A House Divided...Can It Stand?

Process Book

Michael Young, Tom Pace, Fangfei Lan

Table of Contents

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:

. 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):
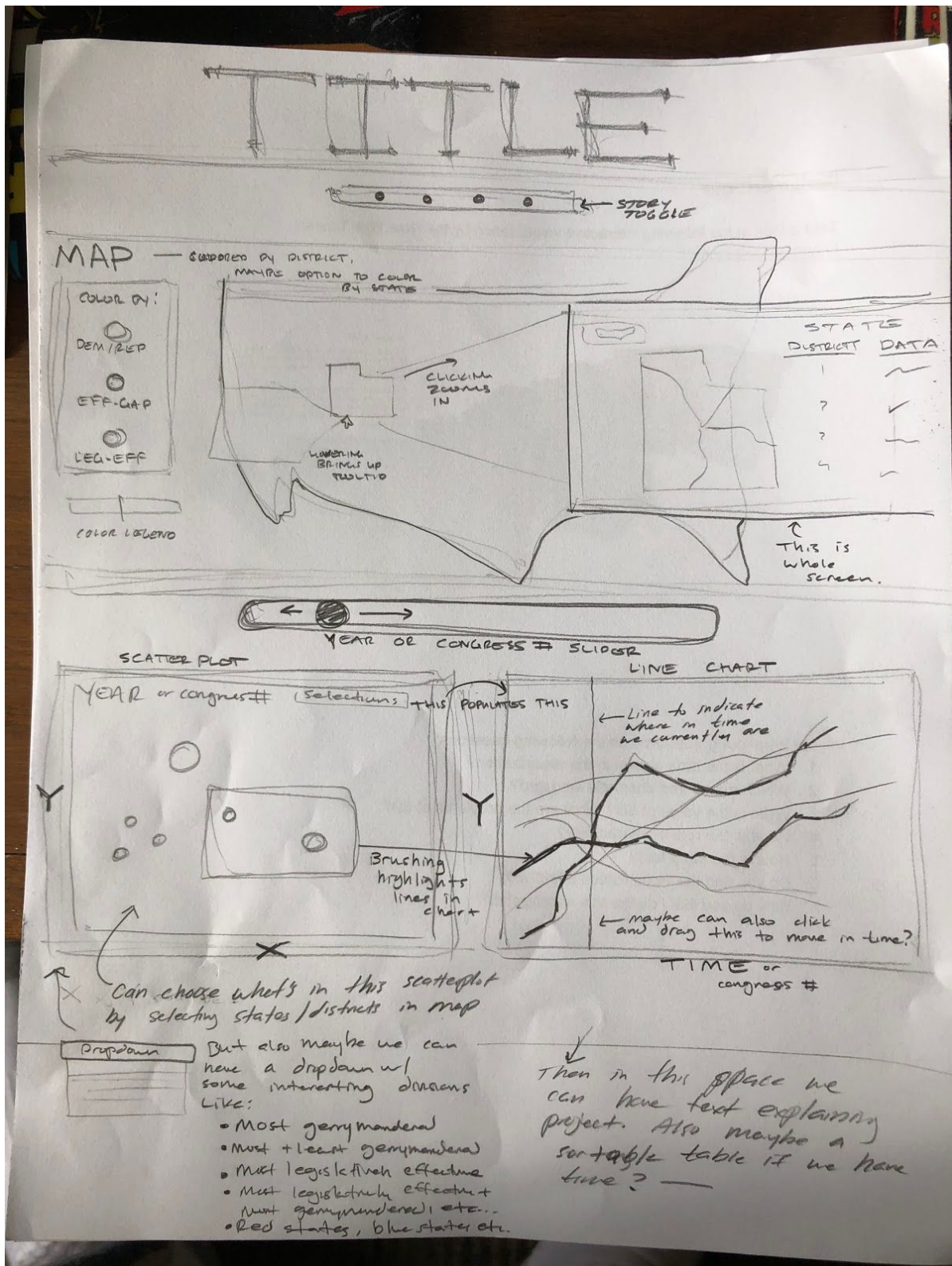- Complete the skeleton of each view

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
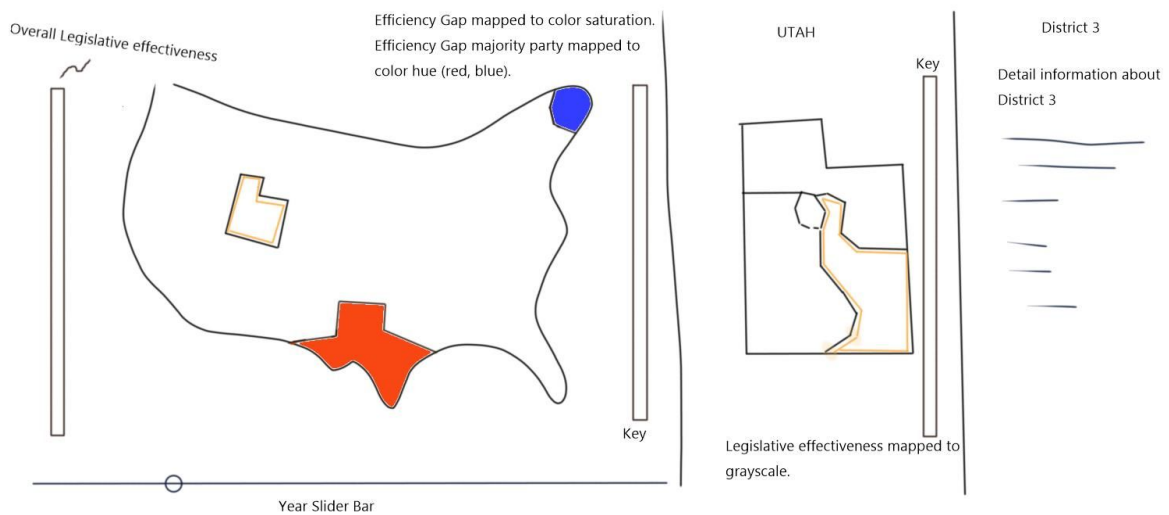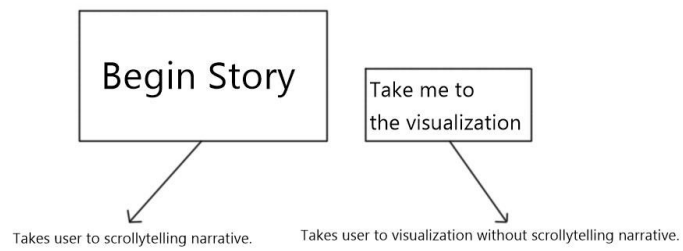- Implement some optional features if time permits
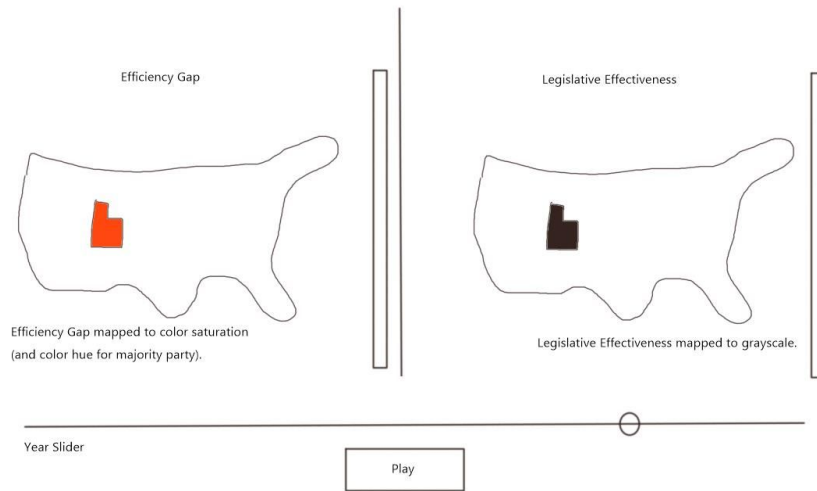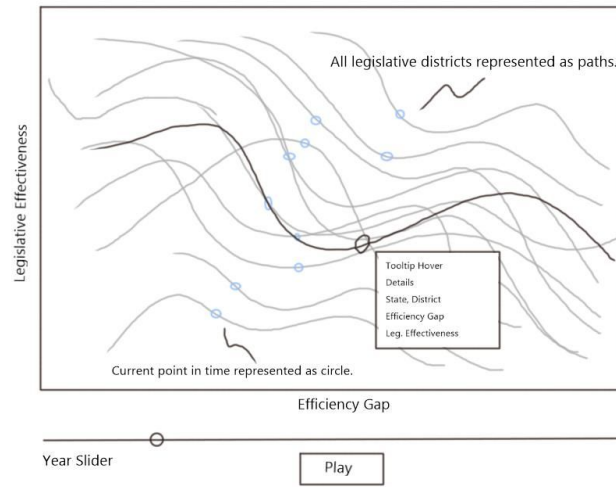
Nov 27: project due.
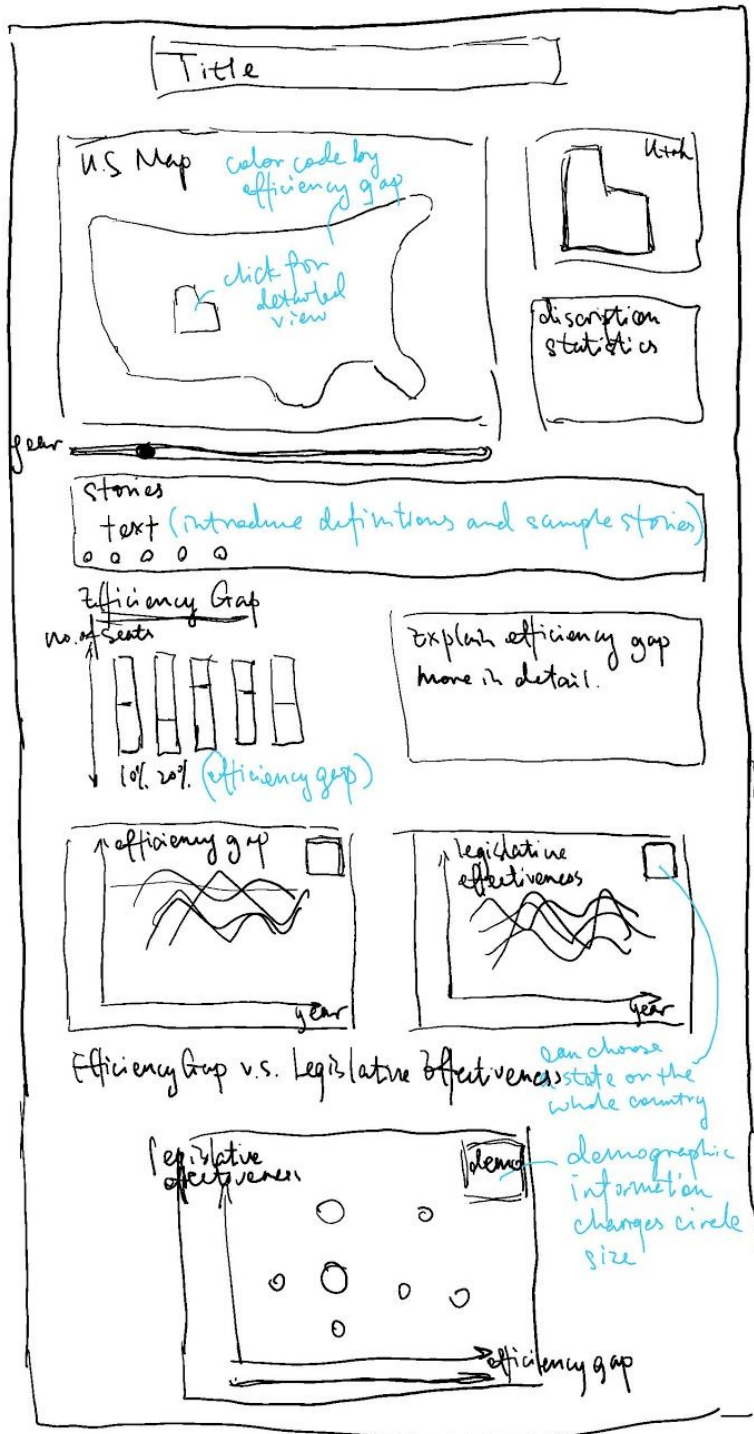
Prototype 1:

Prototype 2:

Introduction Page

Begin Story

Take me to the visualization

Takes user to scrollytelling narrative.

Takes user to visualization without scrollytelling narrative.

Overall Legislative effectiveness

Efficiency Gap mapped to color saturation. Efficiency Gap majority party mapped to color hue (red, blue).

Key

Year Slider Bar

UTAH

Key

Legislative effectiveness mapped to grayscale.

District 3

Detail information about District 3

All legislative districts represented as paths.

Legislative Effectiveness

Tooltip Hover
Details
State, District
Efficiency Gap
Leg. Effectiveness

Current point in time represented as circle.

Efficiency Gap

Year Slider

Play

Efficiency Gap

Legislative Effectiveness

Efficiency Gap mapped to color saturation
(and color hue for majority party).

Legislative Effectiveness mapped to grayscale.

Year Slider

Play

Prototype 3:

Project Sketch
Thursday, October 24, 2019   10:03 PM

Final Design:

Final Design Sketch
Thursday, October 24, 2019   10:25 PM

Title

Introduction to the topic
example story line

• EG
• LE

choose to color code by
efficiency gap / effective resistance

click zoom in
to state details

• EG
• LE

Data
description

Legislative
effectiveness

year

• EG
• LE

this time bar controls the
current time point for all the
views. It stays on top of the page.

choose
the y-axis

brush
highlight

efficiency
gap

year

current time point

Overview and Motivation

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. The goal of the project was to create an interactive visualization with three separate views that would allow the user to explore the efficiency gap and legislative effectiveness and the relationship between the two over the recent history of the United States.

Related Work

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

- Fivethirtyeight's [Atlas of Redistricting](#)
- [Planscore](#)
- Bostock's [Command Line Cartography](#)
- HW4 - Gapminder

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?

Our primary question at the beginning of the project was to explore the relationship between legislative effectiveness and efficiency gap. We wanted to potentially investigate what role demographic information plays in the relationship between the two variables as well, but we later decided that was a more ambitious goal. As the visualization evolved, we added more questions about the states' progressions. How does the efficiency gap change overtime for a certain state? How does the party change for the states overtime?

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:

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

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.
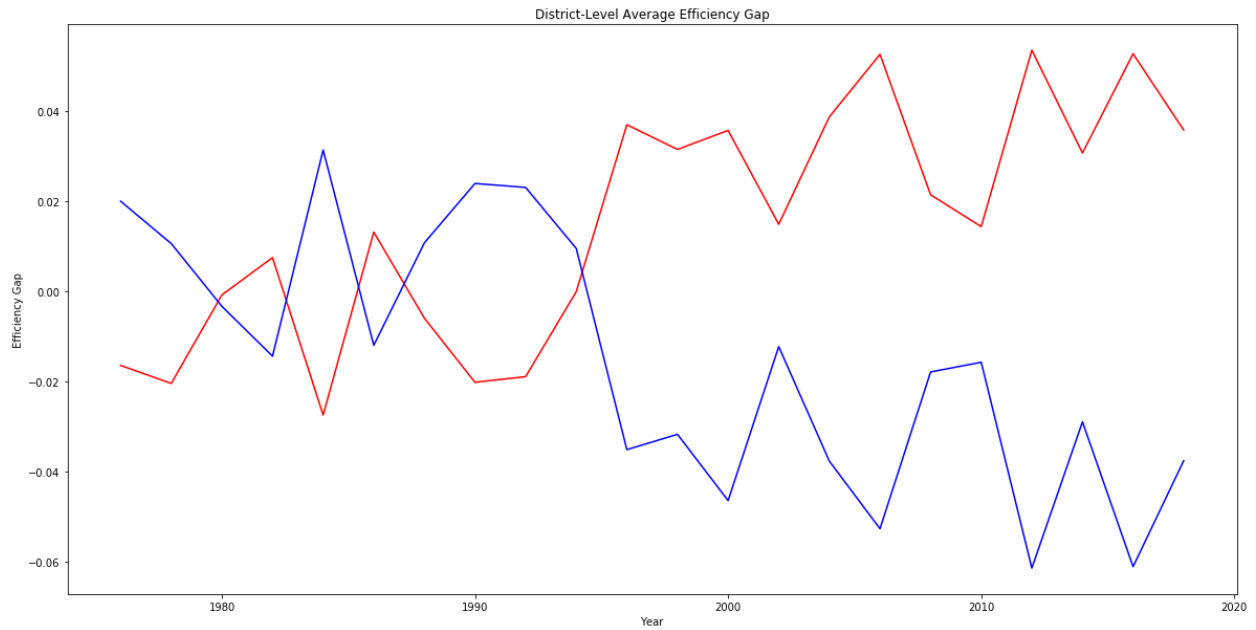
District-Level Average Efficiency Gap: This figure illustrates the average efficiency gap from 1976 to 2016 for the US House of Representatives. The data shows that on average Republicans have had a partisan advantage due to the voting district boundaries since around 1994.
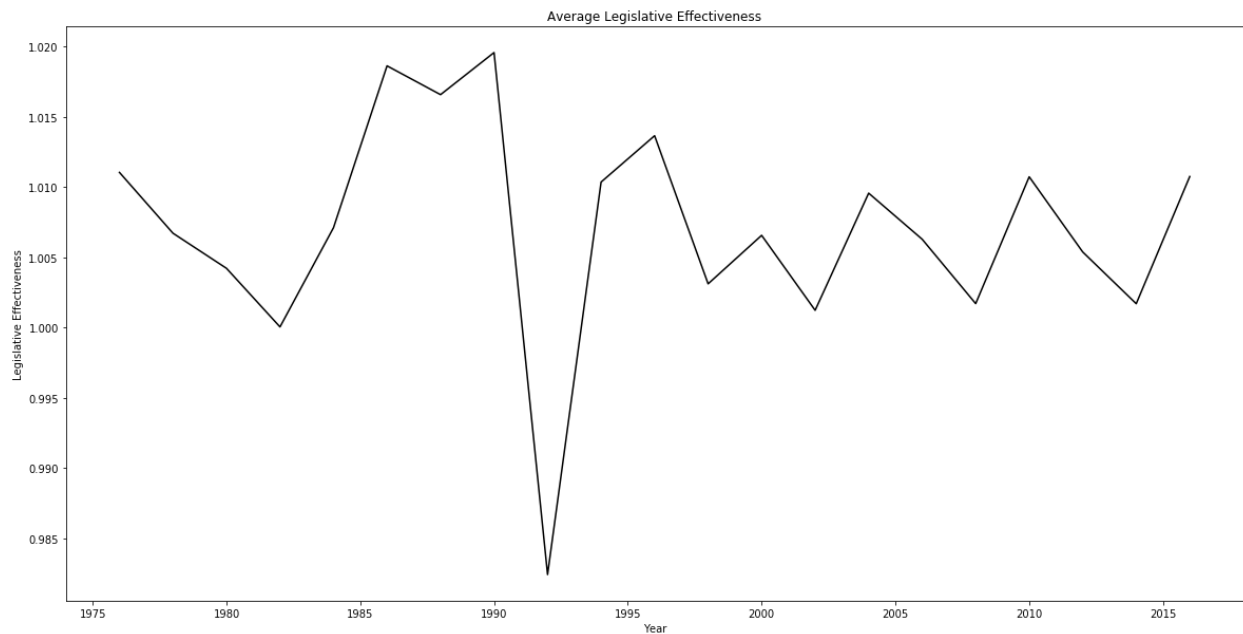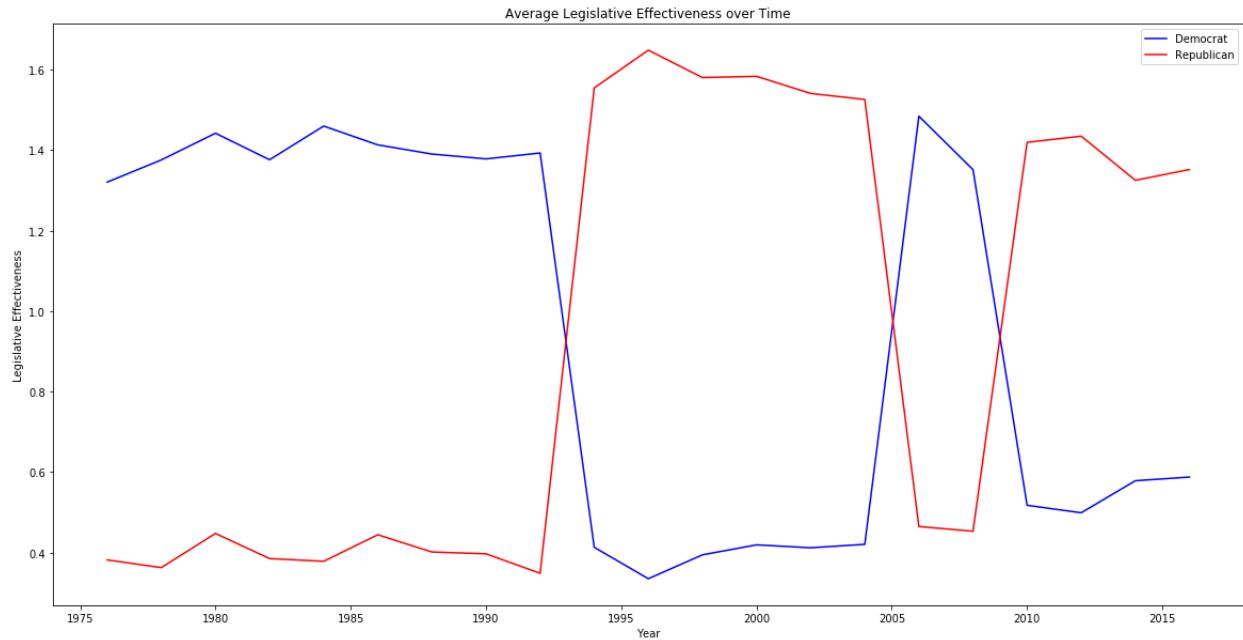

Average Legislative Effectiveness: This figure shows the average legislative effectiveness through time for the US House of Representatives. No clear relationship between the efficiency gap and legislative effectiveness could be detected.
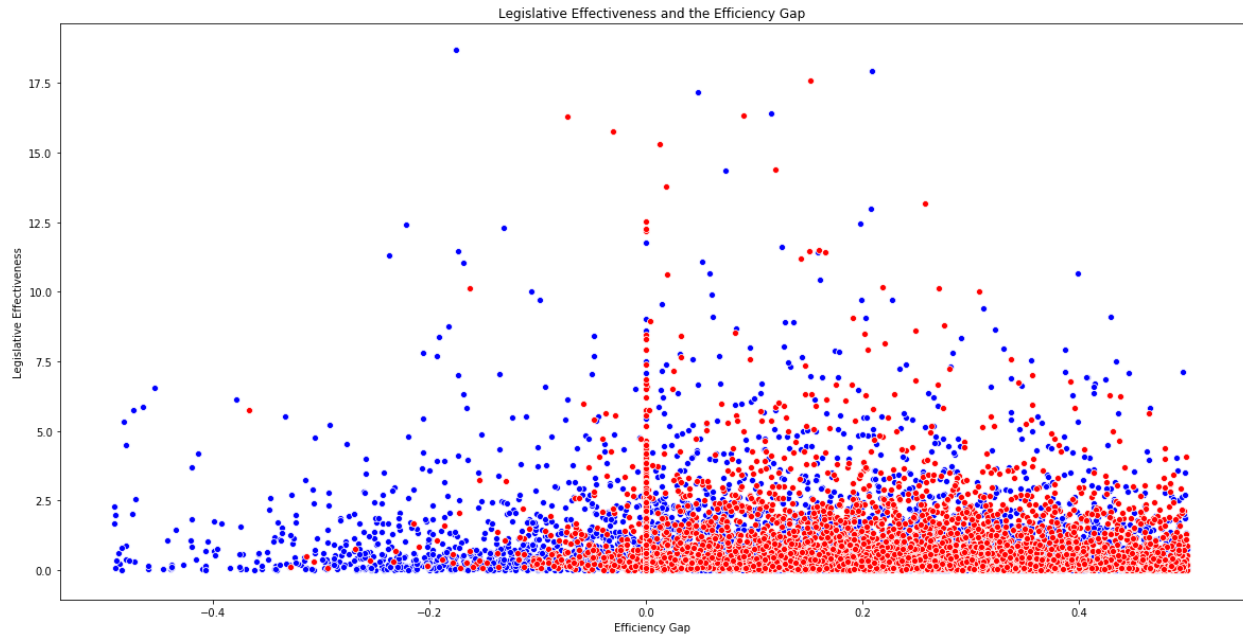
Legislative Effectiveness and The Efficiency Gap: This figure illustrates the relationship between legislative effectiveness and the efficiency gap for all Congressional Representatives between 1976 and 2014. No clear relationship could be detected.
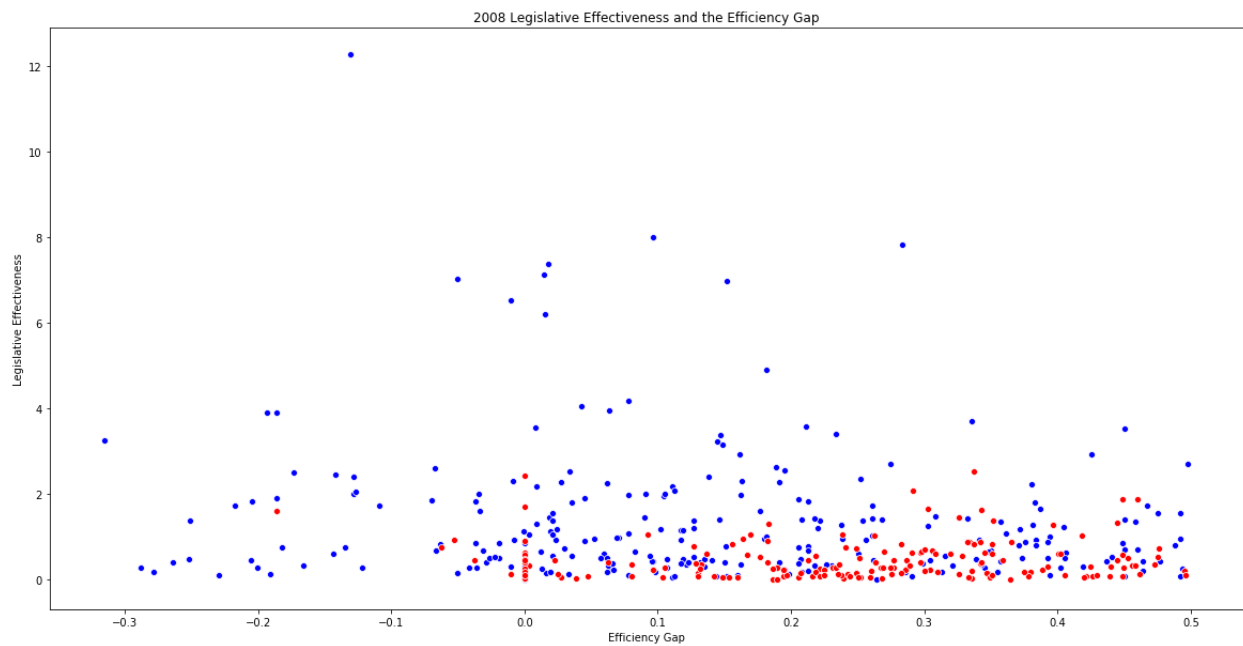
Legislative Effectiveness and The Efficiency Gap for 2008: This figure illustrates the relationship between legislative effectiveness and the efficiency gap for all Congressional Representatives during 2008. No clear relationship could be detected.

Average Partisan Legislative Effectiveness: This figure shows the average Republican and Democratic legislative effectiveness through time. As expected, the data shows that on average the legislative effectiveness of a member of congress is related to which party controls the US House of Representatives at that time.
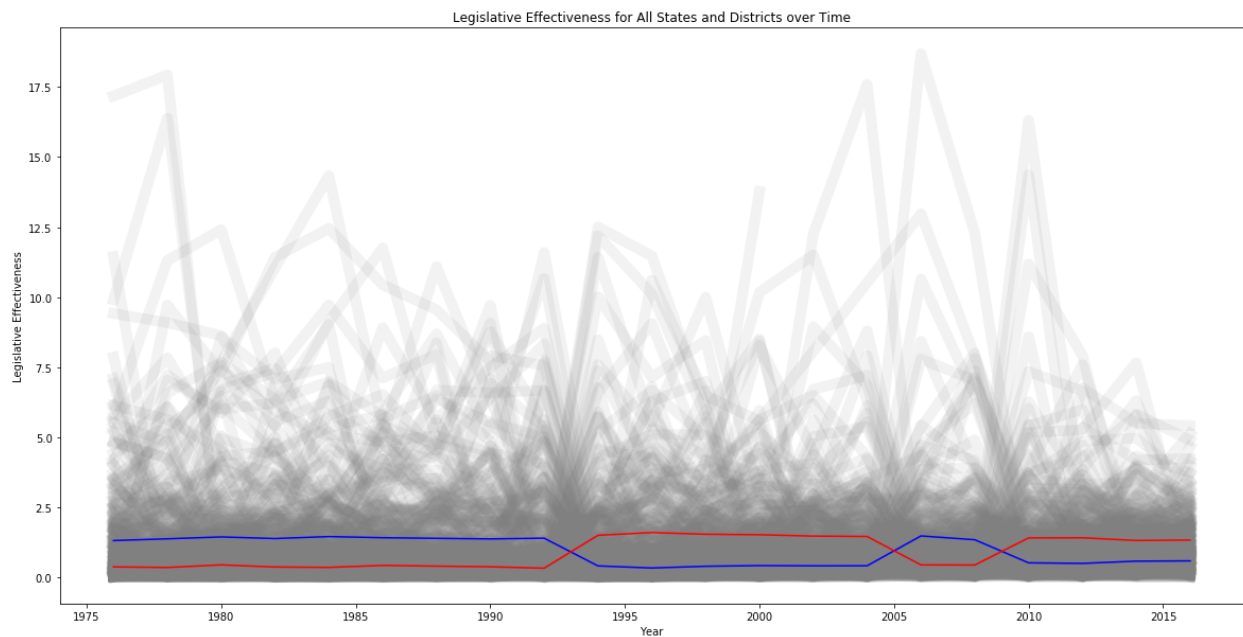
Average Partisan Legislative Effectiveness: This figure is an extension of the figure above with every legislative district represented in gray and the overall average Republcan and Democratic values in red and blue, respectively.

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 stayed pretty close to what we imagined in our initial design described in the proposal, with a few small changes. Initially, the view was an overview of the nation with all of the districts colored by their relative efficiency gap score. However, this became somewhat busy visually, so we adjusted it to show states colored by their average efficiency gap, with the districts visible upon hover. Additionally, as described in the proposal, we offer the user the option to color by the legislative effectiveness.

When the user selects a state, the view zooms into the state and the rest of the map becomes hidden. We implemented it this way because it made the transition render more smoothly, and it reduces visual clutter. There is an infobox in both the nation and the district views that show information about the highlighted state/district. Part of this info box is a donut chart showing the relative composition of republican and democratic candidates in the state. We chose a donut plot because, well, it looks nice. And the information it encodes is simple enough that a lot of the pitfalls that come with using pie/donut charts don't apply.
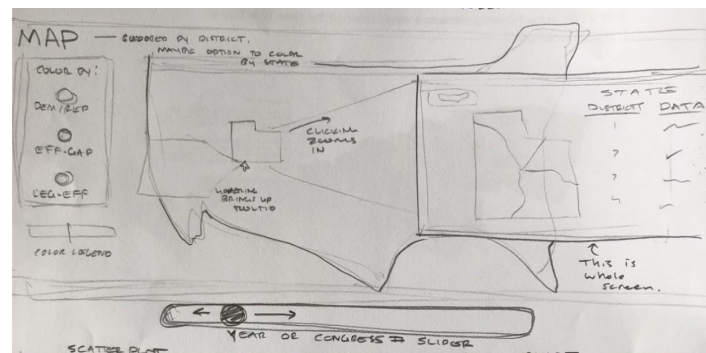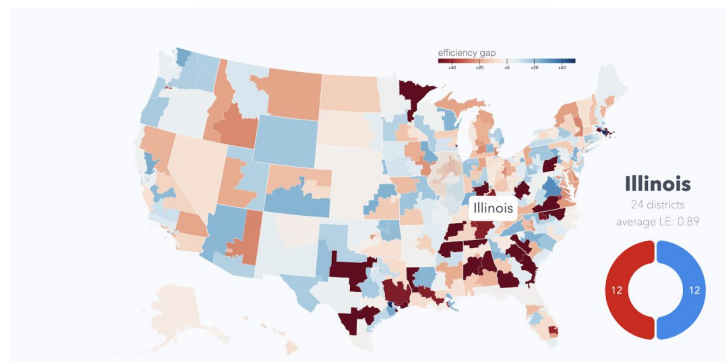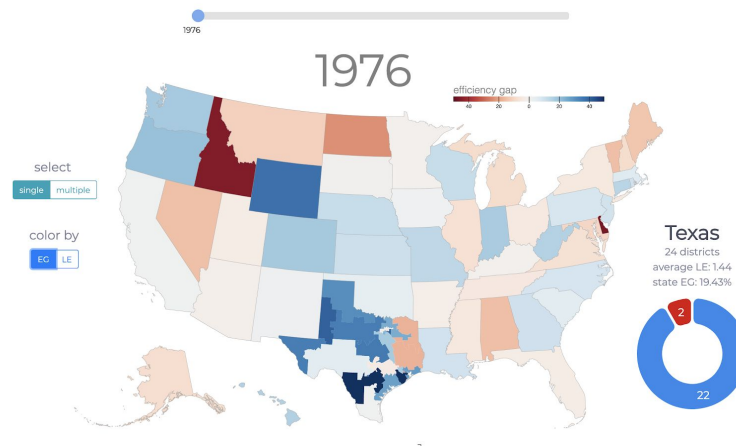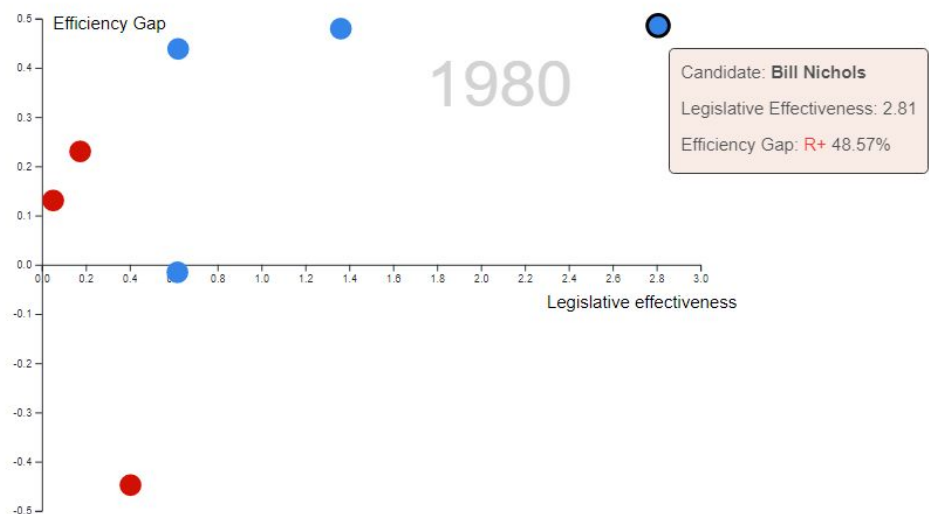
Map evolution from the proposal to the final implementation.

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.

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
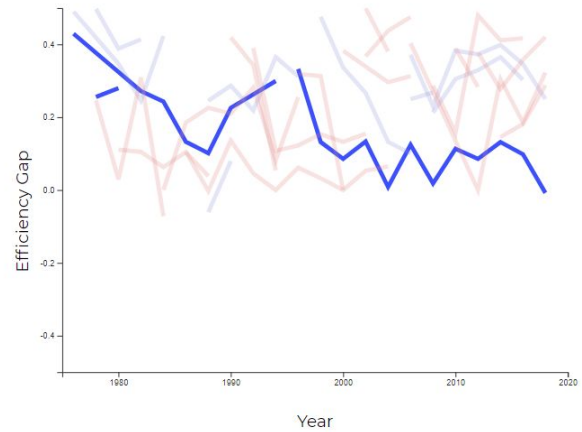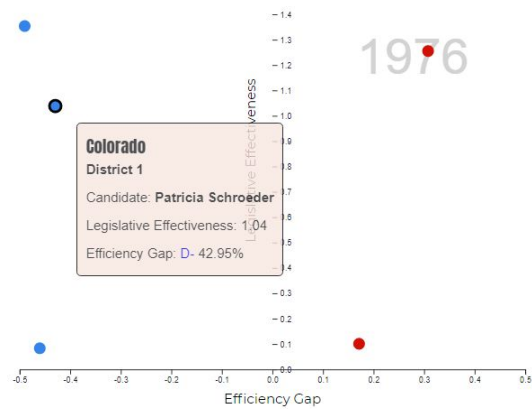
display all the districts simultaneously on the chart. As an optional feature, we would like to add direct interaction of the scatterplot with the map. We would like to have the user select multiple states on the map and show the updated scatterplot as well. Up to this point, we have not deviated from our proposal and we do not anticipate much deviation in the future either.

After meeting with Kiran, I made some changes to our design for the scatter plot. We first of all switched the x and y axes. Now, the x axis is efficiency gap and the y axis is legislative effectiveness. We realized that efficiency gap data is not always symmetrical, so we picked the larger absolute value of the negative and the positive, to center the axis at 0. Since legislative effectiveness is a non-negative parameter, it made sense for it to be the y axis.

The scatter plot has a few different functions. When nothing is selected on the map, the scatter plot displays the state average efficiency gap and legislative effectiveness at the given year, like the screenshot shown to the right. We decided to add the overall view in order to show the distribution of the two variables overtime. Also plotting all the districts of all states would have created visual clutter. We considered adding animation, but with the number of data points and the several other

functions of the scatterplot, the transition would not add more useful information to the scatterplot, instead it would make the plot more chaotic.
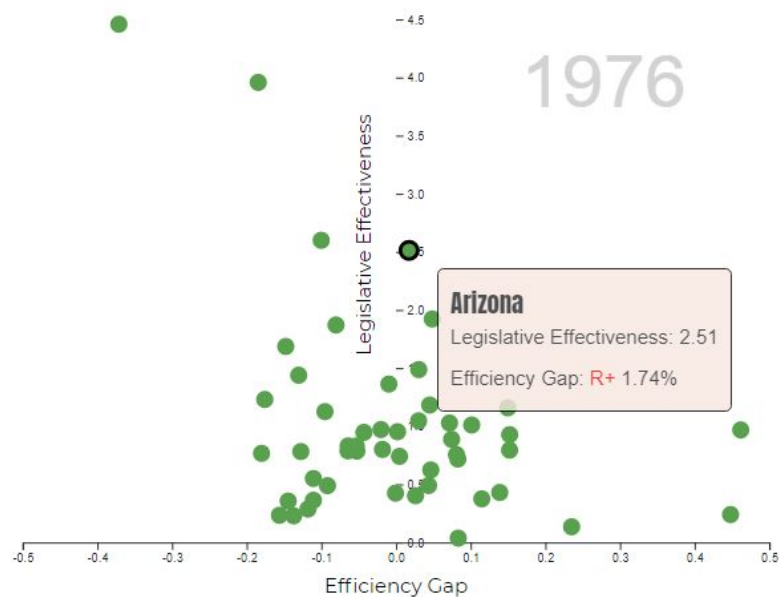
When a single state is selected on the map, the scatter plot displays all the districts in the selected state in a given year. The districts are colored by party. When the user hovers on top of a district, the corresponding district highlights in the line chart. We added the highlighting function, so we can see the progression of the district over the years.

1976

Colorado
District 1
Candidate: **Patricia Schroeder**
Legislative Effectiveness: 1.04
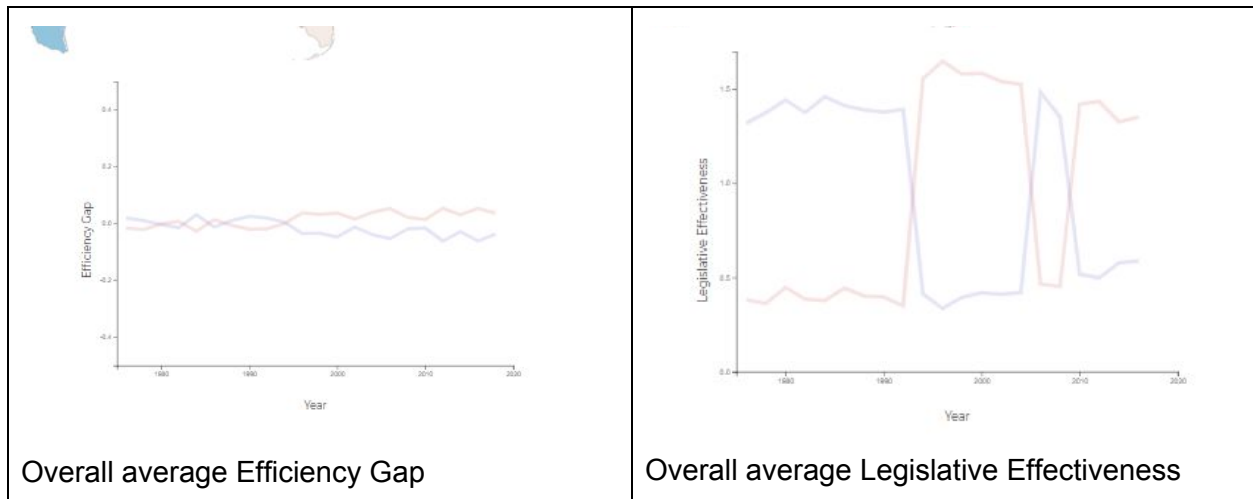Efficiency Gap: D- 42.95%

When the user selects multiple states on the map, the scatter plot displays all the districts of the selected states. The dots are colored by party by default, but the toggle buttons on the side provides another choice of color by state. The screenshot on the right shows all the districts in both Colorado and Kansas, where Colorado districts are in blue and Kansas districts

1982

Party

State

are in orange. We added the color by state feature to compare the selected states at a given year.
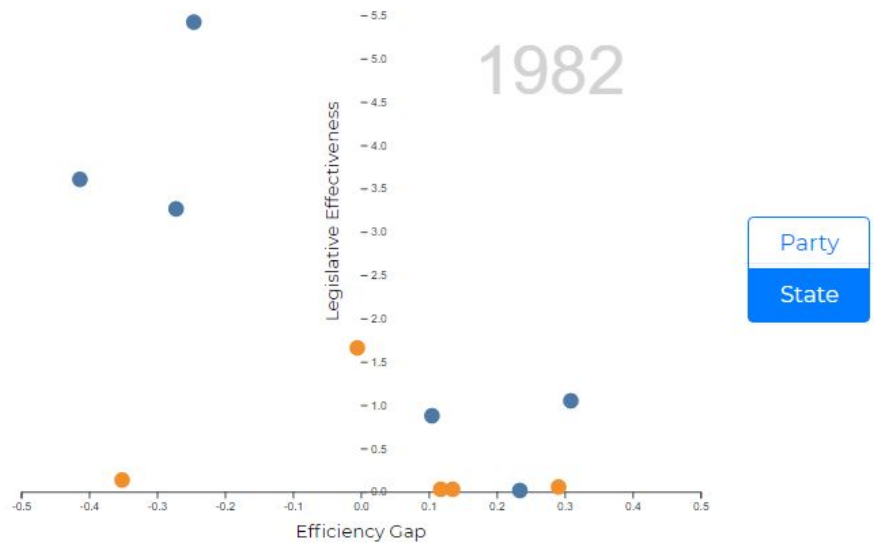
Line Plot Evolution

The goal for the line plot was to show the legislative effectiveness and efficiency gap through time and to provide interactivity with the other views. Initially, the overall average data for Republicans and Democrats were created:


| Overall average Efficiency Gap | Overall average Legislative Effectiveness |

After the overall average data plots were working, we created a path to represent every individual representative in a given state when that state was activated in the map view.
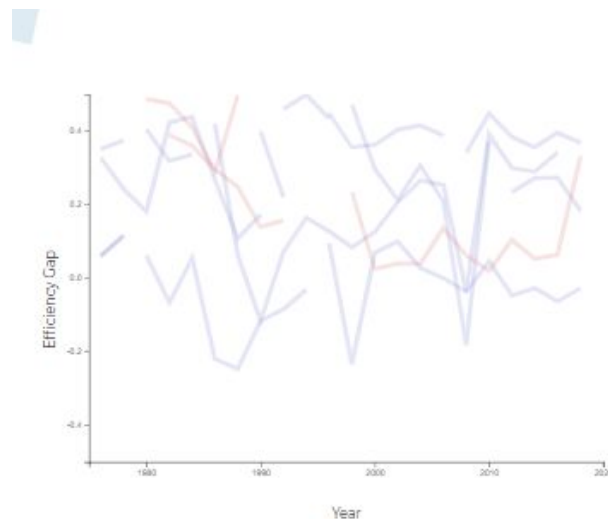
This figure shows the efficiency gap for Oregan through time.

After the individual state functionality was working, a tooltip was added to provide detail about the individual district in the line view.
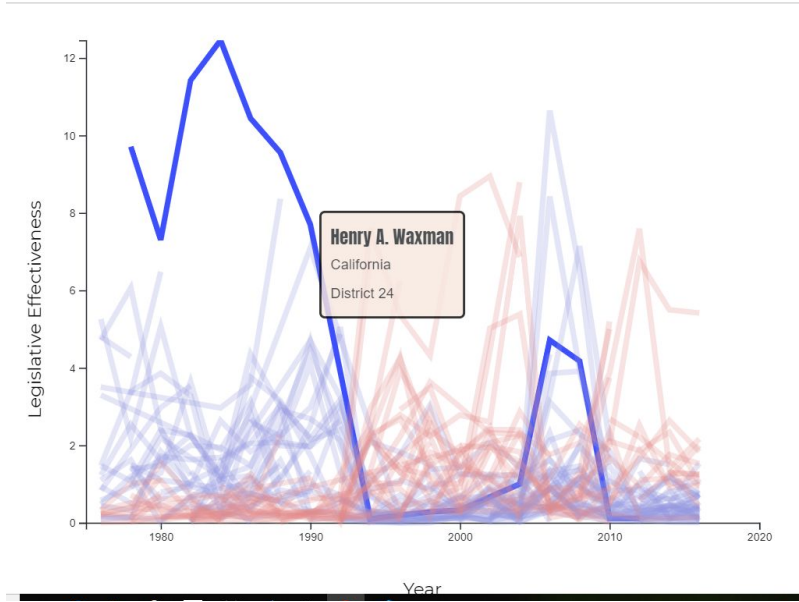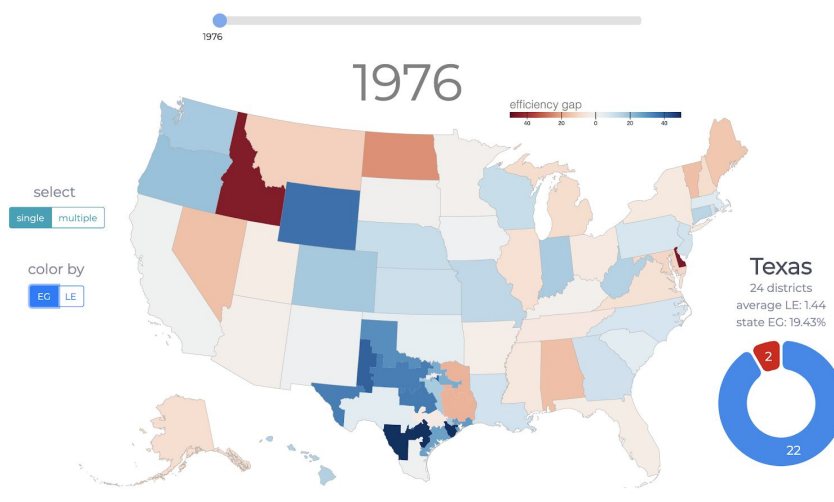
This figure shows the tooltip detail in the legislative effectiveness view for Henry Waxman, who was an extremely effective legislator from California.
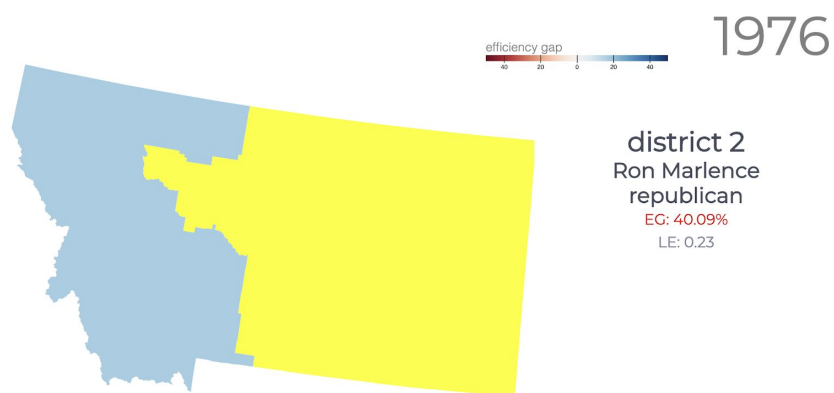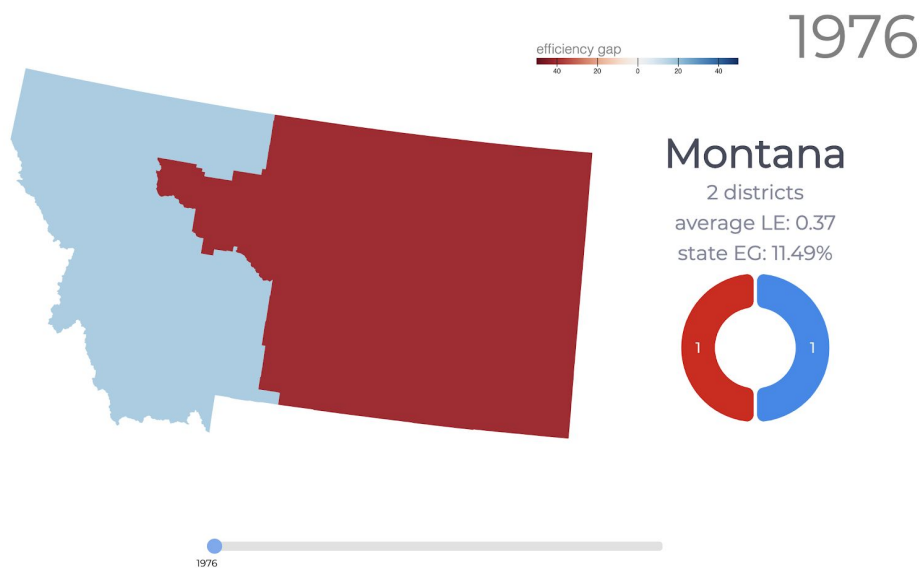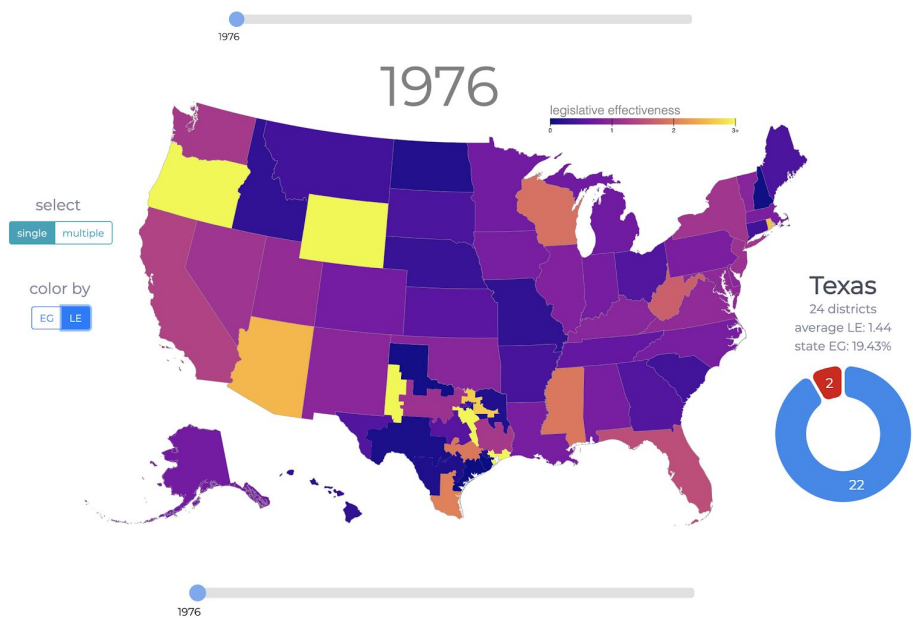
Implementation

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

In the range between a fully guided narrative experience and a sandbox exploratory tool, our vis lies closer to the sandbox paradigm. Our intent, as stated in the overview and motivation, was to create a tool 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."

In order to accomplish this, we implemented 3 interactive views: a map, a scatterplot, and a line chart. The map serves as the focal point of the visualization - it is colored by either legislative effectiveness or the efficiency gap and allows the user to drill down and explore the districts of each state. There is a time bar located at the top of the visualization (and "sticks" to the top of the view bar as the user scrolls) which allows the user to explore the data through the years 1976-2014. As the time bar as adjusted, the district lines and colors (as well as all other time-dependant aspects of the vis) transition seamlessly.

1976

legislative effectiveness
0   1   2   3+

select
single multiple

color by
EG LE

Texas
24 districts
average LE: 1.44
state EG: 19.43%

2
22

1976

efficiency gap
40 20 0 20 40

Montana
2 districts
average LE: 0.37
state EG: 11.49%

1   1

1976

efficiency gap
40 20 0 20 40

district 2
Ron Marlence
republican
EG: 40.09%
LE: 0.23

The map also allows the user to either select a single state (and allow drill-down to districts) or to select multiple states.
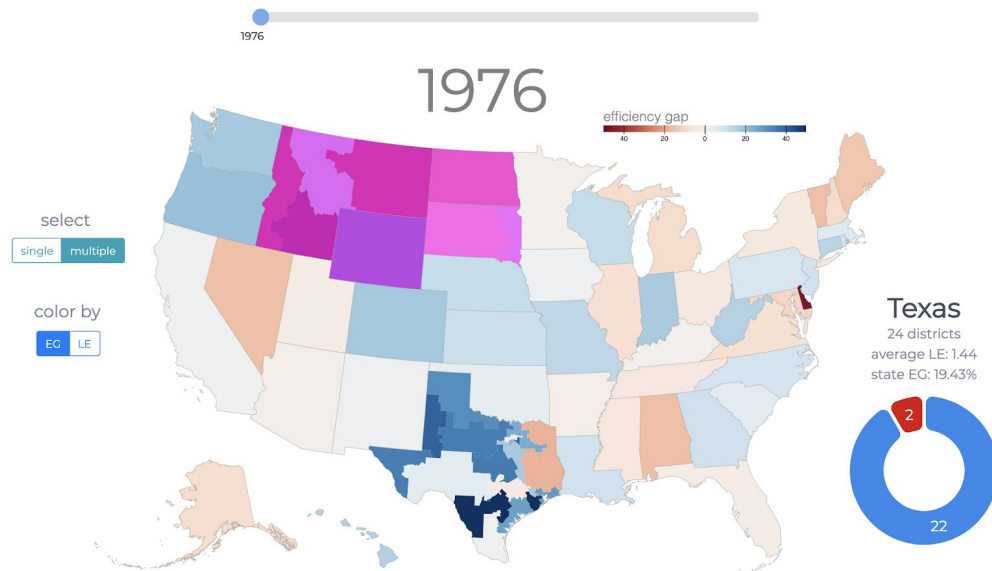

The districts of the states in the multiple state selection populate the scatterplot located below the map. The purpose of the scatterplot is to show the relationship between legislative effectiveness (y-axis) and the efficiency gap (x-axis). A potential question this plot could answer is, "is there a positive linear relationship between efficiency gap and legislative effectiveness in texas in 1978?".
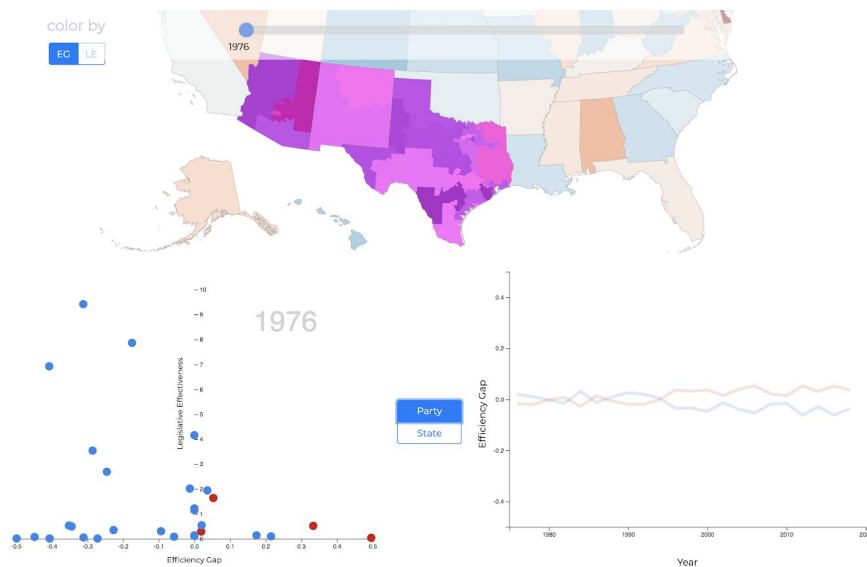
If just a single state is being selected, then all of the districts of that state populate the scatterplot. Additionally, the districts of that state across the entire time set populate the line chart to the right of the scatterplot. The purpose of the line chart is to give the user a sense of how either the efficiency gap or legislative effectiveness (selected by the same button that colors the map) vary over time. This could potentially answer some questions like "how has the efficiency gap of district 9 in texas changed over time?" The lines in the line chart can either be hovered over, or highlighted by hovering over the bubbles in the scatterplot. The interaction between these charts is intended to be a useful and generative one in the sense that the bubble chart may reveal an interesting data point, and then the line chart allows the user a way to see how that data point changes with time. The motivation for this was learning that we should always trust "eyes over memory."
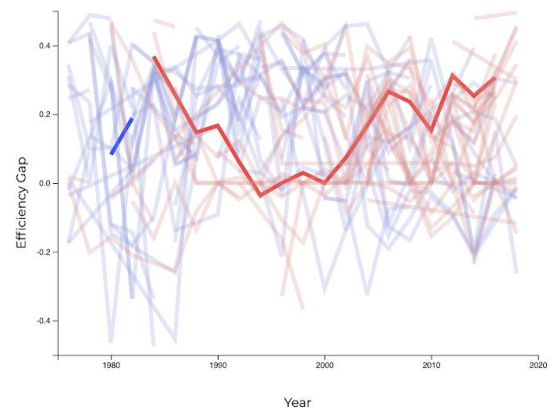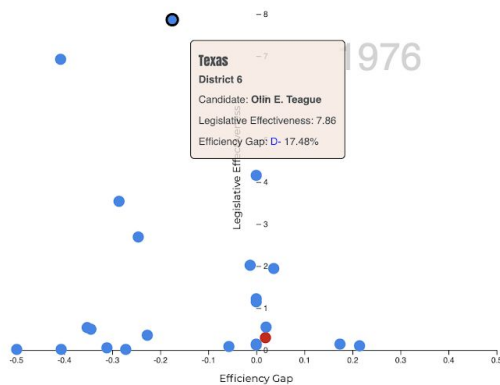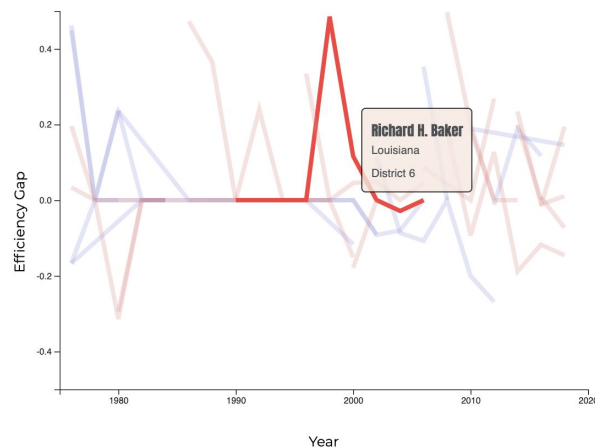
One last aspect of the vis that is important are the tooltips and embedded links which explain more in depth what the efficiency gap and legislative effectiveness are for the interested user.
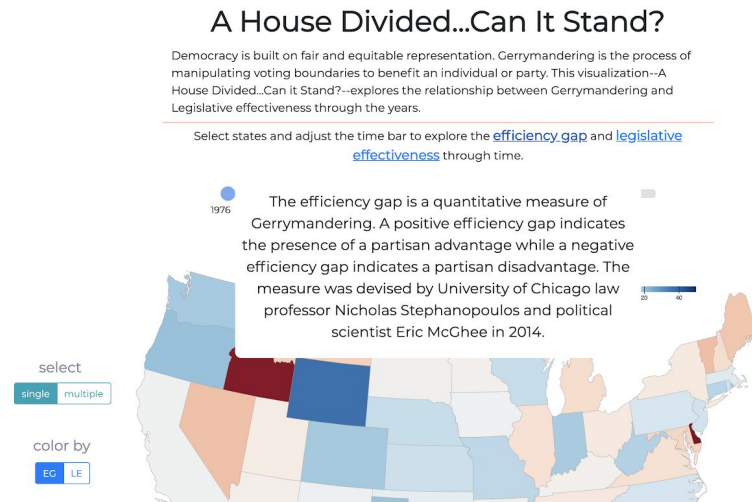
Taken together, all of the features create a tool that allows the user free range to explore both efficiency gap and legislative effectiveness and how the two may or may not be connected.

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?

We learned that there is not a clear relationship between efficiency gap and legislative effectiveness. We built a tool for exploratory analysis. The visualization serves its purpose for users to explore the dataset in detail. The views interact with each other well, which provides more linked information about the states the user is interested in. The time bar shows the time variant components of the data. All the features of the visualization are connected and they display different aspects of the data. Although we have included a variety of functionalities in the visualization, there is still a lot we can do to improve. First of all, the layout of the visualization

could be better if it does not need to be scrolled down to see all three views. Secondly, we can add more pre-computed selections for the users to clearly see, for example, the states with top five legislative effectiveness scores. We can also add more functionalities of the line chart to display state averages when multiple states are selected. Finally, we can add the story-telling component to make the visualization more engaging and self-explanatory for new users.

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.