Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

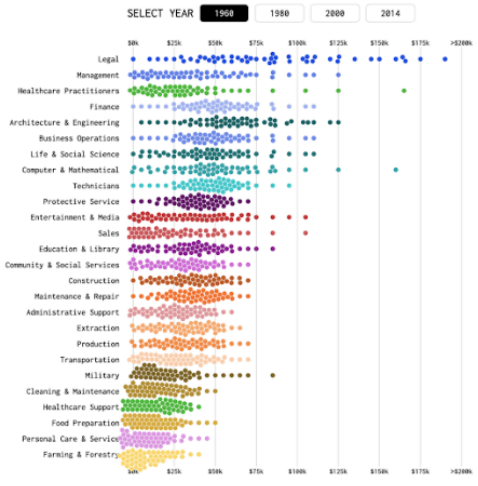
Group 4 Process Book

Overview and Motivation: Provide an overview of the project goals and the motivation for it. Keep in mind that this will be read by people who did not see your project proposal.

The overall goal of this project is to use the data found in the Religious Landscape Study (RLS) to answer the question: Do people who identify themselves with a certain political party align with the classical party views or are their views more influenced by their education or socioeconomic status?

The Religious Landscape Study (RLS), is a national cross-sectional survey conducted for the Pew Research Center by NORC at the University of Chicago. It is a survey that paints a religious portrait of the US and allows the Pew Research Center to examine the religious identities, beliefs, and practices of US adults. Since the U.S. census does not ask Americans about their religion, there are no official government statistics on the country's overall religious composition. Moreover, nongovernmental surveys that ask Americans about religion typically include fewer questions than the RLS, or far fewer respondents, or both. The RLS aims to fill these gaps by employing detailed questions and a large sample size in order to capture representative data about numerous religious traditions.The survey was conducted in English and Spanish from July 17, 2023, to March 4, 2024, among a nationally representative sample of 36,908 U.S. adults.

Related Work: Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.



Our project was inspired by several sources and experiences throughout the course. The class discussions and homework assignments, particularly the one where we analyzed visuals related to job and income levels, helped us understand how to effectively

communicate insights through data visualization. These exercises guided our thinking about which visual formats best convey relationships between variables.

When our group first met, we spent a significant amount of time exploring different datasets to find one that aligned with everyone's interests. We reviewed datasets on topics such as energy consumption, healthcare, and politics. Eventually, we discovered a comprehensive dataset that included political, educational, and demographic (age-related) information. We chose this dataset because it offered a broad range of variables that allowed us to explore multiple dimensions of social data and find meaningful intersections among them.

Our prior assignments and in-class discussions on visualization design principles were instrumental in shaping how we approached presenting our data. These experiences influenced our decisions regarding the types of charts, comparisons, and analytical questions we pursued in our project.

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?

We initially looked into answering the following questions:

- Is one gender more likely to convert to their spouse's religion?
- Is the quality of life tied more directly to religious involvement or financial level?
- How closely do individuals who identify with a certain party align with the classical party views?

After looking through our data, we realized it would be difficult to pursue the first question because the 'current religious identity' responses of people who did not identify themselves as not married were not recorded. This resulted in our dataset being cut in half, so we decided to try to answer a question that would include as many data points as possible.

While looking through our data, we had some difficulty comparing the different categorical variables needed to answer these questions, as the different variables had a different number of response options. Some of these variables had 3 response options, while others had 4 or 5, and there is no accurate way of comparing categorical responses with different scales. So, we decided to look into answering the final question.

Finally, we considered the third question. We had learned that there may be problems when questions have different response options, so we selected questions that were answered with "better, worse, or no difference". Before we decided to look at how socioeconomic status or education level would influence response to political questions, we initially wanted to investigate if people's responses would be more influenced by their religious affiliation or their age group. However, when attempting to plot the first visual, having age groups or religious affiliation resulted in a very large visual that would have made it hard to meet the requirements of three visuals on the screen. The size of the visual was influenced by the many options in these variables.

We finally decided to only include socioeconomic status, political party, and education level to answer the final question. When looking at our data, we noticed that we ended up with 30,258 points after cleanup which indicates that we will not lose a lot of information.

Data: Source, scraping method, cleanup, etc.

- Started from Pew Center Research Religious Landscape Survey
- After making an account, the publicly accessible survey responses and codebook can be downloaded as .csv files.
- Scrubbed dataset to remove replicate ranked weights along with data that is inaccessible through the public database (location, specific birth year, children age). This significantly decreased the size of the file from ~112MB to ~10MB
- Merged redundant datatypes in the csv codebook (no longer a variable name for each response label, only one per question)

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?

We initially attempted to use a scatterplot to visualize political party vs the questions indicating classical party views. However, both these features are categorical, so we ended up with clumps of data. When attempting to plot political parties against socioeconomic status or age, we also had the same issue.

We had not yet learned about force directed graphs, so we began only using aggregated visualizations to make our first visualization, although our ideal was creating something similar to what we saw in class.

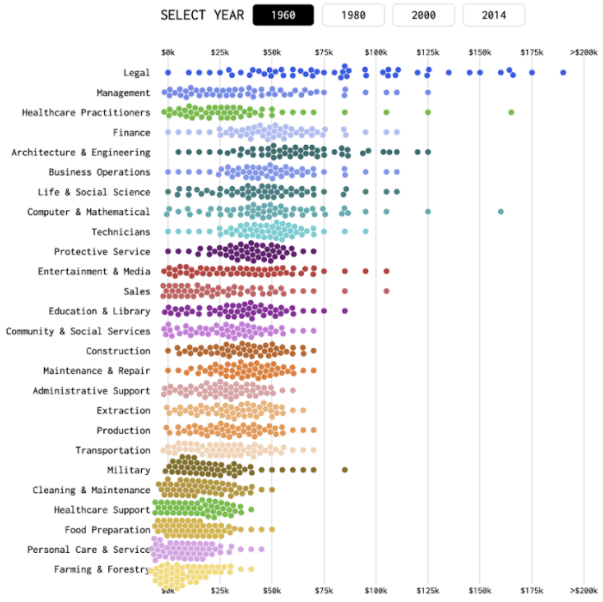
Despite only being able to look at aggregated data, we learned that we had a large number of Caucasian participants which gave us problems in the Sankey diagram later. We also

learned that there was an approximate 50/50 split between people who were married at the time of the survey and those who were not.

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?

Visual 1 Evolution:

Despite not knowing exactly how to re-create the graph shown below, we know we wanted to do something similar for our first visual. We wanted our first visualization to show the difference in responses to certain questions among the political parties/religious affiliation when we were considering it. We wanted the extraction of this information to be simple, and for the reader to understand the trends we observed. Through each iteration, we thought of ways to improve the design.

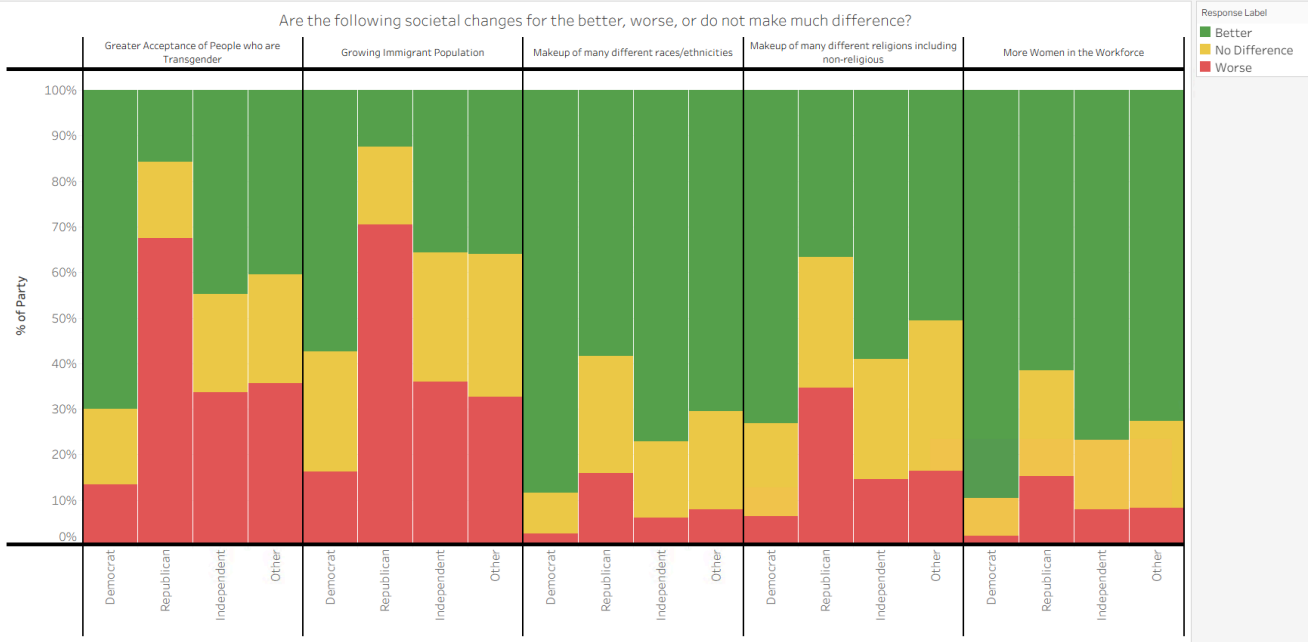


We were able to create the graph shown below which meets the requirement of having one mark per item. The x-axis is at the top of the graph and splits the graph into 5 sections, each corresponding to a question asked. The y-axis divides the graph into 3 sections that contain how many users responded with “better”, “worse”, or “no difference”. We believe a user could be able to extract information about how Republicans and Democrats view these topics relatively easily. However, users will probably not be able to gain information about how Independents and Other political

parties view these questions because their color indicators are drowned out by those of the major political parties. This visual does not make it possible to extract the relative percentage of people in a party with a particular response. Although we do not think the user needs to know, for example, "50% of Republicans think more women in the workplace is a good thing", the ideal visual should allow you to more accurately compare the percentage of people in a certain party with a specific opinion.
Despite this visualization meeting the requirement, we ended up choosing a different one for our proposal because you cannot easily gain information from 2 political parties by looking at this visual, and it is not visually appealing. An improvement that could be made to the graph is the relocation of the x-axis to the bottom of the graph. We could also improve the way the points are distributed, but at this point, we did not know how.

We then decided to create the following visual, despite it not meeting the 1 mark per item requirement, because we were able to see the distribution of response for all parties. The y-axis was the percentage of people belonging to a certain political party who answered with either "better, worse, or no difference". The actual answer was indicated by the color. There are two x-axes in this visual. The primary x-axis is on top of the graph and divides the overall graph into five questions, corresponding to the question being asked. The secondary x-axis is on the bottom of the graph and is further divided into the 4 political parties. Although this visual looks much better than the first attempt, it is a bit difficult to extract information about how the political parties respond to each of the questions because there is no one area to look at and compare 1 political party at a time. This graph provides a lot of certainty to the user because they are able to see the percentage of a particular party's responses. An improvement to the graph could be making the primary x-axis the y-axis since the certainty provided with percentages is not

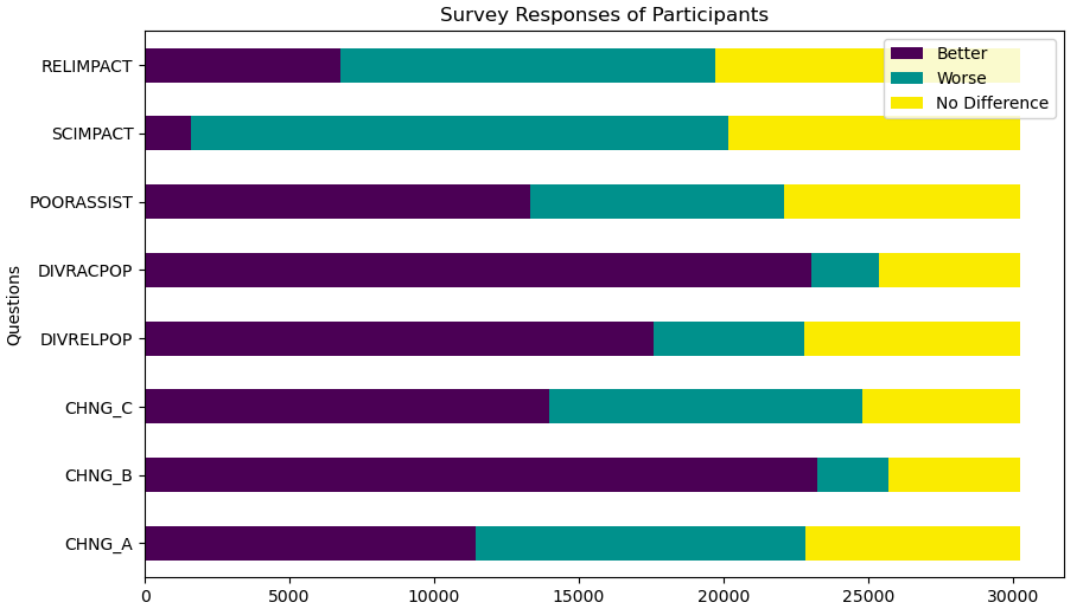
really needed. A method to also change the secondary x-axis and the legend could improve the graph.



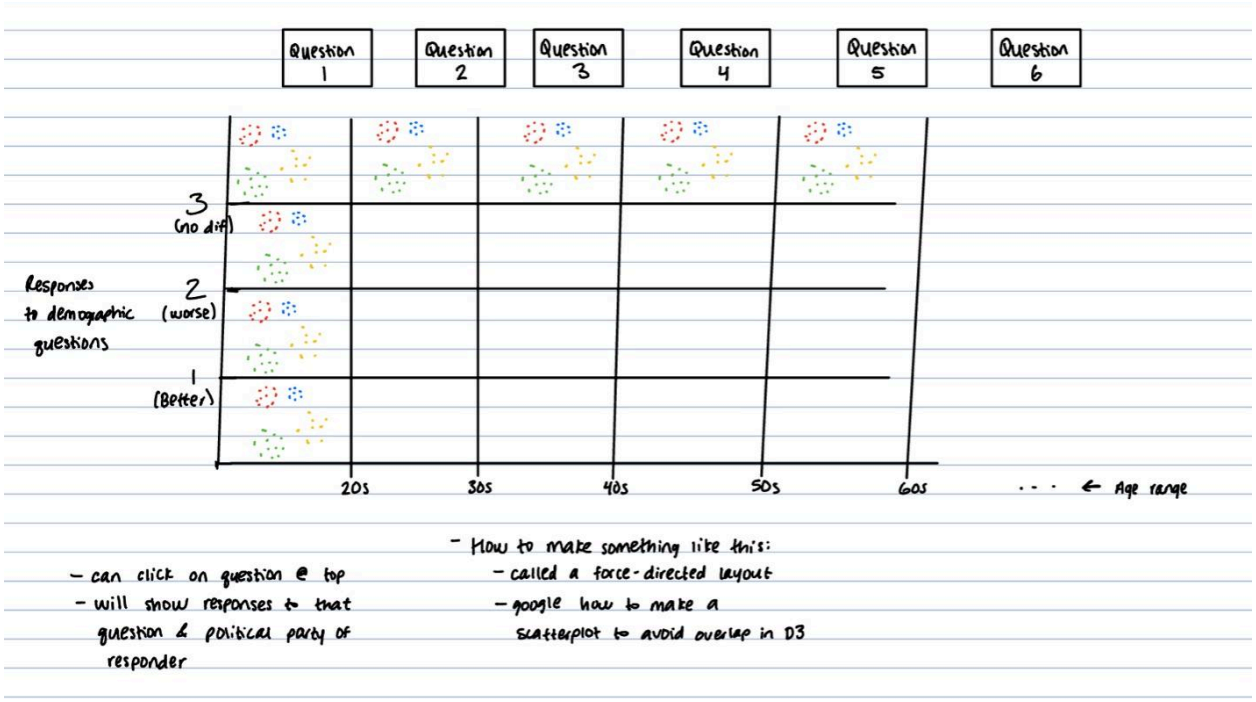
We also considered the visual shown below, which also does not show one mark per item. This visual had the x-axis show how many people responded with “better”, “worse”, or “no difference”. The y-axis was the different questions asked. This visual was made in Python, and we were not able to determine how to add information about the political parties of the respondents. It is easy to gain information about people’s responses, but there is no way to determine how these responses correspond to political parties, so the user cannot use this graph to answer our overall question. This graph provides a lot of certainty to the user because they are able to see the number of responses, but there is no certainty about which party those responses belong to. An improvement to this graph would be adding the option to filter by political parties, which would also require a change to the color scale and overall graph partitions.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025



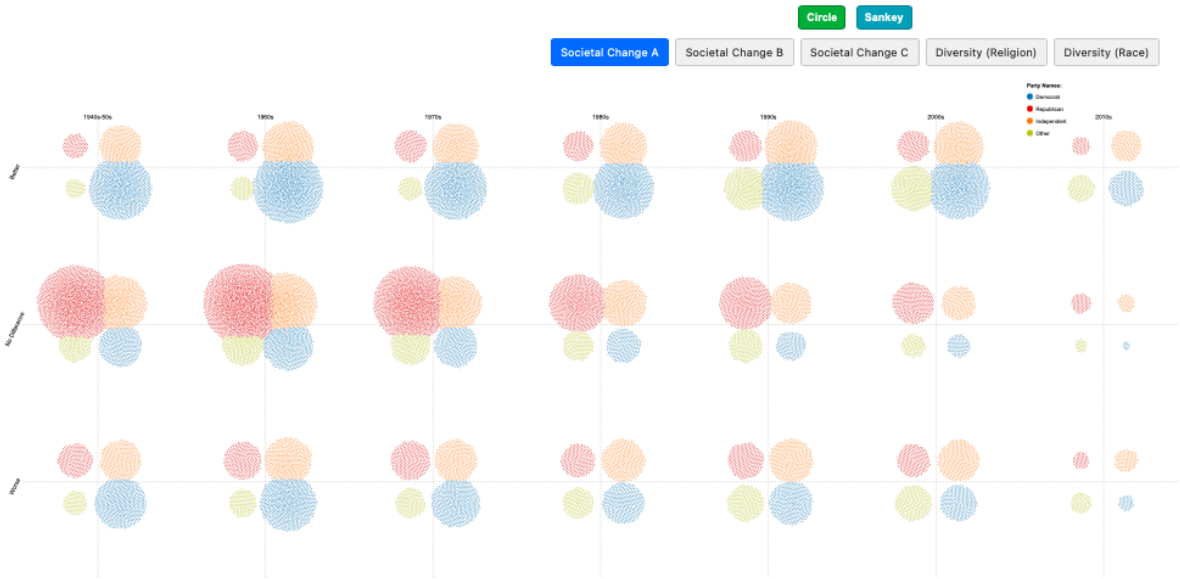
After receiving feedback from Milestone 2, we spoke to Dr. I, and he suggested creating a force directed diagram since the reason we had attempted to make bar charts for visual 1 was because of difficulties viewing all the points. A force directed layout would also enable us to easily click on each respondent and if we group our responses like indicated below, we can easily track a rough percentage of each party with a certain response.



Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

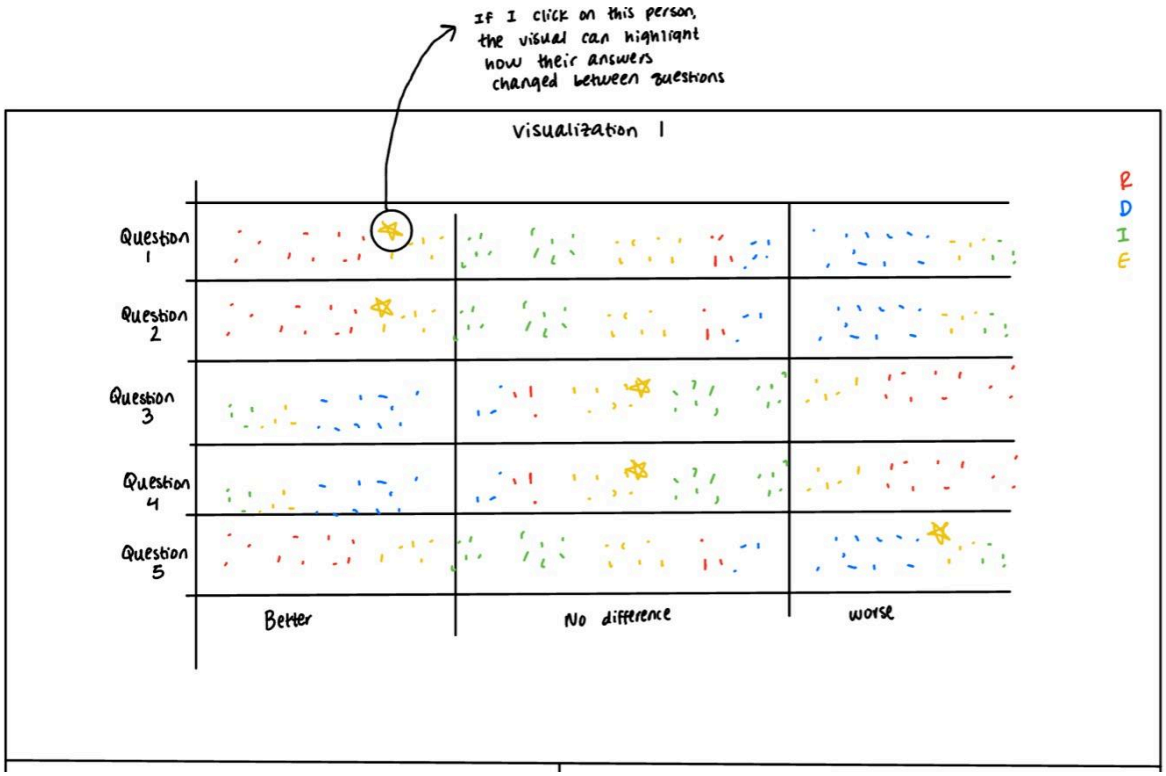
The discussion with Dr. I led to the creation of the following visual. This visual has tabs on the top that the user can click on to see the responses to the different questions. The y-axis is separated into the different response options, “better”, “worse” or “no difference”. The x-axis is divided into the different age ranges, as we were still considering including them at this point. The different colors correspond to the different political parties, as is also shown in the legend. The visual also has buttons on the very top that allow you to switch between this visual and the Sankey and Sunburst plot. It is very easy to gain information about how many people in each age range and political party respond to each question. Although we no longer have a way to determine the count or percentage of responses, the size of the circle encompassing all people of one political party in a certain age range increases as the number of people meeting these criteria increase. This allows the user to easily extract if more/fewer people of each party and age answer a certain question. The certainty provided by this graph is reduced because we are not numerically quantifying the number of responses, however, you are still able to make accurate conclusions. This visualization can be improved by moving the x-axis to the bottom of the graph and by moving the legend so it is outside the boundaries of the main graphing area. This graph also does not allow you to see the responses across different questions at one time because those responses require clicking and “re-creating” a new graph. The graph also emphasizes how responses change across different ages, but the main variable we were interested in investigating is political party. The visual is also really large, which is a reason we created separations for the different questions, but, it makes it very difficult to meet the requirement of having 3 visuals on the screen at the same time without scrolling.



Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

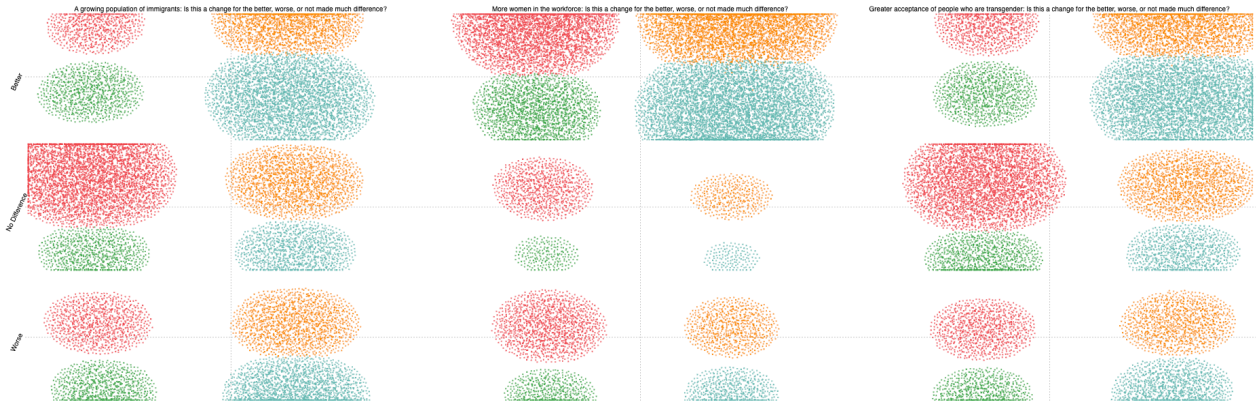
After discussing this visual with Dr. I, he suggested making changes so we can shrink down the size of the graph so 3 visuals can eventually fit. He also mentioned it may be more interesting to create a visual where we can potentially track how one respondent changes their answers to the different questions. This would not be possible in the visual above since you have to load a new graph to view the different responses. Therefore, we made a sketch of a new visual that would have the x-axis divided into 3 sections for the "better", "worse", or "no difference" responses. The y-axis would be the different questions, and the colors would indicate the political party of the respondents.



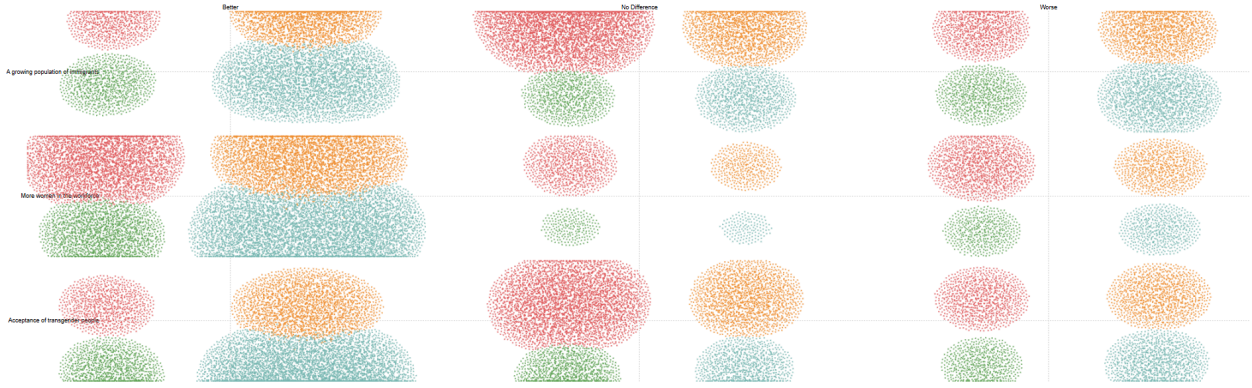
Our most recent update to visual 1 is shown below. The prototype version needed to be reduced due to the massive amount of points that would be included if all five questions, and every response to each were included. By reducing the number of questions and responses considered, the total number of marks was reduced from 150k marks, a value that lead to the website failing to load, to a more reasonable 60k marks, which still took some time to load and complete the force-directed simulation, but resulted in a much more reasonable output.

We believe the user will be able to easily gain information about what people from the different political parties think about these questions. Although there are no counts or percentages, the user will be able to determine if more respondents from a political party have a specific opinion

because the sizes of the circles/areas containing the number of responses corresponds to the number of responses. Our goal was not to leave the user with exact numbers, but a general understanding of trends, so this visual accomplishes that goal. The visual can be improved by making the text containing the questions and responses larger. Another improvement could be the swapping of the x- and y-axis.



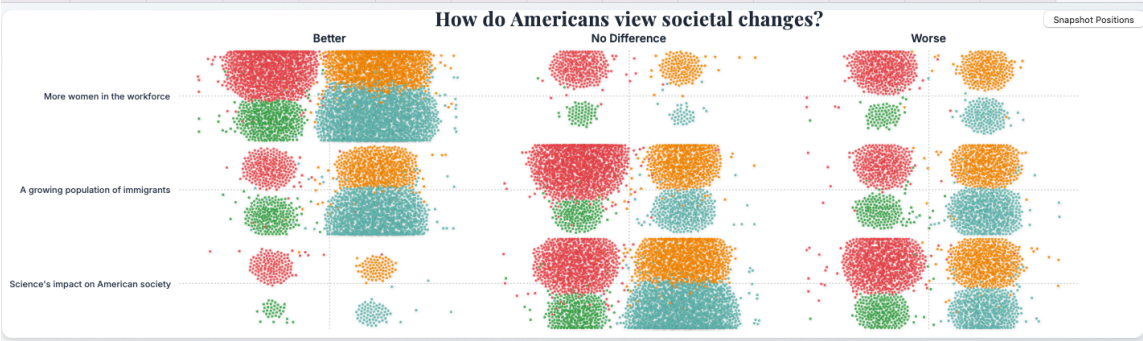
Swapping the X and Y axis was also attempted, such that the visualization matched the suggestions from Dr. I.



From the feedback received from the class and from our own testing, it became clear that the force directed diagram took much longer to settle than originally thought. This led to significant delays in every interaction attempted to be made, and decreased the overall effectiveness of the visualization. To counteract the significant computational burden required by the >50k responses and associated marks, the dataset was subsampled, and a ‘snapshot current position’ button was added. This allows the user the capability to either run the classic force directed diagram and wait for the marks to settle, or skip the force simulation entirely and use the positions obtained from the ‘snapshot current position’ button. This meant the simulation would not run in the

background while the visualization is being interacted with, and overcame the performance bottleneck.

Final Visual 1:



Our final visual 1 includes the addition of a title and larger text on the horizontal and vertical axis headings for clarity. We also switched the location of the women in the workforce and immigrant question because the distribution of points seems to go from most to least in the 'Better' column, and then switch in the 'No difference' column for these questions. We hope the user will be able to pick up on this difference as they use our visualization.

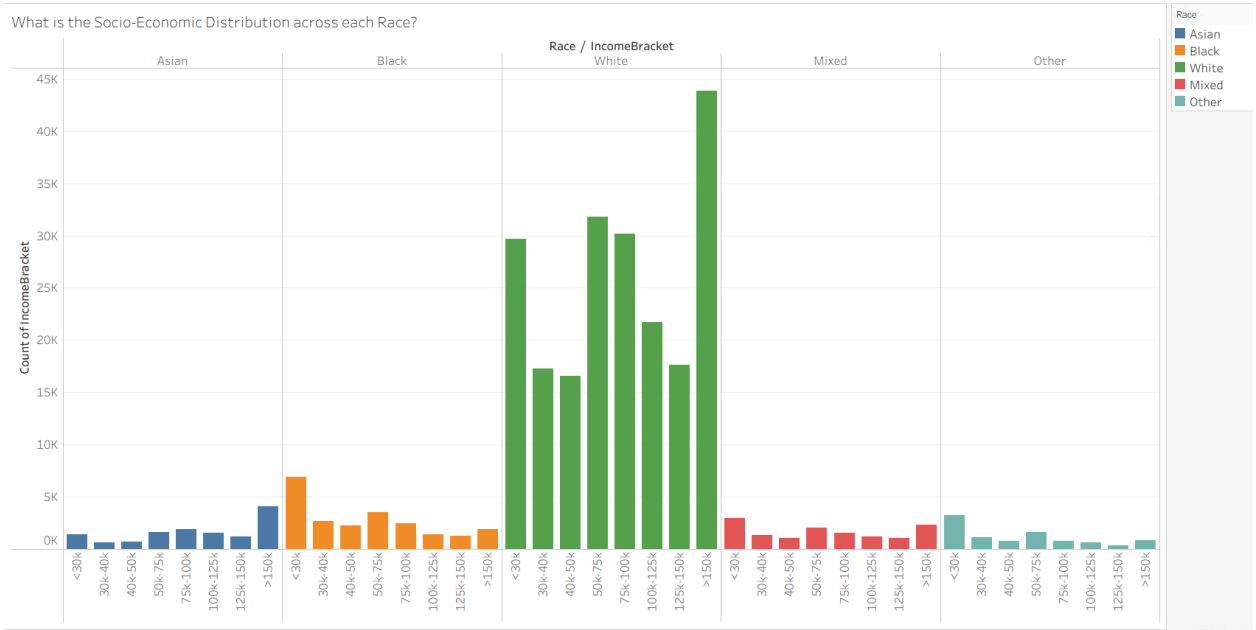
Visual 2 Evolution:

At the beginning of our visualization creation, we were interested in being able to view the changes in responses based on religious affiliation. Our x-axis was going to be the different religious affiliations, and the y-axis was going to be the percentage of people from a certain birth decade. Our visualization was going to look like the one sketched below.

We began attempting to create this visualization, but we realized that there were many Christian respondents and not that many in other religious groups. We decided to pivot

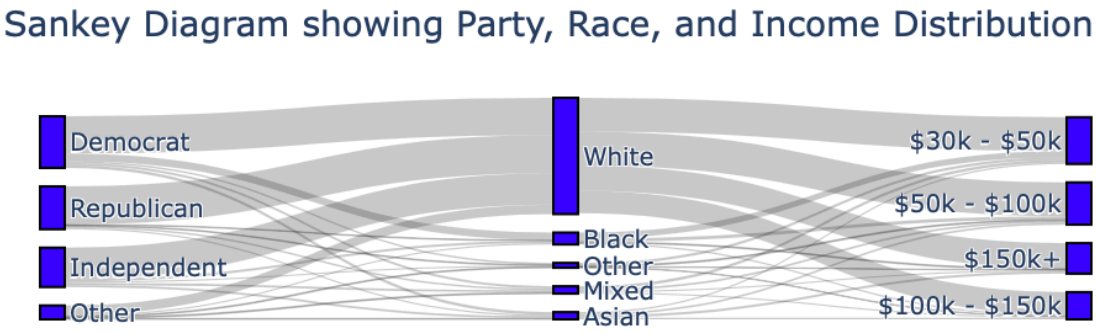
and change our variables of interest to race and socioeconomic status, since there was a varied distribution for the income question especially. Therefore, we wanted to create a visual that would show how different socioeconomic groups vary amongst the political parties.

Our first attempt to create a visual that uses socioeconomic status and race is shown below. The primary x-axis is the racial divisions, which are also indicated by color. The secondary x-axis is further divided into 8 socioeconomic groups. The y-axis shows the number of people belonging to each bar on the x-axis. We are able to gain information about how many people from each race belong to each socioeconomic group. It is relatively easy to gain this information, and a user can be fairly certain about this information since the y-axis shows the count. The visual can be improved by titling the x-axis on the bottom to improve readability. There are also too many people in the white section, this could be due to some people belonging to multiple races or the lack of a Hispanic category, especially because the survey was conducted in the US. This graph also does not have any information on political parties, so inclusion of that information would improve this visual.



The issues encountered during the creation of the first visual led to a different approach. We decided to make a Sankey diagram that would allow for an easier flow of information from political parties to income distribution.. There are 3 branches that correspond to political party, race, and socioeconomic status. It is a bit difficult to extract information from this graph because the White race section is significantly larger than all the others, so the lines flowing out of these

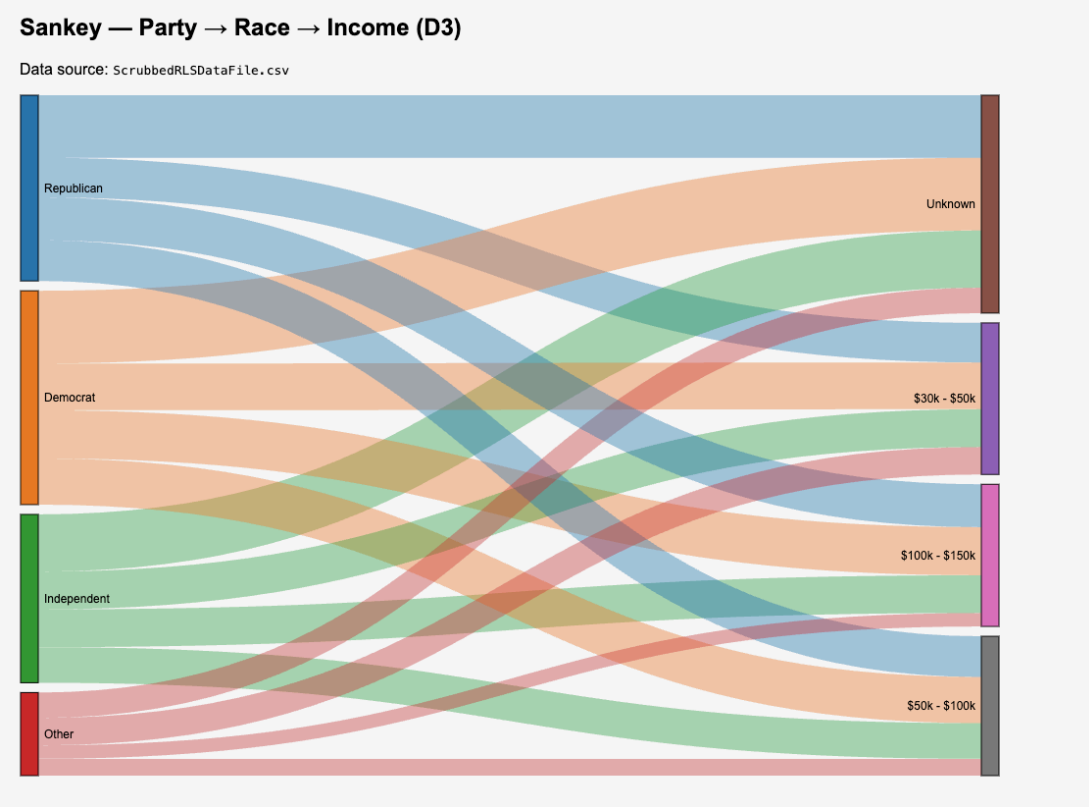
race branches are basically useless. The user is really only able to view the political party and socioeconomic status of White people, which significantly reduces usability of this graph. It is only easy to extract information from the White category. The user can only be certain of information gained from the White category because the thickness of the lines is easy enough to follow. The user would not be able to easily gain information or be certain about information from other race categories. For the white category, there are no numbers being explicitly shown, but a user can look at the distribution and determine which socioeconomic status has more/fewer people. An improvement to this visual could be re-sizing the other races to enable information gain and certainty.



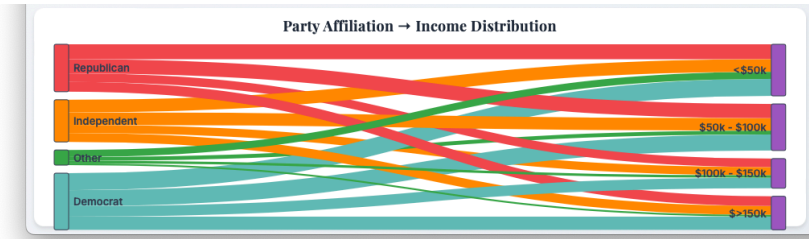
The problems encountered in the creation of visual 2 led to the final form of the visualization shown below. We removed the race branch since there was a very large number of people who identified as White, and this caused the other branches to become small and lacking information and certainty. This visualization only has 2 branches, political party and socioeconomic status. The independent and other political parties have a lower number of respondents, but it is still relatively easy to view how many people in each political party belong to a certain socioeconomic group. It is easy to extract this information, and a user can be certain of the information they gain. This visualization can be improved by reconsidering the color palette selected since some of the Independent links are hard to follow. The axis labels can also be increased in size and maybe placed outside the main graph area.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025



Final visual 2:



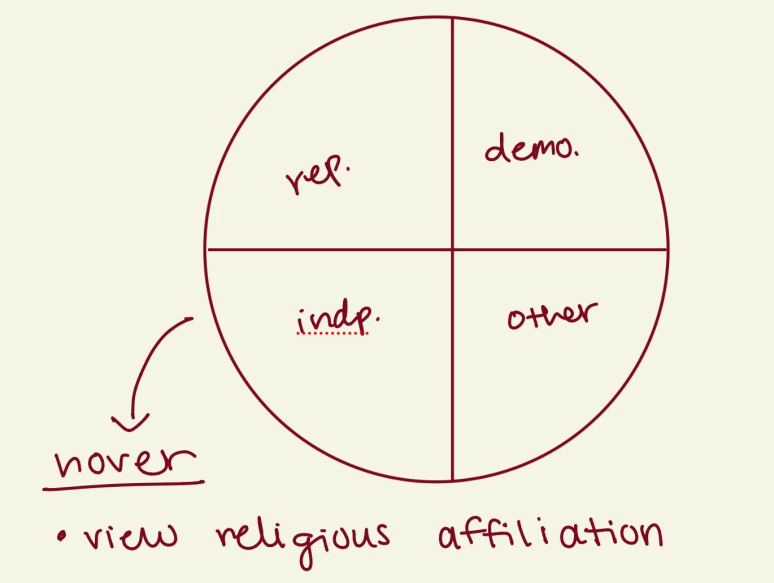
Our final changes included changing the title of the visualization and correcting the ‘Unknown’ error in previous versions. The dataset being used was not the final cleaned version, so some rows with blanks were still present. Updating the file we pulled from fixed this issue completely. We also selected one color as the color to represent all income levels. It is not generally recommended to have more than 8 color channels, and we were right on the border with the previous iteration. The main difference we wanted the user to be able to track visually was political party, so we decided to make all income levels the same color, purple, so that the user would clearly see that the information contained in all the purple rectangles is related, but it not necessarily indicative of political party.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

Visualization 3 Evolution:

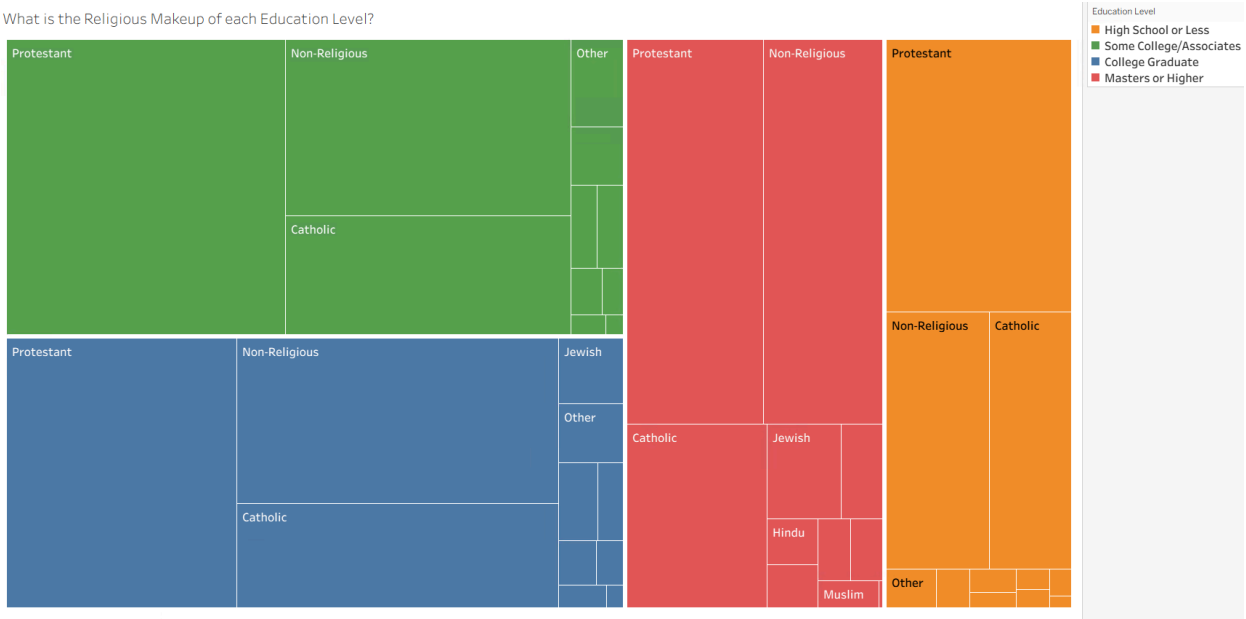
At the beginning of our visualization creation, we were interested in being able to view the changes in responses based on education level and religious affiliation. We wanted to create a visualization similar to the one sketched below.



Our initial idea was to create a Stock Market Style Block/Tree Diagram. This visual used color to distinguish the education levels, and each color was subdivided into the different religious affiliations. The user is able to gain information on education level and religious affiliation, but it is not possible to gain any information about political parties. The graph is a little difficult to use, because the religious affiliations with fewer members are not labeled so the user does not know what those sections on the graph indicate. Although there are no numbers, a user can be fairly certain that a certain religious affiliation has more/fewer members, but only if that section is labeled. There is no way to gain information or obtain certainty about the religious affiliations that are not labeled. The graph could be improved by ensuring all religious affiliations are labeled.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

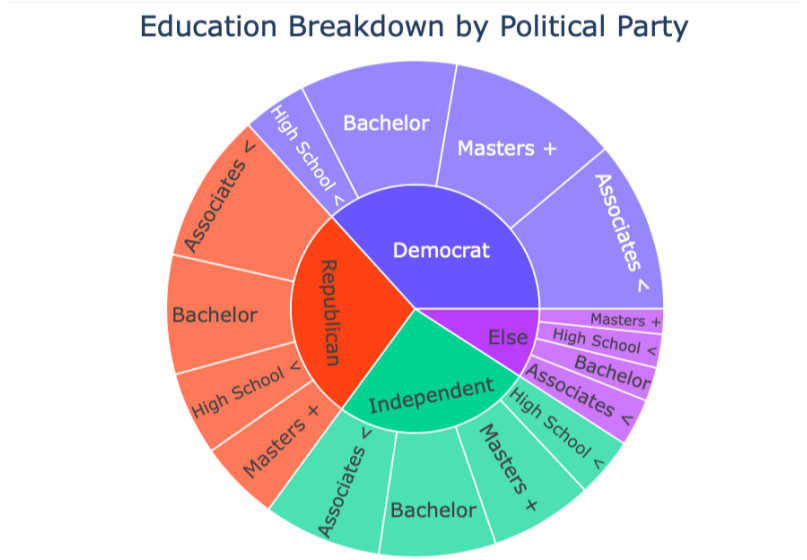


After the difficulties experienced while creating the first version of the visual, we decided to remove religious affiliation since we are mostly interested in information regarding political parties.

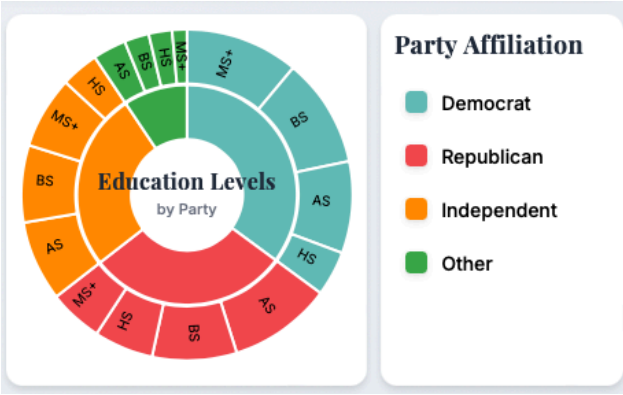
The final version of visual three is a sunburst plot that shows political party and education level information. The inner pie chart contains the different political parties. The outer pie chart contains the number of people in those parties with a certain education level. It is relatively easy to gain information from the graph below. There are no numbers, so certainty is not as great for the smaller categories. This graph can be improved by adding labels to the inner pie chart. It is a bit difficult to determine if the Republican or Democrat section contains more responses, which is an unfortunate feature of pie charts since we now have to look at angles. The visual automatically ordered the outer pie by fewest to most responses, so once we determine which political party overall has more respondents, we can easily determine the education breakdown of each party. However, this feature could also be improved by the addition of labels.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025



Final Visual 3:

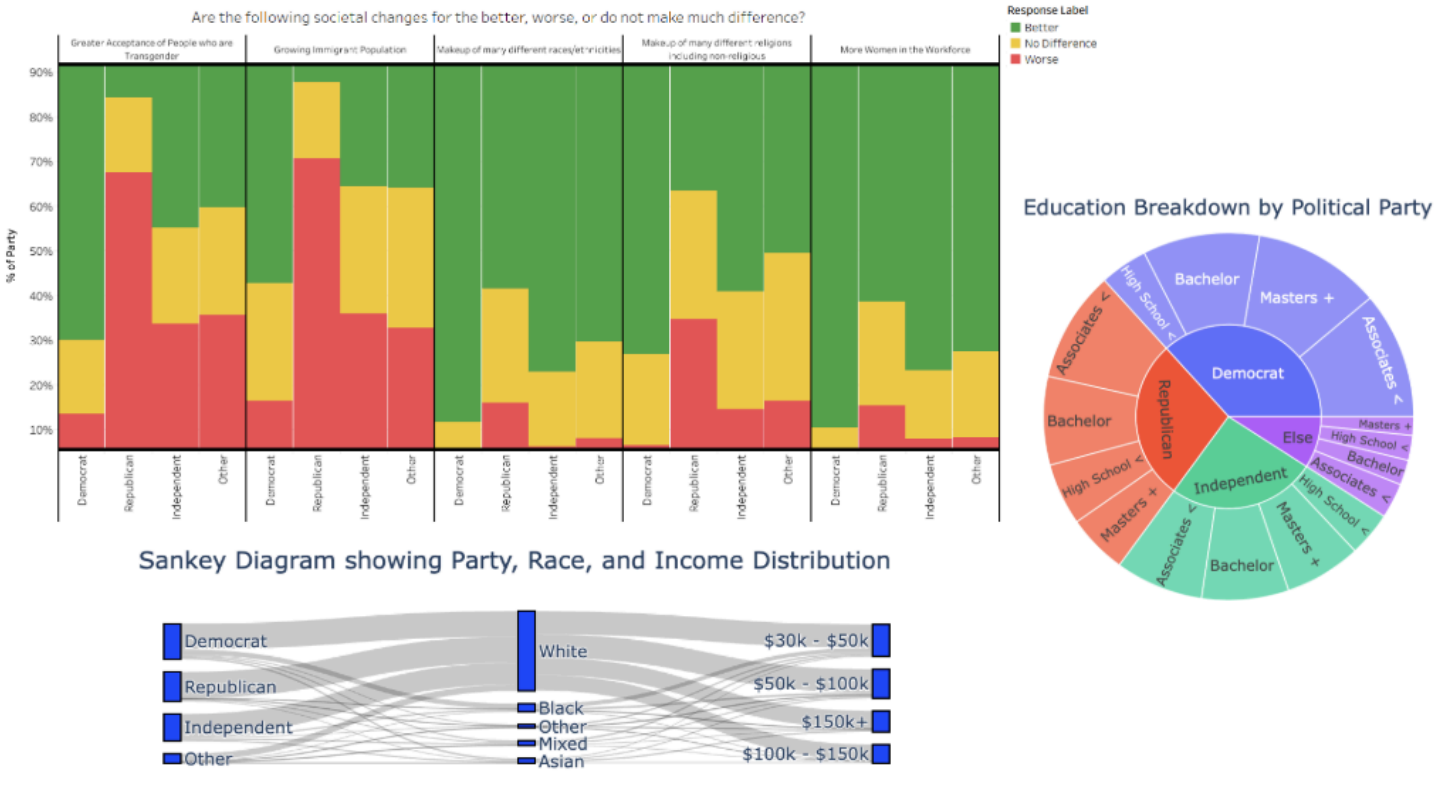


Our final iteration of visual three saw the change of the title from above the graph to inside the graph. We also removed the labels containing the party names because we added a legend on the side of our dashboard titled ‘Party Affiliation’ that one can use as a filtering mechanism.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

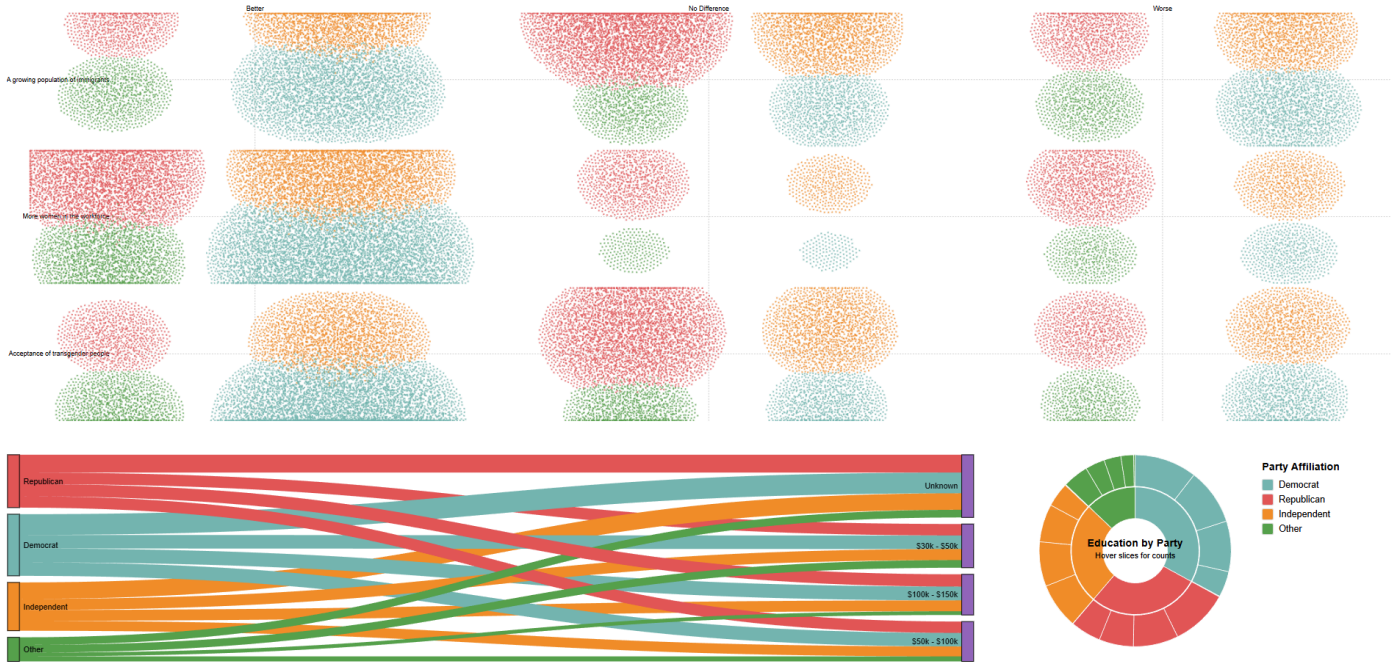
Dashboard at Milestone 2:



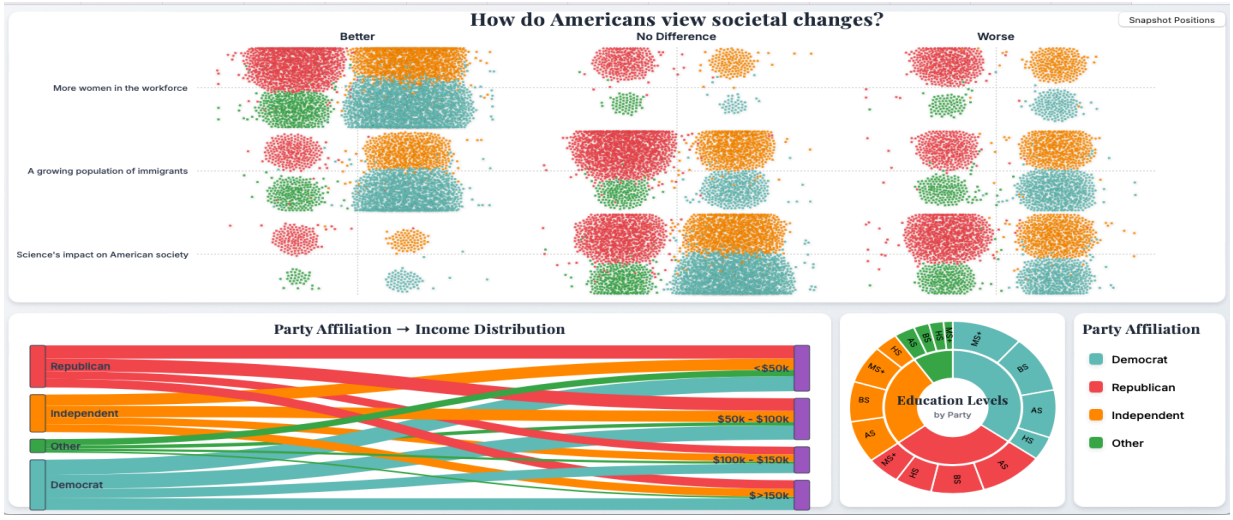
Current Dashboard:

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025



Final Dashboard:

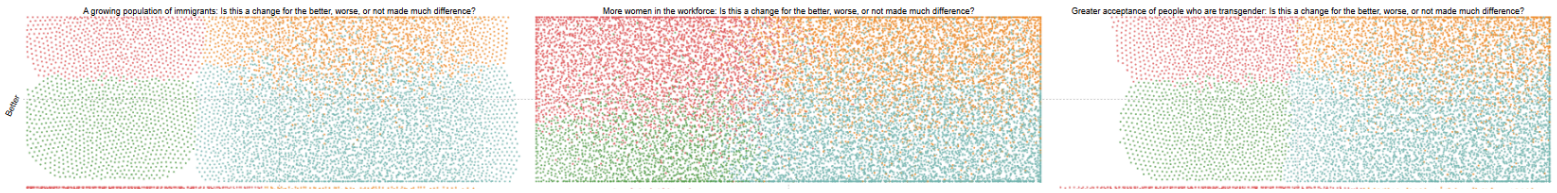


Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

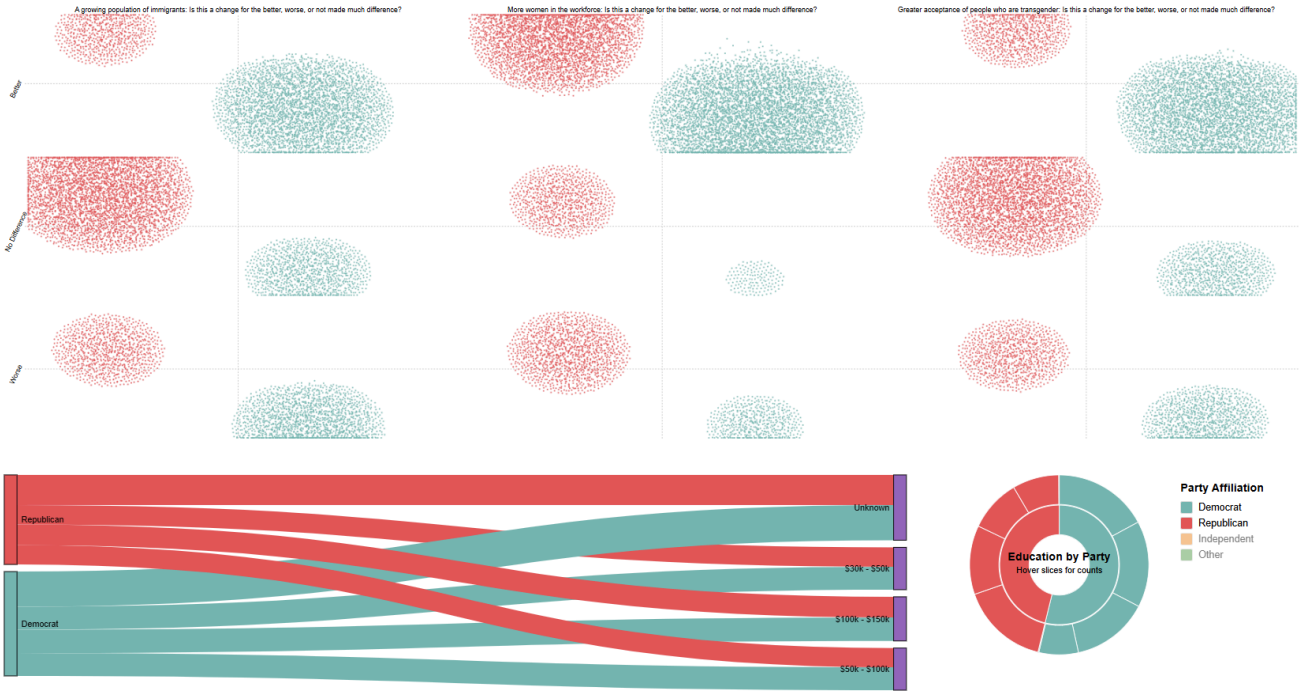
The primary goal of the first design was to explicitly connect each response to an individual mark. This allows the reader to observe how the general population viewed relevant cultural and societal questions, as sections with a large number of marks have more general support by the general population.

Ruth Mwangomo
Daniel Zintsmaster
Keerthi Surisetty

CPSC 6030
Clemson University
12/10/2025

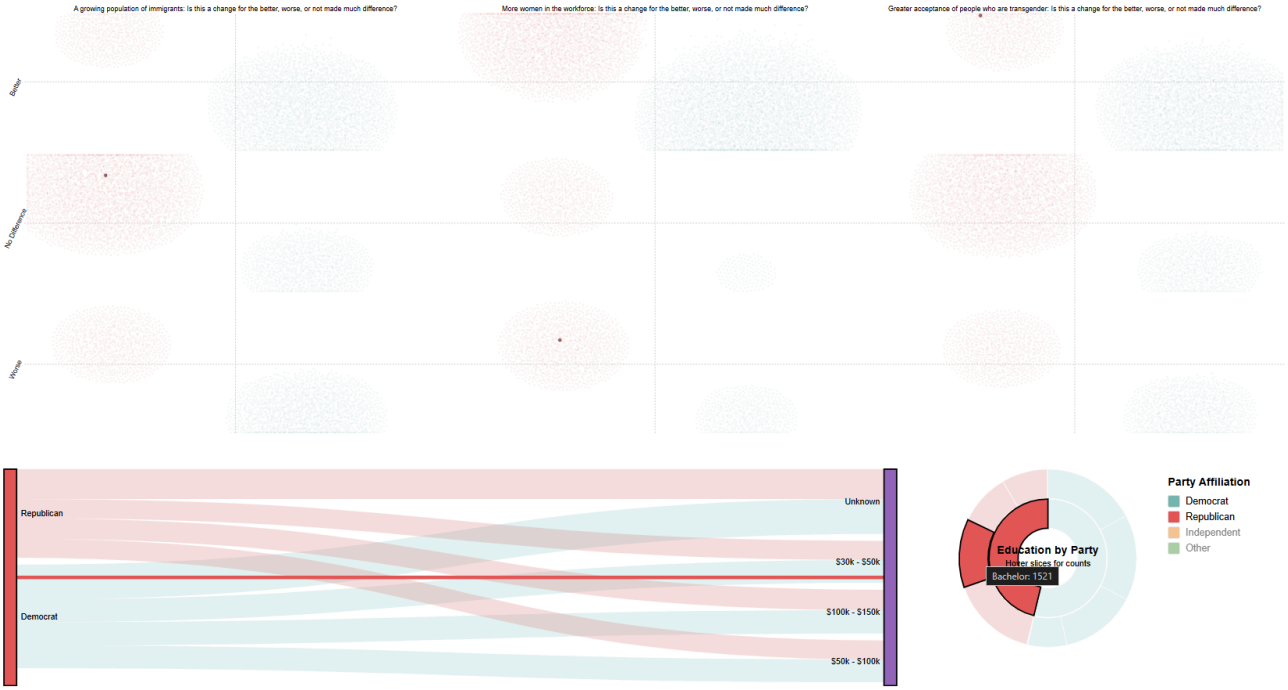


From the plot, it's clear that there is significantly more societal support behind more women in the workforce than acceptance of people who are transgender.
The next step was to separate these responses into each political party, such that each response would show the political distribution behind each response. This presented the opportunity to use political affiliation as the primary filter that would apply to all three plots. When one of the parties in the bottom right corner is selected, the responses, income brackets, and education levels are filtered to only include individuals who identified as that party. This allows for a more direct investigation into how aligned an individual party is without inclusion of distracting irrelevant data. For example, when only considering the Republican and Democratic parties, the other two parties in the legend can be selected to show their distributions for each question, as well as their income and education levels.

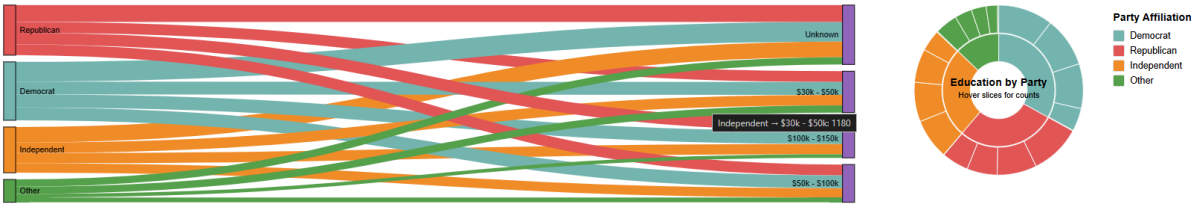


In the first visualization, the individual responses can be clicked on to show what those individual responses are for other questions through highlighting, and decreasing the

opacity of all other marks. Selecting a mark also highlights the sectors that the individual belongs to in the other visuals, clarifying the individual's income bracket, and their education level.



The final piece of interactivity comes from hovering over either of the 2nd or 3rd visualizations, in which it will give an exact count of how many respondents fall within that category. This is shown in the above screenshot for education (1521 respondents were Republicans with a Bachelor's Degree), and below for the income brackets.



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?

Through our visualizations, we uncovered several meaningful insights about how political affiliation intersects with social opinions, income levels, and educational attainment within the dataset. The interactive dashboard enabled us to explore these relationships holistically and gain

a deeper understanding of the factors that may influence individuals' stances on opinion-based questions.

The force-directed visualization clearly highlighted partisan divides on social issues. Even with the reduced prototype dataset, clusters of responses formed in ways that reflected ideological separation, suggesting that political identity remains a strong predictor of opinion patterns.

The Sankey diagram revealed notable trends in income distribution. While members of all political parties span the full income range, Democrats and Republicans tended to concentrate in distinct portions of the middle and upper-middle income brackets. This supported our inquiry into whether socioeconomic factors track with political leanings.

The sunburst chart provided a clear overview of educational differences across party lines. The hierarchical structure made it easy to see how educational attainment is distributed within each political group.

Although the visualizations were effective in helping us address our questions, there are a couple of improvements that would strengthen their analytical value. For the force-directed chart, rendering clarity could be improved, particularly when large numbers of marks overlap, by refining the simulation forces and making the layout more legible.

Another enhancement to the force-directed chart would be the addition of an average party answer, perhaps shown through box plots. This would make it easier to compare alignment within and across parties without relying on the visual interpretation of dense point clusters