CS171 Final Project: Analyzing Congressional Travel Data from 2007-Present

Robert Doles, Steven Kekacs, Kunal Mehta

Background and Motivation:

Until 2007, privately funded travel was allowed for Members of Congress as long as it was not sponsored by a foreign government or a registered lobbyist. International travel could last no longer than seven days and domestic days could last no longer than 3 days (4 for the House). In 2007, both chambers of Congress changed their rules so that all privately funded travel must be approved by the House or Senate. Members of congress are required to disclose data of any privately funded travel (both domestically and abroad) that they undertook while in office that was sponsored by a third party. This data is publicly available for the Senate and the House. Analyzing this data can provide valuable insights into which third parties have closest relationships with certain members, where certain members travel the most, and if and when members traveled together. The project is motivated because the students working on the project are affiliated with Quorum, a student startup based out of Cambridge that provides data analytics for politics. As such, we will have access to a large database with existing data about congressmen which we will integrate with this travel data to make it searchable, filterable and cleaner.

There are many trips that do not have to be disclosed: trips paid from taxpayer allowances provided to members, paid for by the U.S. government or foreign governments, paid for elected representatives or staff for campaign purposes or paid for by personal friends. Thus our data will only provide part of the overall picture but a meaningful part nonetheless.

Project Objectives:

To create a visualization of congressional travel over the past 7 years in the form of a world map with additional graphs and tables that is interactive, searchable and filterable. We will use this visualization to gain insights into how and where congress travels. Some examples of questions we hope to be able to answer:

1. Which Congressman have traveled to Israel in the past 2 years and who sponsored them?
2. Which Congressman has Harvard University sponsored to travel in the past 5 years?
3. Where does Congressman Ben Ray Lujan travel the most?
4. When did two congressman travel to the same place at the same time last year?

Data:

The Data is available for the Senate for Members and Staffers and for the House for Members. We restrict our data to only members of congress (not their staffers). The data is available in XML format with the dates of travel, the member, the destination and the sponsoring organization for the House. There are over 11,000 such trips recorded for the House. For the Senate, the XML data does not provide the destination of the trip or the sponsoring destination. There are only 169 recorded instances of privately funded travel for

the Senate that are publicly disclosed so we will manually add in the destinations and the sponsors.

Data Processing:
The Data from the XML files requires a lot of processing and sanitization to get it to a form where we can use it. We used the Django framework for the backend to create a database that is linked to the overall Quorum database that contains a number of different models.

TravelDestination:
Object for a destination for travel that contains a number of fields. The state and country fields are important for filtering, whereas the latitude and longitude are essential for a map-based visualization.

- name in the disclosure form
- cleaned name (google geocoding api verified)
- state
- country
- latitude
- longitude

TravelSponsor: Sponsor of the travel.

Travel: Actual trip.

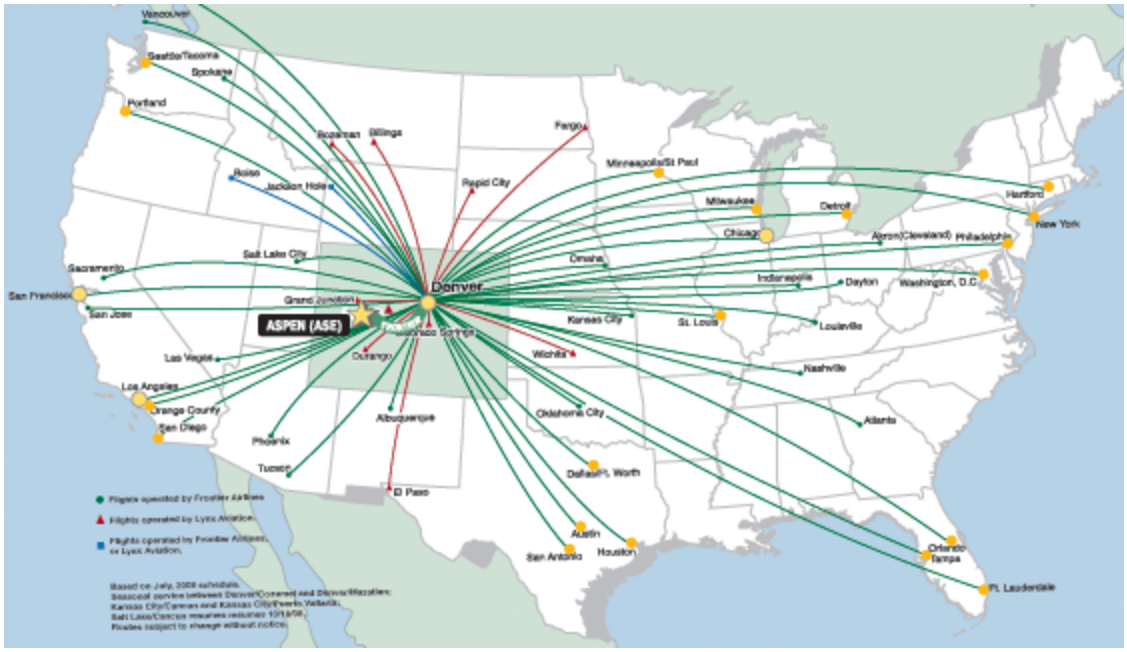
- member - contains a foreign key to the member in the Quorum database. This member object contains various biographical and legislative information about the member which
- destination - contains a foreign key to the aforementioned destination object
- sponsor - foreign key to aforementioned sponsor object
- departure date
- return date

To take the data from the XML and get it into a form accepted by our database and sanitizing the data took a number of steps.

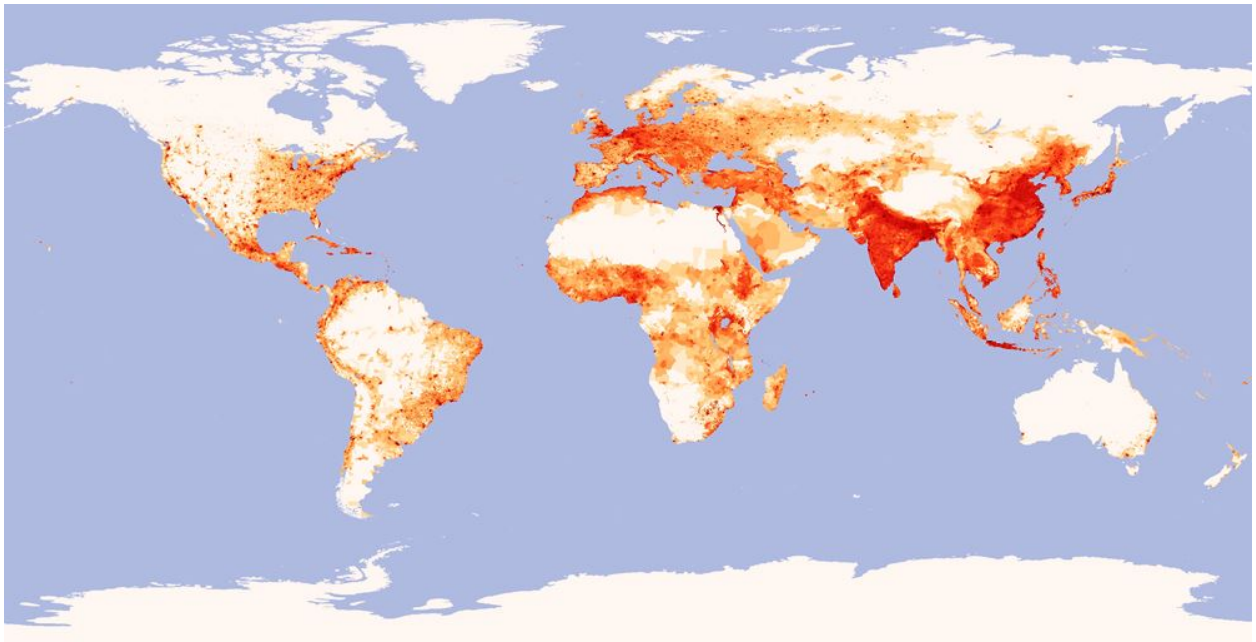
We used the Google GeoCoding API to search each location and get back a cleaned address, along with geocoded information that we used to create the TravelDestination object. We also had to manually sanitize the Travel Sponsor Data to correct for differences in spelling and grammar. For instance, Fellowship Foundation and Fellowship Foundations are clearly the same sponsor, yet our database would treat them as two different sponsors so we had to create a mapping of over 2000 distinct sponsor names to <500 unique sponsors to account for these differences. We also had to match the person's name in the XML data to the person's name in the overall Quorum database, which was very time consuming as well.

Visualization:
We hope to create a flight map of sorts to visualize data as well as a “heat” map to see what areas members of congress are traveling to. We want to make this data searchable and filterable by various parameters: member, country, state, date, sponsoring organization, and any other parameters we can think of that would be valuable.

An example of the flight map is here:



An example of the heat map (this is by population density, but a similar visualization):



Finally we attach a sketch of what we expect our final product to look like:

Must-Have Features:

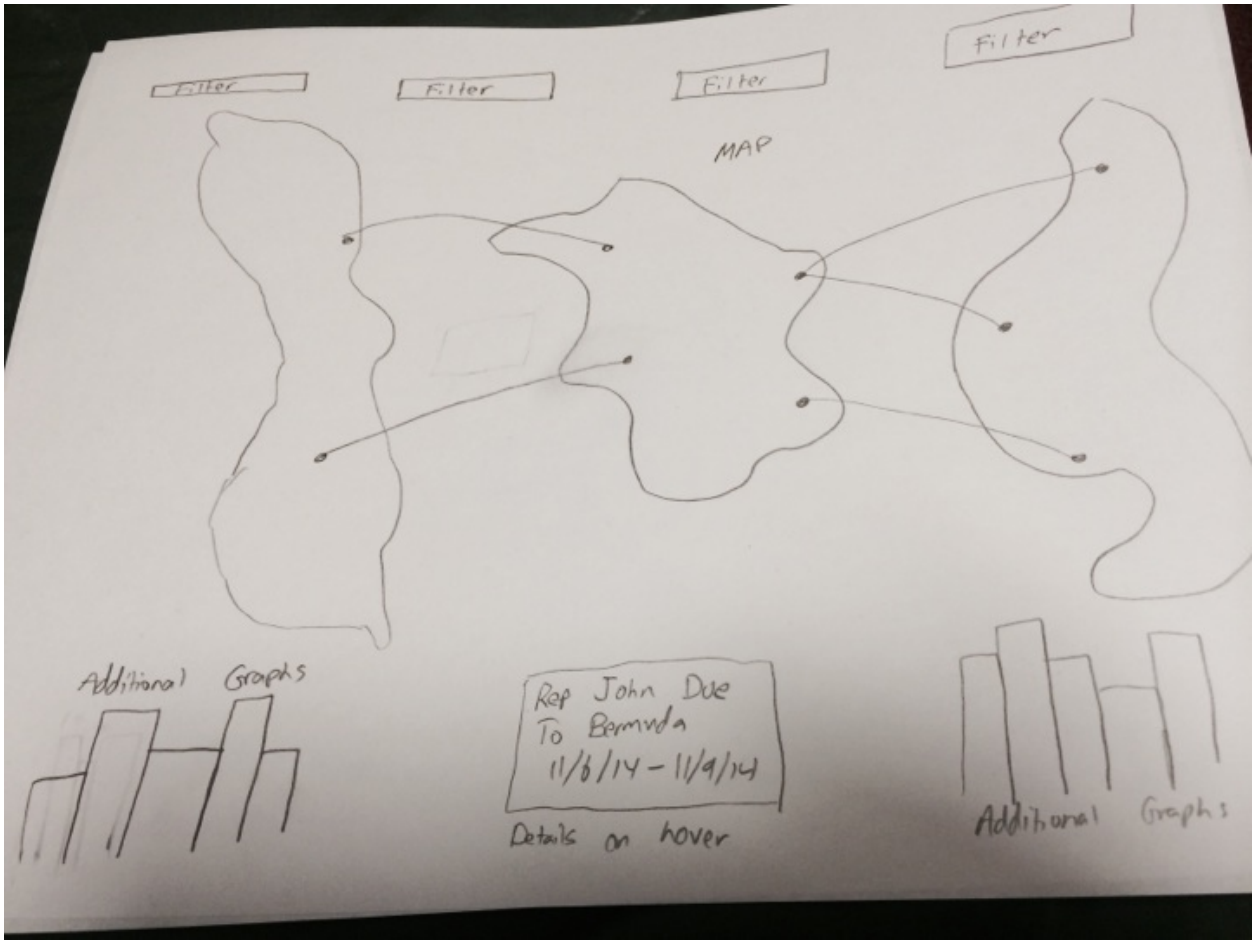
- Be able to search for all travel by an individual Congressman
- Be able to search for all travel funded by a specific source
- Map that allows users to see geographic patterns of travel
- Be able to visualize travel data over time in the form of graphs

Optional Features:

- Be able to search for all travel by groups of Congressmen (like find all travel by members of the House Committee on Appropriations)
- Make graphs interactive (potentially add brushing if relevant)

Project Schedule:

- April 3 - April 10: Acquire and clean data. If done early start researching how to implement the world map
- April 10 - April 17: Begin working on implementation of project prototype. Implement map with travel data included. Begin working on search and graph features, which should be more familiar tasks.
- April 17 (Milestone 1): Submit Milestone 1. Project to this point should have an implemented map and the foundations of the graph and search features.
- April 18 - May 1: Finish implementation. Ensure that map, search, and graphs all work well. If time permits, attempt building optional features.
- May 1 - May 5: Implementation should be complete. Take this time to finish Process Book and create screencast.
- May 5 (Due date): Hope to have project submitted


Filter
Filter
Filter
Filter
MAP
Additional Graphs
Rep John Due
To Bermuda
11/6/14 - 11/9/14
Details on hover
Additional Graphs