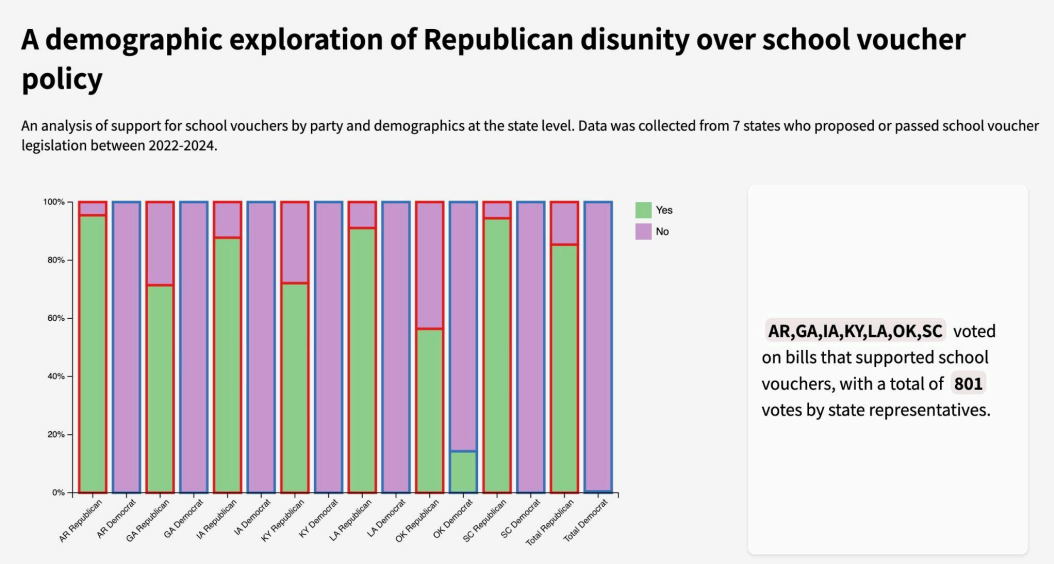
CSC 362 Final Project - Lexi & Maddie



Data

The dataset we used was a combination of U.S. Census Bureau surveys and manually entered data regarding state legislative branches created by Maddie for her major paper in Political Science. Data from the U.S. Census Bureau was organized by State Upper Legislative Chamber and State Lower Legislative Chamber geographies. Data was extracted from Table P2: Urban and Rural dataset within the DEC Demographic and Housing Characteristics from the 2020 Decennial Census to provide the urban/rural population of all state upper and lower legislative districts. Data was extracted from Table S0701: Geographic Mobility by Selected Characteristics in the United States from the 2021 ACS 5-Year Estimate to provide racial demographics for each district. Table B14003: Sex by School Enrollment by Type of School by Age from the 2021 American Community survey records public and private school enrollment rates by district and Table S1901: Income in the Past 12 Months (in 2022 Inflation-Adjusted Dollars) from the 2021 American Community survey records the average income by household in each district. These Census Bureau tables were combined with an original data set detailing each selected state's State House and Senate representatives by district, party affiliation, and voting record on the selected school voucher bill for each state. Seven school voucher bills from seven states are recorded in the data. The selection rationale was three-fold. First, I looked for states that had a history of racial segregation in schooling and/or recent increases in Latino migration, as it was the focus of my major paper. Second, I checked that the school voucher bill was near-universal and that the bill was narrowly tailored such that its passage would only create or expand a voucher system. Many bills were turned down due to the addition of "riders", for example, a North Carolina bill sought to increase funding for Education Savings Accounts while also authorizing local police departments to cooperate with ICE. The third requirement

was that the bill was not passed straight down partisan lines. The majority of hypotheses in my major paper were focused on Republicans who break party ideology, so to collect as much data as possible, I focused on bills that some number of Republicans voted against. Legislative districts whose representatives were absent from the vote were excluded from the analysis.

Below is a table that describes the type of voucher for each state in the dataset. To clarify, voucher funding occurs at the state level, so in order for a voucher to be passed, it is voted on by the state's Lower Representative District (often called the House) and the Upper Representative District (often called the Senate). So, each item in the dataset is one upper or lower representative district, and its vote attribute is based on the vote the district's representative cast in the State House or State Senate for the bill.

State	Bill	Type of voucher	Passage
Arkansas	SB294 (2023)	Universal	Passed
Georgia	SB601 (2022)	Universal	Failed
Iowa	HF68 (2023)	Near-universal	Passed
Kentucky	HB9 (2022)	Universal	Passed, overturned by Kentucky Supreme Court, failed in state referendum
Louisiana	SB313 (2024)	Universal	Passed
Oklahoma	SB1647 (2022)	Near-universal	Failed
South Carolina	SB39 (2023)	Near-universal	In committee

We cleaned this dataset by eliminating any districts whose representatives were absent for the vote (vote = -1 in dataset). Because I had already used the data in the past, there was no missing data. Since we were looking for overall trends we planned to aggregate the data by state and vote choice rather than looking at individual entries. Thus, we chose to subset the data to exclude some attributes, such as representative's name, which would not be used once aggregated. Below is a screenshot of the metadata for the variables we chose to include in the

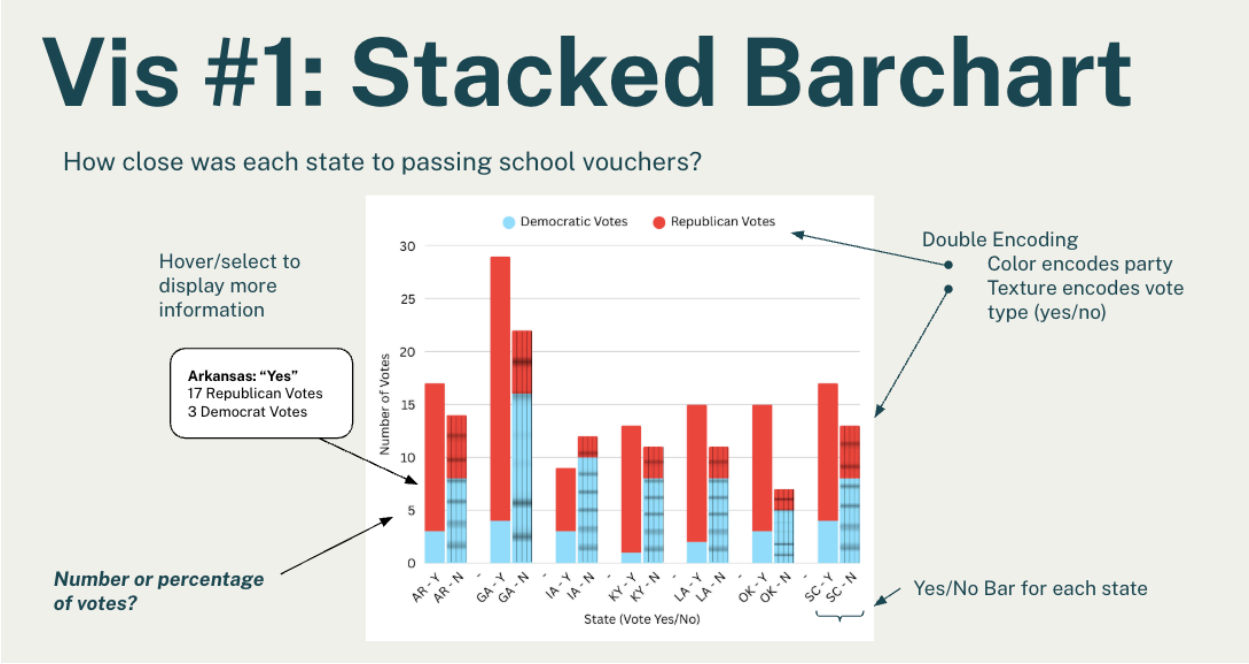
data visualization.

```
STATE,State abbreviation
PARTY_R,"0=Democrat, 1 = Republican"
YES,"0=Nay vote, 1= Yea vote"
PERCENT_RURAL,Percentage of rural constituents
MEAN_INCOME,Mean income by household for the district
PERCENT_PRIVATE,Percentage of total students that attend private school
PERCENT_21_W,2021 percent White population
PERCENT_21_B,2021 percent Black population
PERCENT_21_H,2021 percent Hispanic population
```

Paper Prototypes

We began our process by creating quick sketches of different ways to represent our data, which turned into medium fidelity prototypes created in Canva. These prototypes captured the tasks of comparing voting distribution across states, ranges in attributes such as income between districts that voted yes or no, and showed the order of attributes with the largest differences.

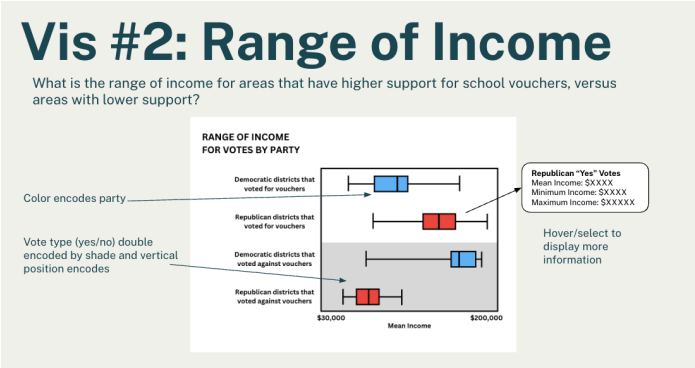
Visualization 1

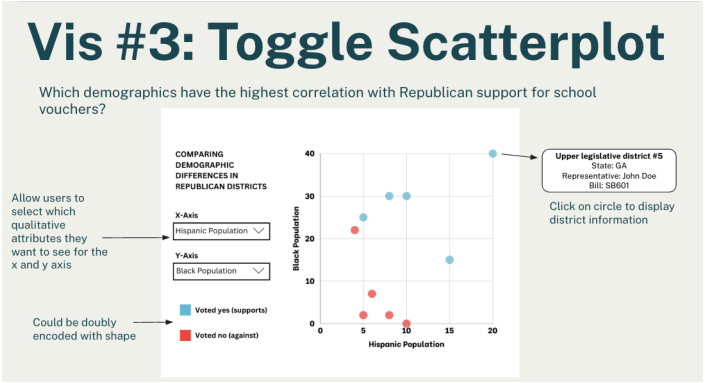


This prototype was initially designed to display the votes cast by each state and party. This prototype was adapted to best accomplish this task by adjusting the x-axis to separate bars by state and party, instead of state and vote type. Another adaptation made was switching the

y-axis to the percentage of votes instead of number, so each state can be proportionally compared. Texture is no longer used to double encode; but, we added a border color to the bars to encode the party.

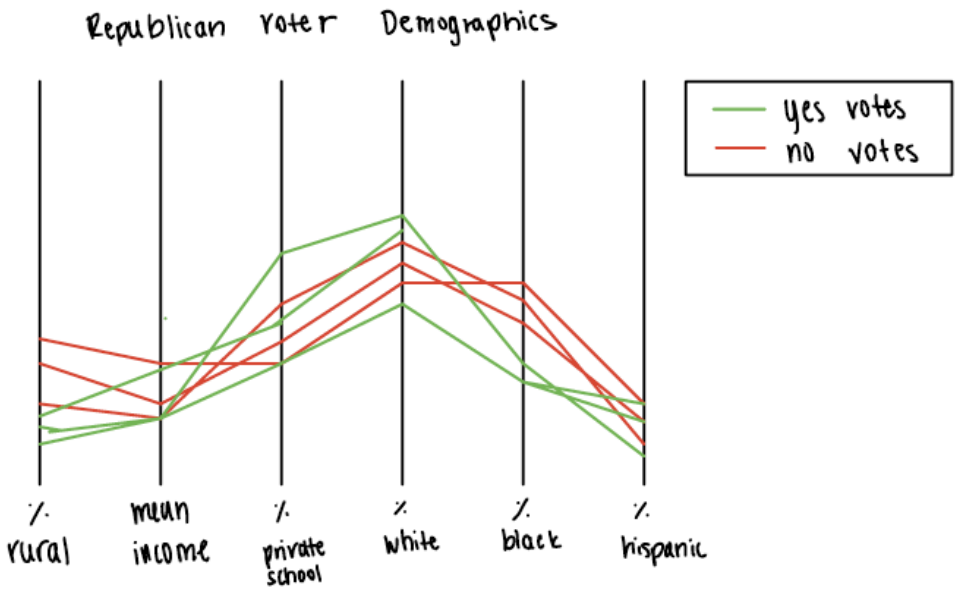
Abandoned Visualizations





Our original visualization 2 and 3 were abandoned and replaced by the visualization 2 seen below. The “range of income visualization” was intended to show the difference between average incomes of republican districts who voted yes versus no and democratic districts who voted yes versus no. We decided the democratic plots weren’t necessary since the focus is on differences among republican districts. Since we didn’t need democrats, income could be represented with the toggle scatter plot instead. The “toggle scatter plot visualization” was meant to show interaction between two attributes at a time. However, the visualization is mostly meant to be exploratory and we worried users would miss important interactions depending on what x and y axes they set. Looking for a little less user input but still the ability for users to compare, we turned to the concept of parallel coordinates, as mentioned next.

Visualization 2



This prototype was designed to display the relationship between various socioeconomic factors and the casted votes within the Republican party, in order to investigate the disparity between political parties as seen in visualization one. Initially, individual items (lines) represented an individual district's vote. To reduce visual noise, all districts were aggregated; solid lines denote the average, dotted lines denote median, and area denotes the interquartile range of all districts, still grouped by vote type. Next, a drop-down menu was added to filter by state, and a second view of the parallel coordinates was implemented so users can directly compare the disparities across states. A toggle was also added to display the information for the mark that the user is hovering over.

Task & Accessibility Analysis

1: Compare

Across all states, which party was more unified in their vote for or against school vouchers?

Task Analysis

To find the party that is more unified in their vote, across all states, the user looks at the x-axis to find the two bars denoting "Total Republican" and "Total Democrat." The user compares the height of the stacks of the "yes" and "no" bars to find the stack with the greatest value (greatest

percentage). The user determines that the democratic party is more unified, with 100% of total votes being “No.”

Accessibility Analysis [Screen-reader]

The user navigates/commands the screen reader to the chart title. Uses arrow keys to move through the chart elements, with the screen reader announcing each element's label and data. When reaching the x-axis, the screen reader announces "Total Republican" and reads the corresponding percentage values: "Yes: approximately 65%, No: approximately 35%." Continuing navigation, the screen reader announces "Total Democrat" and reads "Yes: 0%, No: 100%."

2: Range

What is the range in the percentage of Republicans across states who broke party lines and voted no?

Task Analysis

To find the range in percentage of Republicans who broke party lines (voted no) across states, the user looks at each bar to find the Republican (red-bordered) items. Among these bars, the user finds the shortest and tallest “No” stacks. The user determines the republican group with the smallest percentage of “No” (AR) votes, and conversely, the republican group with the greatest percentage of “No” (OK). votes. Using the tooltip, the user can determine the specific percentage for each, and carry out the required calculations: 44%-5% = **39%**.

Accessibility Analysis

To find the range in percentage of Republicans who broke party lines (voted no) across states, the screen reader user navigates through each state's data using accessible commands, with the screen reader announcing the state abbreviation and Republican voting percentages. The user identifies Arkansas (AR) with 5% "No" votes as the smallest percentage and Oklahoma (OK) with 44% "No" votes as the largest percentage. The user calculates the difference between these values: 44% - 5% = 39%.

3: Order

What is the order of demographic variables with respect to the greatest possible effect on voting disparity?

Task Analysis

To find the order of demographic variables with respect to their greatest possible effect on voting disparity, the user scrolls to the parallel coordinates view. The user can hover over each variable to find the difference between “Yes” and “No” groups. The user can determine through the tooltips and increasing visual disparity between the average lines, that variables increase in greatest possible effect on voting disparity as the x-axis increases. The user returns the order of the variables as they are given on the x-axis.

Accessibility Analysis

To find the order of demographic variables with respect to their greatest possible effect on voting disparity, the screen reader user navigates to the parallel coordinates view using keyboard commands. The user systematically moves through each demographic variable on the x-axis, with the screen reader announcing the variable name and the average values for the "Yes" and "No" voting groups, as well as the difference between the two groups. As the user progresses along the x-axis, the screen reader indicates increasing differences between these values. The user notes that the variables are ordered by increasing effect on voting disparity as they appear on the x-axis, and reports this original sequence of variables as the answer.

Piloting

At this point in the design process, we had implemented mostly-complete visualizations based on sketches 1 and 2 within a scroll-style visualization format. The scroll function gave us the ability to explain the data and usage of our visualizations at each step, as well as create a flowing story with the data.

Unset

Speaker: “We are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don’t need to worry about being “right” as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole pilot session should take under 5 minutes. Do you consent to participate?”

[Wait for yes]
Speaker: “Thank you for agreeing to participate. We will start with the three tasks. Please ‘think aloud’ as you complete the task, meaning voice what you are thinking as you work through the task. Your first task is: _____”
* You may answer clarifying questions, such as “what do you mean by x?” or “are you looking for an exact number or an approximation?”
[Pause to allow the participant to complete the task]
[Repeat question with tasks 2 and 3]
Speaker: “That is the end of the third task. For this last bit, we welcome any feedback you may have about the visualization or about your process for completing the tasks.”
[Allow participant to speak first, then informal discussion]

Task Completion

Task 1: Across all states, which party was more unified in their vote for or against school vouchers?

We asked this question at step one of the scroll, when the stacked bar chart was visible. We gave the user a hint that they could scroll if necessary.

- User asked for clarification on the meaning of a party being “unified in their vote”
- User found the two answers with no problem upon clarification
- User noted the helpful text-callout as they scroll, noting the answer is highlighted

Task 2: What is the range in the percentage of Republicans across states who broke party lines and voted no?

We asked this question when the user was at step 3 of the scroll, viewing the highlighted bar for total republican votes, which the user had arrived at to answer task 1.

- The user scrolled to see if the next step would highlight the information for the question
- The user landed on the corresponding view, highlighting the republican bars.
- The user found the correct items and information, calculating the correct range.

Task 3: What is the order of demographic variables with respect to the greatest possible effect on voting disparity?

We asked this question at the first step of the parallel coordinates visualization.

- The user scrolled to view of parallel coordinates
- The user explored dropdown menu
- The user was able to determine the order by looking at the order of x-axis variables.

Participant Feedback

- Difficulty determining order within parallel coordinates (task 3).
    - At this point in our design process, the attributes of the parallel coordinates that made up the x-axis were not sorted by order of difference. Instead they were grouped by race-based demographics and non-race-based demographics. Thus, the participant had to scan the whole graph and even do some mental math to estimate which attribute had the greatest differences in averages between “yes” and “no” votes.
- Clearer wording
- Reduce tooltips, some information appears multiple times, unnecessarily
    - At this point in our design, the parallel coordinates visualization had a tooltip that showed information when the user hovered on the “yes” or “no” vote region and a separate tooltip that would appear when the user hovered over the x-axis titles, displaying the information for “yes” and “no” votes.
- It is difficult to see both parallel coordinate visualizations at once because it requires zooming out.

Next Steps

- Reorder parallel axes based on order of disparity (order based on the difference in average value for each group)
- Embed more concise, understandable language and explanations in the scrolly sidebars
- Remove tooltip for items in parallel coordinates to reduce visual clutter, move all information into the tooltip triggered by hovering over the x-axis

After making these changes, we presented our completed visualization to three more students to test the usability of the visualization.

User Testing

We used the same prompt on these new users as we did for our initial participant feedback. Each participant was asked to complete each task: compare, range, and order, while we observed. Overall, the users were able to successfully complete the tasks but asked for clarity on some wording. The parallel coordinates graph can take some time to acclimate to because of its size and business, so it would be helpful to give users a “walkthrough” of it with the scroll sidebar of text. One user did not initially realize the visualization was scrollable, but the others noticed immediately.

User	Major/Minor	Task 1	Task 2	Task 3	Feedback
1	Econ	No issue	Did not scroll	Explored different	Group bars

	/Applied Math		to determine the answer, relied on tooltips and personal knowledge	state views before finding disparity	for each state so they are visually grouped (vis 1)
2	Art History /Comm.	Asked for clarity on technical terminology	Asked for clarity on technical terminology and approach	Asked for clarity on the visualization meaning, and was able to complete the test successfully upon understanding th ea	Add more (digestible) explanations of technical terminology, such that the tasks an be completed more independentl y. More harmonious design.
3	English /Hispanic Studies	Scrolled through website before returning to vis 1, no issue	No issue	Asked for clarity “greatest possible effect on voting disparity”	Revise for clarity

Next Steps

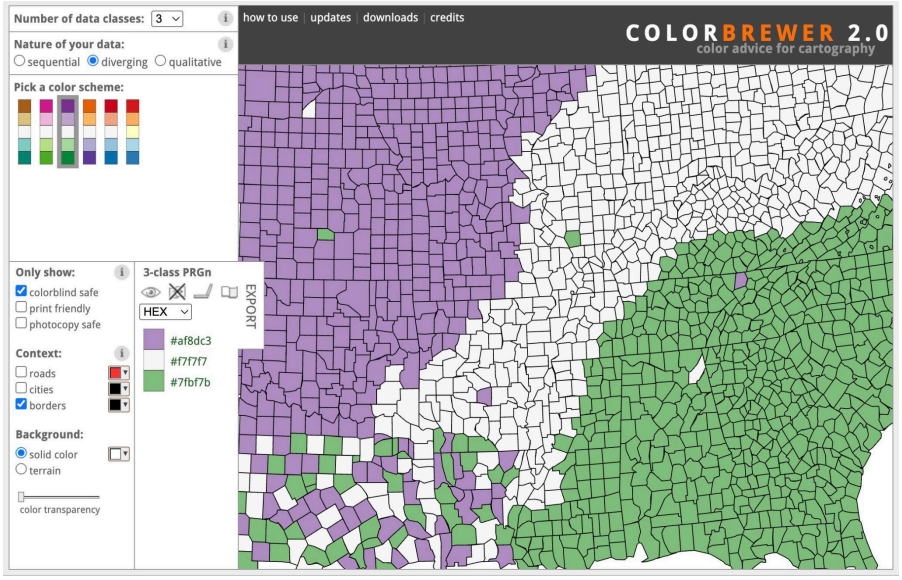
- Add more levels of detail within scrolling steps to contextualize and clarify terminology
- Redesign color palette to appeal to more accessibility needs (such as color blindness)
- Adjust the range of axes in parallel coordinates to scale for all percentage-based variables. This way, the user can more easily find the growing disparity, visually.
- Add more detailed ARIA labels for screen reader accessibility
- Resize the visualization components to fit in the user’s window, specifically so that both parallel coordinates can fit on the screen without zooming out.

Testing accessibility checks

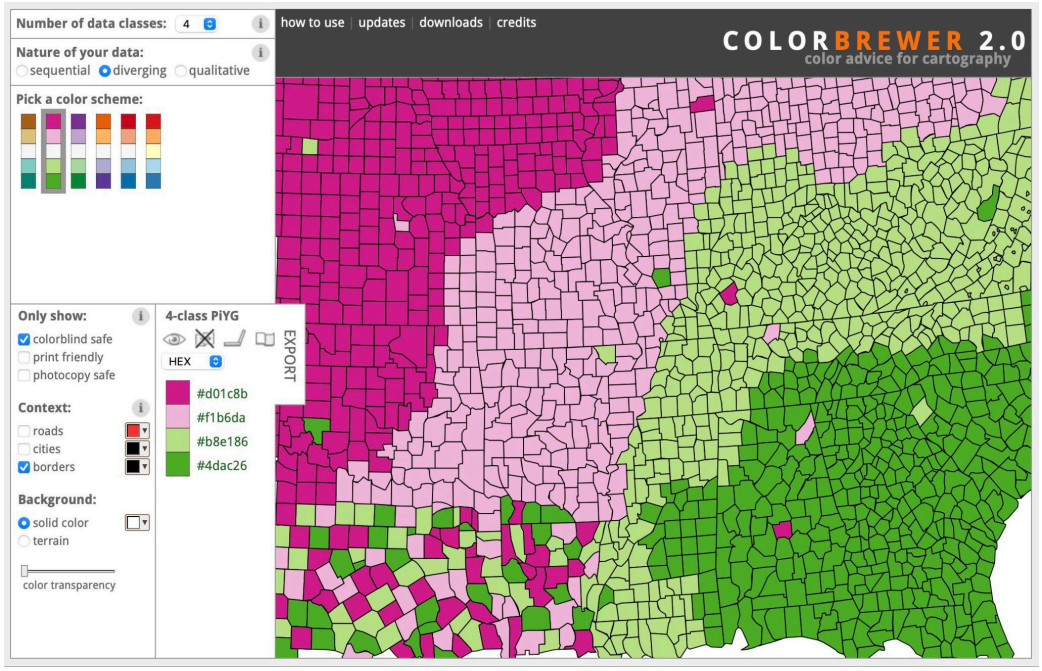
Throughout the design process, we were conscious of accessibility features, while acknowledging that we are still learning more about designing accessible visualizations. We did not have a participant in user-testing who needed to use the accessibility features, but we tried our best to imagine the experience for someone in the accessibility space. Below are the initiatives we took to make our visualization more accessible.

Colorblind-safe

One accessibility check was to ensure our visualization was color-blind safe. This was a bit difficult on the first visualization, the stacked barchart, because we wanted four colors to encode “Republican”, “Democrat”, “Yes”, and “No.” Since party affiliation was also labelled on the x-axis, we focused on ensuring the “yes” and “no” colors were colorblind safe. We chose purple for “no” and green for “yes”on ColorBrewer. Then we encoded the parties with red and blue because we felt that was most intuitive to users.



For the second visualization, the parallel coordinates, we chose a colorblind safe pink and green combo on color brewer and represented “yes” and green and pink as “no.” We felt green intuitively makes sense for “yes,” and since pink is close to red, it also makes intuitive sense for “no.”

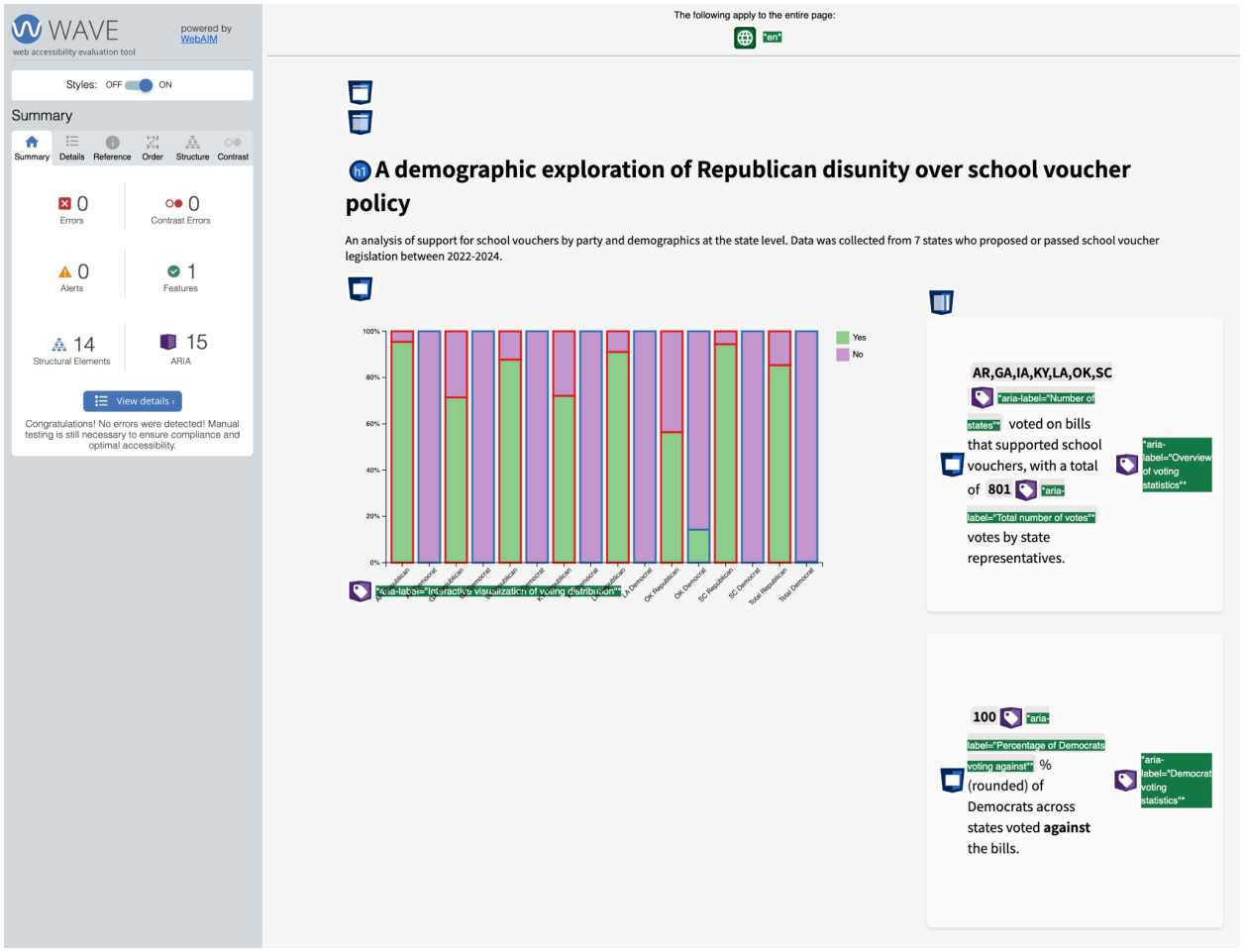


Screen Reading - ARIA

We improved the information for screen-readers in our visualization. We added ARIA labels in our html file to describe the graphs being shown and we ensured the language in the html was set to English.

WAVE Test

We ran the WAVE plugin to evaluate our visualization for accessibility. It did not report any errors or alerts.



Final Visualization

Here is a link to our final visualization:
https://maddiecreed.github.io/finalproject/index.html

The final visualization includes a stacked bar chart to show the differences overall and by state between republicans and democrats and those in each party who voted “yes” or “no” for school vouchers. We include highlight features to guide users to important trends in the graph. Next, there is a parallel coordinate visualization that allows users to explore differences in demographics between Republican districts with a “yes” vote versus a “no” vote. These demographics are in no way all the factors that could be at play in voting behavior, but they are an interesting start to further questions and research. There are interesting differences between states as well, which we allow users to explore by positioning two parallel coordinate visualizations beside each other for comparison. With more time to troubleshoot, we would have scaled the visualization to responsively fit different screens and zooming (based on Lab 8). We would have also liked to continue to explore the trends seen in different subsets of the data, but unfortunately, we should not make our user scroll forever.

Personal Reflection

Overall, I really enjoyed working on this project, despite some difficulty. Having just come off my capstone last semester, I found myself wishing I’d had this class earlier, especially when it came to designing dynamic visualizations with multiple views. Back then, I built a mostly on instinct and less formal testing/feedback, but the process we followed for this final project made everything much more coherent and intentional. Rather than only thinking about what looked good or felt intuitive, we moved through each design decision with purpose, and the result was not only more effective but also far easier to execute.

The collaboration also came together naturally. Early on, I struggled to fully grasp the dataset. My partner, who had worked with it before, was really helpful in breaking it down in a digestible way. That background knowledge helped me generate new ideas for different visualizations that complemented hers. It felt like a genuinely collaborative process; my more distanced view helped provide new perspective for my partner, which is how we were able to go through various vis designs for the same tasks.

We were lucky to avoid any major roadblocks, but one of the trickier parts was writing the user tasks. We understood what we wanted people to do, but putting that into clear, non-technical language proved harder than expected. That challenge was actually what led us to adopt the scrollytelling format. It gave us space to provide users with context and gradual guidance, which really helped with clarity.

Given more time, I would have liked to incorporate additional details and accessibility features. For instance, we had plans to group the bars by state in the first visualization, using horizontal position to encode that information. Before taking this class, I probably would have spent too much time on tasks like that early in the process. But through this project, we learned how to prioritize effectively within our deadlines.

This project— and the course as a whole— has been one of the most meaningful parts of my academic and professional journey so far. I feel far more confident navigating the end-to-end design process, and I know that the skills I've gained here will be essential as I move forward in my career.