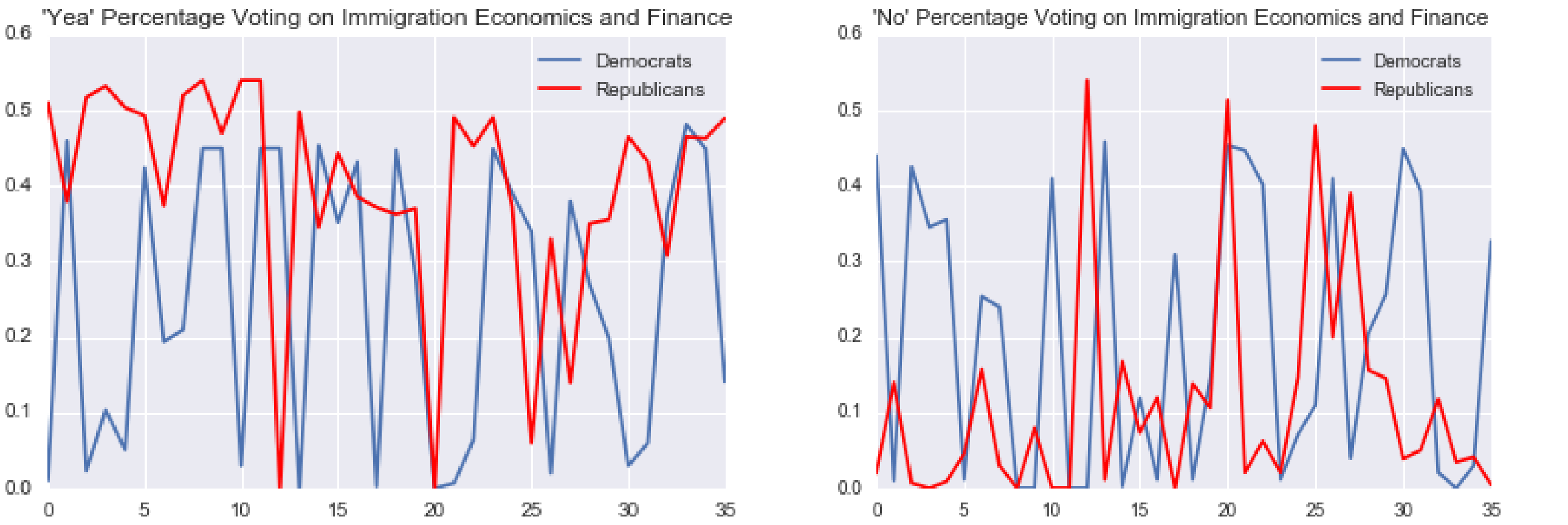
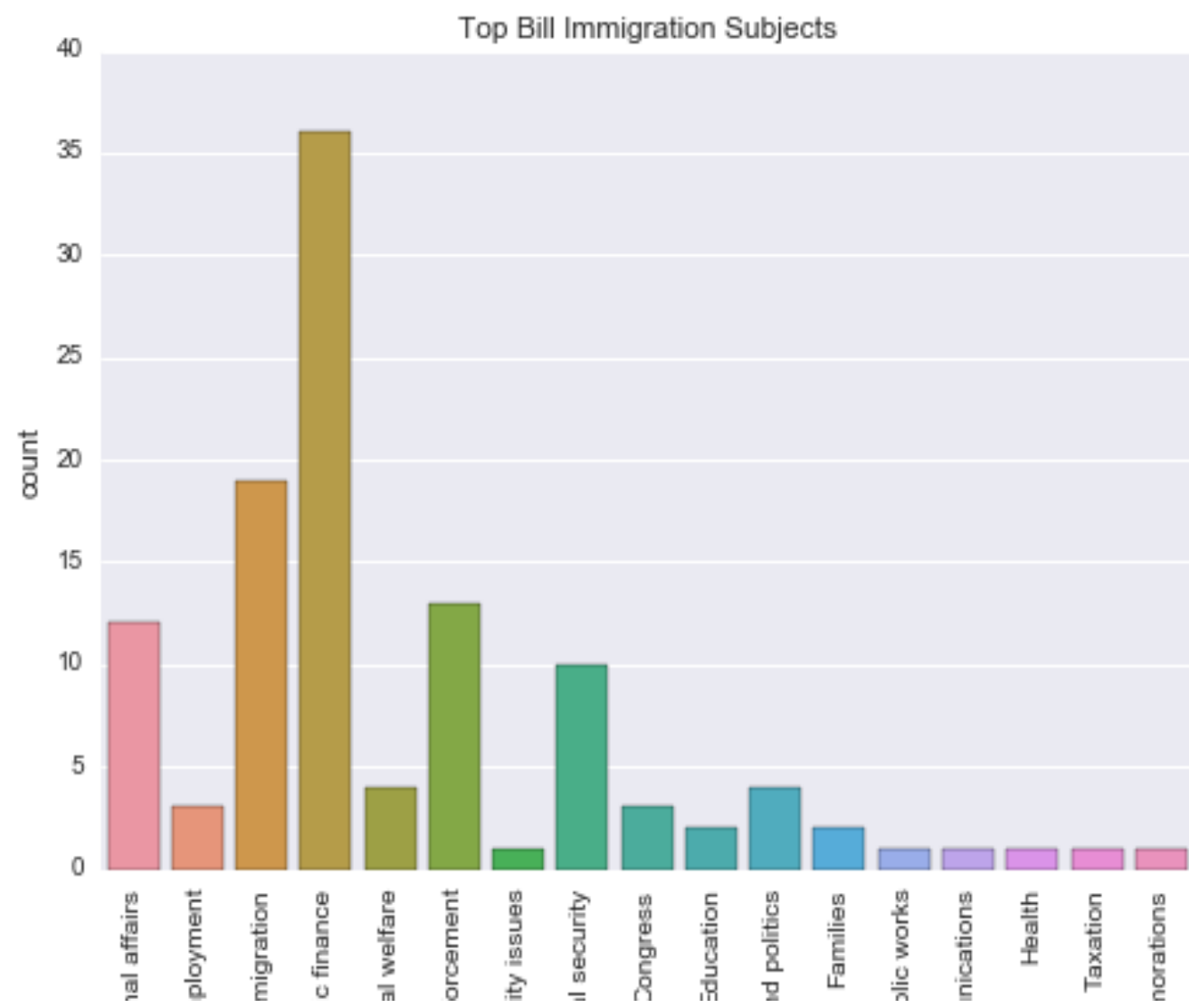
After clustering, we grab the primary words from each cluster along with their importance score (from TF-IDF) and used a word cloud to visualize them. On the left, we show the important words in the “kosovo” (second in order) cluster. The interesting words here are “kosovo”, “albanian”, “serbian” and such.

Congressional Bills



We performed two main functions on our data: clustering based on bill content and voting percentage analysis. The former allowed us to identify patterns and groupings present in the verbose dataset and point out patterns and events that significantly improve our understanding of the government’s treatment of the topic of Immigration.

We came to realize that the topics of those bills varied significantly, ranging from the topic of “Family” to “Economics and Public Policy”. As you can see below, a large percentage of the bills were involved in the later topic alluded to.



Looking into the specific vote inclinations of parties at the time 1995-2005 on the most “famous” and discussed topic of Economics and Public Finance related to Immigration, we realized that political parties to a great extend disagreed and oscillated away from each other. After all, it was continuously brought up in sessions of both the house and the senate enhancing its contradictory nature.

Clusters were very interesting, as we identified peculiar groupings around “terrorism”, “prison” and “founding” that enforced how realistic and rational our model was, indicating nice results on both of our KMeans implementations on congress and press data.