Milestone 2: Design

1. Basic Info

Project Title: Mapping the Dark: An Interactive Visualization of U.S. Power Outages (2019–2023)

Group members:

Name	Student ID	Email
Alethea Kramer	15219141	aletheakramer@outlook.com
Muna Ibrahim	35763119	muna.ibr08@gmail.com
Mo Fardinzaman	57779167	mfardinzaman@gmail.com
William Ho	42307041	who64@students.cs.ubc.ca

2. Overview

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
Population of county (3221 county attributes in total)	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 Derived Variables

Derived Variables	Category	Cardinality/Range
Month	Ordinal	12
Year	Ordinal	5
Total Power Outage Events	Quantitative	[0, ~35 000]
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]

3.5 Data Preprocessing

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

3.5.1 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.

3.5.2 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.

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 ‘Geographic View’, which displays a timeline and choropleth map (Figure 1), and a ‘Tile View’ with a cartogram in which each state is represented by a square tile (Figure 2). The subsequent sections provide details on these views and the data they visualize.

In all views, 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 ‘Geographic View’) or a state (in ‘Tile View’) will display a tooltip containing its name and relevant statistics – such as the number of outages in the Geographic View, and number of affected people in the Tile View.

5.2 Timeline & Bar 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 bar 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 bar 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. A bar chart is ideal for representing discrete aggregated outage counts over time, allowing users to compare discrete outage counts across different months more effectively. Unlike line charts, which emphasize trends but may obscure event magnitudes, bar charts provide a direct and interpretable view of monthly outage frequencies. 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 bars encode monthly outage counts in their 1-directional vertical size channel while the 1-directional horizontal position channel encodes time. Hue is used to distinguish the selected time range, with bars outside the range appearing in a lighter shade to indicate they are currently not in focus.

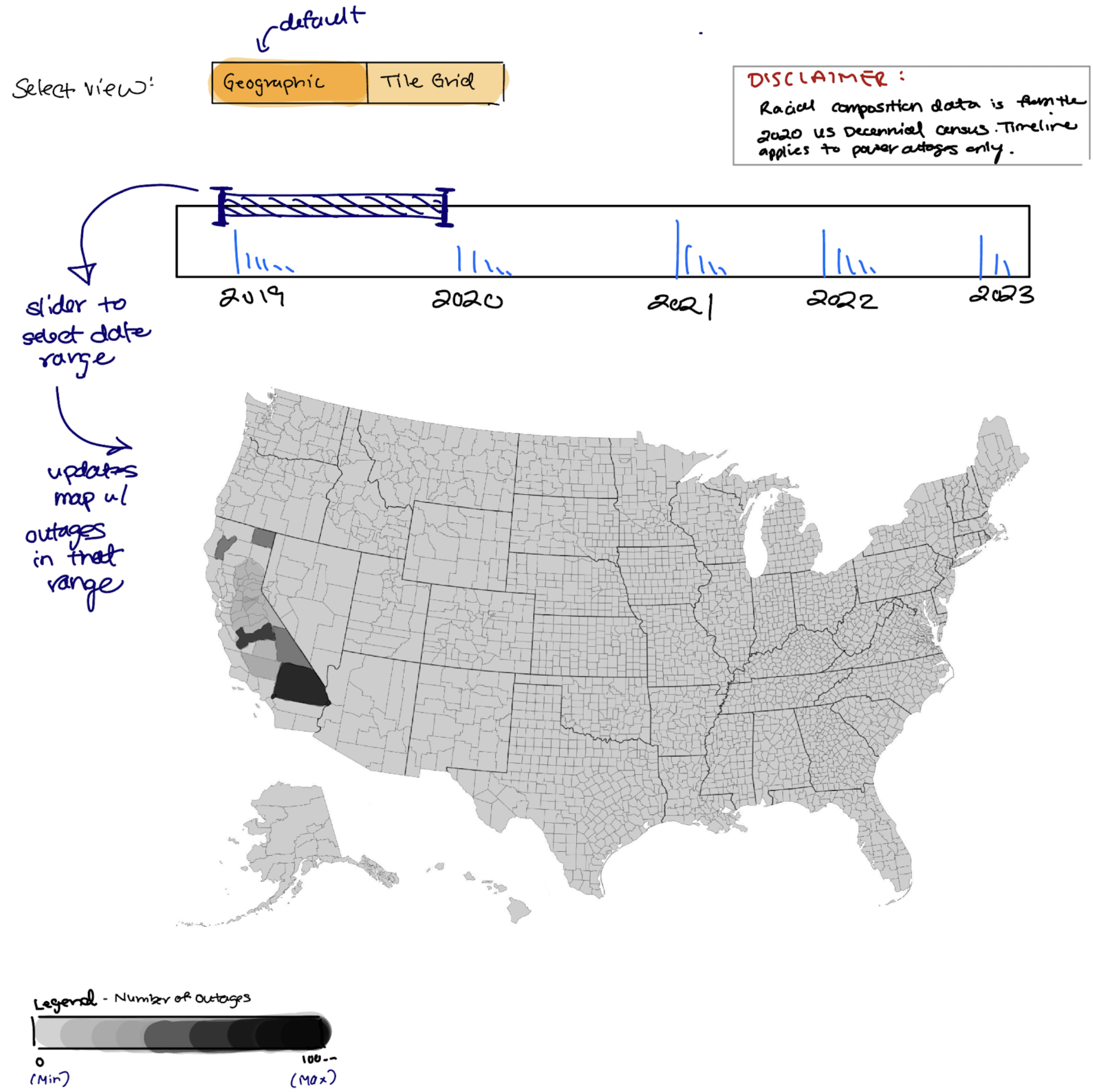

Figure 1: Sample sketch of timeline/barchart (view 1) and choropleth map (view 2). Map of the US is adapted from Wikipedia [2]. A portion of California has been filled in as an example. Note that this sketch is provided as a demonstration only, and does not reflect accurate data.

5.3 Choropleth Map

The choropleth map allows users to explore the distributions of power outages between counties. The colour channel, ranging from white to black, 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 bar chart 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. 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.

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.

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 number of people affected by power outages.
   - **Colour** indicates the ethnic majority of affected individuals (ie 'white' vs 'non-white').

   Tiles are positioned in proximity to their geographic location. 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.

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 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 seven racial groups described in Section 3.5.2. We recognize that displaying seven groups may pose challenges in accurately interpreting the data. Therefore, we plan to test this approach, and if it proves ineffective, the pie chart will be simplified to show only the two main racial categories: White and Non-White.

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. **Tooltip**: When the mouse hovers over a state, a tooltip displaying the state name, total number of power outages and the total number of affected people will appear.

A detailed marks and channels analysis is provided in section 5.4.2

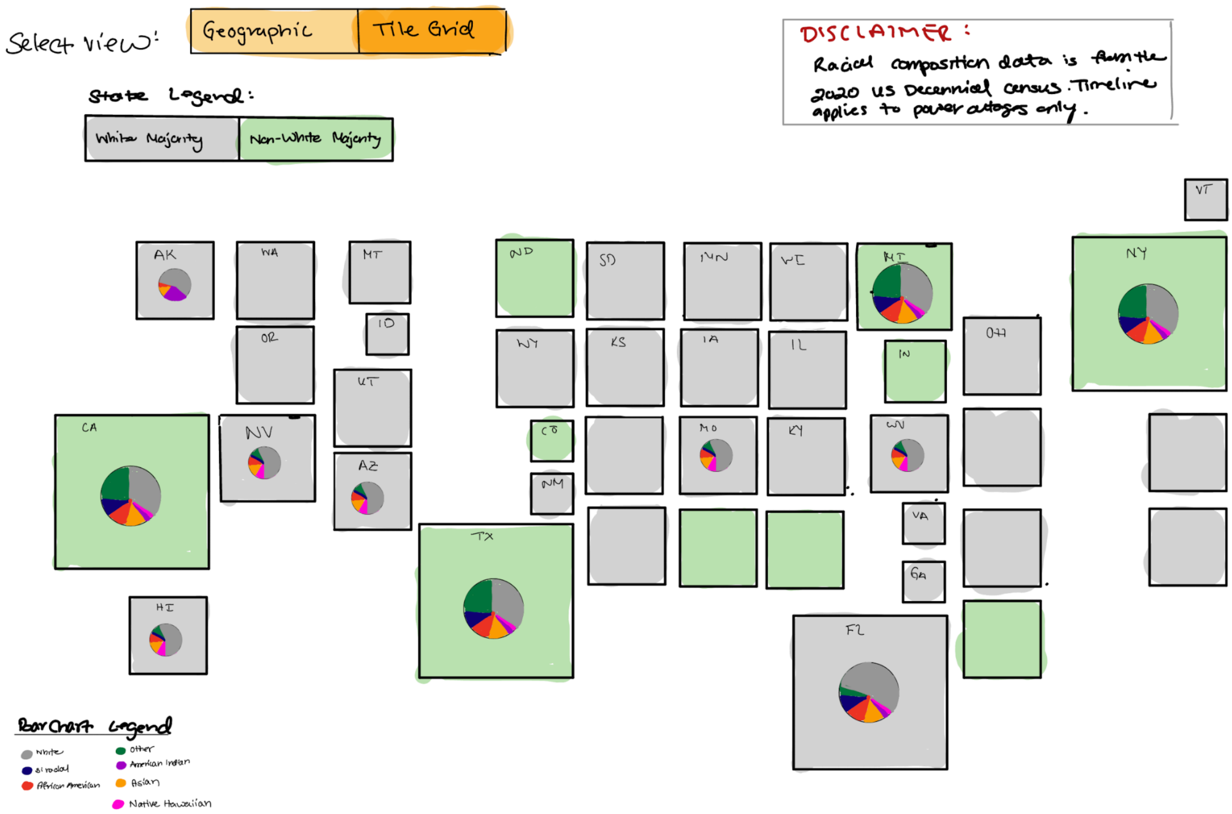

Figure 2: Sample sketch of the tile grid map with embedded pie charts (view 3). Not all states and pie charts are depicted in the sketch. *Legend text and colour encoding are subject to change; pie charts are not to scale*. Note that this sketch is provided as a demonstration only, and does not reflect accurate data.

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	no
	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	no
	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	yes	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 barcharts 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 bar 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:** Bar 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 zooms in on Texas, scanning for her county. Right away, she notices that her county is shaded darker than some of the surrounding areas, suggesting it has had more outages. She wonders if this is just a recent trend or if her county has always had frequent outages, so she uses the timeline selector filter above the map. She selects a time period from the slider and the map updates to show the number of outages during that time.

As she scans the timeline, Jessica sees that outages in her area seem to follow a pattern—spiking in the summer and again in the winter. This makes sense to her, given the extreme weather Texas sometimes experiences. But she's curious whether her county is just unusually bad compared to other areas. She zooms out slightly and looks at neighboring counties, noticing that some of them have lighter shading. Hovering over one of them, 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 largest tiles, 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.

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 the states with fewer outages, and she wonders if certain communities within her state experience more frequent outages than others. She decides to switch back to the choropleth map, focusing on counties with different racial compositions to see if any patterns emerge. 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

[leave blank for now]

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 & bar 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 to be Completed:

Component	Contributor(s)	Estimated Hours	Estimated Date	Actual Hours	Actual Date
Data preprocessing - cleaning racial data by dropping unnecessary columns and grouping non-white data together	Muna	2 hour	2025/02/28	2	2025/03/07
Data preprocessing - merging racial and outage data	William	3 hour	2025/02/28		
Data preprocessing - grouping outage data together by month	Alethea	3 hour	2025/03/03		
Repo structuring and initialization - transform the template to a structure intended for our project	Mo	4 hours	2025/03/03		
Static version of timeline & bar chart	Alethea	5 hours	2025/03/07		
Static choropleth map	Mo	6 hours	2025/03/07		
Static racial demographic pie chart	Muna	4 hours	2025/03/07		
Static version of tile grid map	William	6 hours	2025/03/07		
Combining demographic pie charts and tile grid map together	Mo	4 hours	2025/03/09		
M3 - WIP DUE			2025/03/10		
Add interactivity for timeline	Alethea	4 hours	2025/03/14		
Add interactivity for choropleth map	Mo	4 hours	2025/03/16		
Add bidirectionality between choropleth map and timeline	Alethea	4 hours	2025/03/18		
Adding interactive widget to select between choropleth map and tile grid map view	William	3 hours	2025/03/18		
Buffer period I		2025/03/19 to 2025/03/21			
Initializing overall chart and adding disclaimer for racial data	Muna	2 hour	2025/03/22		
Add tooltips	Muna	2 hours	2025/03/25		
Finishing touches to chart	Mo	6 hours	2025/03/28		
Buffer period II		2025/03/29 to 2025/04/05			

Testing/Bug Fixing	All	3 hours	2025/04/06		
Finalize Report for M4	All	10 hours	2025/04/07		

Hourly contribution per team member:

Name	Milestone 1	Milestone 2	Total contribution per person (hr)	Estimated future work (hr)
Alethea Kramer	**4.25 hrs**	**5.75 hrs**	**10 hrs**	**29 hrs**
Muna Ibrahim	**6 hrs**	**10 hrs**	**16 hrs**	**20 hrs**
Mo Fardinzaman	**4.25 hrs**	**5.5 hrs**	**9.75 hrs**	**37 hrs**
William Ho	**6 hrs**	**6.25 hrs**	**12.25 hrs**	**25 hrs**
Total Team	**22.5 hrs**	**27.5 hrs**	**50 hrs**	**111 hrs**

9. Credits

[leave blank for now]

10. Bibliography

[1] 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

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