Process Book

Analyzing Congressional Travel Data from 2007-Present

by Steven Kekacs, Robert Doles, and Kunal Mehta

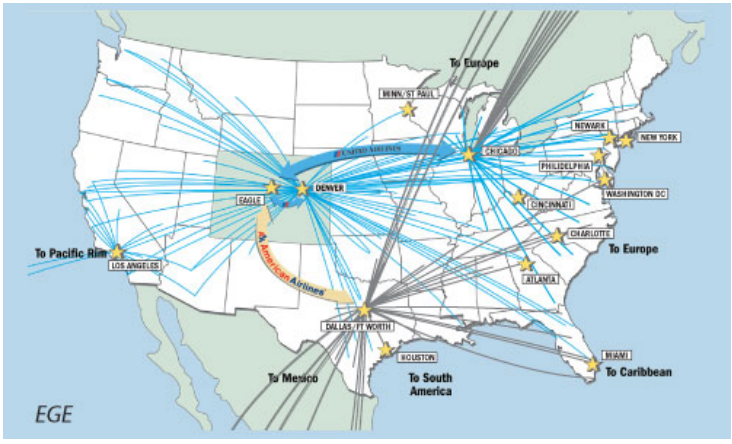
Can analyzing where congressmen travel give insights into what is important to them?

Part 1: Overview 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 House (the Senate is available as well, but the interesting data is tiny comparative to the House, so we decided to focus our analysis on 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.

Part 2: Related Work



The most important inspiration for the data analysis part of the project was the startup that the members of the project are affiliated with, Quorum Analytics. This gave us access to an enormous amount of information about each member of Congress, so we were able to match the travel to those members and then use the rest of the Quorum database to provide even more insights into the travel patterns of the members of Congress. In terms of our visualizations, we were inspired by flight maps, such as the one to the right. We were also inspired by the CS171 project from last year, A Century of Corn, in terms of the overall look and feel of the visualization. For the “dots” part of the visualization, which we will explain further in the next sections, we were inspired in part by the following CS171 visualization, as well as the Datamaps library we used liberally in our project. From a interest perspective, through our work at Quorum, the members of the team are deeply interested in finding and

visualizing information about our legislators and we found the travel data to be one that had not been looked into by Quorum previously and also one that provided very interesting data to be visualized.

Part 3: Questions

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?

The data made the answers to these questions fairly easy to ascertain. The difficulty was in creating a visualization that allowed a user not acquainted with the dataset to understand and make the same insights we were able to with a detailed analysis of the dataset. To that extent, the initial questions we had upon first deciding on our dataset did not change much. When we successfully matched the Travel Data to the Quorum Database, we were able to answer even more interesting questions, related to the ethnicity and religion of the members of congress:

1. When did Catholic Congressmen visit Rome in the past year and who sponsored them?
2. When did Jewish Congressman visit Israel?

Part 4: 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 did not bother adding in the Senate data. An example of the House XML Data:

```
<Travel>
    <DocID>500005076</DocID>
    <FilerName>Bobby Cornett</FilerName>
    <MemberName>Franks, Trent</MemberName>
    <State>AZ</State>
    <District>8</District>
    <Year>2011</Year>
    <Destination>Las Vegas, NV</Destination>
    <FilingType>Original</FilingType>
    <DepartureDate>1/7/2011</DepartureDate>
    <ReturnDate>1/9/2011</ReturnDate>
    <TravelSponsor>Consumer Electronics Association</TravelSponsor>
</Travel>
```

We used the parser attached to this file (travel_parser.py) to parse the travel data. While much of it is simple XML parsing, there are several processing functions we had to write to clean the data and get in in a source we wanted.

Overview of 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.

Part 5: Exploratory Data Analysis

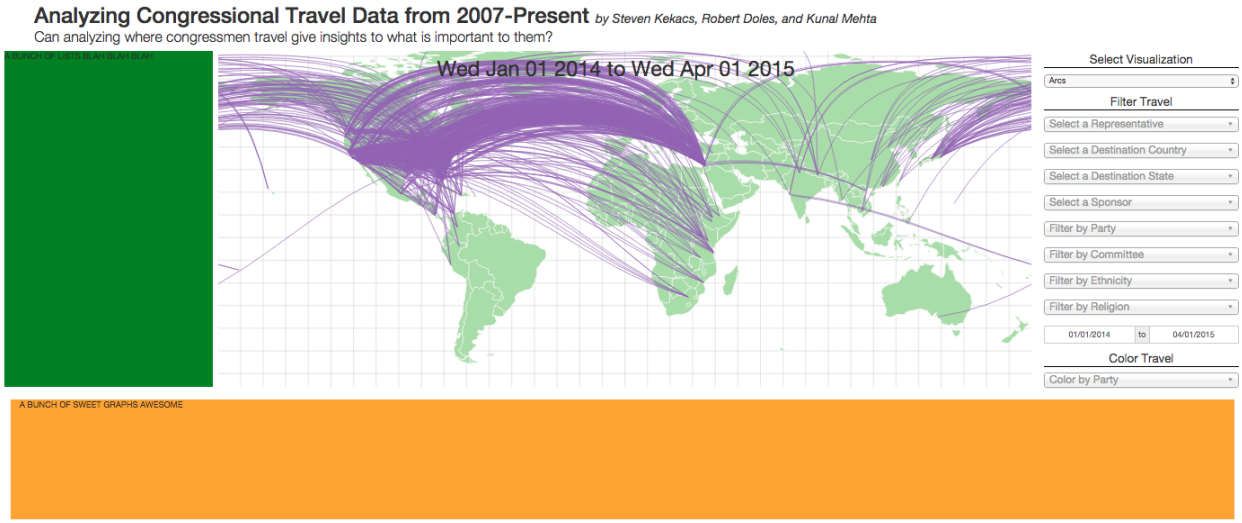
Part 6: Design Evolution

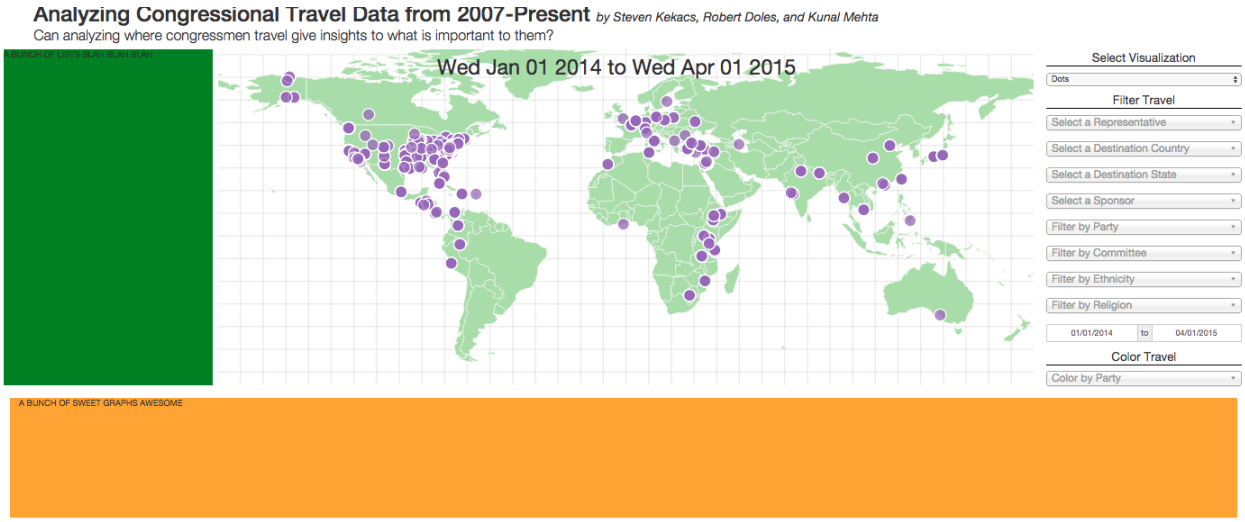
From the beginning we knew that our main visualization would be our world map, with lines that show where from and where to congressmen had traveled. We knew this wasn't enough, however, to gain useful insights into the motivation behind travel and notice interesting patterns in the data. Thus, from there we decided that a way to filter the travel based on a number of different options would be useful to be able to narrow in on the data and really analyze it at a closer level. For example if I want to see where exactly Catholic Democrats were traveling, I could filter the travel based on congressmen who are Democrats, and who are Catholic. From there we thought it would also be very helpful to be able to color the travel by party, religion, sponsor, or ethnicity. This would be useful for showing how Democrats travel differently from Republicans, among many other examples.

We also thought that it would be useful to show distributions of the number of times traveled based on party, ethnicity, religion, sponsor, and committee. For this we thought the best way was to include a set of histograms below our map and filters that would allow one to examine these distributions and notice any outliers or patterns.

Finally, we thought that a storyboard that contained preset filters and coloring to show interesting and insightful findings would be very helpful. If you scroll through this storyboard, you will see several different interesting patterns arise in our map and graphs, and we will provide some background explanation as to why you see what you see. Here is an initial sketch of what we planned to design:

Current Status





We have implemented the arcs and dots features on the map and have all of the intended features working. We have also implemented the zooming feature on the map, have the dots visualization showing details of each trip on hover. We still have to implement the ancillary lists, storyboards and graphs as well as a table that shows the raw data of each trip updating with the filters.

We believe we are well on track to accomplish our goals by the due date.