

Political Mapping of Massachusetts

CS 504 Spark Project

Yufeng Chen, Ci Chu, Yuwan Xiao, Ruihong Zhu
Boston University Department of Computer Science

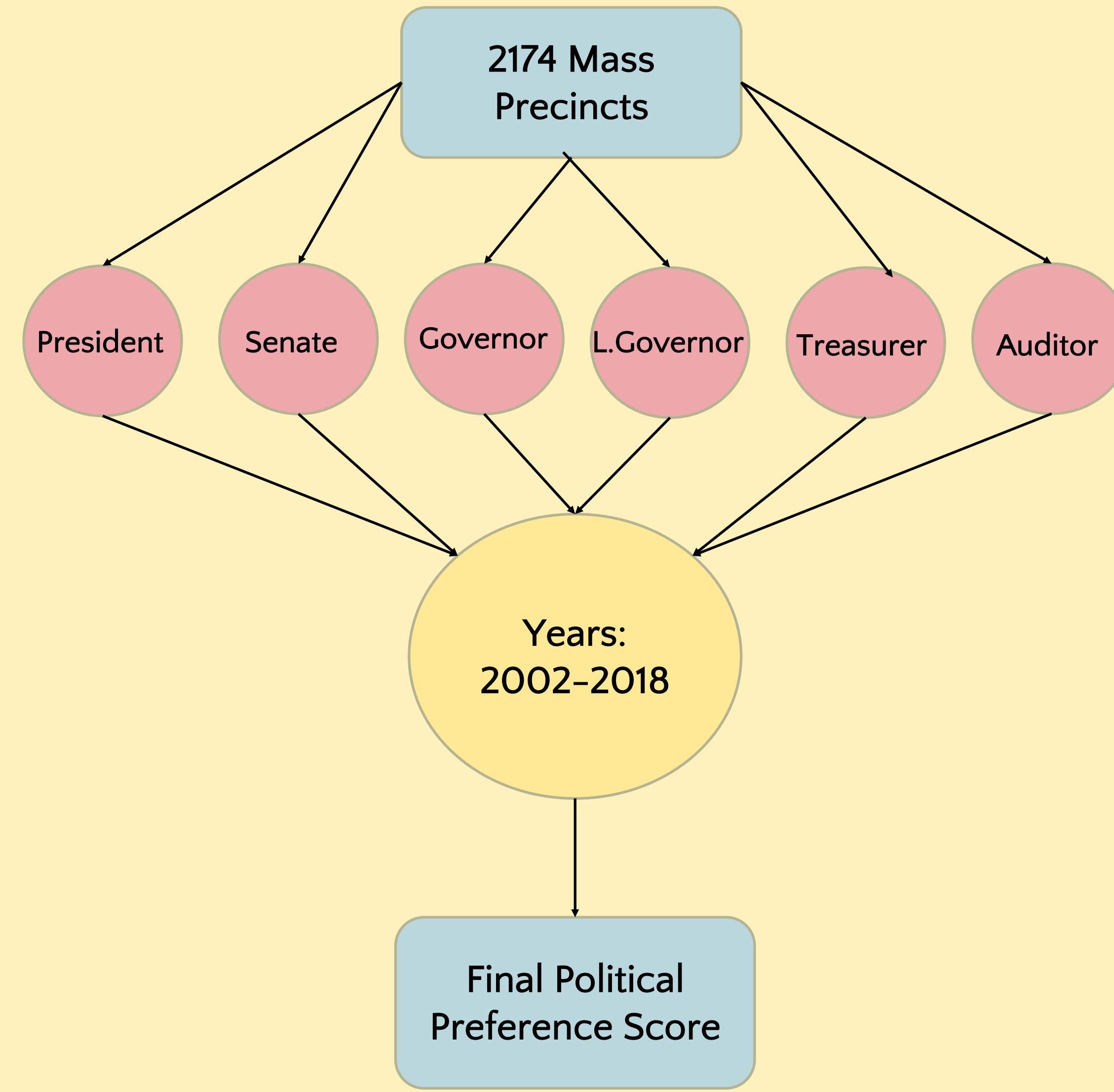
Abstract

Mass Alliance is a coalition for political and advocacy organizations that work together to build a politically aware Massachusetts. Our mission is to collect and organize voter data on Massachusetts House and Senate Districts, score districts' ideologically, and map districts to creatively present data for this organization.

This project has three steps:

- ① Scrap all candidates of President, Governor, Lieutenant Governor, Senate, Treasurer, and Auditor elections from 2002 – 2018 and score them on the level of progressiveness.
- ② Calculate the score of each Mass precincts based on their voting cast to each candidate.
- ③ Visualize the results and generate political preference map based on House and Senate voting districts.
- ④ Further improve the result by using other ways to find regions with a high level of progressiveness.

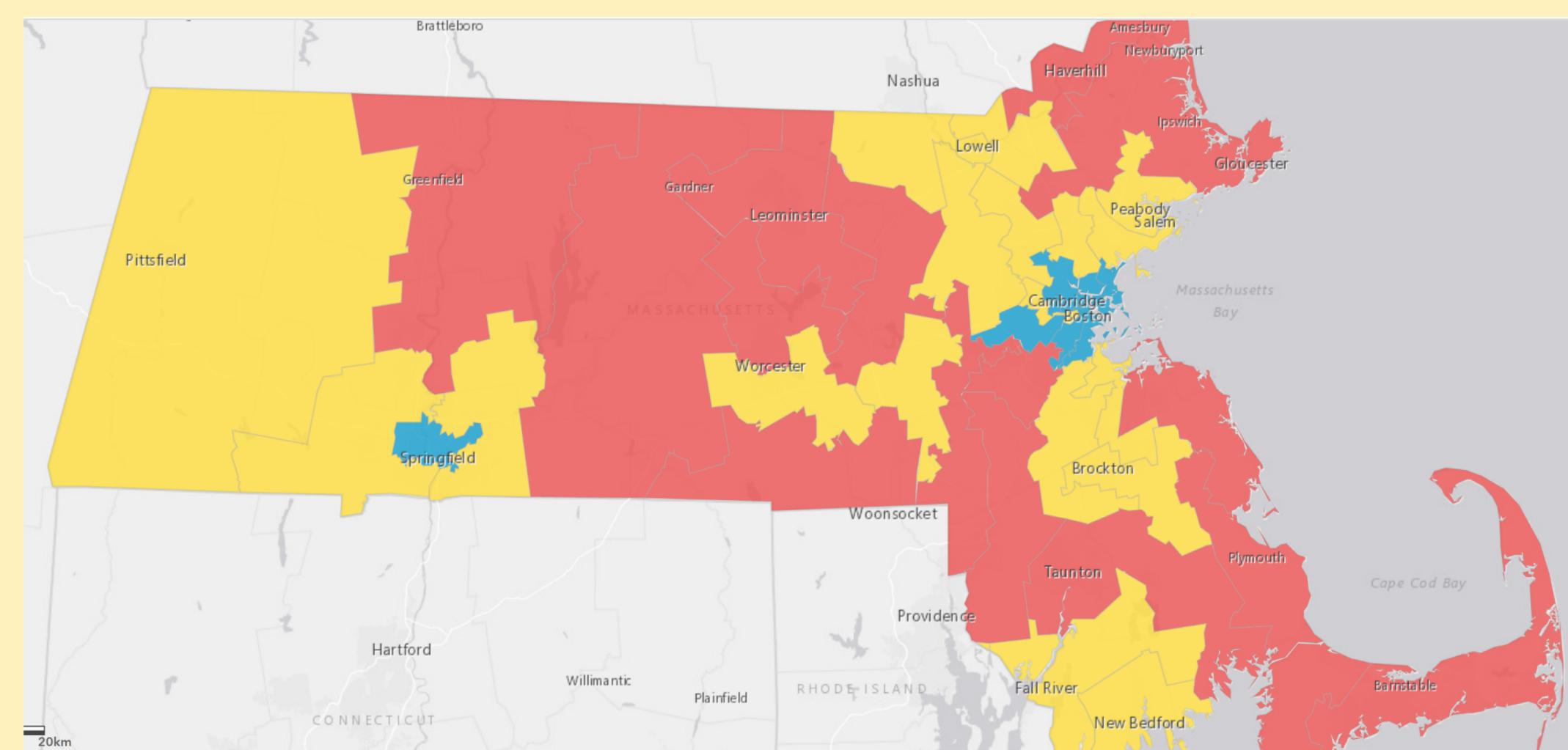
Process



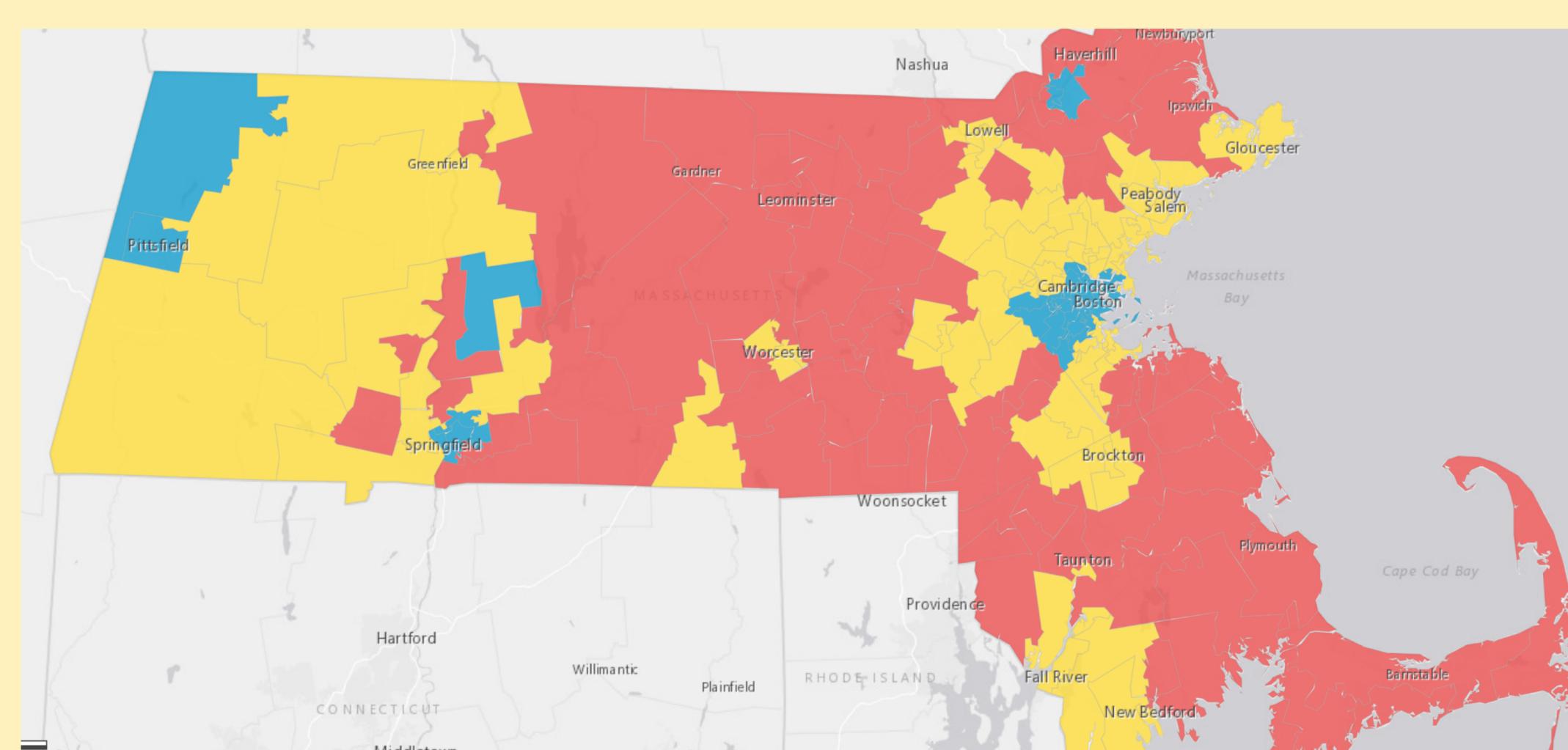
Political Map

We used the voting data from 2002–2018 of each Massachusetts precincts to calculate the level of progressiveness. Our Spark Partner Mass Alliance gave us the progressiveness score for every candidate and we used them to calculate the score of each precinct based on the weighted election results of the following six positions: President (25%), Governor (25%), Lieutenant Governor (10%), Senate (20%), Treasurer (10%), and Auditor (10%). The score of each election is also weighted on the total voting casts. Finally, we calculate the final score for each precinct by combining scores of each position from each year and used these scores to build our map. The House and Senate voting districts are made up by using a different combination of precincts, so we can easily visualize these two maps by using our precincts' scores. The following two maps show the progressiveness of each Senate and House voting districts.

Senate Map



House Map



Score: 1 – 5
Higher Score means higher level of progressiveness.
Blue: > 3.4
Yellow: > 3.0 and <= 3.4
Red: <= 3.0

Data process

① Retrieve data

All the election results are available online at PD43 website, to get these result, we use Python's request, beautiful soup, and pandas libraries to develop a program to automatically convert all the information from the PD43 into .csv file for further use.

② Calculate score

To calculate the politic score for each town and city, we load all the .csv file we have obtained into pandas' data frame and calculate the score based on the formula:
$$\text{sum(candidate's vote number * candidate's politic score)}/\text{sum(candidate's vote number)}$$
, which is in fact a weighted average.

Find Progressive Regions by using Uber

Based on the two maps we created, we observed that city districts where a large amount of population lives usually have the highest level of progressiveness.

Therefore, we decide to use another method to find districts with potentials to have a high progressiveness score. From the internet, we found the dataset of Boston's uber travel record, including each travel's starting place, destination, mean travel time, etc. We want to know the main area that people most frequently go to based on these data, and these regions are what we are looking for.

Process:

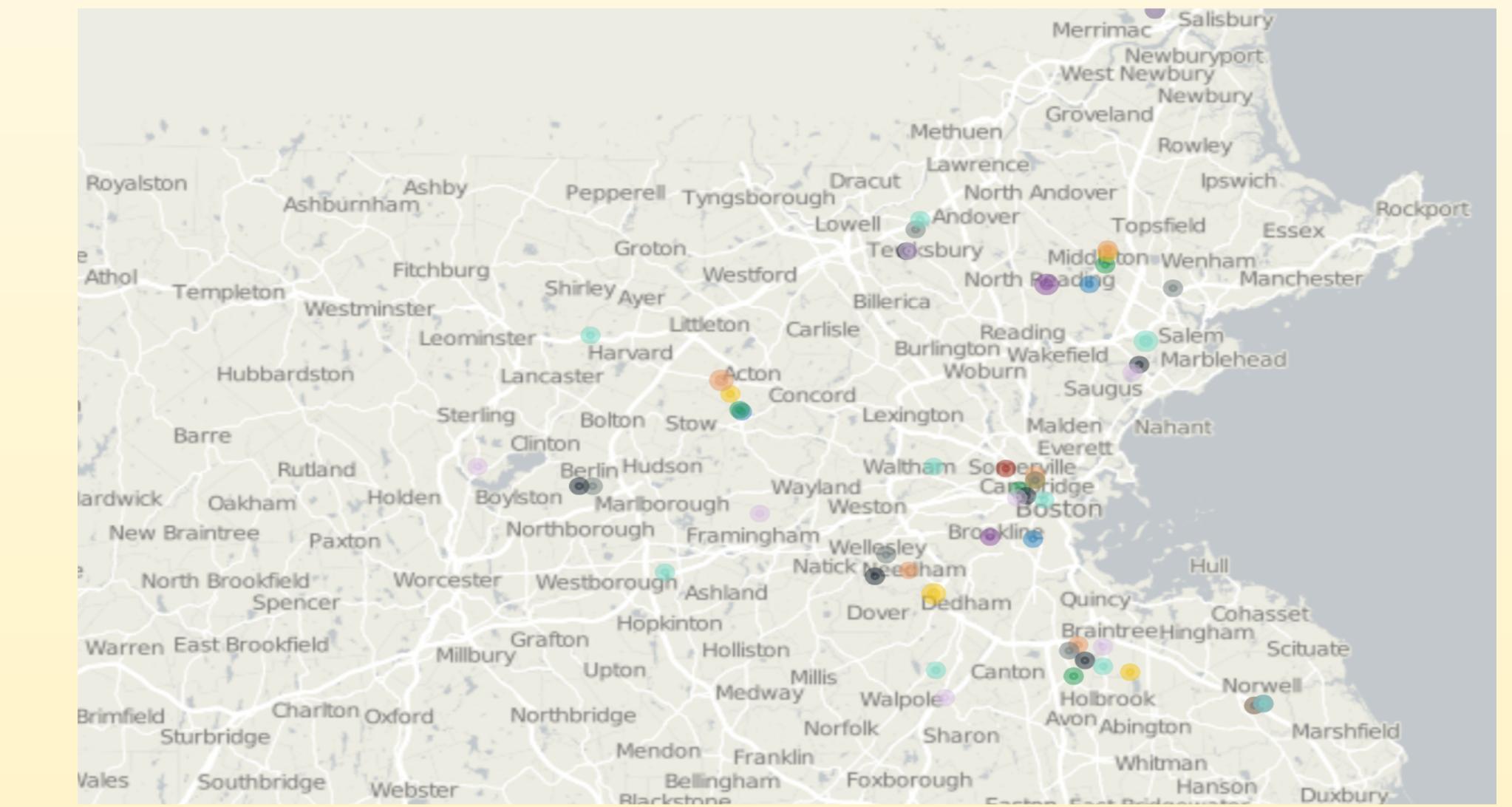
1. data preprocessing :
Since the dataset already contain the destination's name, we can get the longitude and latitude by requesting Google Map API.

2. running the k-mean

With the geolocation, we can run K-mean algorithm to divide the destination into k related areas.

Result:

We assign a unique color to each cluster, so we can see clearly the main k region center in the map and get a sense of where is the popular place that people go frequently. We also visualize the result in a web-based map so that we can conveniently check the location of each cluster and get an instinctive sense of the data.



By comparing this uber map to our political preference map, we can clearly see a positive correlation popular regions and progressive regions, which provides us another way to find progressive districts.

Score Distribution

We used ECharts to create the score distribution of each of the six positions. Each node represents a precinct. We can see the clear difference between the level of progressiveness between city and rural areas.

Below is the score distribution of Governor election..



Acknowledgements

- ① PD43 <http://electionstats.state.ma.us/>
- ② Arcgis <https://www.arcgis.com>
- ③ Leafletjs <https://leafletjs.com>
- ④ Echarts <https://echarts.baidu.com>