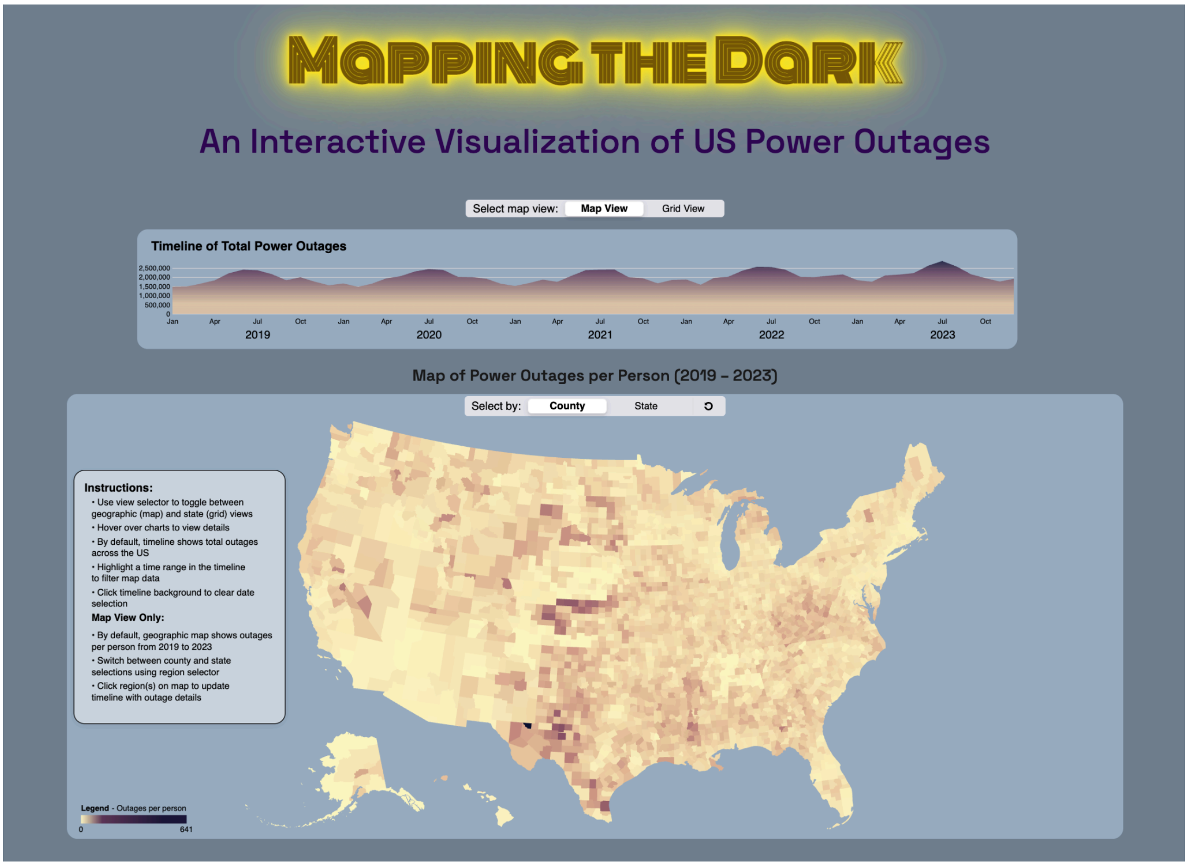
Milestone 4: Final Project Submission

1. Basic Info

Project Title: Mapping the Dark: An Interactive Visualization of U.S. Power Outages

Group members: Alethea Kramer, Muna Ibrahim, Mo Fardinzaman William Ho

2. Overview



DEMO LINK: https://youtu.be/pC0ikUwIQN4

Power outages are a widespread issue in the United States, affecting one in four households in 2023 [1]. However, the scale, frequency, and distribution of outages may vary significantly across regions and demographics. This project aims to provide an interactive data visualization platform to explore power outages across U.S. counties from 2019 to 2023. Through an interactive choropleth map, time-based exploration, and statistical charts, users will be able to examine outage trends, identify anomalies such as during the COVID-19 pandemic, and analyze correlations between outages and racial demographics.

This visualization is designed with policymakers, infrastructure planners, and advocacy groups in mind, hoping to provide insights that could help identify regions most in need of power grid improvements. Additionally, it serves as a resource for the general public to understand the intersectionality between power outages and their impact on different communities. By making complex outage data more accessible, this project helps drive informed discussions around energy infrastructure and identify imbalances in consistent access to energy among various populations.

3. Data

3.1 Power outage dataset

Source: Power Outages in the United States from 2019 to 2023
https://figshare.com/articles/dataset/The_Environment_for_Analysis_of_Geo-Located_Energy_Information_s_Recorded_Electricity_Outages_2014-2022/24237376?file=44574907

Attributes	Category	Cardinality/Range
fips_code (county identifier)	Categorical	~3000
run_start_time (Date of Data Recorded)	Quantitative	[2019-01-01 00:00:00, 2023-12-31 23:45:00]
County	Categorical	~2000
State	Categorical	53
sum (Customers Without Power)	Quantitative	[0, ~500 000]

3.2 Map of the US Counties Dataset

Source: Map of the US Counties in geojson format
https://gist.github.com/sdwfrost/d1c73f91dd9d175998ed166eb216994a#file-counties-geojson

Attributes	Category	Cardinality/Range
Geometry of county	Categorical	~3000
GEOID (fips_code)	Categorical	~3000

3.3 US Decennial Census for Race (2020)

Source: Government Decennial Census Data for Race

https://data.census.gov/table

Attributes	Category	Cardinality/Range
County	Categorical	~3000
Population of county	Quantitative	[0, ~10 000 000] depending on the county

The census dataset has over 3000 county attributes (columns) and 71 race groups (items/rows). The row labels include but are not limited to: 'Total', 'Population of One Race' (which is then further broken down into specific races such as 'White' or 'Black/African American'), and 'Population of Two or More Races'. The 'Population of Two or More Races' is also broken down into multiple categories, depending on the race combination. Overall, this dataset indicates the total count of each ethnicity in each county.

3.4 Population Data by County (2019 - 2023)

Sources: County Population Totals: 2010 - 2019, County Population Totals: 2020 - 2023, Annual Town and County Population for Connecticut

Attributes	Category	Cardinality/Range
Population of county (3221 county attributes in total)	Quantitative	[0, ~10 000 000] depending on the county

3.5 Derived Variables

Derived Variables	Category	Cardinality/Range
Month	Ordinal	12
Year	Ordinal	5
Total Power Outage Events	Quantitative	[0, ~35 000]
Power Outages Per Person in a County	Quantitative	[0, 641]
Customers Without Power (within a State)	Quantitative	[0, ~20 000 000]
Racial Demographics for Each County for Each Race (Percentage)	Quantitative	[0, ~100]
Percentage of Non-white Population	Quantitative	[0, ~100]
Average Proportion of People Affected by Outages Within State	Quantitative	[0, ~80]
X and Y-coordinates of State for cartogram loosely based on longitude and latitude.	Quantitative	[0, 11]

3.6 Data Preprocessing

All data pre-processing including dataset joins, clean up, imputation, and calculations, will be conducted using a python script.

3.6.1 Power Outage & Geometric Map Dataset Processing

The power outage data is logged once every 15 minutes which is too specific for our purposes. Thus, the outage data will be aggregated by month. The values will be grouped together by counting the number of outage events to highlight how disruptive the outages can be precisely because of their frequency. We will also be summing up the number of customers without power per state to highlight the number of people affected by power outages.

Furthermore, this dataset provides each year's outage records in a separate file, so each of the 5 years of data (i.e. 2019 to 2023) will be merged for further convenience. However, much of the power outage data is inconsistent across years, and these inconsistencies must be handled prior to merging the data. For example, the variable name for the number of customers missing power in 2023 was changed from `customers_out` to `sum` which must have consistent naming before merging. Additionally, the county names in the power outage data have incomplete naming. For example, Baltimore City and Baltimore County are both listed as "Baltimore" in the county attribute, causing a gap between the cardinality of county and FIPS codes. As a result, all joins must be done on the FIPS code.

The geometric map data's GEOID column exactly matches the FIPS code values, allowing for easy joins. Only the geometry attribute is needed for the map data; all the other attributes from the map data can be ignored.

3.6.2 Population data preprocessing for choropleth map

Our time range required us to aggregate the data from two sources. I first preprocessed the 2019 data by cleaning, standardising column formats, and merging it with the FIPS code for unique identification. The 2020-2023 data is similarly cleaned and standardized. However, The county designations were changed, namely in Connecticut and Alaska in the middle of this time frame. Alaska's counties were merged since larger counties were split into smaller ones. Connecticut, on the other hand, restructured the counties altogether. Therefore, the population data from the previous year was carried forward for the years where the counties were restructured. Finally, both datasets are merged on FIPS codes to create a unified dataset and saved as a CSV file.

3.6.3 Census Race Dataset Processing

The government census data, on the other hand, doesn't have a variable for FIPS codes clearly linked. It has the racial demographic data in the form of rows and each column denotes a single county while the power outage data has an attribute denoting the county and each power outage is represented by a row. This means the government census data needs to be transposed to be consistent with the power outage data, where each row represents a single event. The census and outage data can be subsequently merged.

Furthermore, the population column is read as a string object when imported using pandas and must be converted to integer.

Finally, due to the mixed race individuals comprising a small portion of the population, individuals identifying as more than one race will be aggregated into a mixed-race category. Single race individuals will be counted as-is. The racial makeup which will be considered are as follows:

For populations identifying as one race only:

- White
- Black or African American
- American Indian and Alaska Native
- Asian
- Native Hawaiian and Other Pacific Islander
- 'Other'

For populations of mixed races:

- 'Population of Two or more races

All other groups (e.g. different combinations of mixed races) will be dropped for simplicity. Furthermore, due to the prevalence of white-majority counties, non-white populations will be aggregated and compared to white-populations. We will also be filtering counties that do not correspond to the 50 American States, such as those in Puerto Rico and American Samoa. However, we also decided to include the District of Columbia and treated it as a State in the cartogram, as we noticed it had significant demographic and outage differences compared to most States.

3.6.4 Dataset Preprocessing for Cartogram

The layout of the cartogram was inspired by this Observable notebook [link: https://observablehq.com/@mcmcclur/simple-grid-cartogram] and was adjusted to appear cleaner, especially when incorporating size encoding (described in section 5.4.1). The cartogram dataset can be merged with the demographic data using the State name as a key. This merged dataset can then be further combined with the power outage data, which also uses State names as a common field. To reduce duplication, the data was separated into two distinct datasets: one containing the grid layout and demographic data (with one row per State), and the other containing monthly outage data per State to reduce duplicate attributes.

4. Tasks

#	Domain-specific task	Abstract task	Datasets/attributes
1	A policy maker or energy provider wants to identify which counties experience the highest number of power outages and whether certain counties are more prone over time.	{analyze distribution} (spatial) {identify trends} over time	Dataset: power outage records, map of US counties Attributes: `fips_code`, `Total Power Outage Events`, `year`, `month`, `geometry`
2	An advocacy group wants to compare the trends between power outages and racial composition to see if minority groups are disproportionately affected	{compare trends}	Dataset: power outage records, census data Attributes: `fips_code`, `Total Power Outage Events`, `Racial Demographics`
3	A public health advocacy group wants to compare power outage patterns during the COVID-19 pandemic against normal years and identify anomalies.	{compare anomalies} to baseline	Dataset: power outage records,, map of US counties Attributes: `fips_code`, `Total Power Outage Events`, `year`, `month`, `geometry`

5. Visualizations

5.1 Overview

The main page will feature a timeline of total power outages in the United States, and a choropleth map displaying the number outages per county. At the top left, users can toggle between a ‘Map View’, which displays a timeline and choropleth map (Figure 1), and a ‘Grid View’ with a cartogram in which each State is represented by a square tile (Figure 4). The subsequent sections provide details on these views and the data they visualize.

In the grid view, a disclaimer will indicate that the racial data applies only to 2020 and the racial group names follow the naming convention of the 2020 US Decennial Census. Additionally, hovering over a county (in the ‘Map View’) or a State (in the Grid View’) will display a tooltip containing its name and relevant statistics – such as the number of outages per person in the Map View, and number of affected people in the Grid View.

5.2 Timeline & Area Chart

The timeline view primarily serves to provide interaction to the user and allow them to select a date range for the choropleth map. However, it also enables users to explore the temporal distribution of power outages from January 2019 and December 2023. The area chart represents the total number of outages recorded in a given month, with the x-axis spanning the entire five-year period and the y-axis indicating the total number of outages. By default, the chart aggregates the outages across all US counties; however, users can refine the data by selecting specific counties in the choropleth map, which dynamically updates the area chart to reflect only the chosen regions. This component is bidirectionally linked to the choropleth map: while the choropleth map restricts the regions, the timeline restricts the time interval.

This view was chosen to support the first task, temporal analysis of power outages, and was inspired, in part, by a similar view on archive.org. An area chart is ideal for representing discrete aggregated outage counts over time, allowing users to compare discrete outage counts across different months more effectively. Bar charts could provide a direct and interpretable view of monthly outage frequencies, but the area chart is able to emphasize trends. We also added a colour gradient ranging from yellow for low values and purple for high values to make the larger values stand out more. Additionally, this view supports the third task, comparing anomalies, by allowing users to view periods of unusual activity at a glance, such as during the COVID-19 pandemic.

The area encodes monthly outage counts in their 1-directional vertical size channel while the 1-directional horizontal position channel encodes time. A brush is shown to show what time range is selected, with areas outside the range appearing in a lighter shade to indicate they are currently not in focus.

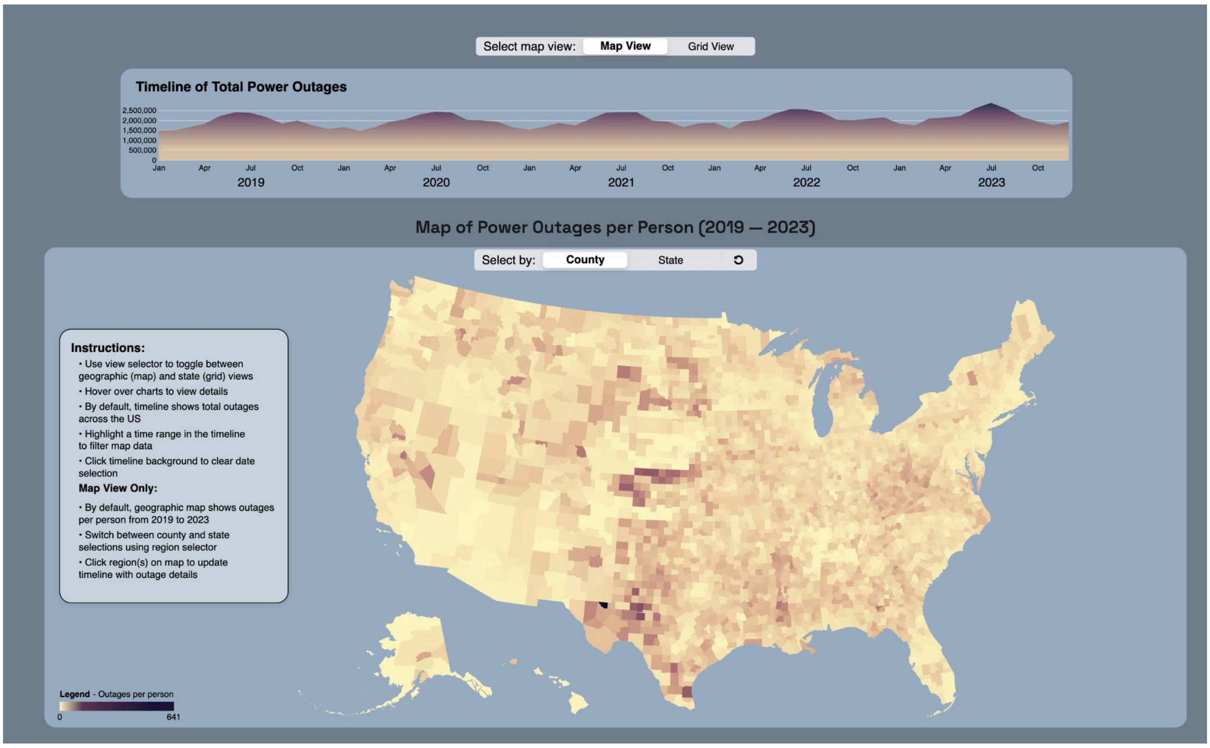


Figure 1: Sample image of timeline/area chart (view 1) and choropleth map (view 2).

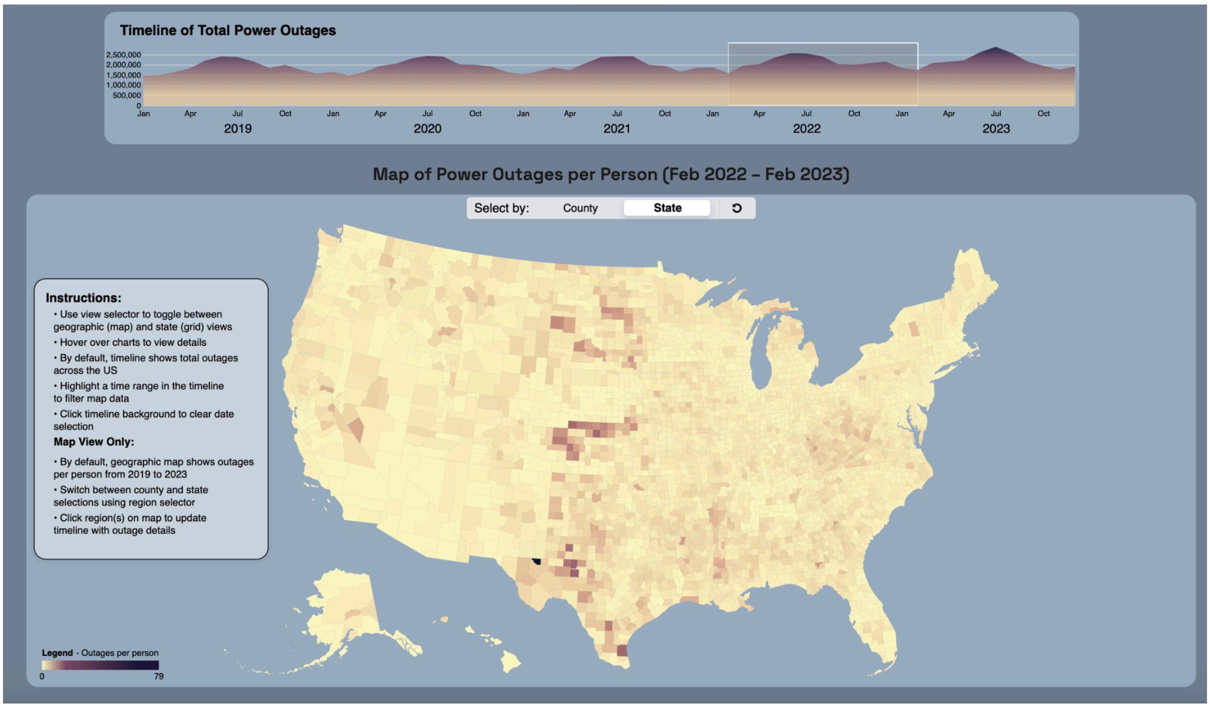


Figure 2: Filter by time

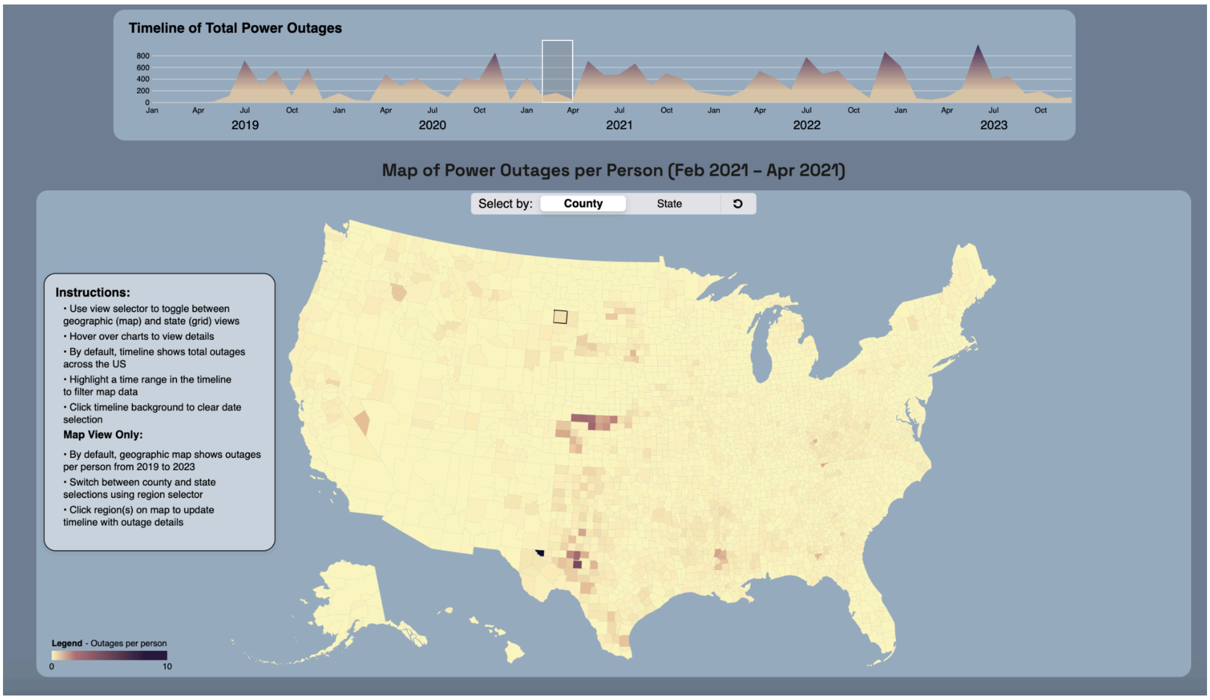


Figure 3: Filter by county and time

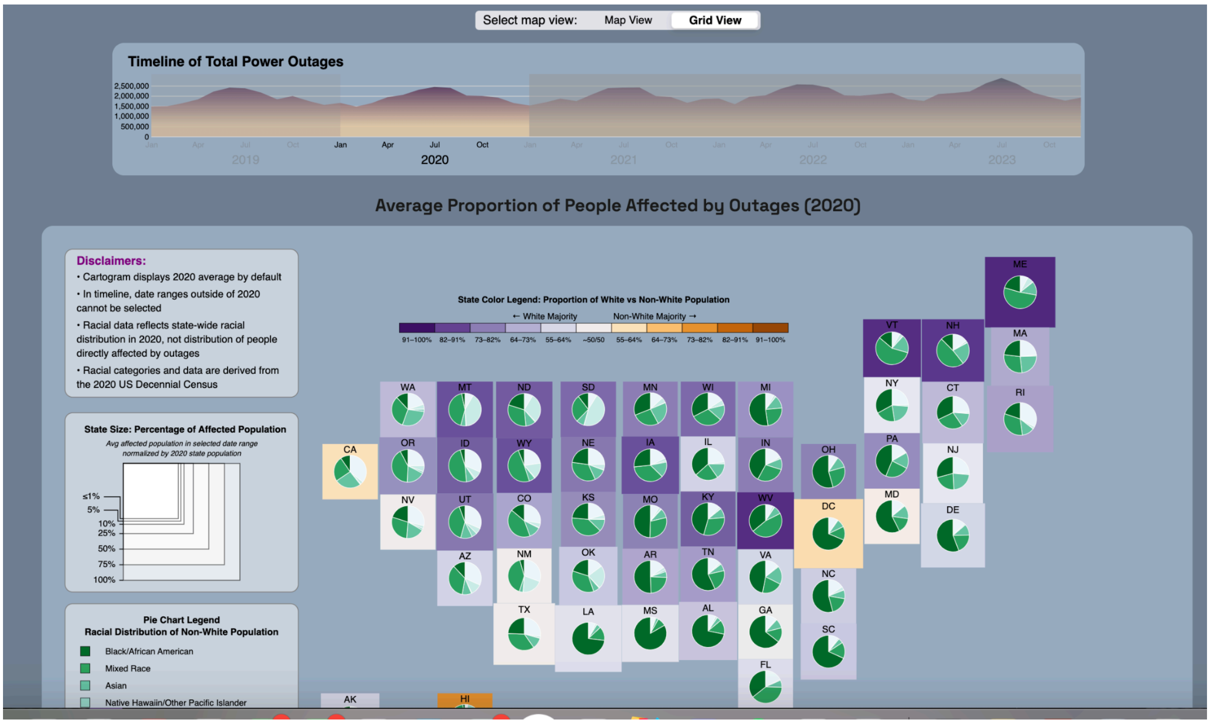


Figure 4: Timeline visually limited to shorter time span

5.3 Choropleth Map

The choropleth map allows users to explore the distributions of power outages between counties. The colour channel, ranging from yellow to purple, will indicate the lowest and highest outage counts, respectively.

By default, the data for the gradient will aggregate all the outages in a county for all 5 years of data. Using the bidirectionality of the timeline and map, users can restrict the time period via the sliders on the timeline. Each county on the map can be clicked which will then highlight or unhighlight the county on the map. The selected counties will be highlighted with a noticeably thicker outline. Multiple counties can be highlighted at once. This will then restrict the timeline to only show the outages of the counties highlighted. If no counties are highlighted, the timeline will show the total outages of all the counties. A radio button was also added to change the selector with 3 options: County, State, and Reset. By default, clicking a county will restrict the timeline to the selected county. If the radio button is set to State, then clicking a county will select every county that shares the same State as the selected one.

The aggregated outage count will use a colour encoding while the horizontal and vertical 1D position and order will be indicated via the longitude and latitude of the geometry values respectively. Selected counties will be indicated by a bolded border. A tooltip will popup whenever you hover over a county which will summarize information such as the county, State, and exact number of power outage events

This view supports the first and third task. The gradient is effective for showing regional patterns and variation. For example, noticeable differences should be observed between nearby counties if an anomaly such as a major outage occurred. Some counties may experience more severe weather, or have higher population densities which can lead to being more prone to power outages. This can help policy makers identify areas which experience outages frequently, or which regions are at risk to major outage events.

The map view is also very accessible to other people, correlating the actual information to a geospatial distribution. Alternatives that don't use a map will make it a lot harder to understand what regions or states are most affected by power outages.

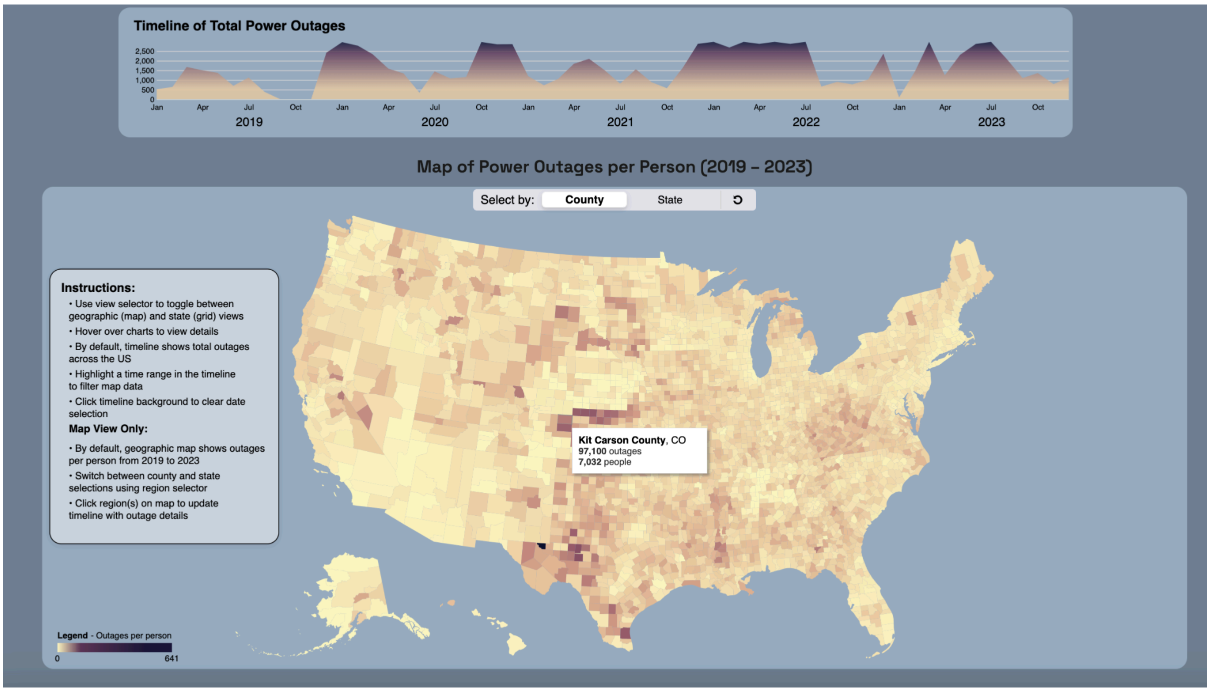


Figure 5: Tooltip for county data and click to filter timeline

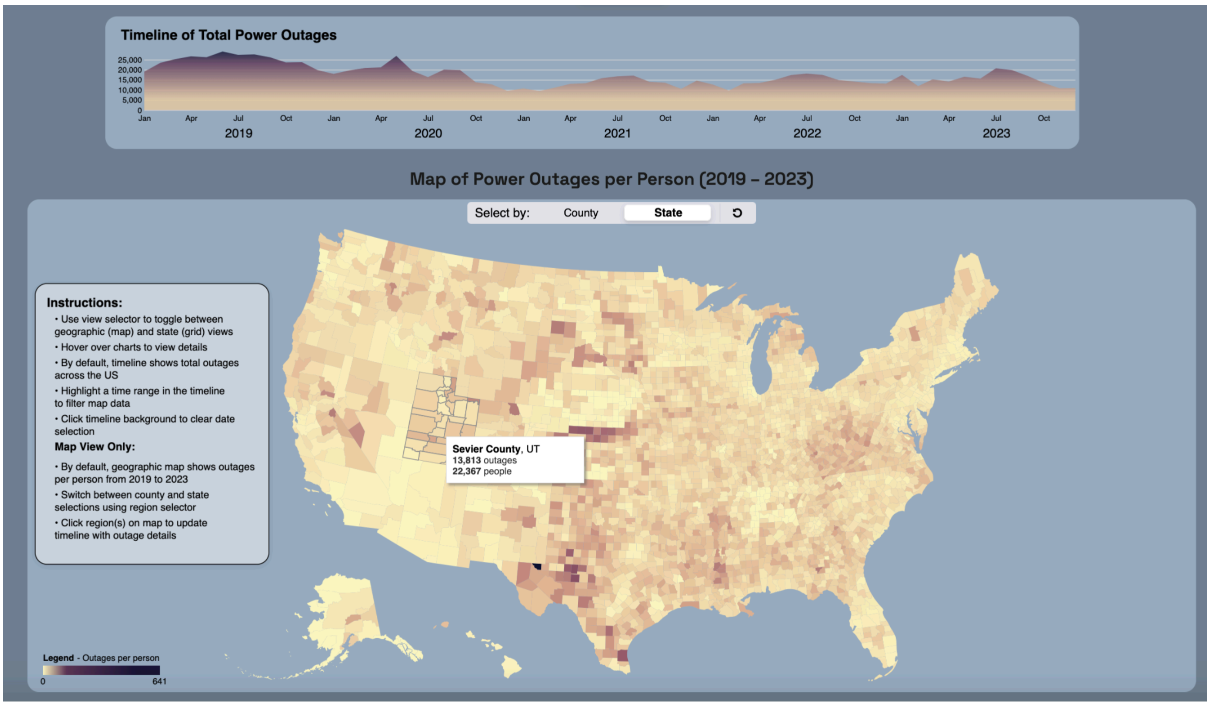


Figure 6: Select by state to filter timeline data

5.4 Cartogram

5.4.1 Description

The purpose of the tile grid view is to demonstrate (i) the total population affected by power outages in each state, and (ii) the racial groups most affected by power outages. This view addresses task 2, which explores whether minority groups are disproportionately affected by power outages across the United States. Since the racial demographic dataset comes from the 2020 US Census (with no equivalent data available for other years), we will focus solely on power outages from that year to ensure accuracy.

This view will have two levels of representation. It's key features:

1. **Top-Level (State Tiles)**:

   Each state is represented by a square tile where:

   - **Size** is proportional to the average number of people affected by power outages, normalized by the state's total population
   - **Colour** indicates the ethnic majority of the state (ie 'white' vs 'non-white'). A diverging color scale of purple (white population) to orange (non-white population) is used. Color scheme rationale is provided in point 4 below.

   Tiles are positioned in proximity to their geographic location. The implementation of the cartogram makes it so that states that are vertical from each other are blessed directly above and below each other. This makes it have 1D vertical sharing Given that the majority population of most counties is 'White', it stands to reason the same is true at the state-level. Thus, as described in 3.5.3, all non-white groups will be aggregated into a single category for clarity. This simplifies interpretation by reducing the number of colours users must decode to just two, and more succinctly demonstrates whether the people who are most affected by outages are an ethnic minority (i.e. non-white) or majority (i.e. white) group.

   Furthermore, the timeline/area chart is linked to the cartogram so that selecting a valid date range (i.e., dates within 2020) updates the tile sizes to reflect the average proportion of affected individuals during that period.

2. **Low-Level (Pie chart)**:

   To supplement the colour encoding of each state tile, pie charts will provide a more detailed breakdown of the racial distribution of non-white populations affected by power outages. Each state tile will contain a pie chart, where each wedge represents the percentage of affected individuals belonging to one of six non-white racial groups described in Section 3.5.2. We initially considered representing the white population in the pie chart, but because it constituted such a large proportion, the other segments became difficult to distinguish. Instead, we chose to visualize the distribution of non-white populations in the pie chart and encoded the proportion of the white population through the state tile color, as described above in Section 5.4.1.

   All pie charts will be uniform in size, constrained by the smallest tile. However, the minimum tile size will be set to ensure that the pie charts remain legible. This standardization will maintain consistency in scale, allowing for easy and accurate comparisons across states.

It is important to note that the pie charts serve as a supplementary visual aid to enrich the existing visualization. The primary focus remains on the tile grid's size and color encoding, which provide the key insights into power outage impacts on 'White' and Non-white' groups.

3. **Tooltips**: When the mouse hovers over a state, a tooltip appears displaying the state name, the proportion of people affected, and a detailed racial distribution for that state. Additionally, hovering over a specific wedge in the pie chart reveals a tooltip showing the number of individuals belonging to that racial group within the selected state.

4. **Color scheme**: To avoid overwhelming users with excessive color, we limited the palette to three distinct hues with varying levels of saturation. The base color scheme consists of green, orange, and purple, selected from https://colorbrewer2.org to ensure accessibility and color-blind safety. For the cartogram, we used a diverging color scale of purple and orange, implemented using D3's built-in scale functions. In the pie charts, we applied varying saturations of blue-green hues, also utilizing D3's built-in blue-green color scale.

A detailed marks and channels analysis is provided in section 5.4.2

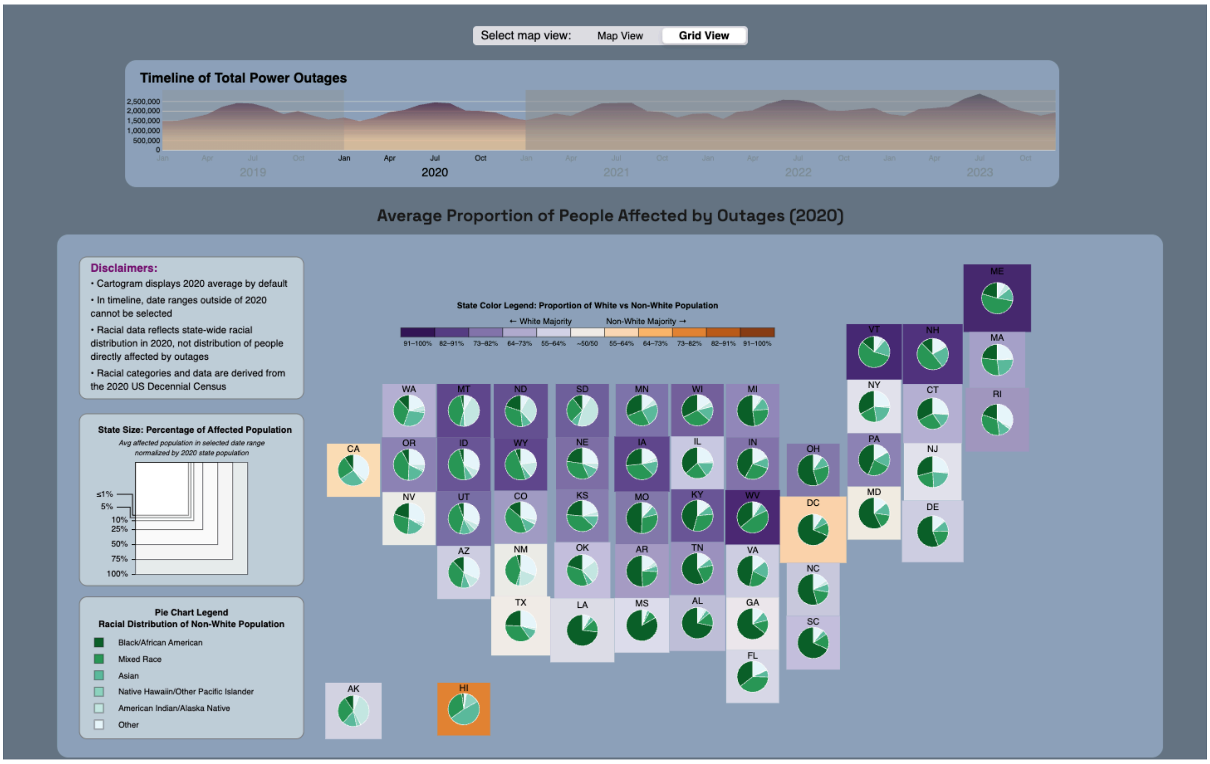


Figure 7: Sample image of timeline/area chart (view 1) and cartogram (view 3).

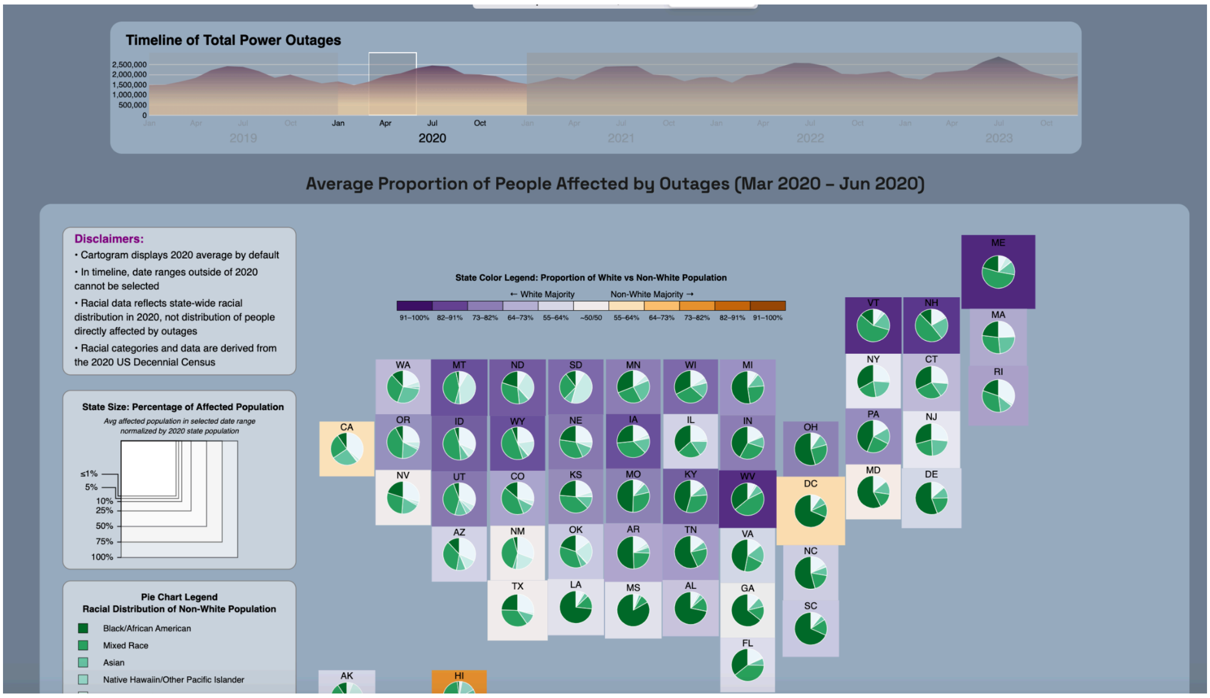


Figure 8: Cartogram filtered by time

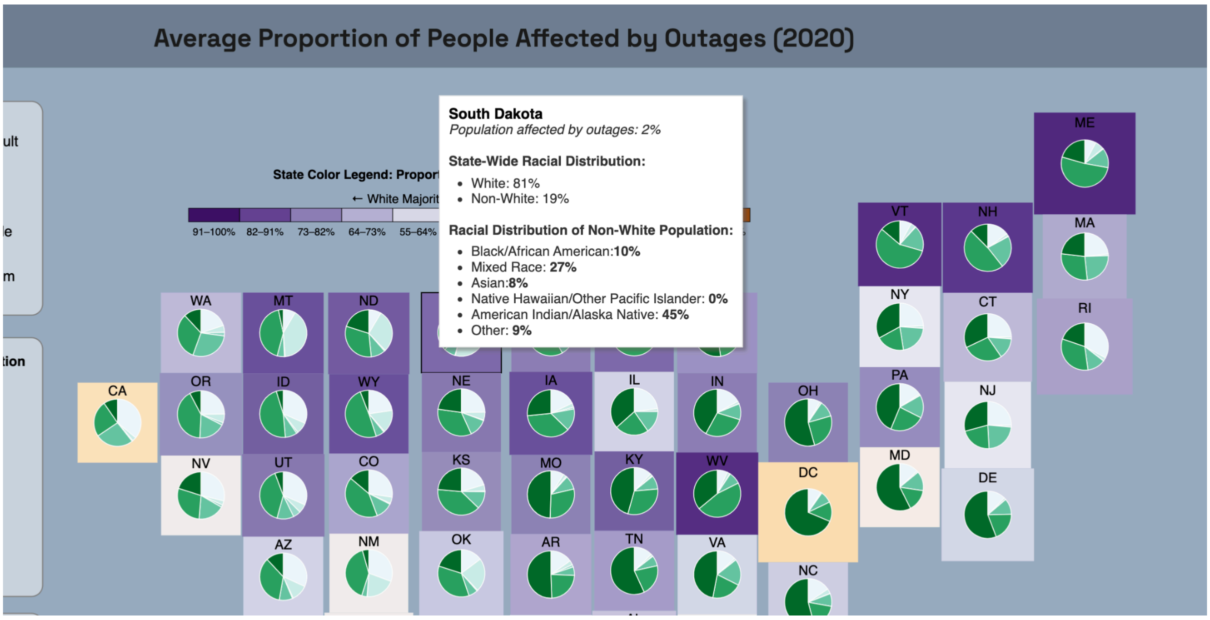


Figure 9: Tools tip for racial distribution by state

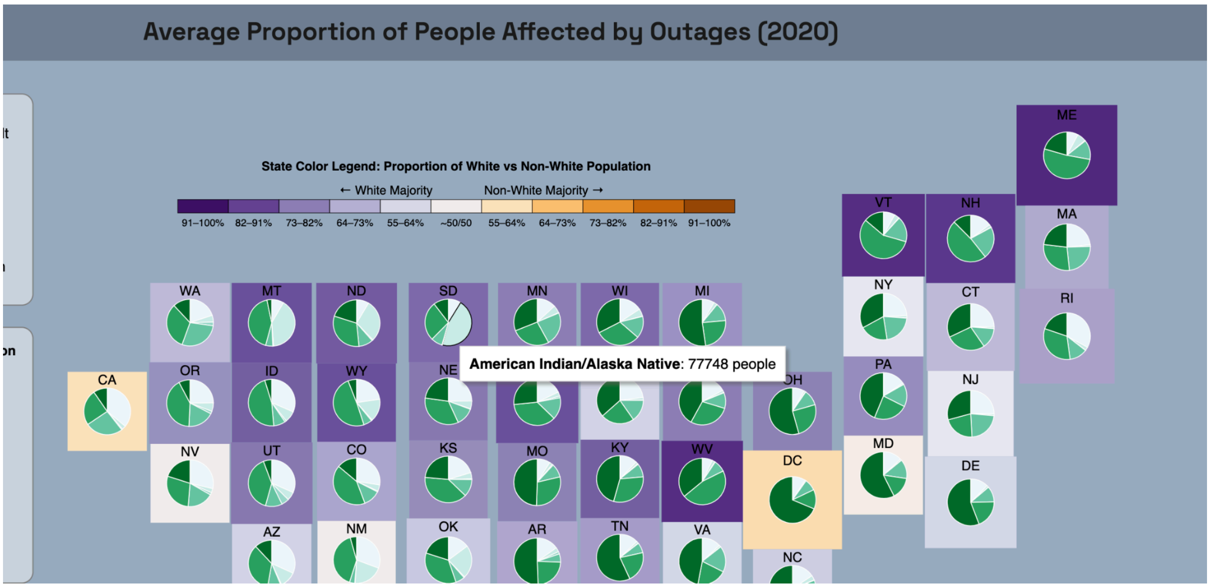


Figure 10: Pie chart tooltips

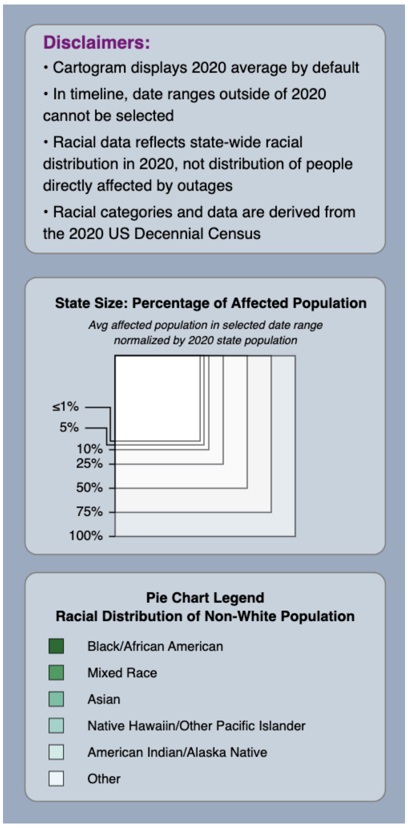


Figure 11: Disclaimers and instructions

5.4.2 Marks & Channels Analysis

		Level 1: Pie Chart	**Level 2: Square in tile grid**
Item		Racial demographics	State
Mark type		poly	Point
1D Cartesian Horizontal	Shared	no	yes
	Position	Roughly based on geographic location (longitude)	Roughly based on geographic location (longitude)
	Order	<same as position>	<same as position>
	Size	unavailable	unavailable
1D Cartesian Vertical	Shared	no	yes
	Position	Roughly based on geographic location (latitude)	Roughly based on geographic location (latitude)
	Order	<same as position>	<same as position>
	Size	unavailable	unavailable
1D Polar Angle	Shared	no	no
	Position	Racial group	unavailable
	Order	<same as position>	unavailable
	Size	Proportion of people affected by power outages	unavailable
1D Polar Radial distance	Shared	no	no
	Position	Not used	unavailable
	Order	Not used	unavailable
	Size	Not used	unavailable
2D	Shared	no	no
	Size	Not used	Total number of people affected by power outages
Orientation		unavailable	Not used
Colour		Racial Group	Racial majority

5.4.3 Rationale

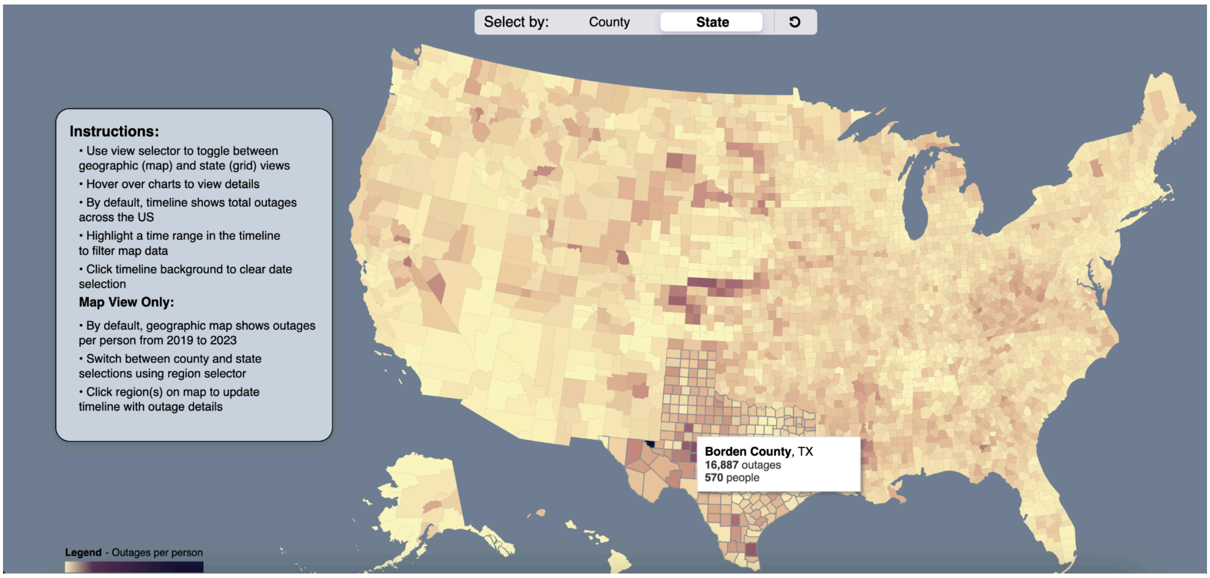
The proposed tile grid map with embedded pie charts provides a number of benefits:

1. **Preserves geographic awareness**: While not as precise as the geographic map, the tile grid maintains relative positioning of states. This reduces the cognitive load on users and allows them to more easily relate power outage events in the geographic view, to the proportion of affected individuals in the tile grid view. This is further reinforced by the presence of a tooltip which will include data on both total power outage and number of people affected.
2. **Clearer majority information through color coding**: The two-color system (White vs Non-white) provides an immediate visual cue about which racial group was most affected in each state
3. **Tile grid is more effective than alternative visualizations**: Unlike a heatmap with nested treemaps or a matrix of embedded pie charts, which are constrained by pixel limitations and may result in low resolution, the tile grid maintains clarity and immediate readability. This makes it easier to identify patterns and trends at a glance.
4. **Pie charts offer better legibility than alternative chart types:** Pie charts can become difficult to interpret when misaligned, while tree maps require a higher cognitive load to compare non-uniform areas. Pie charts provide a clear and intuitive representation of racial distributions while maintaining a consistent format across states.

6. Usage Scenarios

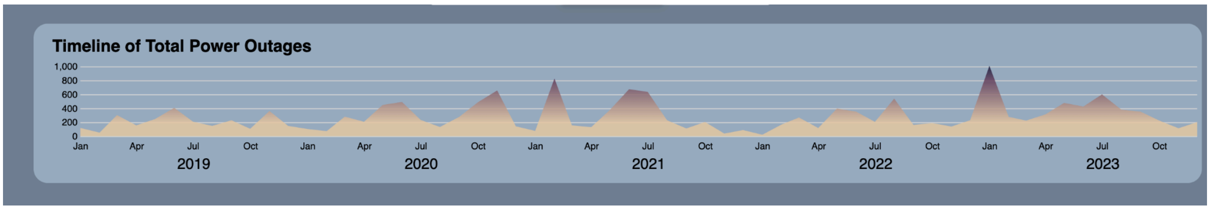
6.1 Usage Scenario 1

Jessica is a working professional living in a small town in Texas. Over the past few years, she's noticed that power outages seem to happen more often than they should. She's lost power during heat waves in the summer, unexpected cold snaps in the winter, and even on random clear days. She's starting to wonder whether her county actually experiences more outages than other places or if it just feels that way. To get some perspective, she visits an online power outage visualization tool, *Mapping the Dark*, that helps users explore outage patterns across the country. When she opens the tool, she sees a choropleth map of the U.S., where counties are shaded based on the total number of power outages recorded over a period of five years. She identifies Texas, visually scanning for her county by hovering over the area and seeing tooltips. Right away, she notices that her county is shaded darker than some of the surrounding areas, suggesting it has had more outages.



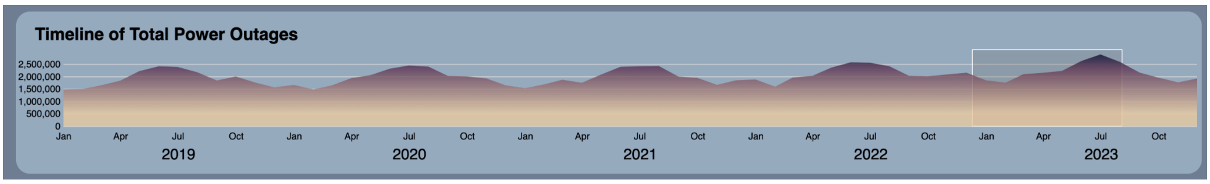
Jessica's initial view of the choropleth map, hovering over her county

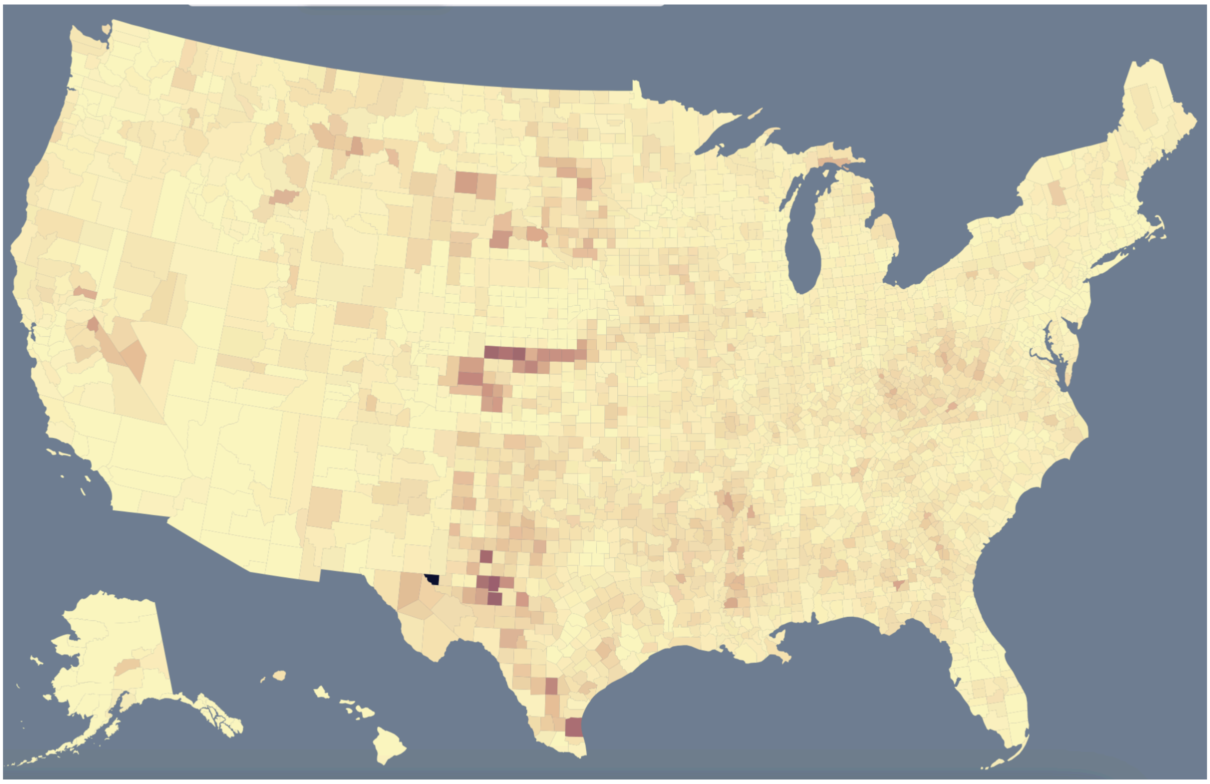
She wonders if this is just a recent trend or if her county has always had frequent outages, so she clicks on her county, which filters the distribution of power outages timeline so she can investigate trends. She sees that there are pretty significant peaks when most outages happen.



Timeline filtered by clicking on her county

She's curious whether her county is just unusually bad compared to other areas. She clicks the reset button to deselect her county, then selects a time period from the slider and the map updates to show the number of outages during that time.

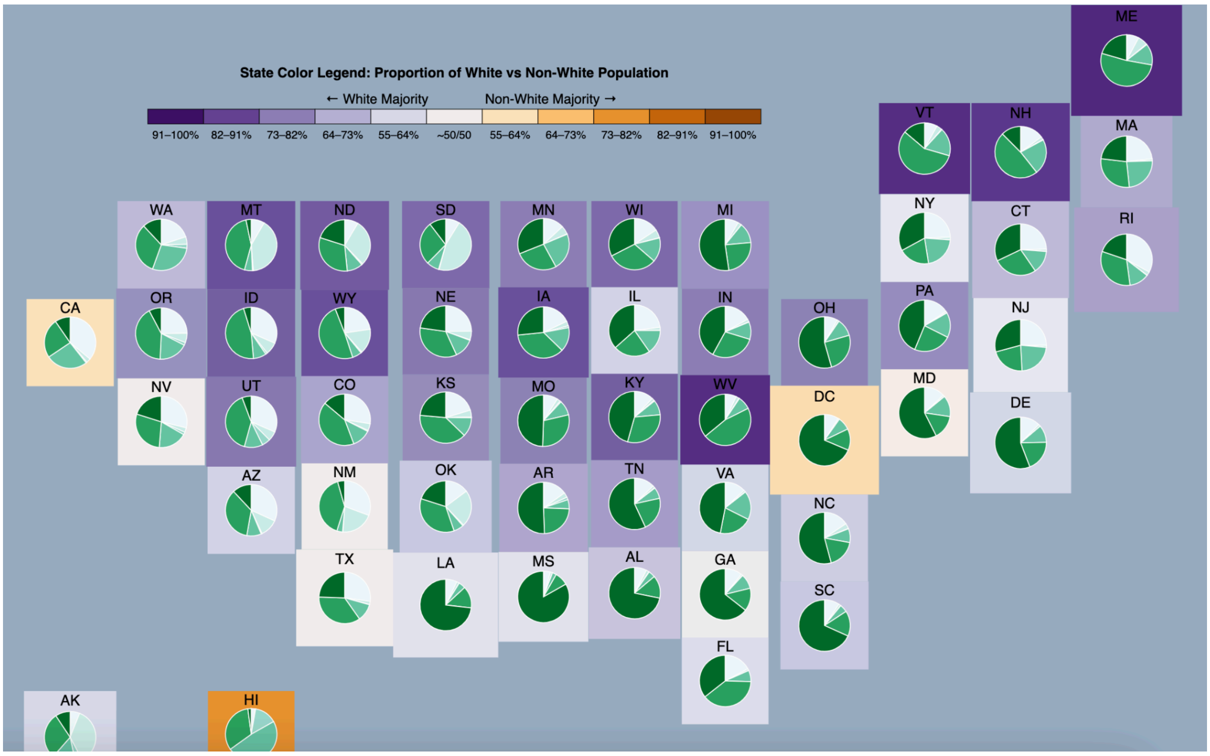




The choropleth map, filtered by timeline selection

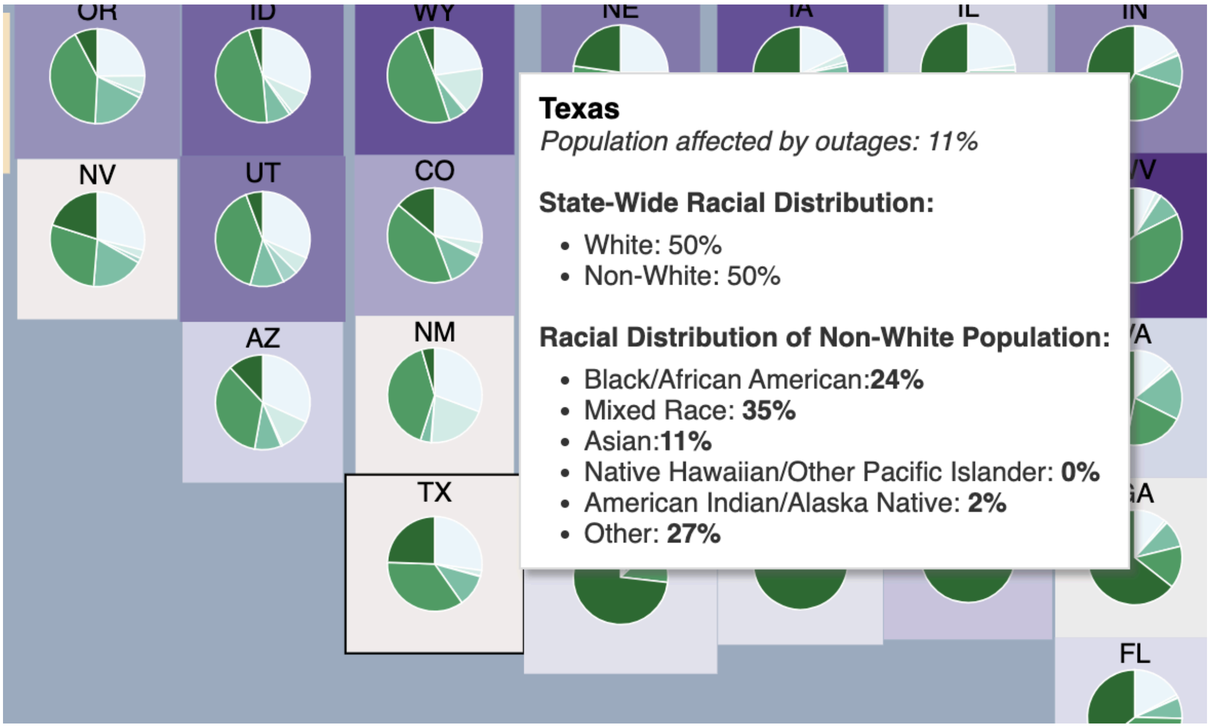
Jessica notices that in a more recent period, many counties outside Texas are lighter so have experienced fewer power outages, which suggests they've improved their infrastructure, while the shading of her county and those surrounding it have stayed dark. Hovering over various counties, a tooltip appears, showing the county name, state, and the total number of recorded power outages. She moves her cursor across different counties, comparing their outage counts to her own. Some nearby counties have noticeably lower numbers, suggesting that her area may indeed be experiencing more frequent disruptions. Wanting to get an even bigger picture, she switches from the geographic map to the tile map view.

Now, instead of counties being shaded, each state is represented by a tile, with the size of the tile indicating the total number of people impacted by outages in that state. Jessica immediately notices that Texas has one of the larger tiles, but not the largest, meaning that more people located there are impacted by outages than many other states. This confirms her suspicion that Texans experience a particularly high number of outages, but also helped her visualize that although her county is particularly impacted, the state as a whole is somewhat comparable to others, and certainly not the most impacted.



Jessica's initial view of the cartogram

She also notices that the tile map includes pie charts of racial distribution within each state's tile. She wasn't originally thinking about demographics, but now she's curious. She sees that Texas has a more racially diverse population than some of its surrounding states with fewer outages, and she wonders if certain communities within her state experience more frequent outages than others.



Tooltip while hovering over Texas, her state

She decides to switch back to the choropleth map, focusing on discovering counties with particularly high rates of power outages. Now that she has a clearer sense of how her county and state compare to others, Jessica starts thinking about what she can do with this information. She takes a screenshot of the visualizations and drafts an email to her local representatives, asking about investments in the power grid. She also shares her findings with neighbors in a community forum, hoping to start a conversation about improving energy reliability in their town.

7. Reflections

Our visualization goals evolved significantly over the course of the project. Initially, we aimed to uncover potential correlations between racial demographics and power outages at a fine-grained level, hoping to surface patterns of infrastructural inequality. However, the limitations of the racial dataset—only available for 2020 and overwhelmingly skewed toward white-majority counties—led us to reframe our goal. Instead of proving a hypothesis, we shifted toward creating an exploratory tool that encourages users to interact with the data and form their own insights about geographic and demographic patterns in power outages. This led us to focus on clarity, accessibility, and visual storytelling, rather than dense comparative analysis.

Before we considered creating a cartogram, our initial design proposal used a treemap inside a heatmap, where each tile in the heatmap would be a state, and the treemap would show the racial demographics. After receiving feedback from TAs and trying to visualize the result ourselves, we concluded that the resolution would likely be too small to meaningfully interpret individual treemap values—especially for smaller race groups—so we scrapped the idea early on.

When we pivoted to the cartogram, we initially planned to use bar charts inside each tile to show racial demographic data. However, this introduced a new problem: bars could be highly inconsistent in size depending on the racial makeup of the state. For example, one state could have 98% of its population in a majority group, while another state's largest group might only represent 50%. This created misleading visual differences, as some bar charts would visually dominate while others looked sparse. To account for this, we would have needed to normalize bar chart height across all tiles, which added complexity and reduced readability. We ultimately transitioned to pie charts, which allowed for a fixed, consistent size across all states, improving layout, reducing visual noise, and making comparisons easier.

Even with pie charts, we faced further challenges. In most states, the white population formed the clear majority, which meant the non-white portions were often too small to be visible. This made the chart less effective at highlighting variation across minoritized groups. To address this, we removed the white category from the pie chart entirely and instead encoded white-majority vs. non-white-majority via the tile color. This prevented redundant encoding and allowed the pie charts to focus on the relative distribution among non-white groups, which better served the exploratory intent. We also moved from a hue-based color scheme to a gradient-based saturation scale (in blue and green tones), which preserved distinctions even at smaller proportions and enhanced readability for users with color vision impairments.

Our timeline visualization also went through a transformation. Originally built as a bar chart, we found it too plain and uninspired—it didn't invite exploration. After seeing a project from a past CS436V cohort, we switched to an area chart, which looked smoother, more engaging, and better suited to show volume and trend over time. The area chart also became an important interaction mechanism, letting users filter data by time and see how patterns changed across the 5-year window.

On the technical side, our early assumptions about interaction and rendering performance proved overly optimistic. For example, we initially implemented a brush tool for the choropleth map that would allow users to select multiple counties by dragging. However, the performance was extremely poor—D3 struggled with lag, and tooltips became delayed or unresponsive. We tried optimizing with caching, but the lag persisted. We considered a multi-state dropdown, but with 50+ options, it became too unwieldy. In the end, we opted for a radio button toggle: one mode allowed individual county selection, and the other allowed users to select all counties within a clicked state. This preserved functionality while vastly improving usability and speed.

Another technical discovery came when we noticed that outage counts were highly correlated with county population, skewing results in favor of more populated areas. To correct this, we normalized outage values by population, and applied a symlog (signed logarithmic) scale to balance the data. The symlog scale helped compress extreme values while retaining mid-range variability, offering a clearer picture of which counties were disproportionately impacted by outages on a per-capita basis.

One of the most time-consuming and unexpected challenges was introduced late in the project, when we decided to make the cartogram reactive to the timeline selection. Originally, the cartogram was static and intended to show 2020 demographic data only. But we later introduced interaction with the timeline so users could filter outage data by time and see the cartogram update accordingly. Because we size-encoded the tiles by the number of people affected by outages, adjusting the time range also altered each tile's size. This made it very difficult to preserve a stable layout. In some filtered views, tile sizes shrank or expanded significantly, disrupting alignment and visual balance. We spent a lot of time adjusting the layout and spacing to make the cartogram look natural for as many cases as possible, but it never felt completely reliable across all ranges.

Looking back, our original proposal was ambitious to the point of being unrealistic in some areas. We underestimated the performance limits of D3, the complexity of handling multiple encodings simultaneously, and the visual challenges of small-multiple design. If we were to start again, we would invest more time in early prototyping, conduct layout stress tests sooner, and simplify the coupling between views. Still, the iterative process was immensely valuable: we learned to adapt our design to technical constraints, reconsider how to encode data most meaningfully, and prioritize usability and interpretability over maximal complexity. Despite the compromises, we're proud of the final product, which supports deep engagement and reflection on how outages affect communities differently across the U.S.

8. Work Breakdown and Schedule

Each project component had designated contributors/owners responsible for ensuring its accuracy and completeness. Contributors were also responsible for writing the corresponding sections in the report. Any additional report components not explicitly listed below were completed by their respective contributors (e.g., William wrote the 'Data' section, and Mo wrote the 'Task' section). All team members participated in editing the report. The table below outlines the primary contributors for each component, along with completion dates and hours spent:

Milestone 1:

Component	Contributor(s)	Time Spent	Completion Date
Dataset selection	All	0.5 hr per person	2025/01/30
Brainstorm purpose/ domain tasks	All	1.5 hours	2025/01/30
Task Abstraction	Mo	1 hour	2025/02/02
Data Characterization *(including starting on data preprocessing)*	William	3 hours	2025/02/02
EDA for power outages dataset	Alethea	2 hours	2025/02/01
EDA for map in tandem with outages dataset	William	1 hour	2025/02/03
EDA for census data *(including preliminary data transformation)*	Muna	3 hours	2025/02/02
Report Writing - Overview	Mo	45 mins	2025/02/02
Report Writing - Team communication Plan	Alethea	15 mins	2025/02/01

Report Writing - Work Breakdown	Muna	15 mins	2025/02/03
Report Editing	Muna	45 mins	2025/02/03
	Mo	30 mins	2025/02/03

Milestone 2:

Component	Contributor(s)	Time Spent	Completion Date
Brainstorming Visualization Ideas & Addressing TA Feedback	All	3 hours	2025/02/18
Starting on Work Breakdown and Schedule	William	45 mins	2025/02/20
Finishing Work Breakdown and Schedule	All	30 mins	2025/02/24
Clean up data section based on TA comments & new visualizations	William	15 mins	2025/02/24
Usage Scenario	Alethea	1.5 hrs	2025/02/24
Sketches for Visualizations (including drafts & revisions)	Muna	2.5 hrs	2025/02/24
High-level description of the visualization interface	Muna	15 mins	2025/02/24
Rationale for timeline & area chart	Mo	1 hr	2025/02/23
Rationale for choropleth map	William	1 hr	2025/02/24
Rationale for most complex view (tile grid map) & analysis of visual encoding	Muna	2 hrs	2025/02/24
Report Editing, Reviewing & double checking analysis of visual encoding	William	1.25 hrs	2025/02/24
	Mo	1 hr	2025/02/24
	Muna	1 hr	2025/02/24
	Alethea	45 min	2025/02/24
Meeting/communicating with TAs for feedback	Muna	1 hr	2025/02/20 to 2025/02/24

Work Completed for M3 and M4:

Note: Components marked as 'NA' were not accounted for in our initial M3 estimation

Component	Contributor(s)	Estimated Hours	Estimated Date	Actual Hours	Actual Date
Data preprocessing - cleaning racial data by dropping	Muna	2 hour	2025/02/28	2	2025/03/07

unnecessary columns and grouping non-white data together					
Data preprocessing - merging racial and outage data	William	3 hour	2025/02/28	2	2025/03/06
Data preprocessing - grouping outage data together by month	Alethea	3 hour	2025/03/03	3	2025/03/06
Data preprocessing - filling in missing outage data for all months and counties	Mo	N/A	N/A	1	2025/03/08
Data preprocessing - aggregating racial data by state + **EDA** + **colour scheme** research	Muna	N/A	N/A	5.5	2025/03/08
Repo structuring and initialization - transform the template to a structure intended for our project	Mo	4 hours	2025/03/03	1	2025/03/07
Static version of timeline & area chart	Alethea	5 hours	2025/03/07	6	2025/03/06
Static choropleth map	Mo	6 hours	2025/03/07	6	2025/03/08
Static racial demographic pie chart	Muna	4 hours	2025/03/07	6hrs	2025/03/10
Static version of tile grid map	William	6 hours	2025/03/07	7	2025/03/08
Combining demographic pie charts and tile grid map together	Mo	4 hours	2025/03/09	N/A	N/A
Bug Fixing/Testing for M3	William	N/A	N/A	3	2025/03/10
M3 - WIP DUE			2025/03/10		
Data preprocessing - population data by county by year	Mo	N/A	N/A	2	2025/03/17
Choropleth map - normalize outages by county population	Mo	N/A	N/A	3	2025/03/17
Add interactivity for timeline/slider + grey out when cartogram selected	Alethea	4 hours	2025/03/14	4	2025/04/05
Update title to switch between map views and dynamically show time periods	Alethea	N/A	N/A	2	2025/04/03
Change Timeline from Bar Chart into Area Chart	Mo	N/A	N/A	3	2025/04/05

Add interactivity for choropleth map	Mo	4 hours	2025/03/16	1	2025/03/17
Add bidirectionality between choropleth map and timeline	Alethea	4 hours	2025/03/18	5	2025/04/05
Debugging slow map	Alethea	N/A	N/A	6	2025/04/05
Adding interactive widget to select between choropleth map and tile grid map view	William	3 hours	2025/03/18	2	2025/04/04
	Mo	N/A	N/A	1.5	2025/04/06
Add interaction from timeline to cartogram	William	N/A	N/A	3	2025/04/05
Add consistent order for racial data in pie chart	Muna	N/A	N/A	1 hr	2025/04/03
Write steps to create cartogram dataset for data preprocessing	William	N/A	N/A	1 hr	2025/04/06
	Muna	N/A	N/A	1	2025/04/07
Buffer period I		2025/03/19 to 2025/03/21			
Pie chart colour and data fixes	Muna	N/A	N/A	2 hrs	2025/04/03
Initializing overall chart and adding disclaimers	Muna	2 hour	2025/03/22	7 hrs	2024/04/06
Add tooltips cartogram	Muna	2 hours	2025/03/25	3 hrs	2024/04/06
Legends for cartogram + pie charts	Muna	N/A	N/A	5 hrs	2024/04/05
Touchup: Adjusting Layout of Cartogram to look neater	William	N/A	N/A	2	2025/04/07
	Muna	N/A	N/A	1	2025/04/07
Choropleth: Add state code to tooltip	Mo	N/A	N/A	0.5	2025/04/03
Choropleth: Add ability to select by state or by county	Mo	N/A	N/A	6	2025/04/05
Timeline: Add snapping to timeline selection	Mo	N/A	N/A	1.5	2025/04/07
Choropleth: Update tooltip to selected date range	Mo	N/A	N/A	2	2025/04/07
Finishing touches to chart (combined)	Mo	6 hours	2025/03/28	6	2025/04/05
Buffer period II		2025/03/29 to 2025/04/05			
Creation of Demo Video	Muna	N/A	N/A	5	2025/04/07

Bug Fixing/Testing for M4	All	3 hours	2025/04/06	3	2025/04/07
Finalize Report for M4	All	10 hours	2025/04/07	3	2025/04/07
Start on Reflections + Finish Credits	William	N/A	N/A	4	2025/04/07
Finish Reflection	Alethea	N/A	N/A	1	2025/04/07
Update Usage scenario to actual implementation	Alethea	N/A	N/A	1	2025/04/07
Adding screenshots and editing report	Alethea	N/A	N/A	3	2025/04/07

Hourly contribution per team member:

Name	Milestone 1	Milestone 2	Milestone 3 + 4	Total contribution per person (hr)
Alethea Kramer	**4.25 hrs**	**5.75 hrs**	**34 hrs**	**44 hrs**
Muna Ibrahim	**6 hrs**	**10 hrs**	**41.5 hrs**	**57.5 hrs**
Mo Fardinzaman	**4.25 hrs**	**5.5 hrs**	**37.5**	**47.25 hrs**
William Ho	**6 hrs**	**6.25 hrs**	**30 hrs**	**42.25**
Total Team	**22.5 hrs**	**27.5 hrs**	**130**	**192 hrs**

9. Credits:

ChatGPT was used for assistance to create the legends, disclaimer boxes and for help bug fixing. Specifically, it used to assist in the creation of the following:

- text-wrapping in disclaimer/instruction boxes for both the choropleth and cartogram
- Creating text labels for the State colour legend in the cartogram
- Creating bent ticks for State size legend in the cartogram

https://observablehq.com/@mcmcclur/simple-grid-cartogram

The initial grid cartogram was based on the simple grid cartogram created by Mark McClure. Because we also added size encoding to the cartogram, we decided to move around some States to make it look better. The States that are only diagonally adjacent to another State in particular (Florida, and Maine) were moved to be directly beside another State. We also moved Alaska and Hawaii to match the location where they would be seen in an AlbersUSA Projection to match with the choropleth map. We also added the District of Columbia as an extra gridtile, because while it isn't technically a State, it has significant differences in terms of demographic data and outage events compared to other States, so it looked noteworthy to include it.

https://www.students.cs.ubc.ca/~cs-436v/22Jan/fame/projects/project_g03/index.html

The project “Visualizing YouTube Trending Videos” was used as inspiration for the timeline. We originally created a bar chart for the timeline, but after seeing this project, we were inspired to use an area chart for the timeline as it looked a lot cleaner. We didn’t have the source code for this project, so we created our own area chart from scratch. We also made the y-ticks so the amount of data being shown at each point is much clearer.

https://observablehq.com/@d3/focus-context?collection=@d3/d3-brush

The brush on the timeline was based on Mike Bostock’s brush for D3. We decided to make this brush filter the data to only include data that is within the brush. Because we are filtering data and rerendering the maps, the rendering of the maps can get fairly slow because of the amount of data processed. To reduce this lag, the brush was made to only update the map when you release the brush instead of updating live.

https://codepen.io/trevald-the-scripter/pen/eYpYgMp

The radio buttons for the map view selector and county selector were inspired by Benjamin Holfve’s segmented controls created in CSS. We changed the text within the controls to relate to our project. We also added an image for a reset button to reset the county selection.

10. Bibliography

[1] Bostock, Mike. "Focus + Context," Observable. December 21, 2022. https://observablehq.com/@d3/focus-context?collection=@d3/d3-brush

[2] Holfve, Benjamin. "iOS 13 Segmented controls CSS only", April 9, 2020. https://codepen.io/trevald-the-scripter/pen/eYpYgMp

[3] Maurer, Annie, Julian Platel, and Jonathan Hirsch. “Visualizing Youtube Trending Videos” April 12, 2022 https://www.students.cs.ubc.ca/~cs-436v/22Jan/fame/

[4] McClure, Mark. “Simple grid cartogram” Observable, Jan. 6, 2024. https://observablehq.com/@mcmcclur/simple-grid-cartogram

[5] P. Madamba, “About 1 in 4 Households Experienced a Power Outage in the Span of a

Year," census.gov. Accessed: Feb. 02, 2025. [Online]. Available: https://www.census.gov/library/stories/2024/10/power-outages.html

[6] Wikipedia contributors. "County (United States)," Wikipedia. Accessed: Feb. 22, 2025. [Online]. County (United States). Available: https://en.wikipedia.org/wiki/County_%28United_States%29