

A House Divided...Can It Stand?

Process Book

Michael Young, Tom Pace, Fangfei Lan

Table of Contents

| | |
|----------------------------------|-----------|
| CS 6630 PROJECT PROPOSAL | 3 |
| Overview and Motivation | 14 |
| Related Work | 14 |
| Questions | 14 |
| Data | 14 |
| Exploratory Data Analysis | 14 |
| Design Evolution | 14 |
| Implementation | 14 |
| Evaluation | 14 |
| Meetings | 15 |
| Peer Feedback | 15 |

CS 6630 PROJECT PROPOSAL

Fall 2019

Basic Info. *The project title, your names, e-mail addresses, UIDs, a link to the project repository.*

Project Title: A house divided, can it stand?

Team Members

- Thomas Pace, u0047764@utah.edu, u0047764
- Michael Young, u0742759@utah.edu , u0742759
- Fangfei Lan, u1266471@utah.edu, u1266471

Project Repository:

<https://github.com/mkcyoung/dataviscourse-pr-egap>

Background and Motivation. *Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.*

Democracy is built on fair and equitable representation. Gerrymandering is as old as representative democracy, however, when left unchecked it represents an existential threat to the health of democratic systems. We were motivated by an interest to explore the effects of Gerrymandering on the US legislative process. Specifically, we wanted to create a visualization that would explore the relationship between the efficiency gap--a common metric used to characterize Gerrymandering--and legislative effectiveness. We were also motivated to explore the potential relationship between the efficiency gap and demography. Many visualizations exist that explore the efficiency gap, however, we found none that explore the relationship between Gerrymandering and the legislative process.

Project Objectives. *Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.*

Our objectives include:

- Create an engaging interactive visualization that explores the efficiency gap and its relationship to legislative effectiveness and demography across the United States.
- Create a novel visualization exploring a data relationship that has not been investigated and it of critical importance to the US democratic system.
- Incorporate at least three primary views that explore the data.
- Incorporate scrollytelling or other narrative features.

Data. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

All of the data we'll need is readily available to the public for free and online at various sources. The specific data we'll need for this visualization consists of three parts:

- Legislative Effectiveness (LE): <https://thelawmakers.org/data-download>
- Efficiency Gap (EG): <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IG0UN2>
- US state district Maps: <http://cdmaps.polisci.ucla.edu/>

Data Processing. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

We don't expect that the data-sets linked above will require anything approaching nightmarish levels of clean-up — in fact, they appear to be complete and organized very nicely. However, there will be some pre-processing involved to obtain the desired quantities from the data.

For the Legislative Effectiveness portion of our visualization, we're after a "legislative effectiveness score." Conveniently, the LE dataset linked above contains this legislative effectiveness score in rows corresponding to the associated district # for the years 1973-2017.

For the Efficiency Gap portion, we'll of course need the efficiency gap measure for each district. Unfortunately, we couldn't find a data set with this measure calculated for us as in the case of LE, but luckily, EG is relatively simple to calculate if you know how many votes were cast for each party in any given election (More info on EG calculation:

https://www.brennancenter.org/sites/default/files/legal-work/How_the_Efficiency_Gap_Standard_Works.pdf). This information is contained in the data set linked above.

For the US district map, we'll need the GeoJSON or TopoJSON files for the US districts. The dataset linked above contains ERSI shapefile's for every district since the beginning of the United States in the year of our blessed lord 1789. We'll only be using the data from 1973 onward. Here's a conversion tool that purportedly converts ERSI data to geoJSON data:

<https://www.statsilk.com/maps/convert-esri-shapefile-map-geojson-format>. We're keeping our fingers crossed that this'll work, but if not, there are several visualizations of US districts online so finding a back-up plan shouldn't be too difficult.

We'll implement all of this pre-processing using javascript and d3. Our general approach will be to create something like a JSON object that has all of the relevant data we'll need attached to each district.

Visualization Design. How will you display your data? Provide some general ideas that you have for the visualization design. Develop three alternative prototype designs for your visualization. Create one final design that incorporates the best of your three designs.

There are two main components in our data, efficiency gap and legislative effectiveness. We would like to present each component individually, and then analyse the potential relationship between the two. Since both are geographic data, we plan to display the data on a US map with state and district details on demand. We will show the analysis of the data in two linked views, a scatterplot and a time series line plot.

Prototype 1:

In this design, we first have a story-line section to lead users through a series of definitions and examples. Then we draw the map. The states can be color-coded by efficiency gap, legislative effectiveness, or political party. When the user clicks on a state, a smaller window pops up, showing the details of the state. Below the map, we have two charts, a scatterplot and a line chart. The scatterplot shows the relationship between efficiency gap and legislative effectiveness for all the states. For the line chart, the x-axis is time and the efficiency gap is the y-axis. By brushing on points on the scatterplot, the line chart updates by highlighting lines corresponding to the brushed points. The time bar in the middle of the page controls the time point for the entire page in order to make sure that all views are in sync.

Prototype 2:

The introduction page gives the user the choice to view the visualization with or without a storyline. The map in this design is slightly different from the one from Prototype 1. This map incorporates a bar on the left for the overall legislative effectiveness across country at a given time point. This feature enables the simultaneous display of both the efficiency gap and legislative effectiveness. We have a detailed state map in the middle section with some descriptions on the right. The chart is similar to the scatterplot shown in prototype 1 but this one connects the points with lines, and as the user moves the time bar, data for the corresponding year will be highlighted. Lastly, we have two maps displayed side by side, comparing the levels of efficiency gap and legislative effectiveness at a certain time. This design focuses on the simultaneous display of the two components of the data.

Prototype 3:

The map and storyline sections of this design have mostly the same features as prototype 1. In order to introduce the concept of efficiency gap in more details, we add a section dedicated for efficiency gap. The chart demonstrates how efficiency gap indicates the level of gerrymandering and what it means to have zero efficiency gap. Then we have two time series chart showing changes of time for both variables. The buttons on the right corner allows the user to plot state or country data. The scatterplot is also similar to the one discussed in prototype 1. One additional feature is that the size of the circles is governed by some demographic data chosen by the user.

Final Design:

We decided on a design that is similar to the layout of prototype 1. Some of the views are combined to avoid visual clutter. We also added more interactive components for a more engaging visualization.

The final design includes three sections: storytelling, map view and chart view. We position the storyline on the top of the page, so users can explore the variables functions before diving into the visualization. For the map view, the user can choose from color coding by efficiency gap or by legislative effectiveness. instead of showing extra windows for state details, we plan to zoom in to the state and display details on the side. The two charts below are the same as discussed in prototype 1. The only difference is that for the time series chart on the right, the user is able to choose from efficiency gap and legislative effectiveness as the y-axis on the right corner of the chart. The map view and the charts are linked. The time bar in the middle of the page ensures that both views are in sync. We want the user to always be able to move the time bar without scrolling the page at all times, so we plan to fix the bar on the top of the page when scrolling down. Our goal is to have a concise, clean, engaging and informative visualization and we think this design will achieve this goal.

Must-Have Features. List the features without which you would consider your project to be a failure.

- Map view
 - Change district lines over time
 - Color code the map based on efficiency gap or legislative effectiveness
 - Toggle between the two color encodings
 - Zoom into a state on click, and add state detail descriptions on the side
 - Tooltip displaying state details
- Chart view - scatterplot
 - A dropdown menu to select the states to be displayed
 - Tooltip showing the state or district details on hover
 - A year indicator showing the current time point
 - Brushing certain points in the scatter plot highlights the corresponding lines in the line chart
- Chart view - line chart

- The y-axis can toggle between efficiency gap and legislative effectiveness
- A vertical line indicating the current time point
- Linked to the scatterplot and display only data filtered by the selection in the scatterplot
- Time bar
 - Move the time bar and all views update to the corresponding time point
 - Fixed at the top of the page when the map is scrolled off the screen

Optional Features. List the features which you consider to be nice to have, but not critical.

- Story-telling
 - Add terminologies and samples stories
 - Move the visualization to the correct position and display corresponding components of the story
- Map
 - Select multiple countries in the map and update the plots with the countries selected
- Chart view - scatterplot
 - Add a third dimension for demographic information
- Chart view - line chart
 - Drag the vertical line to change the time point on all views

Project Schedule. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Nov 1st:

- Organize the data in the appropriate format
- Load it correctly in javascript for future access
- Assign responsibilities for each team member

Nov 8 (project milestone):

- Complete the skeleton of each view

Nov 15:

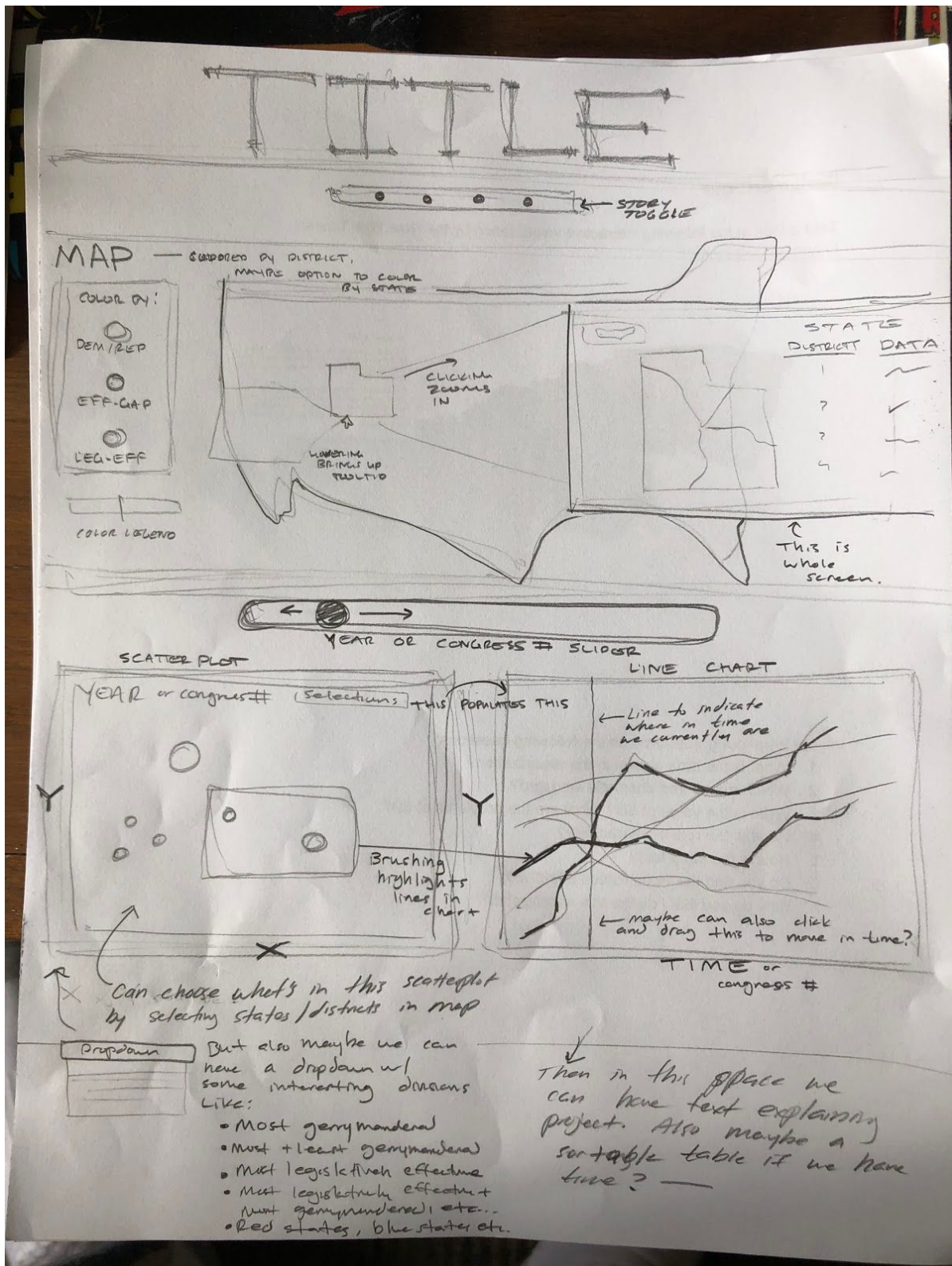
- Mostly finish implementing all of our views individually
- Set up the time bar, to vary the time for all the views

Nov 22:

- Complete all must-have features
- Make sure all views interact smoothly
- Implement some optional features if time permits

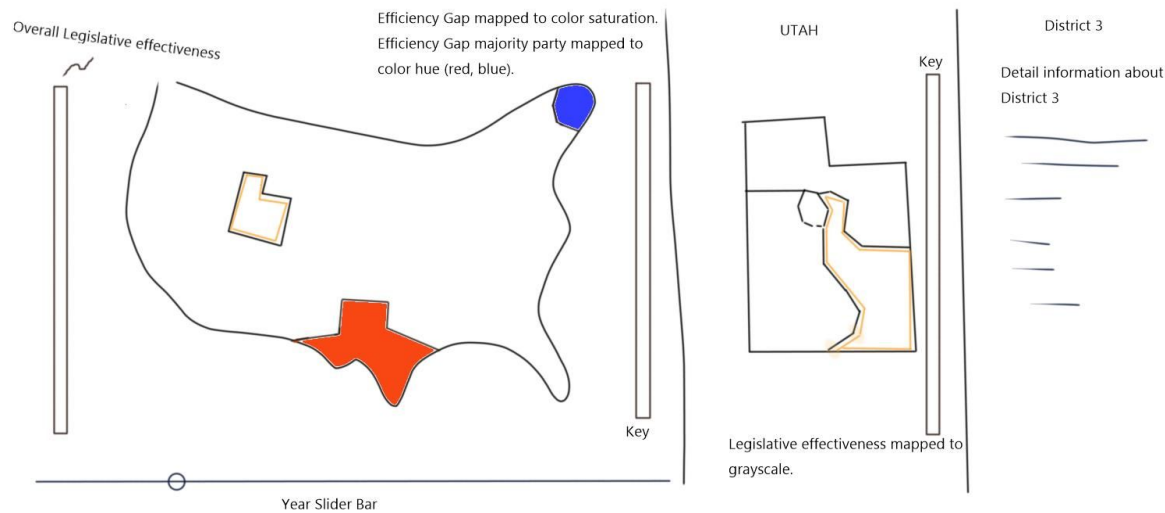
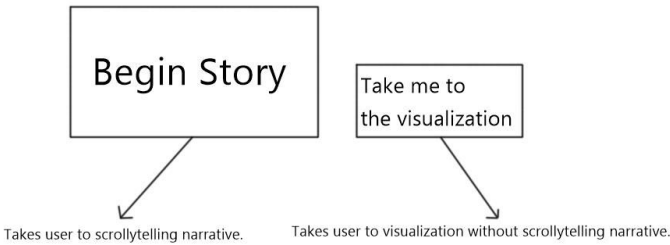
Nov 27: project due.

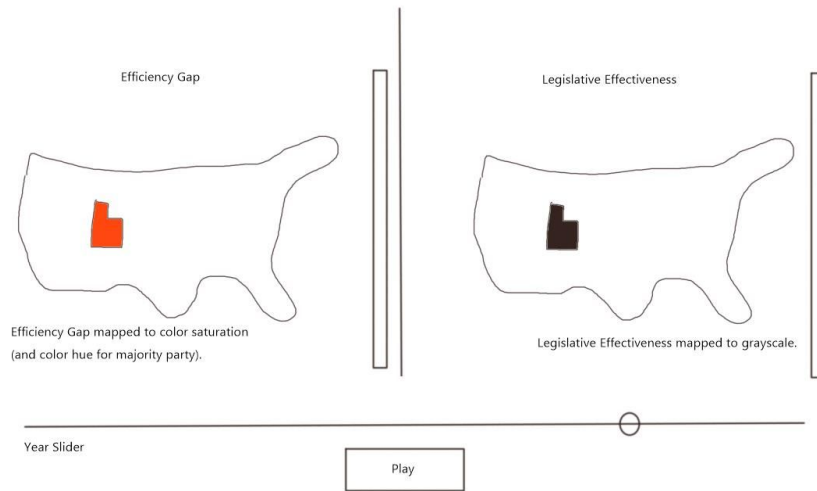
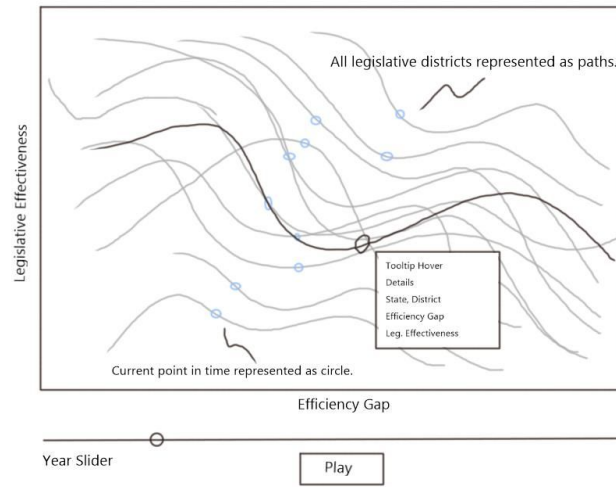
Prototype 1:



Prototype 2:

Introduction Page

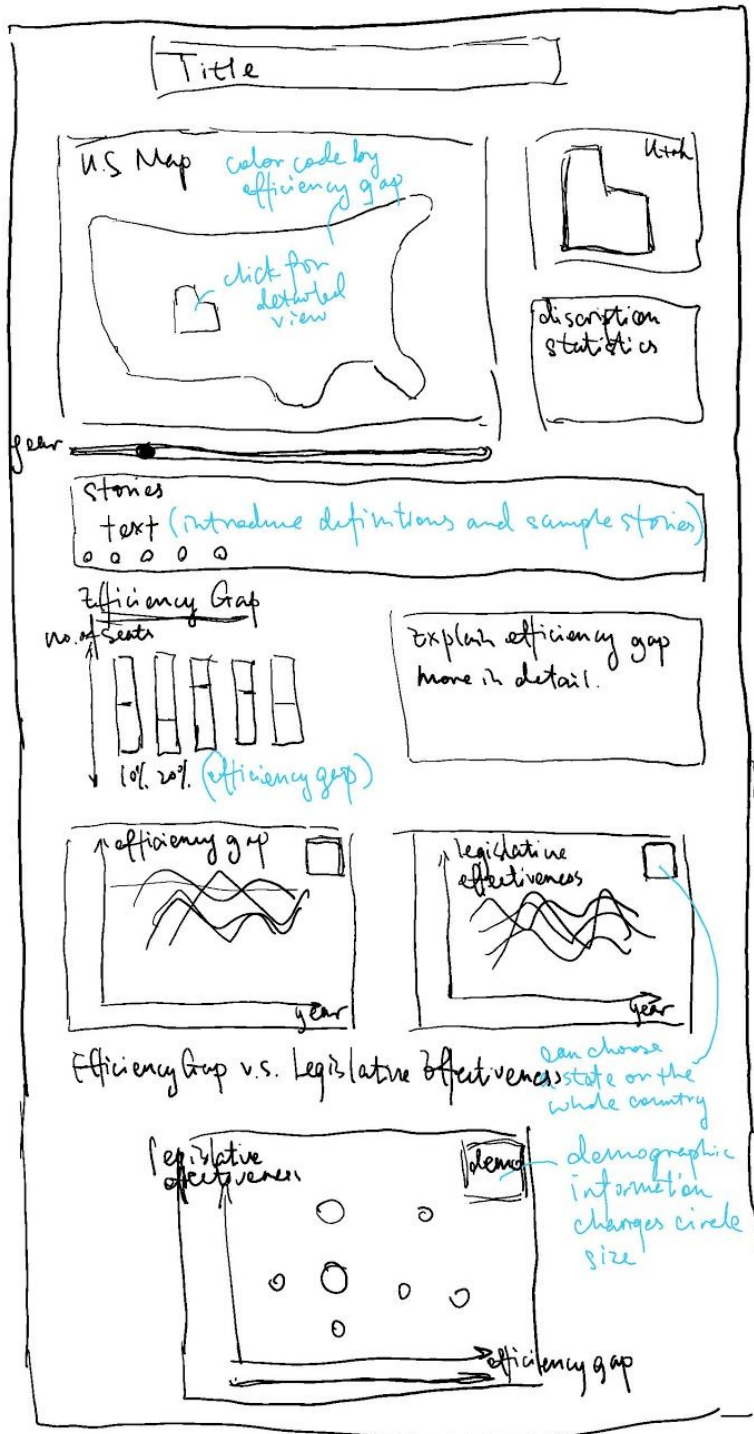




Prototype 3:

Project Sketch

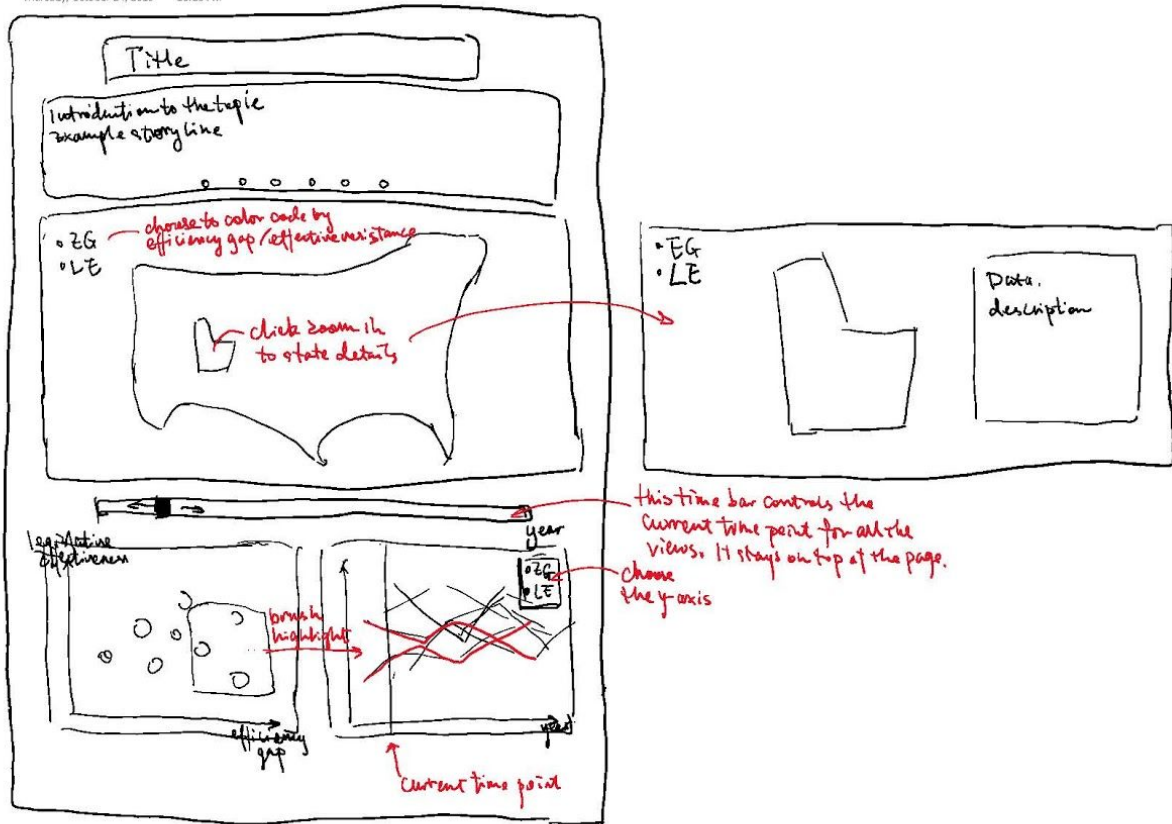
Thursday, October 24, 2019 10:03 PM



Final Design:

Final Design Sketch

Thursday, October 24, 2019 10:25 PM



Overview and Motivation

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

Related Work

A collection of visualizations and other sources which we drew inspiration from

- Fivethirtyeight's [Atlas of Redistricting](#)
-

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

Data

The data relating to the efficiency gap was obtained from <https://thelawmakers.org/data-download> and the data relating to legislative effectiveness was obtained from <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IG0UN2>. The efficiency gap was calculated according to the formula described here: https://en.wikipedia.org/wiki/Wasted_vote. The Republican-Democrat efficiency gap was calculated according to:

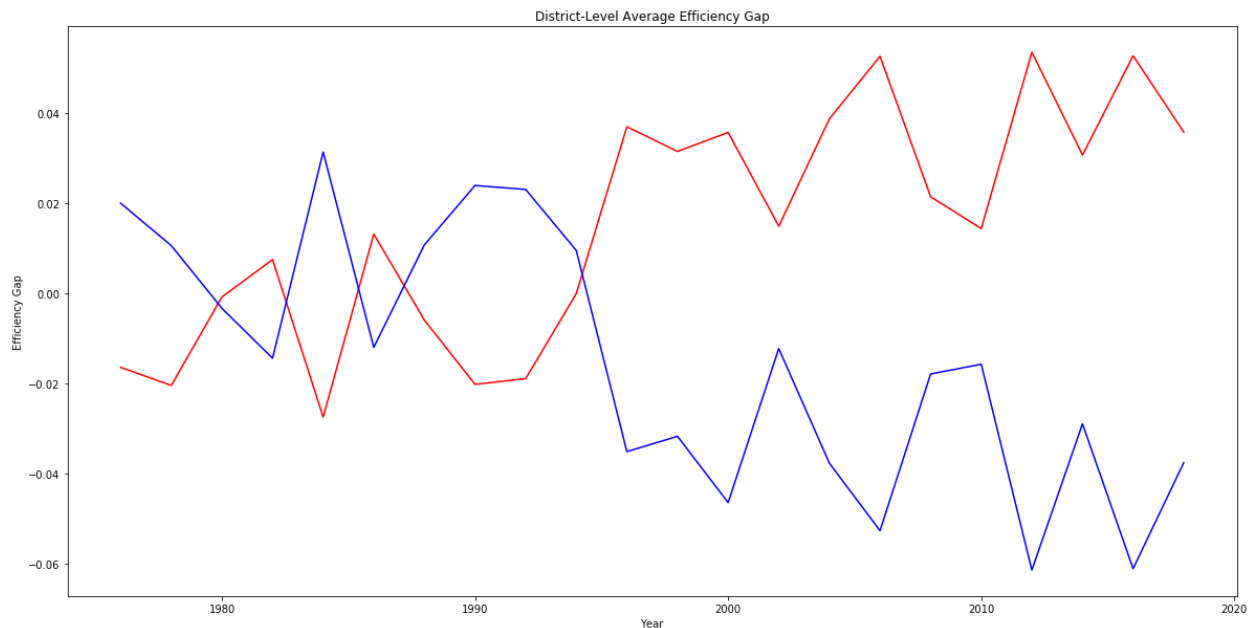
$$\text{Democratic Efficiency Gap} = (\text{Wasted Democratic Votes} - \text{Wasted Republican Votes}) / \text{Total Votes}$$

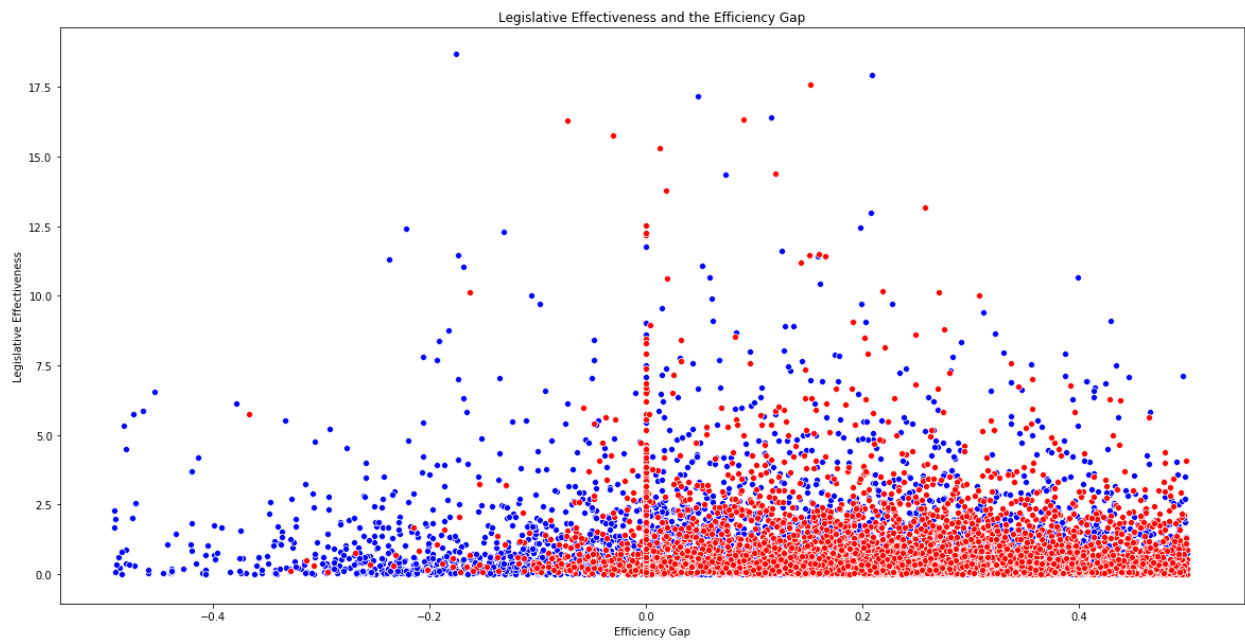
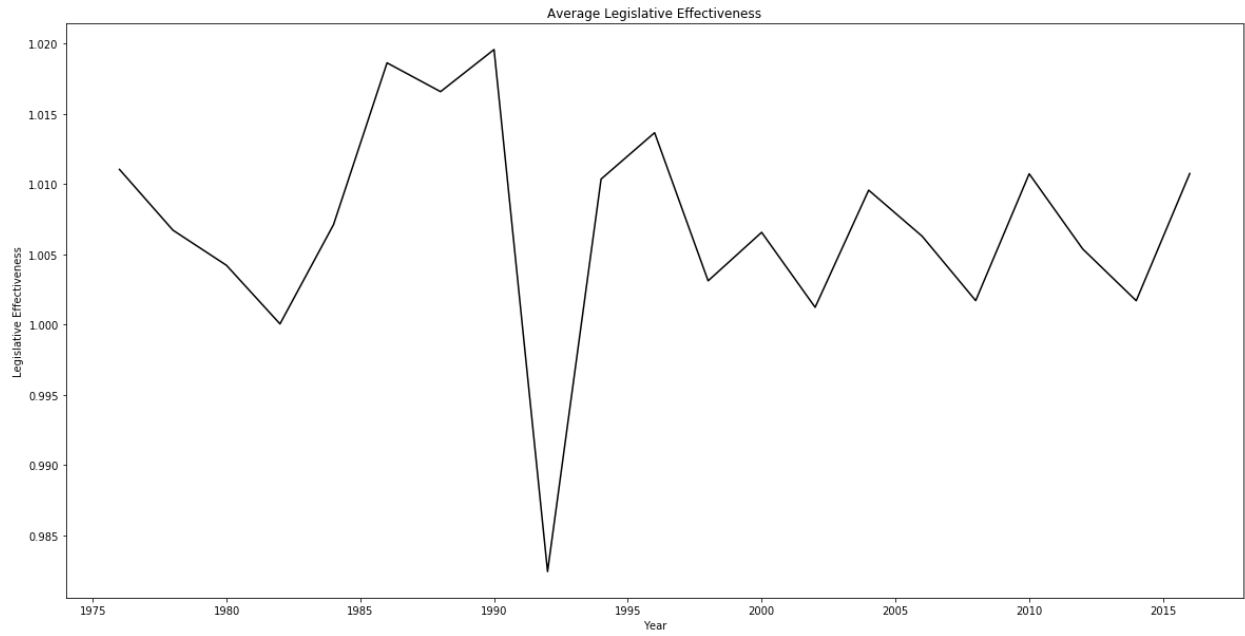
The legislative effectiveness data was used directly from the source. The data required minimal cleaning, which was performed in Python.

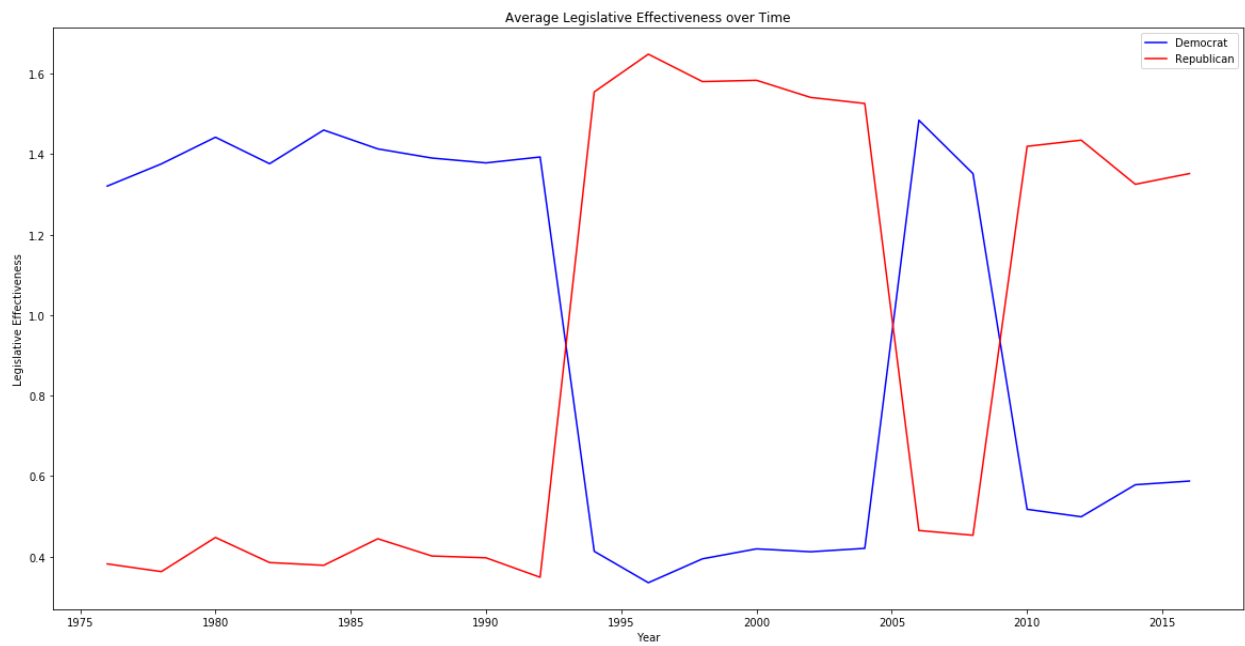
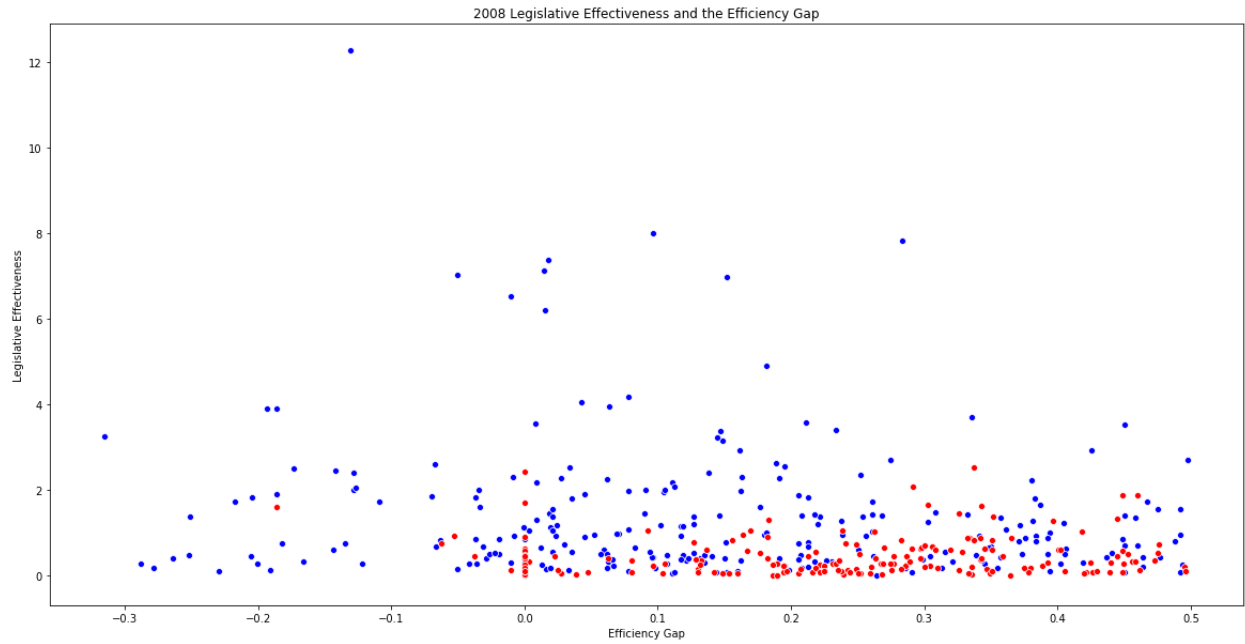
Exploratory Data Analysis

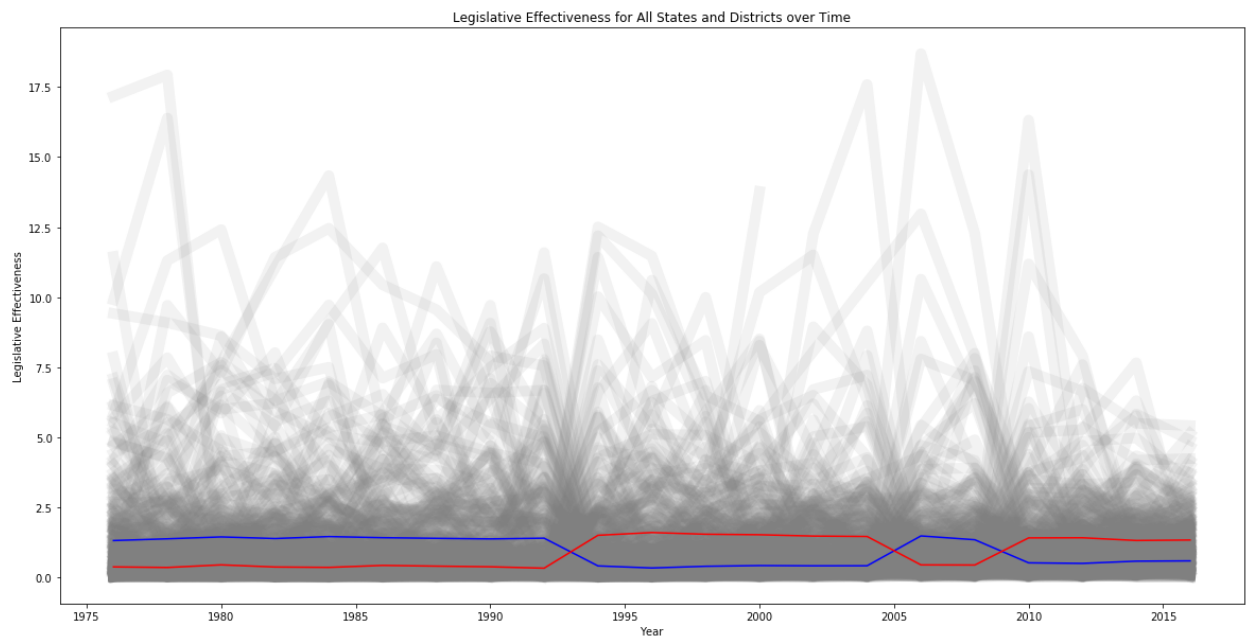
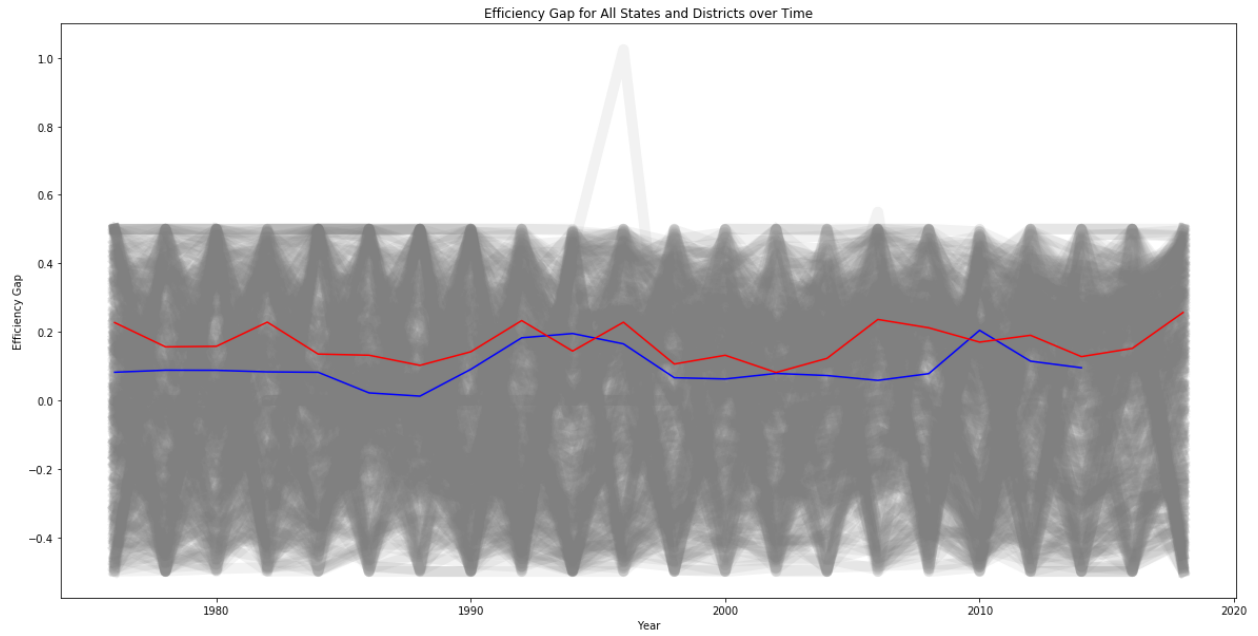
What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

Many visualizations were generated that explored the relationships between the efficiency gap, legislative effectiveness, party, state and time. Many of those visualizations are presented in the following Python plots. Originally, we hoped to see a clear and interesting relationship between legislative effectiveness and the efficiency gap, however, the relationship was more complex than we anticipated. We decided that a visualization that explored the relationship would still be interesting and valuable.









Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

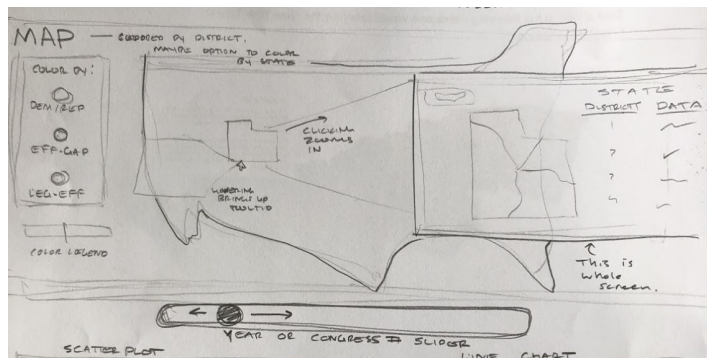
Map Evolution

The design of the map thus far is staying close to what we imagined in our initial design described in the proposal. The initial view is an overview of the nation with all of the districts colored by their relative efficiency gap score. When the user selects a state, the view zooms into the state and the rest of the map becomes hidden. There is an infobox in both the nation and the district views that show information about the highlighted state/district.

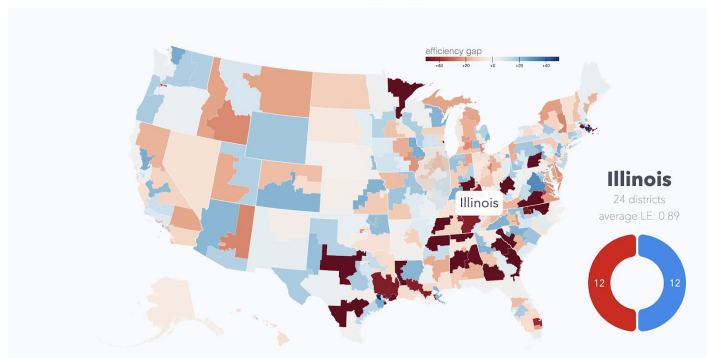
Something that we're considering implementing which wasn't necessarily something we included in our initial design, is to color the nation view by the state's efficiency gap score rather than showing all of the districts. Then, when the user's mouse hovers over a state, the district's are revealed and colored by their efficiency gap. This may be interesting because a state's overall efficiency gap score is arguably more informative than what each state's individual district's efficiency gaps are, so it may be better to first show the user which states have extreme efficiency gap values, then allow them to drill down into the districts.

Remaining features to implement:

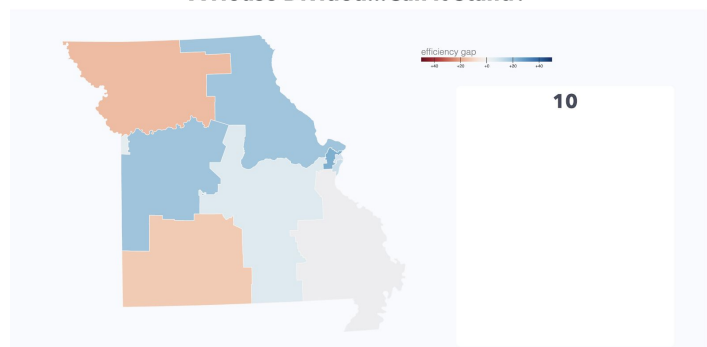
- Coloring by legislative effectiveness



A House Divided...Can It Stand?



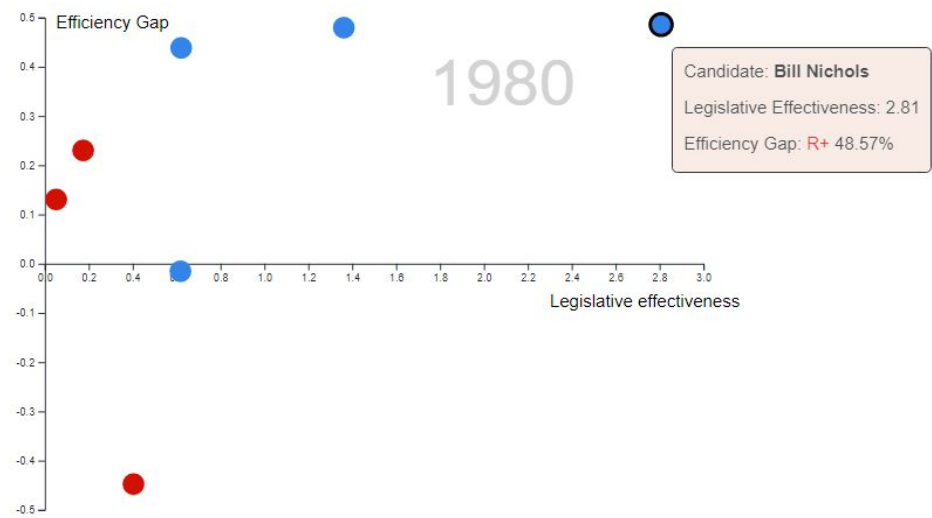
A House Divided...Can It Stand?



- Populating district infobox
- Implementing time bar to show multiple years
- Including buttons that allow the user to select multiple states at once to populate the charts
- Implementing interactivity between views

Scatterplot Evolution

Initially we drew the scatterplot with efficiency gap as the x axis and legislative effectiveness as the y-axis. Then we soon realized that since efficiency gap can go both the positive and negative directions, depending on which party it is leaning towards, it was more appropriate to have the legislative effectiveness axis positioned at 0 of efficiency gap. We color coded the circles based on the winning party of the district. We added a tooltip for each district showing the detailed data. We added a background year text to show the year that is currently active. One thing we need to clarify is that the top half of the efficiency gap axis is republican leaning and the bottom is democract leaning. We need to add this information near the axis to avoid confusion.



efficiency gap. We color coded the circles based on the winning party of the district. We added a tooltip for each district showing the detailed data. We added a background year text to show the year that is currently active. One thing we need to clarify is that the top half of the efficiency gap axis is republican leaning and the bottom is democract leaning. We need to add this information near the axis to avoid confusion.

Ultimately, the scatterplot should update as we slide the time bar. We have set a default active state to initiate the scatterplot, but we would like to add an option to select multiple states and display all the districts simultaneously on the chart. As an optional feature, we would like to add direct interaction of the scatterplot with the map. We would like to have the user select multiple states on the map and show the updated scatterplot as well. Up to this point, we have not deviated from our proposal and we do not anticipate much deviation in the future either.

Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Meeting Notes

Meeting 1 (10/18/19)

- Discussed the main topic, definitions of relevant terminology
- Set up a date to meet and discuss our design prototype

Meeting 2 (10/22/19)

- We each explained our design prototype in detail
- Decided on a final version of the design
- Divided tasks for the proposal

Meeting 3 (10/31/19)

- Looked at Tom's results from exploratory data analysis
- Discussed in detail the tasks we need to complete before the milestone deadline
- Divided the initial visualization tasks among team members

Peer Feedback

Reviewer: Patrice Nicholes

Patrice commented that the scope of the project seems large, but she thought that our division of labor was effective. She liked the interactions across all our views in the design. She said that the time bar would be beneficial for seeing the changes overtime. Regarding visual components, she said that although the story-telling feature is currently optional, it would be helpful for users to have a better understanding of the quite complex concepts. We have gotten similar feedback from our TA. After we finish all the must-have features, we will implement the story-telling component as our first option feature.