Power Outage Trends in California, Florida, Pennsylvania, and Texas from 2015 to 2023

https://github.com/sarahj-hall/Antonucci_Hall_Huang_Kuehn_ENV872_FinalProject.git

December 12th, 2025

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Write 1-2 paragraph(s) detailing the rationale for your study. This should include both the context of the topic as well as a rationale for your choice of dataset (reason for location, variables, etc.). You may choose to include citations if you like (optional).

At the end of your rationale, introduce a numbered list of your questions (or an overarching question and sub-questions).

Electric power outages are costly, for both grid operators and customers. (expand/give examples and cite). With increasing extreme weather events, the U.S. grid is more vulnerable to outage events. However, there have also been advancements in electric power systems planning, operations, and outage detection. Understanding what trend is occurring in the U.S. with regards to power outages could support.... In additional to the general trend, understanding if power outages follow a seasonal trend, that aligns with extreme weather periods (such as hurricane season), could help states focus their planning (resiliency?) for specific months. Furthermore, power outages may be increasing (or decreasing), but are their impacts the same? Are outage durations increasing? The outage duration helps quantify actual impact, as the longer duration, the more costly.

- damage of outages (health, financial)
- increasing climate related weather events, vulnerable power system infrastructure
- could be interesting trend/hard to estimate, could be increasing with more grid demand and aging power system + weather events or decreasing with increase grid planning and monitoring technology (or outage length should be decreasing)
- selected four states, tried to spread out geographically but also selected states with the highest power outages/most impacted by recent extreme climate events. Texas, California, Florida, and Pennsylvania

Four U.S. states were selected for analysis: California, Florida, Pennsylvania, and Texas. These states were selected to cover different geographical areas of the U.S., and due to their large populations. (and vulnerability to climate change?). For all states the following questions were explored:

Question 1: How has the frequency of power outages changed over time? Question 2: Is there a seasonal trend? Are certain months more prone to outages? Question 3: How has the length of power outages changed over time?

2 Dataset Information

Provide information on how the dataset for this analysis were collected, the data contained in the dataset, and any important pieces of information that are relevant to your analyses. This section should contain much of same information as the metadata file for the dataset but formatted in a way that is more narrative.

Describe how you wrangled your dataset in a format similar to a methods section of a journal article.

The Event-correlated Outage Dataset in America was downloaded from the Open Energy Data Initiative (OEDI) (https://data.openei.org/submissions/6458). The dataset includes an aggregated and event-correlated analysis of power outages in the United States. The specific dataset selected for this analysis is the Aggregated Outage Data which integrates data from the Environment for the Analysis of Geo-Located Energy Information (EAGLE-I), and Annual Estimates of the Resident Population for Counties 2024 (CO-EST2024-POP). The EAGLE-I dataset, provides county-level electricity outage estimates at 15-minute intervals from 2014 to 2023. It encompasses over 146 million customers, but this coverage has increased over time from 137 million in 2018. EAGLE-I only start providing . The Aggregated Outage Dataset provides monthly outage data at the state level, including total number of outages, the duration of outages, and the customer weighted average of outages.

The data was wrangled by combining the yearly data from 2015 to 2023 was combined into one data frame. The year 2014 was removed from the analysis because it did not have monthly data, only the yearly summary. From this file, four datasets were created by filtering for each state (CA, FL, PA, and TX). For each state, the monthly value equal to 0 was filtered out, which represented the yearly summary. Additionally, a date column was added that combined the monthly and yearly columns into a date object.

Add a table that summarizes your data structure (variables, units, ranges and/or central tendencies, data source if multiple are used, etc.). This table can be made in markdown text or inserted as a kable function in an R chunk. If the latter, do not include the code used to generate your table.

This data structure in Table 1 applies to all four state files:

Table 1: Summary of Outage Dataset Structure

Variable	Description	Units
state	Two-letter state abbreviation	N/A
year	Year of outage	N/A
month	Month of outage	N/A
outage_count	Number of outages per month	N/A
max_outage_duration	Longest outage duration in a month	Hours
customer_weighted_hours	Customer-weighted outage hours	N/A
date	Date of Outage	N/A

3 Exploratory Analysis

Insert exploratory visualizations of your dataset. This may include, but is not limited to, graphs illustrating the distributions of variables of interest and/or maps of the spatial context of your dataset. Format your R chunks so that graphs are displayed but code is not displayed. Accompany these graphs with text sections that describe the visualizations and provide context for further analyses.

Each figure should be accompanied by a caption, and each figure should be referenced within the text.

Scope: think about what information someone might want to know about the dataset before analyzing it statistically. How might you visualize this information?

3.0.1 California

Initial data exploration of California power outage data suggest a slight increasing trend (Figure 1). Furthermore, Figure 2 shows.

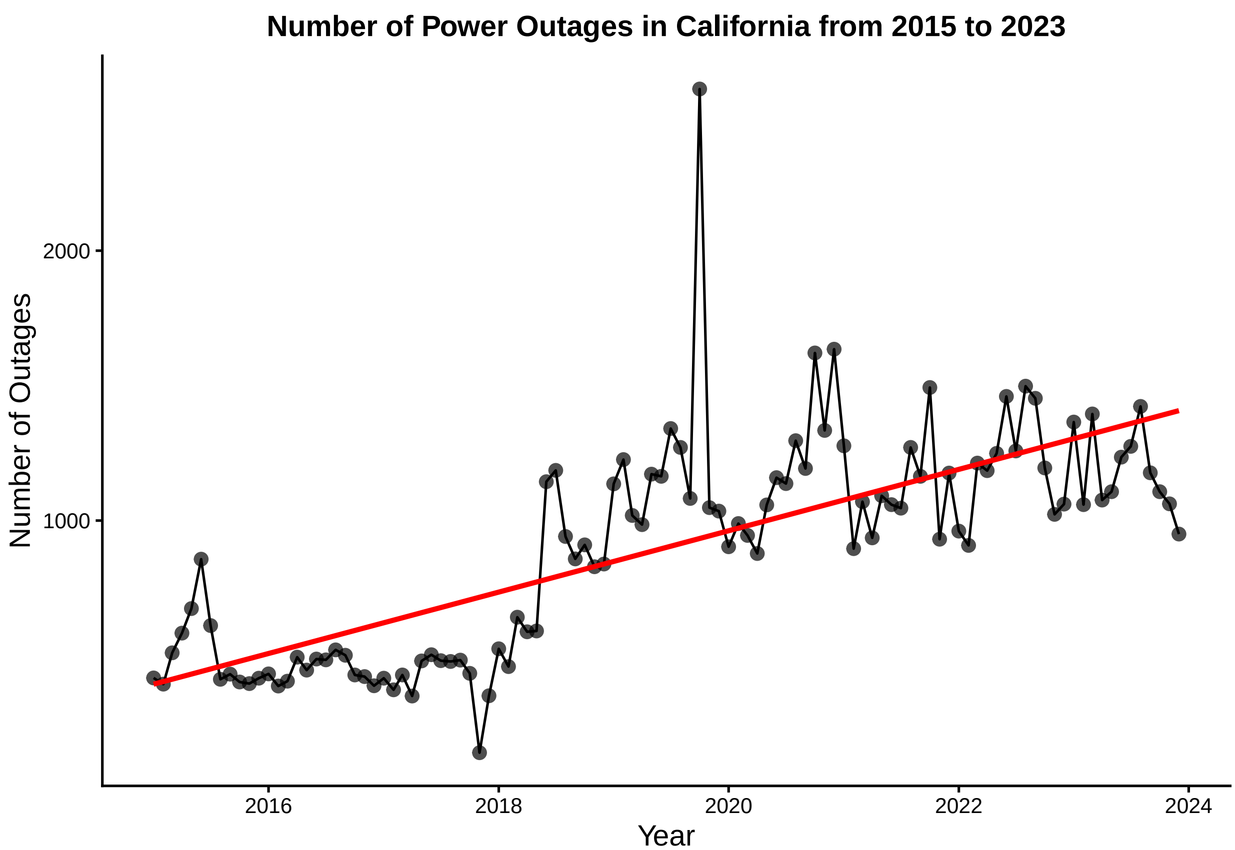

Figure 1: Yearly plot of power outages in California from 2015 to 2023.

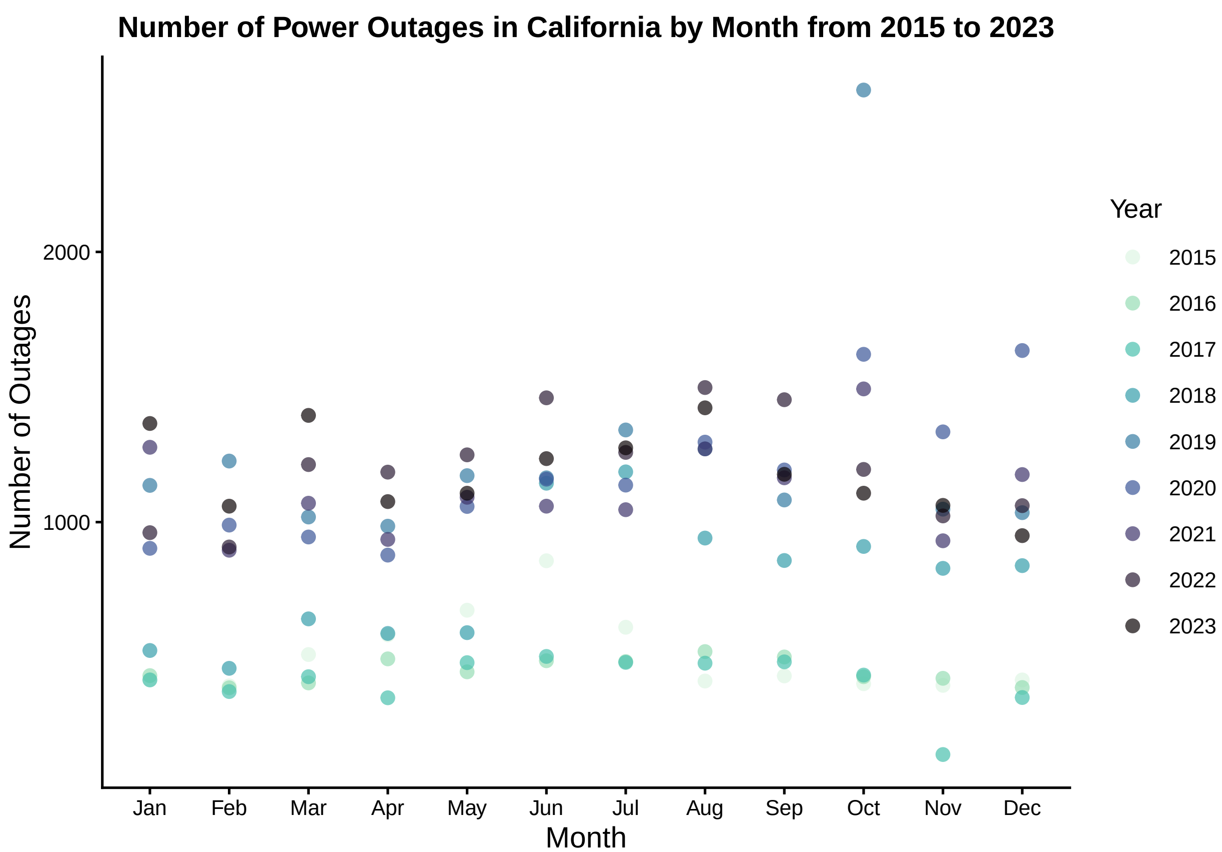

Figure 2: Monthly plot of power outages in California from 2015 to 2023.

3.0.2 Pennsylvania

From the initial data exploration of Pennsylvania, it is clear that power outages have gradually increased over the past decade. On a monthly basis, there is a noticeable rise in outages during the summer months, suggesting a seasonal pattern likely related to weather or energy demand. Overall, the data points to both long-term growth in outage frequency and predictable seasonal fluctuations.

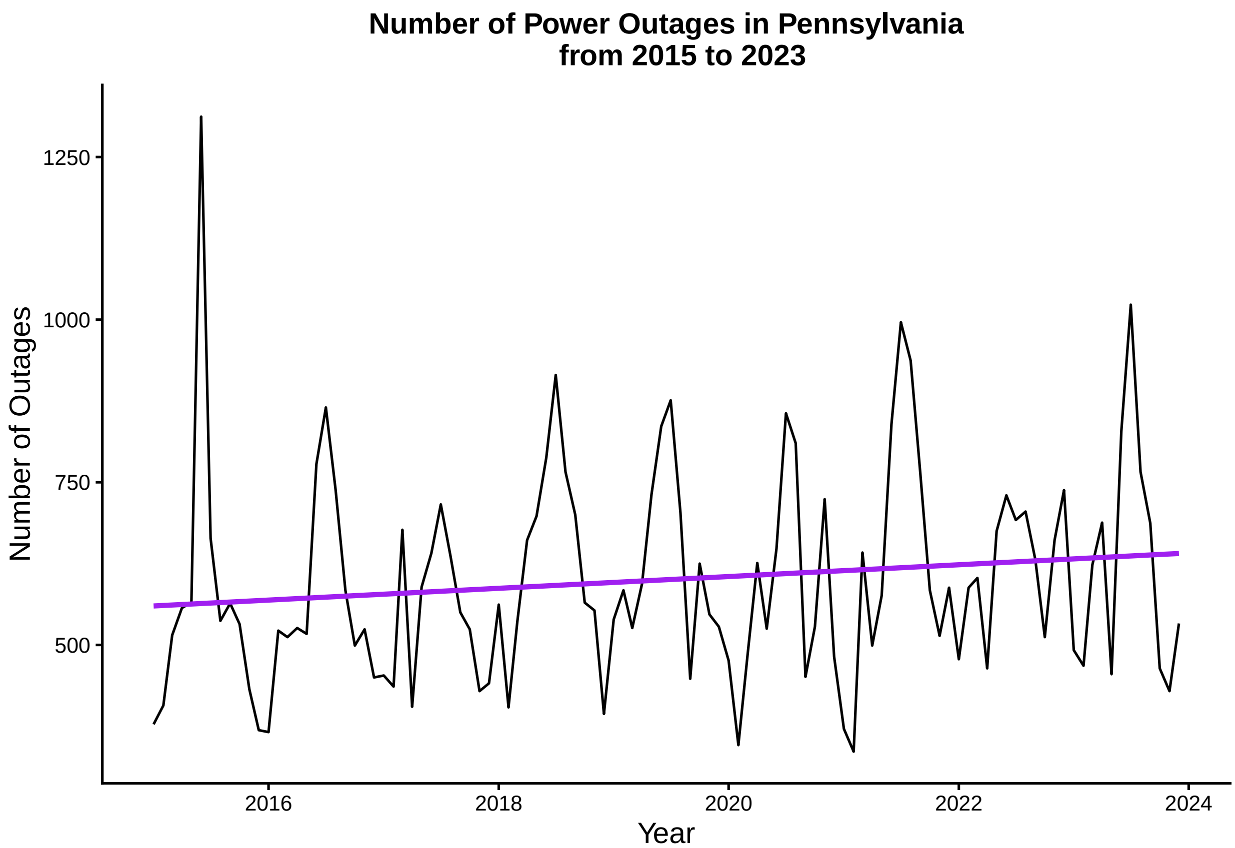

Figure 3: Yearly plot of power outages in Pennsylvania from 2015 to 2023.

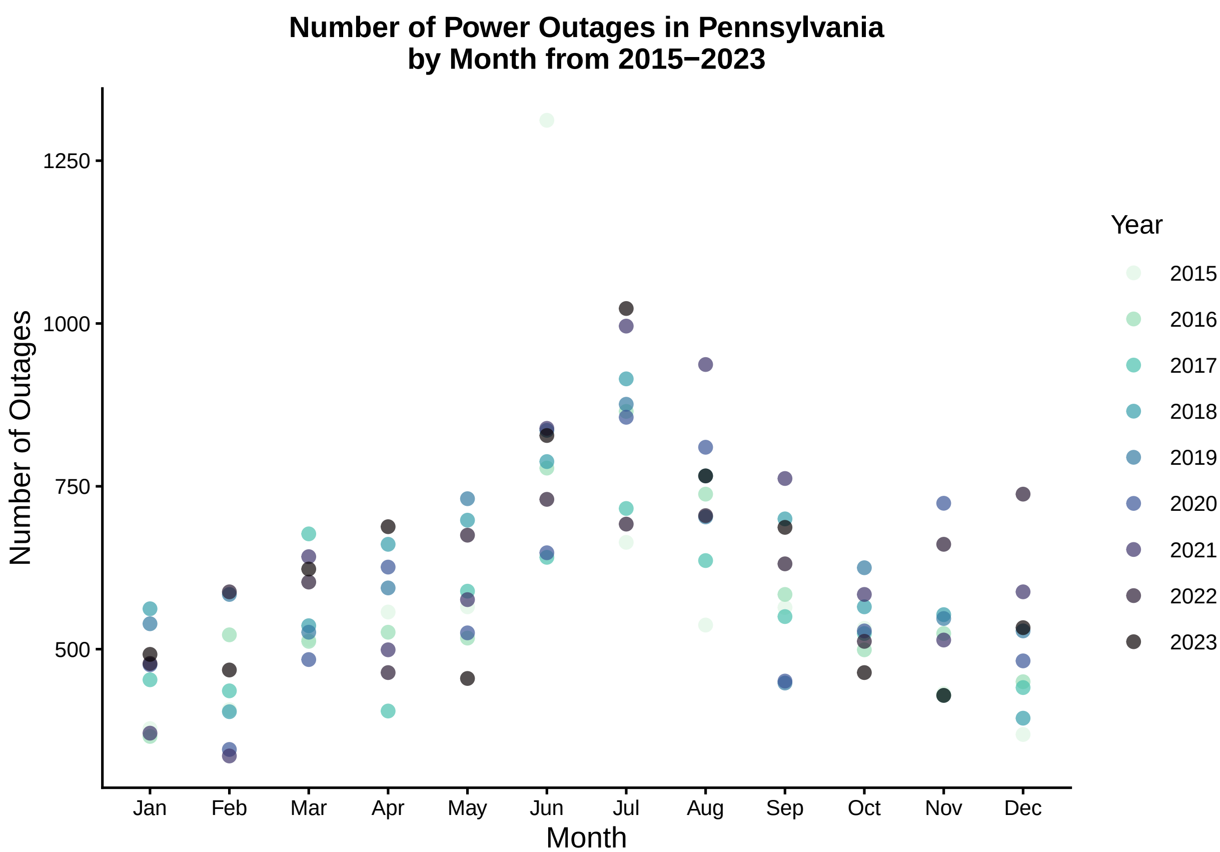

Figure 4: Monthly plot of power outages in Pennsylvania from 2015-2023.

3.0.3 Texas

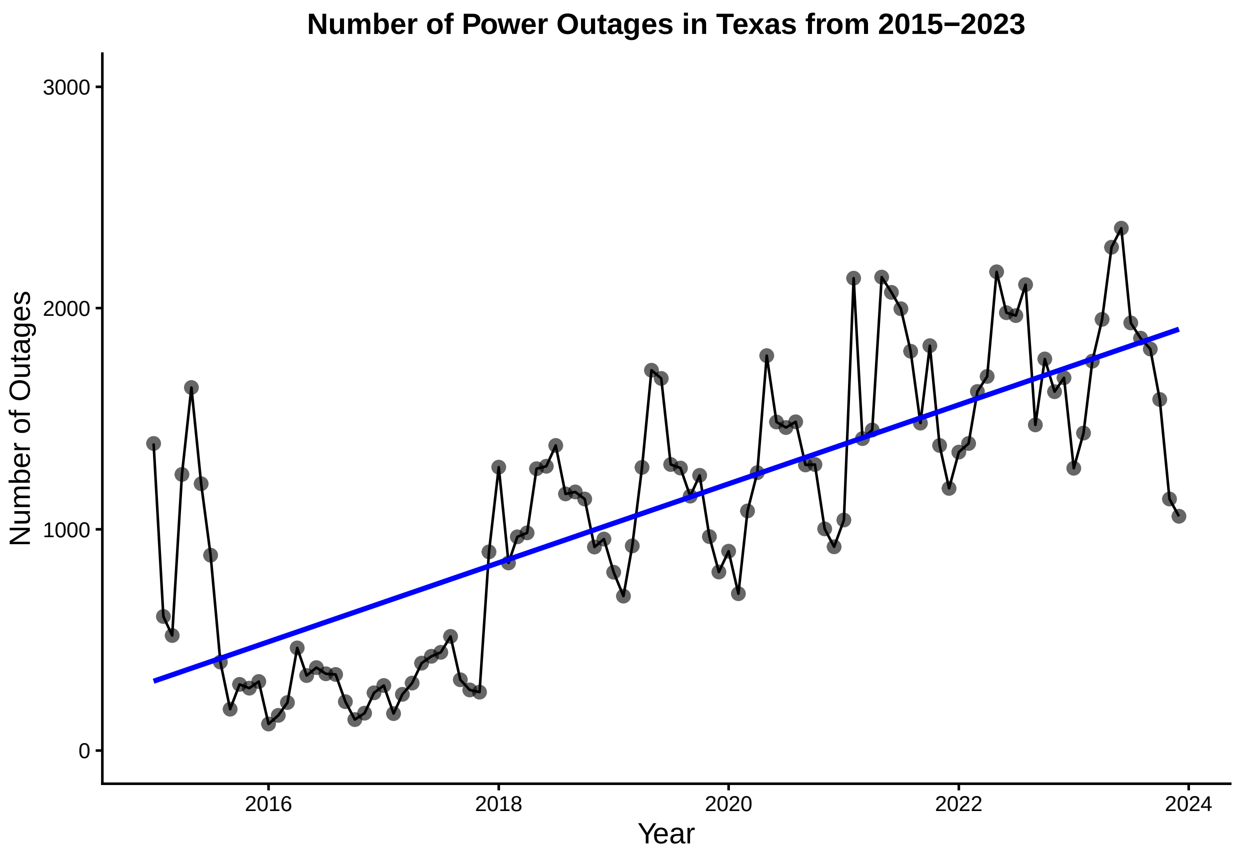

Figure 5: Yearly plot of power outages in Texas from 2015-2023.

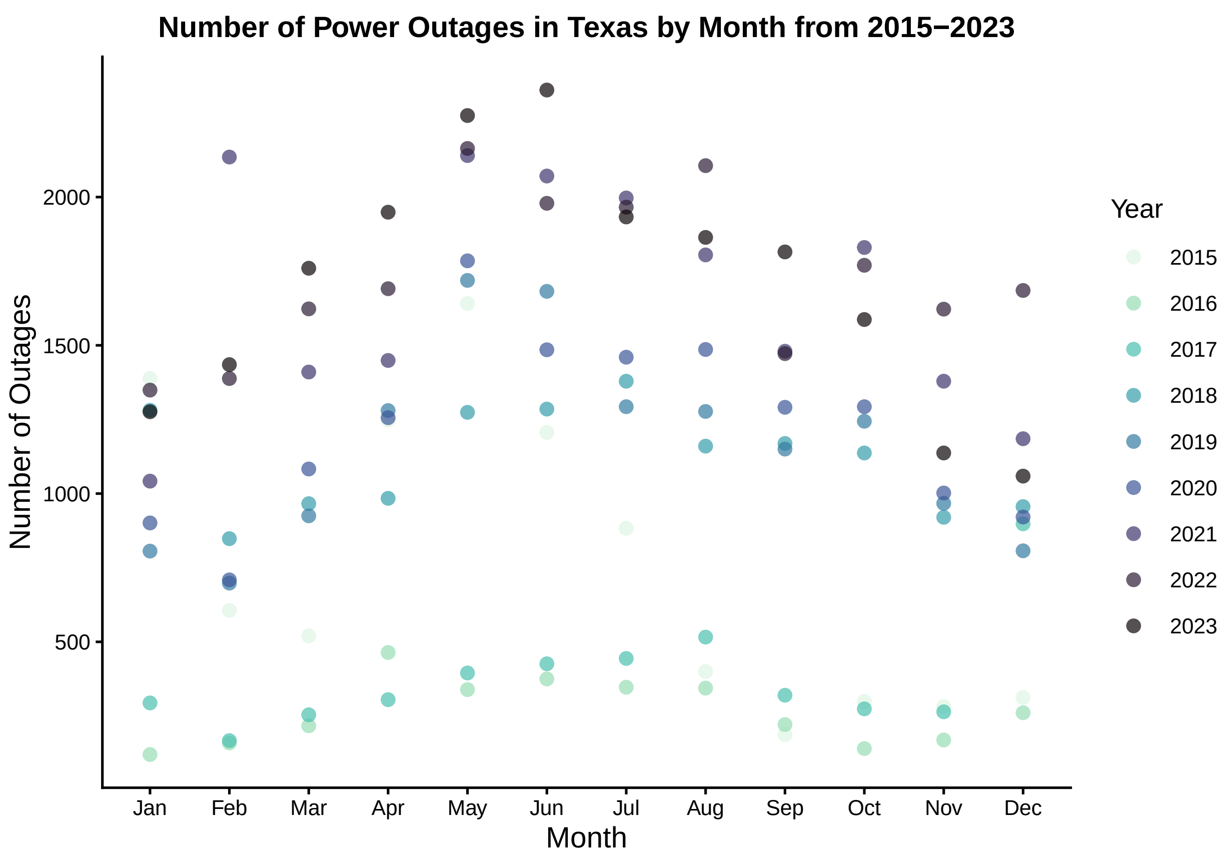

Figure 6: Monthly plot of power outages in Texas from 2015 to 2023.

3.0.4 Florida

4 Analysis

Insert visualizations and text describing your main analyses. Format your R chunks so that graphs are displayed but code and other output is not displayed. Instead, describe the results of any statistical tests in the main text (e.g., "Variable x was significantly different among y groups (ANOVA; df = 300, F = 5.55, $p < 0.0001$)"). Each paragraph, accompanied by one or more visualizations, should describe the major findings and how they relate to the question and hypotheses. Divide this section into subsections, one for each research question.

Each figure should be accompanied by a caption, and each figure should be referenced within the text

**trying to figure out the best organization for this.. where is the best place for the seasonality component?

4.1 Question 1: How has the frequency of power outages changed over time?

4.1.1 California

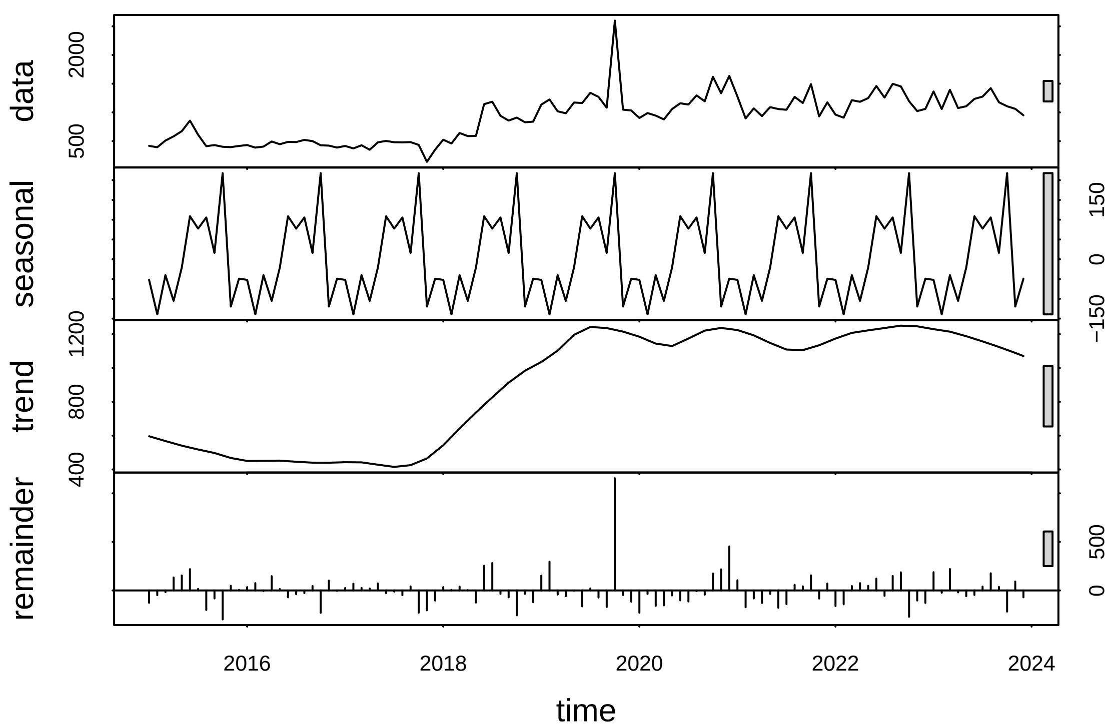

Figure 7: Decomposed components of the California power outage count time series.

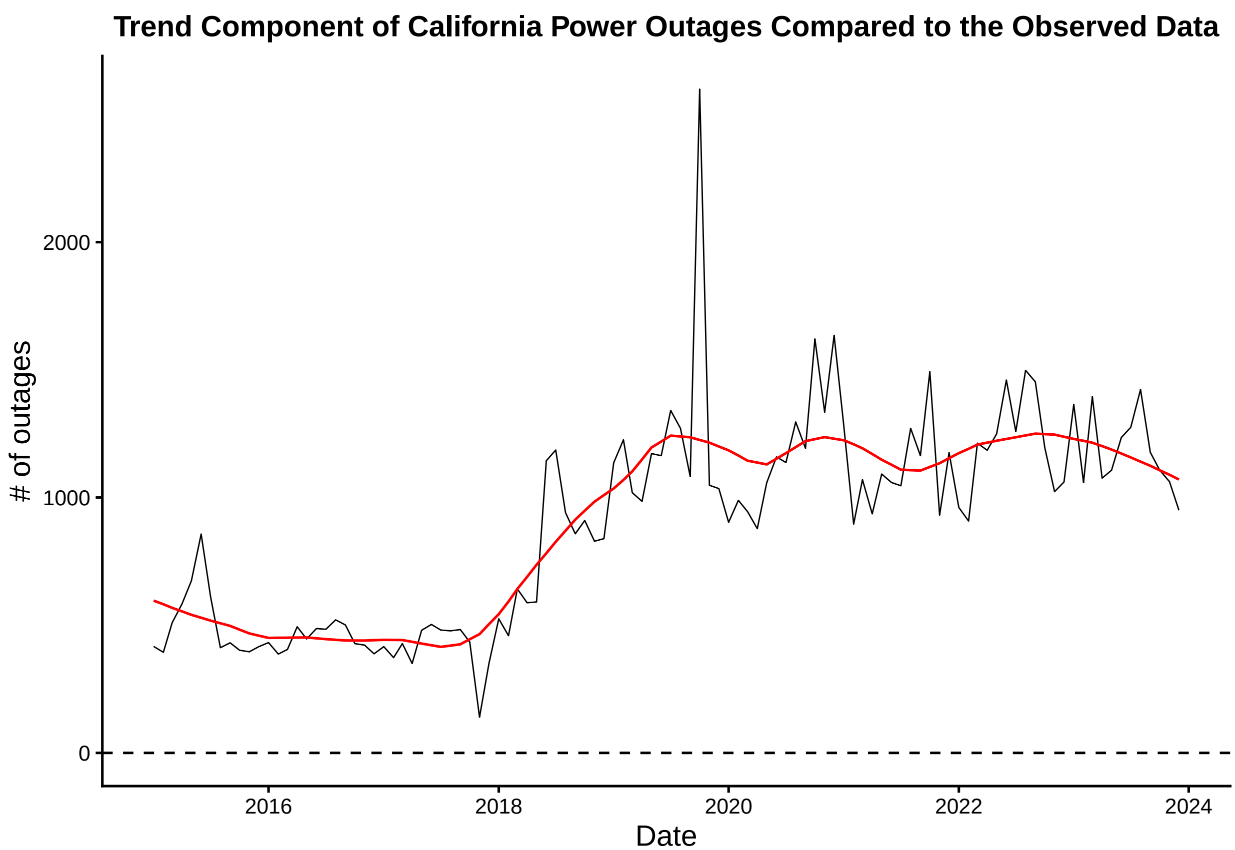

Figure 8: Trend versus observation of power outages in California (2015-2023).

4.1.2 Florida

4.1.3 Pennsylvania

After removing the seasonal component from the Pennsylvania outage time series, the Mann–Kendall trend test revealed a significant positive trend ($z = 3.02$, $n = 108$, $p = 0.002$), indicating that outage frequency has increased from 2015 to 2023. This result supports the hypothesis that outages have become more common during this time, independent of seasonal patterns. The decomposed time-series visualization shows this trend clearly, revealing that even after accounting for monthly fluctuations, the long-term component continues to rise.

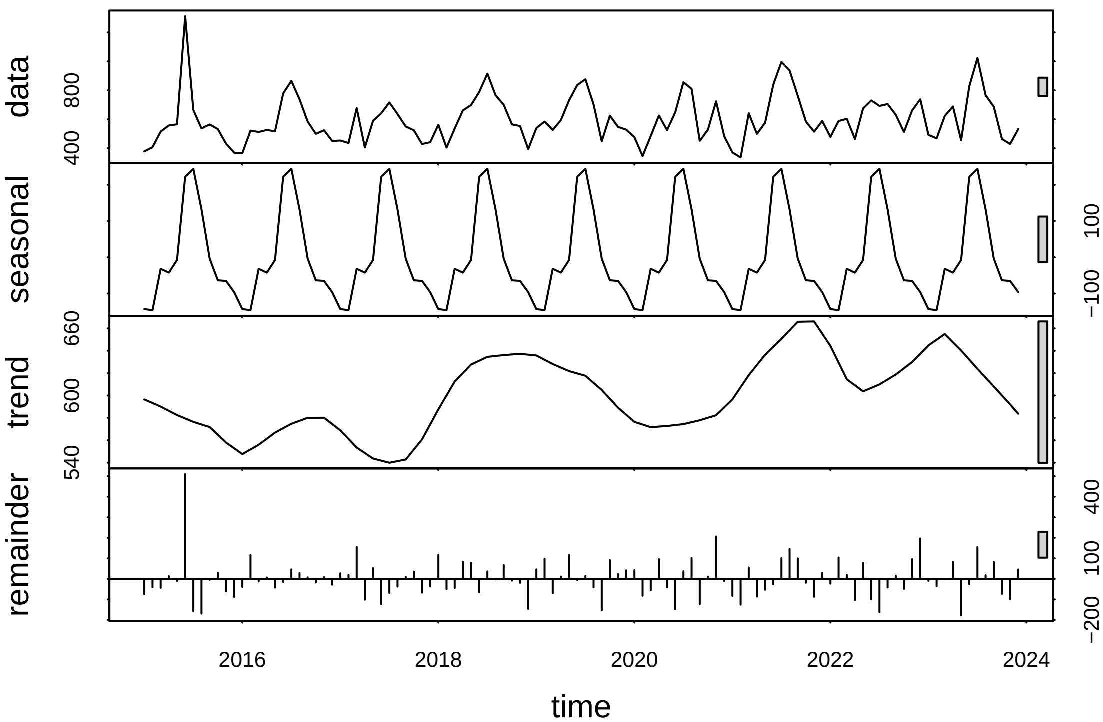

Figure 9: Decomposed components of the Pennsylvania power outage count time series.

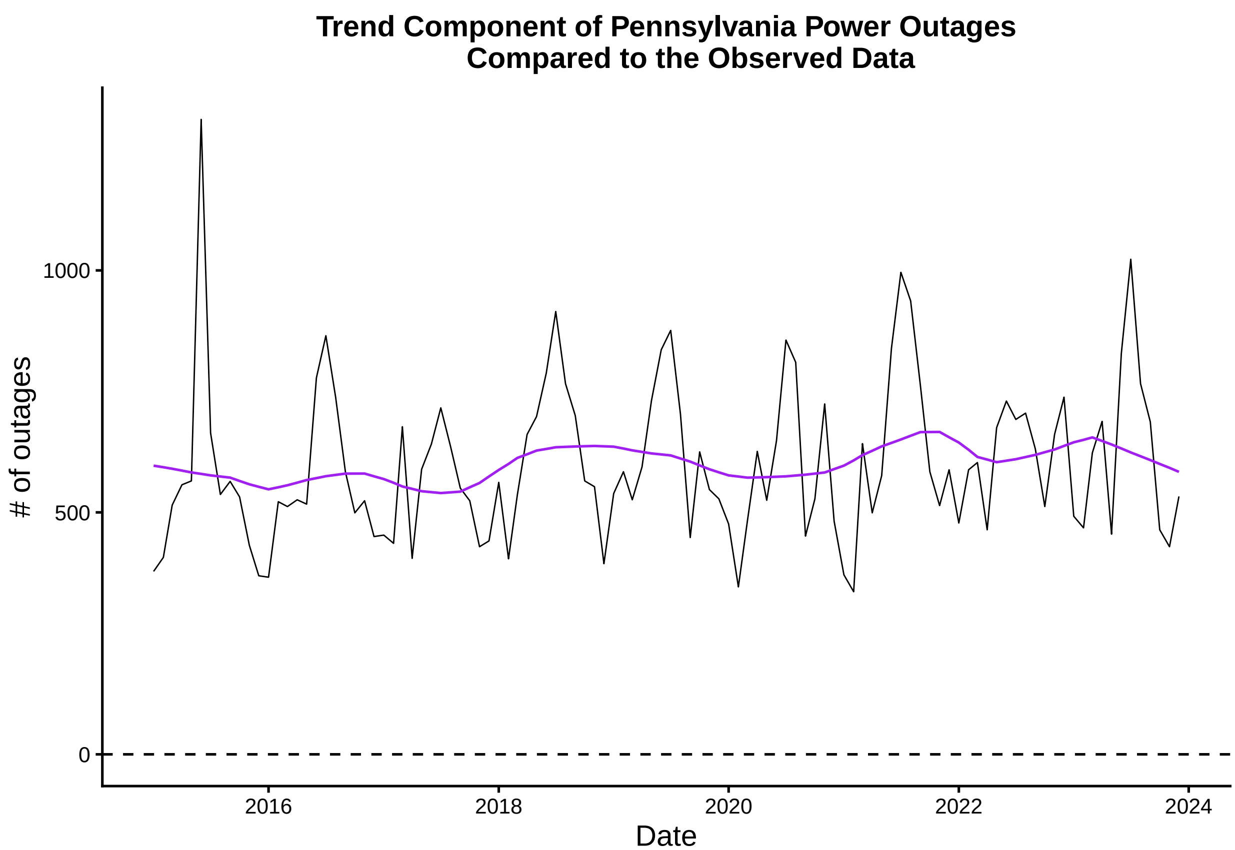

Figure 10: Trend versus observation of power outages in Pennsylvania (2015-2023).

4.1.4 Texas

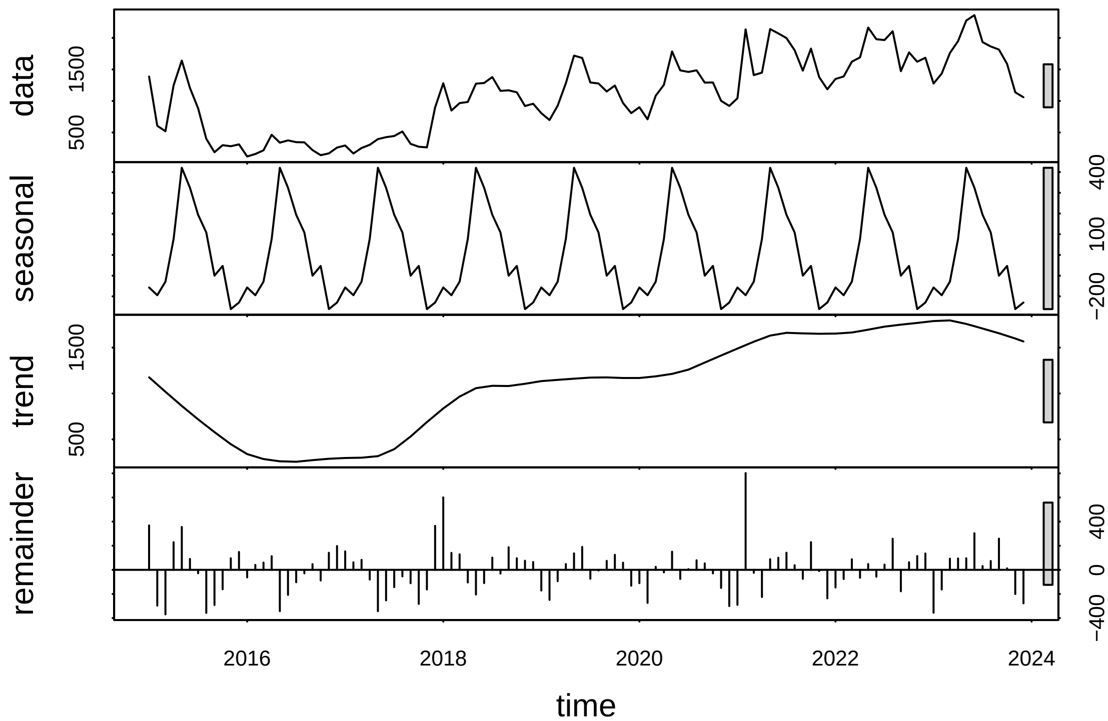

Figure 11: Decomposed components of the Texas power outage count time series.

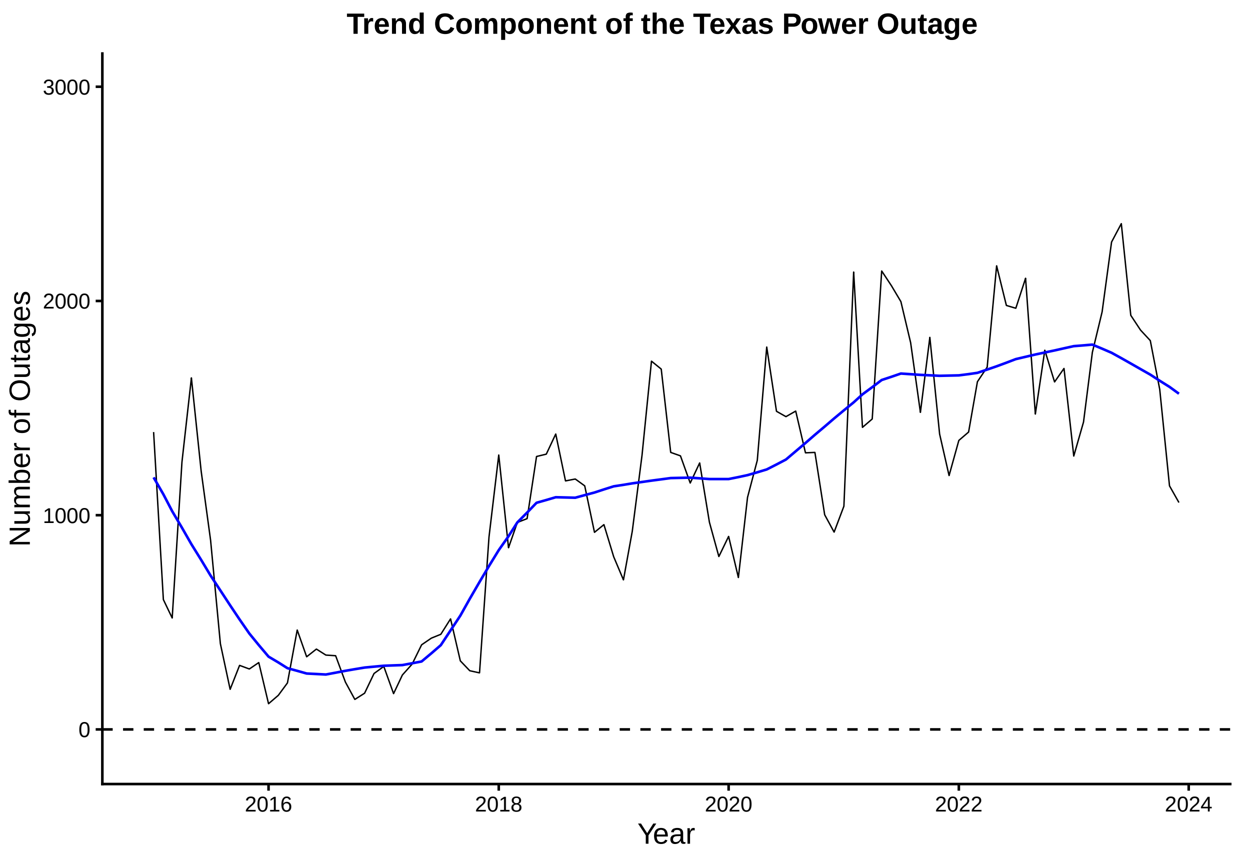

Figure 12: Trend versus observation of power outages in Texas (2015-2023).

4.1.5 Comparision between States

4.2 Question 2: Is there a seasonal trend? Are certain months more prone to outages?

4.2.1 California

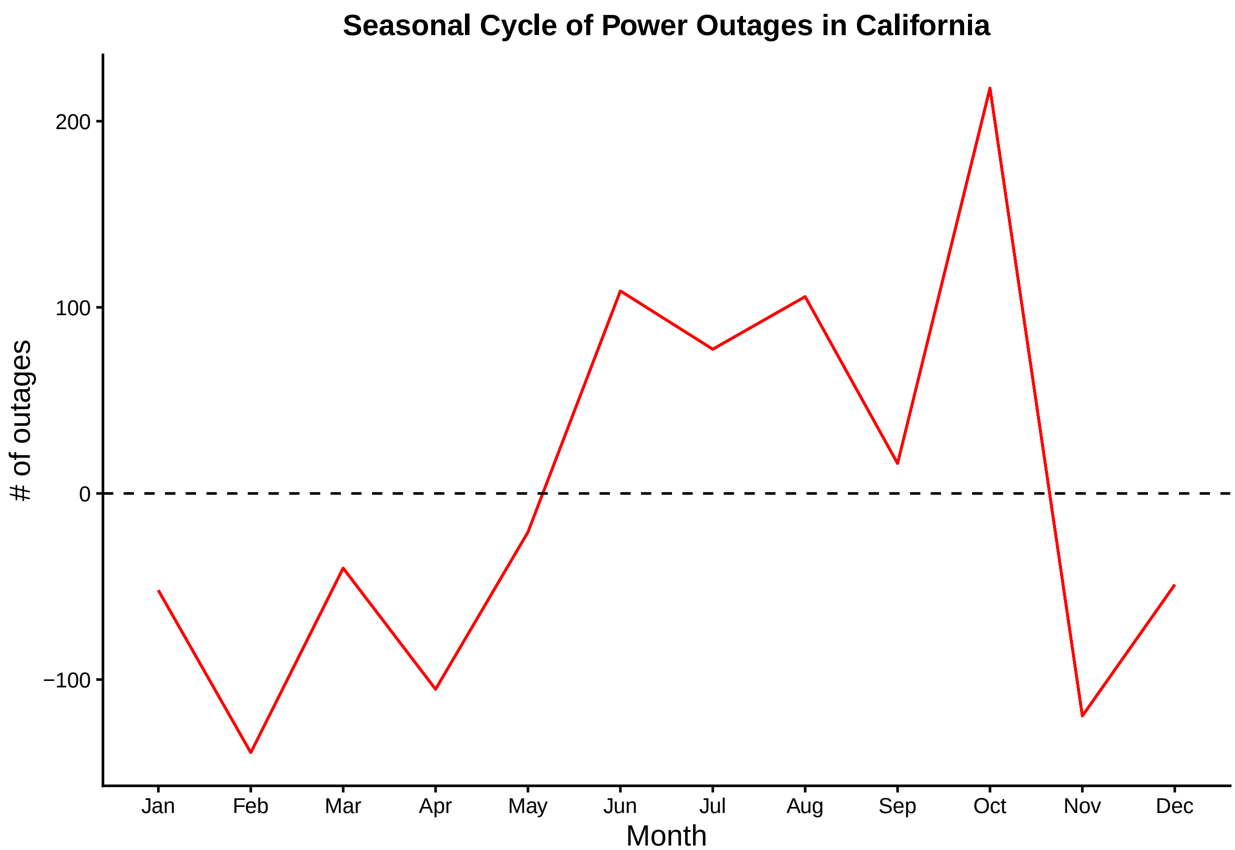

Figure 13: Seasonality component of the California power outage count time series analysis where yearly data is grouped by month.

4.2.2 Florida

4.2.3 Pennsylvania

The Mann–Kendall seasonal trend test for Pennsylvania outage counts indicated a significant seasonal signal in the data ($\tau = 0.20$, p = 0.009), demonstrating that outage frequencies vary across months. This finding supports the hypothesis that outages are not evenly distributed throughout the year but instead follow a seasonal pattern. The visualizations further highlight this pattern, showing clear peaks in June, July, and August, when outage counts are highest. Together, these results suggest that summer months consistently experience elevated outage activity, potentially driven by seasonal weather conditions or increased energy demand.

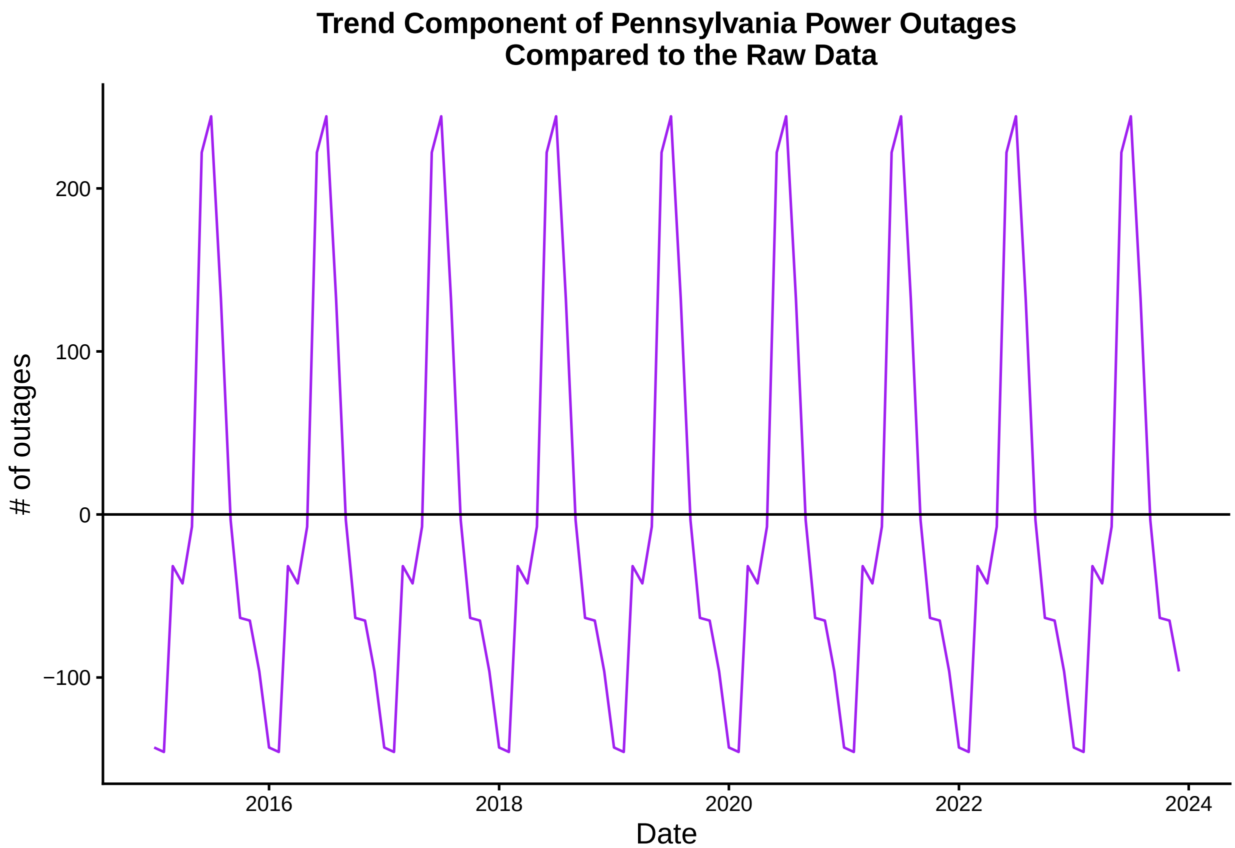

Figure 14: Isolated seasonal component of PA power outages.

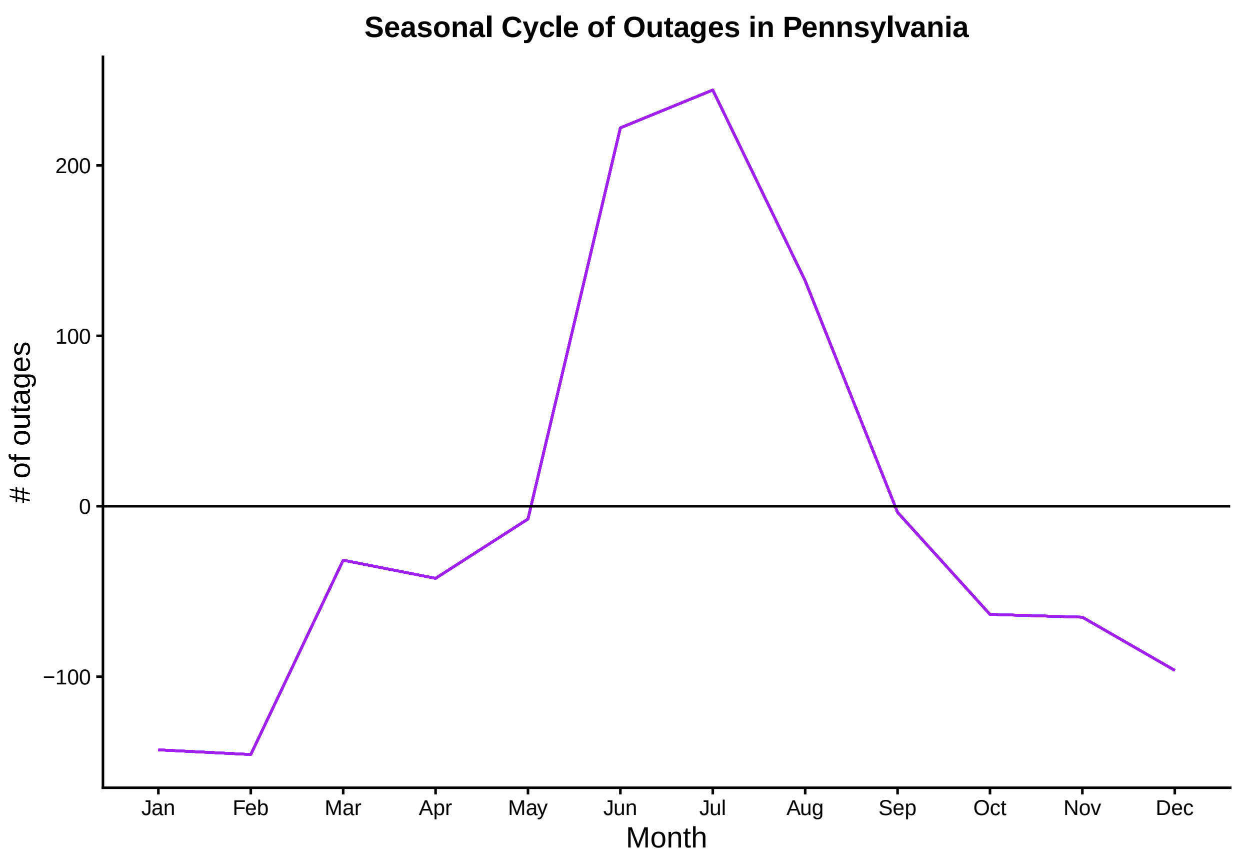

Figure 15: Single year breakdown of seasonal component of outages in PA.

4.2.4 Texas

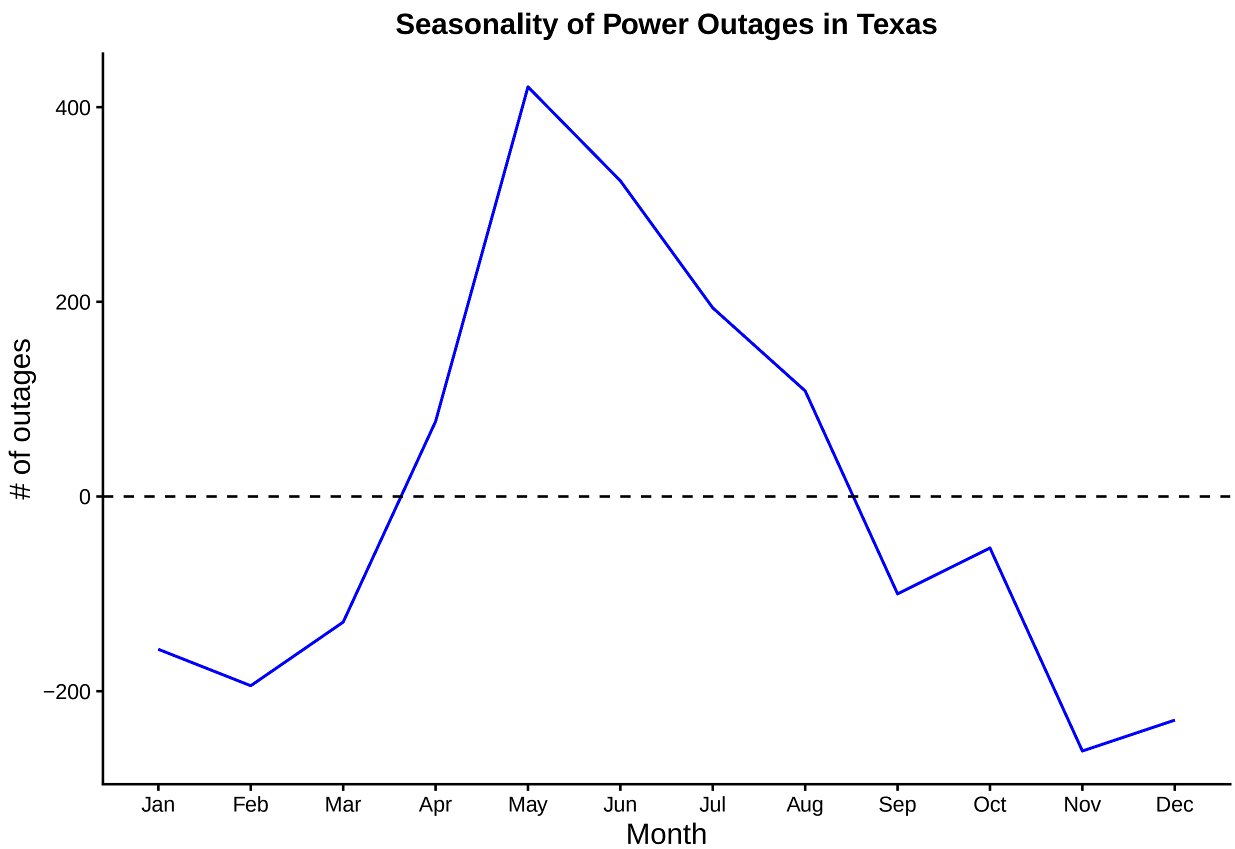

Figure 16: Seasonality component of the California power outage count time series analysis where yearly data is grouped by month.

4.2.5 Comparision between States

4.3 Question 3: How has the length of power outages changed over time?

4.3.1 California

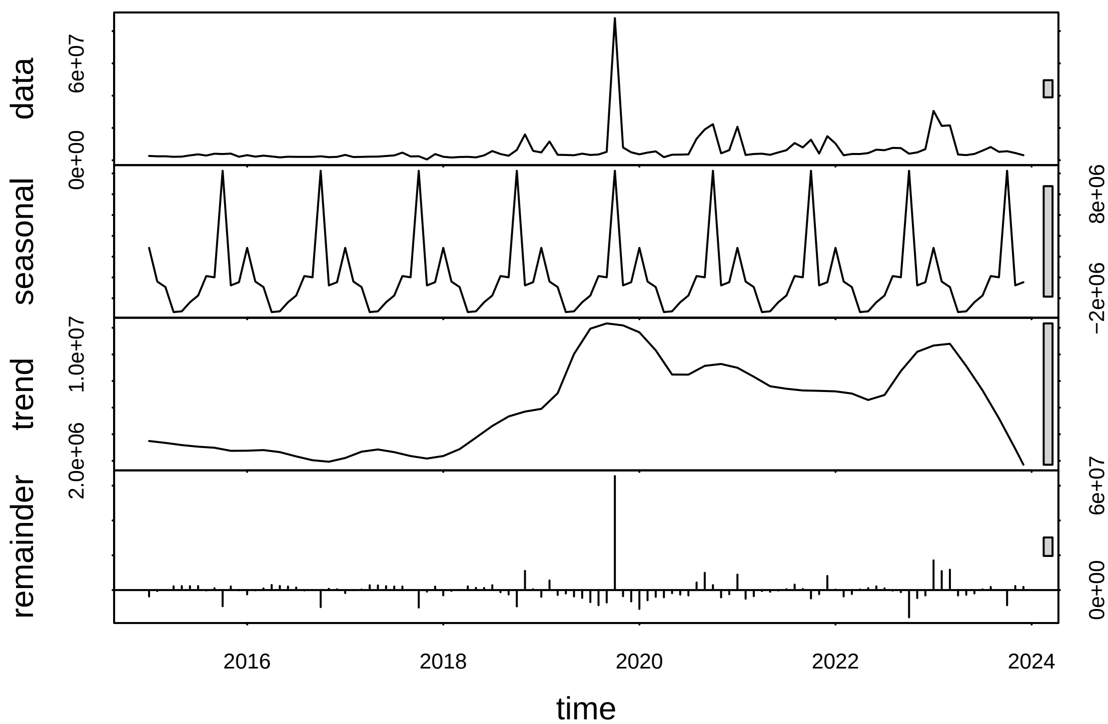

Figure 17: Decomposed analyis of customer weighted outage time in California

4.3.2 Florida

4.3.3 Pennsylvania

For customer-weighted hours, the seasonal test was not statistically significant (tau = 0.13, p = 0.09), indicating no consistent monthly pattern in outage impacts on customers. However, after de-seasonalizing, the Mann-Kendall trend test suggested a marginally increasing trend in customer-weighted hours over time (z = 1.89, n = 108, p = 0.06). Overall, these results suggest that while outage counts are increasing and show clear seasonality, the impact on customers is also trending upward, though less strongly.

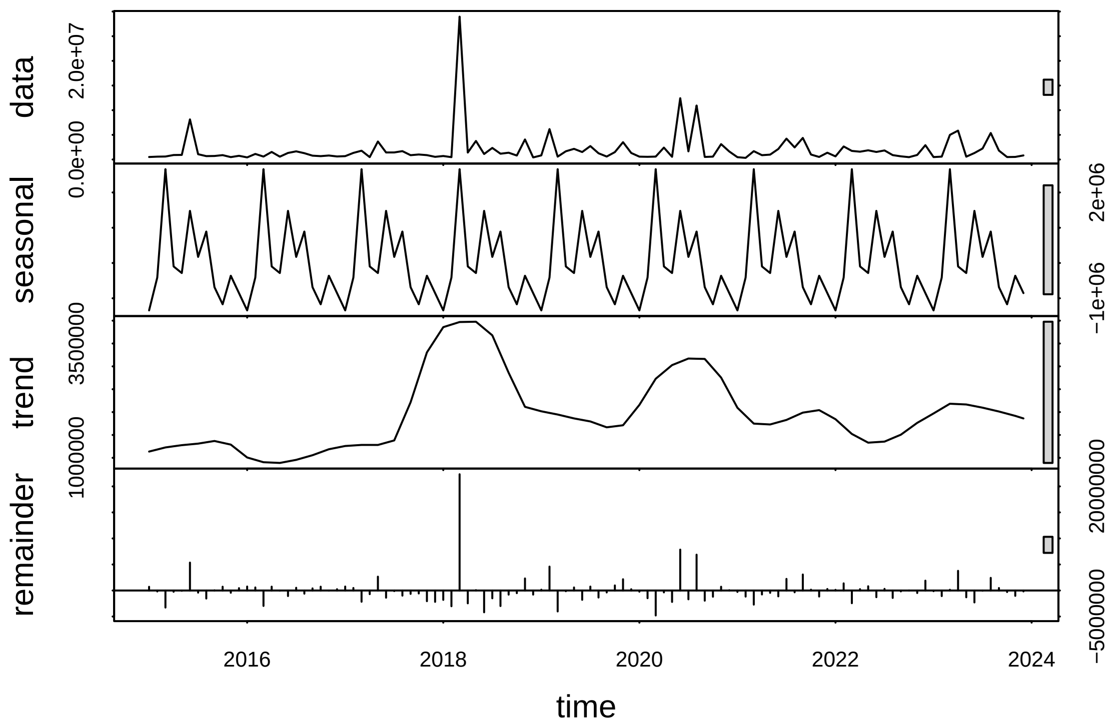

Figure 18: Decomposed analyis of customer weighted outage time in Pennsylvania.

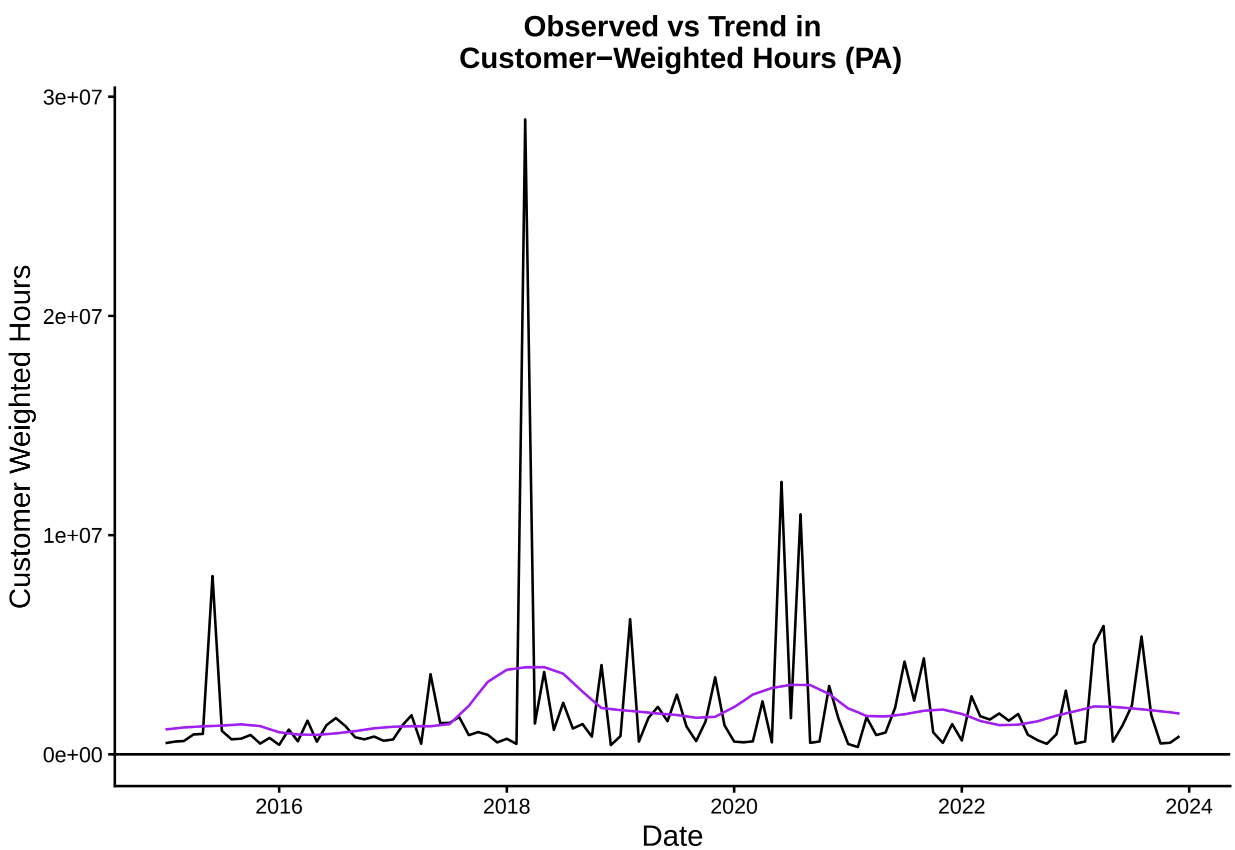

Figure 19: Trend versus observation of customer-weighted hours in PA (2015-2023).

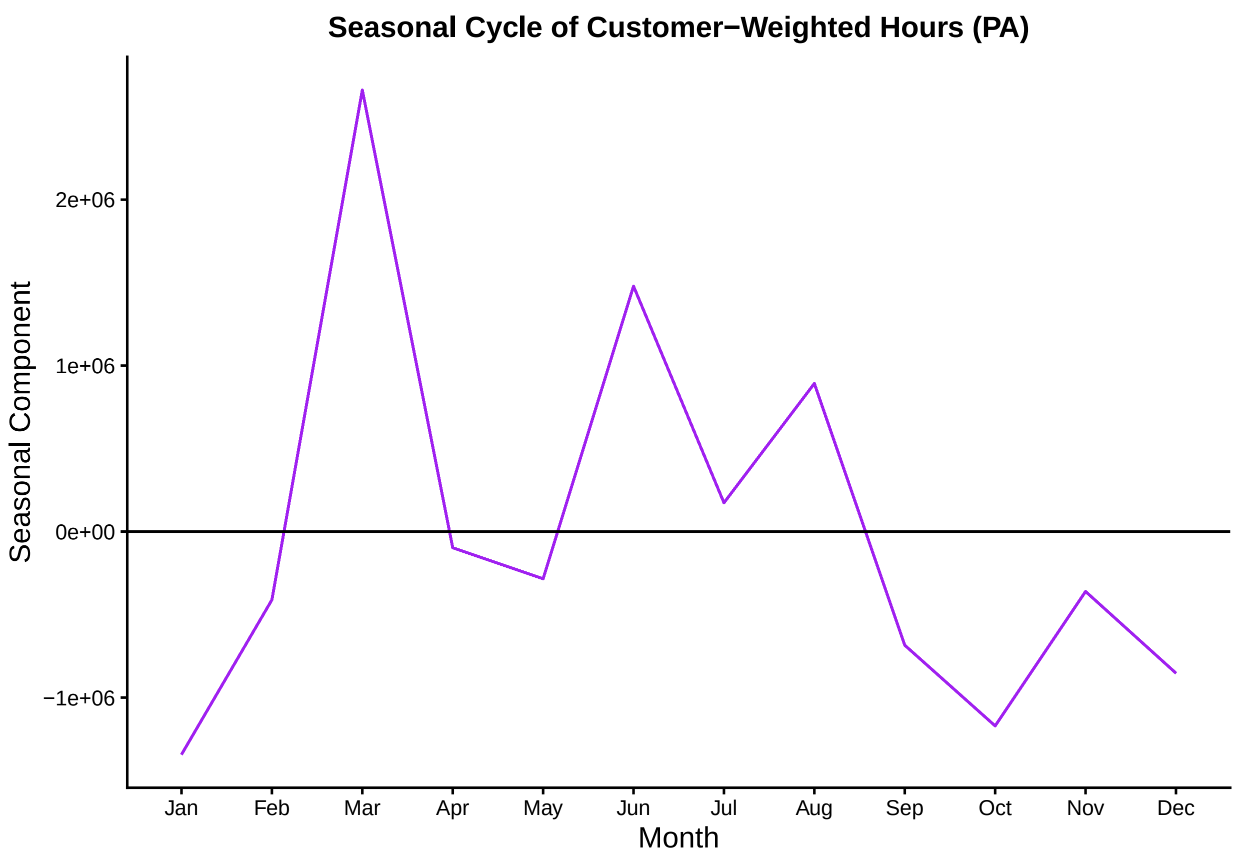

Figure 20: Single year breakdown of seasonal component of customer-weighted hours in PA.

4.3.4 Texas

4.3.5 Comparison between States

5 Summary and Conclusions

Summarize your major findings from your analyses in a few paragraphs. What conclusions do you draw from your findings? Relate your findings back to the original research questions and rationale.

After accounting for seasonal patterns, the analysis shows that the frequency of power outages in Pennsylvania has increased from 2015 to 2023. Outage counts display a clear seasonal trend, with the highest numbers occurring in the summer months of June, July, and August. In contrast, customer-weighted hours do not show a strong seasonal pattern, but there is a slight upward trend over time, indicating that the impact of outages on customers is also gradually increasing. Overall, while outages are becoming more frequent and seasonally concentrated, the effect on customers is rising more moderately.

6 References

<add references here if relevant, otherwise delete this section>