

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

[https://github.com/sarahjhall/Antonucci\\_Hall\\_Huang\\_Kuehn\\_ENV872\\_FinalProject.git](https://github.com/sarahjhall/Antonucci_Hall_Huang_Kuehn_ENV872_FinalProject.git)

December 12th, 2025

## Contents

<b>1 Rationale and Research Questions</b>	<b>5</b>
1.1 Question 1: How has the frequency of power outages changed over time? . . . . .	5
1.2 Question 2: Is there a seasonal trend? Are certain months more prone to outages? . . . . .	5
1.3 Question 3: How has the length of power outages changed over time? . . . . .	5
<b>2 Dataset Information</b>	<b>6</b>
<b>3 Exploratory Analysis</b>	<b>7</b>
<b>4 Analysis</b>	<b>9</b>
4.1 Question 1: How has the frequency of power outages changed over time? . . . . .	9
4.2 Question 2: Is there a seasonal trend? Are certain months more prone to outages? . . . . .	11
4.3 Question 3: How has the length of power outages changed over time? . . . . .	12
<b>5 Summary and Conclusions</b>	<b>14</b>
<b>6 References</b>	<b>15</b>

## **Contents**

## **List of Tables**

## List of Figures

1	Power outages in California from 2015 to 2023. . . . .	7
2	Power outages in California from 2015 to 2023 by month. . . . .	8
3	Decomposed components of the California power outage count time series. . . . .	9
4	Trend component of the California power outage count time series compared to the raw data. . . . .	10
5	Seasonality component of the California power outage count time series yearly data group by month. . . . .	11
6	Decomposed components of the California customer weighted outage time series. . . . .	12

# 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 addition 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:

- 1.1 **Question 1: How has the frequency of power outages changed over time?**
- 1.2 **Question 2: Is there a seasonal trend? Are certain months more prone to outages?**
- 1.3 **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 applies to all four state files:

### 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?

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

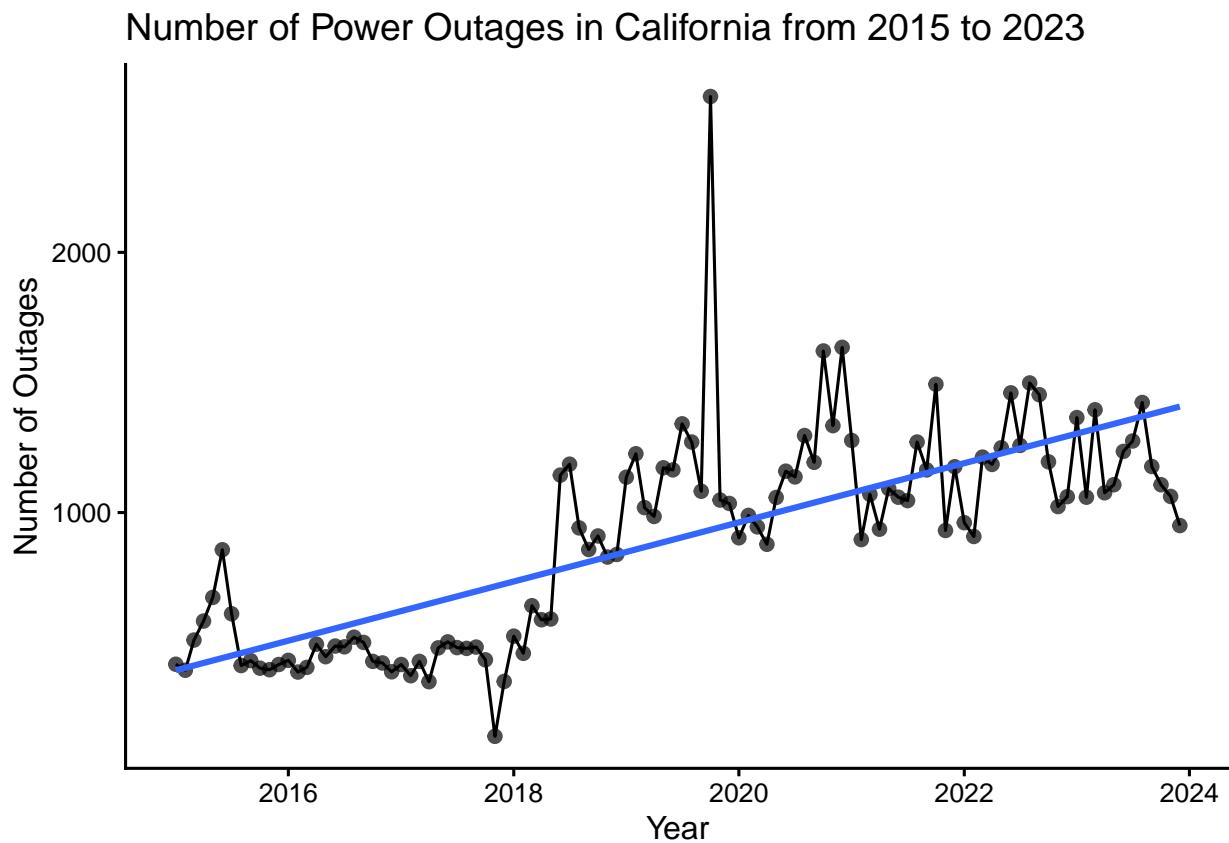


Figure 1: Power outages in California from 2015 to 2023.

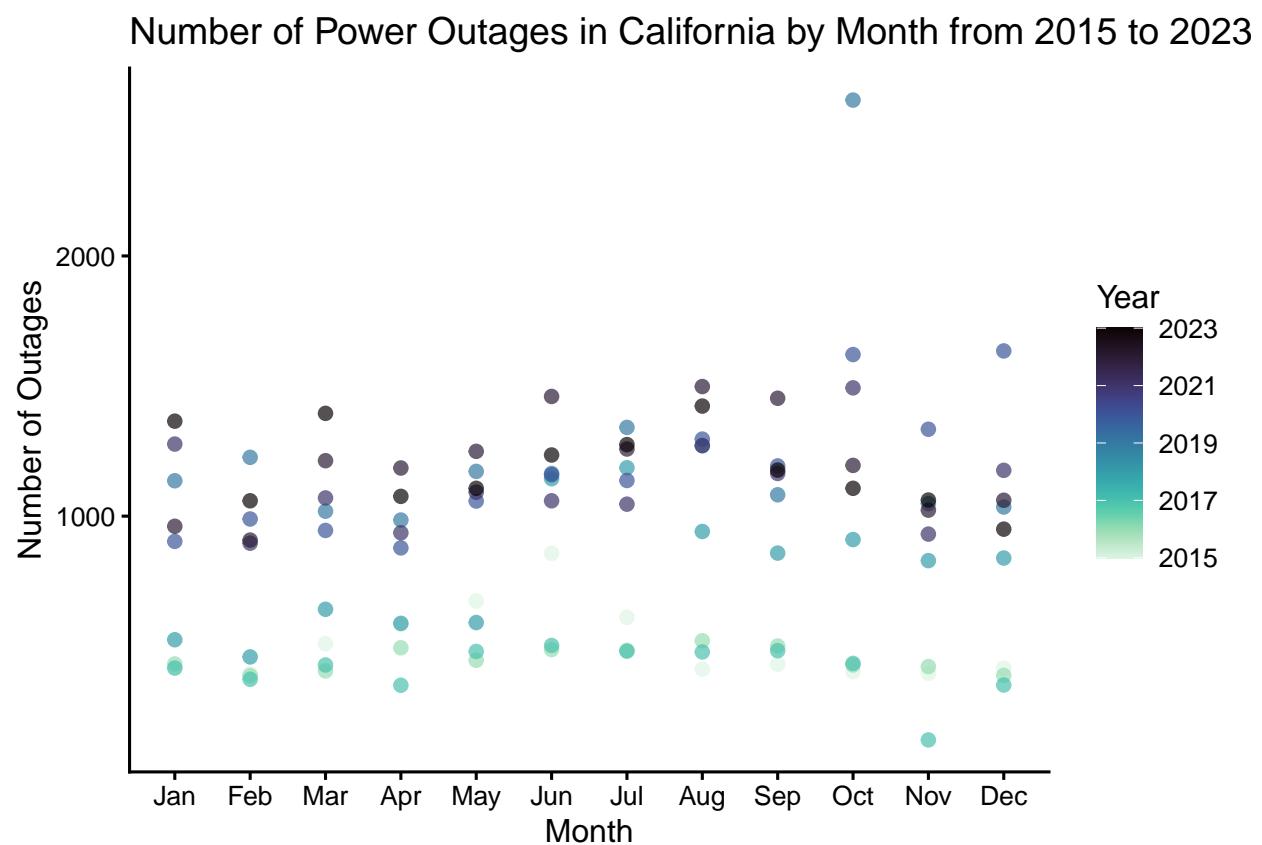


Figure 2: Power outages in California from 2015 to 2023 by month.

## 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 1.1 California

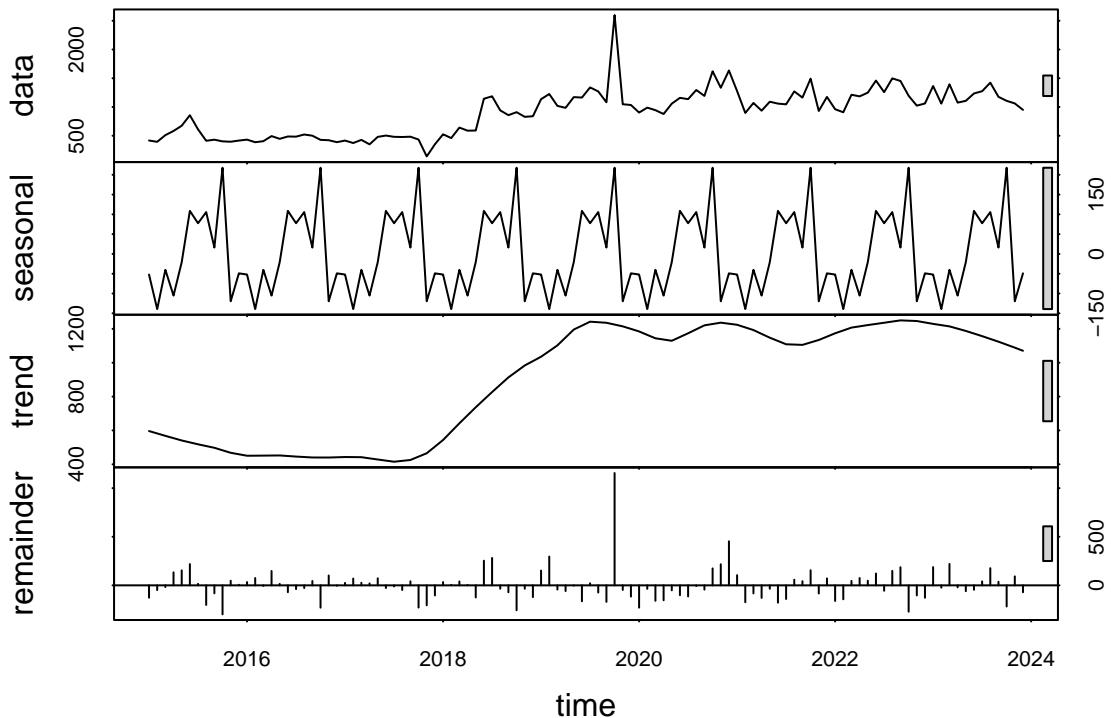


Figure 3: Decomposed components of the California power outage count time series.

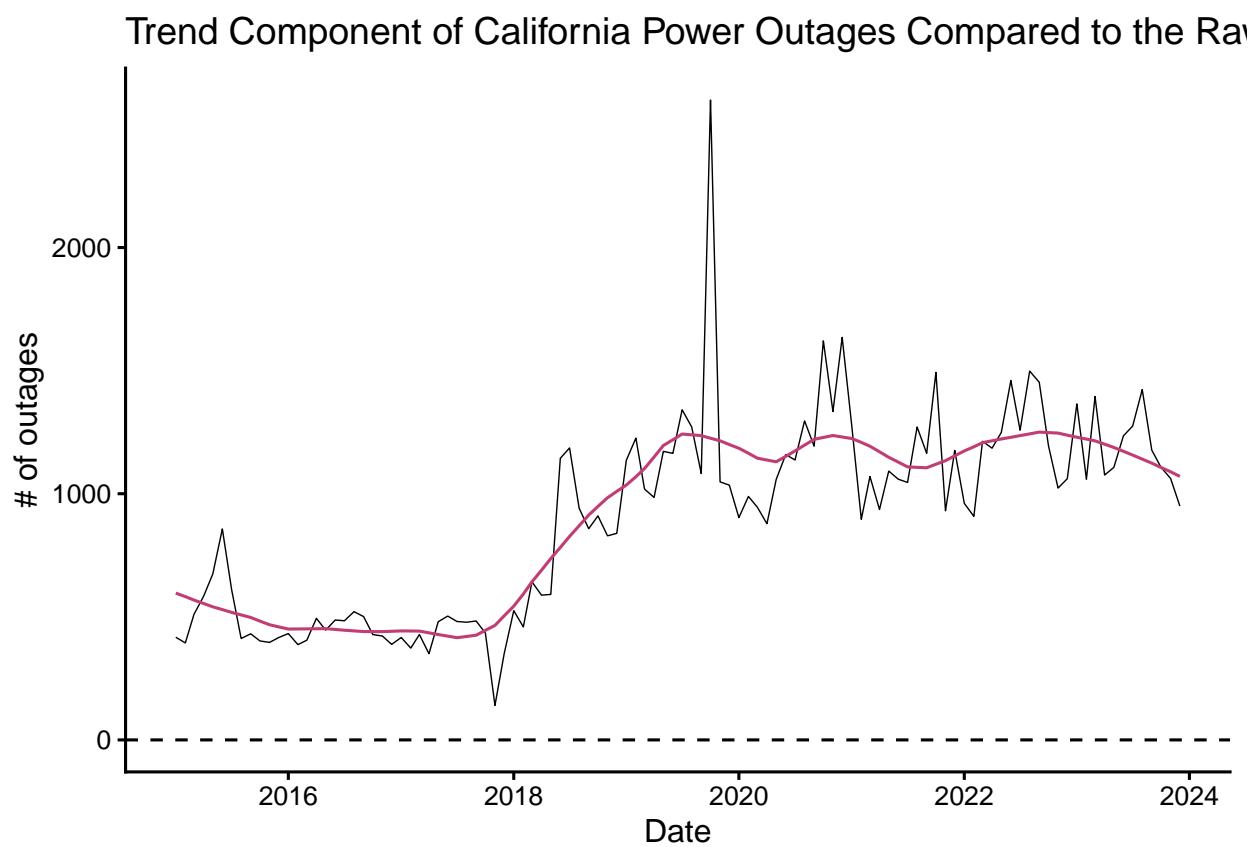


Figure 4: Trend component of the California power outage count time series compared to the raw data.

4.1.2 1.2 Florida

4.1.3 1.3 Pennsylvania

4.1.4 1.4 Texas

4.1.5 1.5 Comparision between States

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

4.2.1 2.1 California

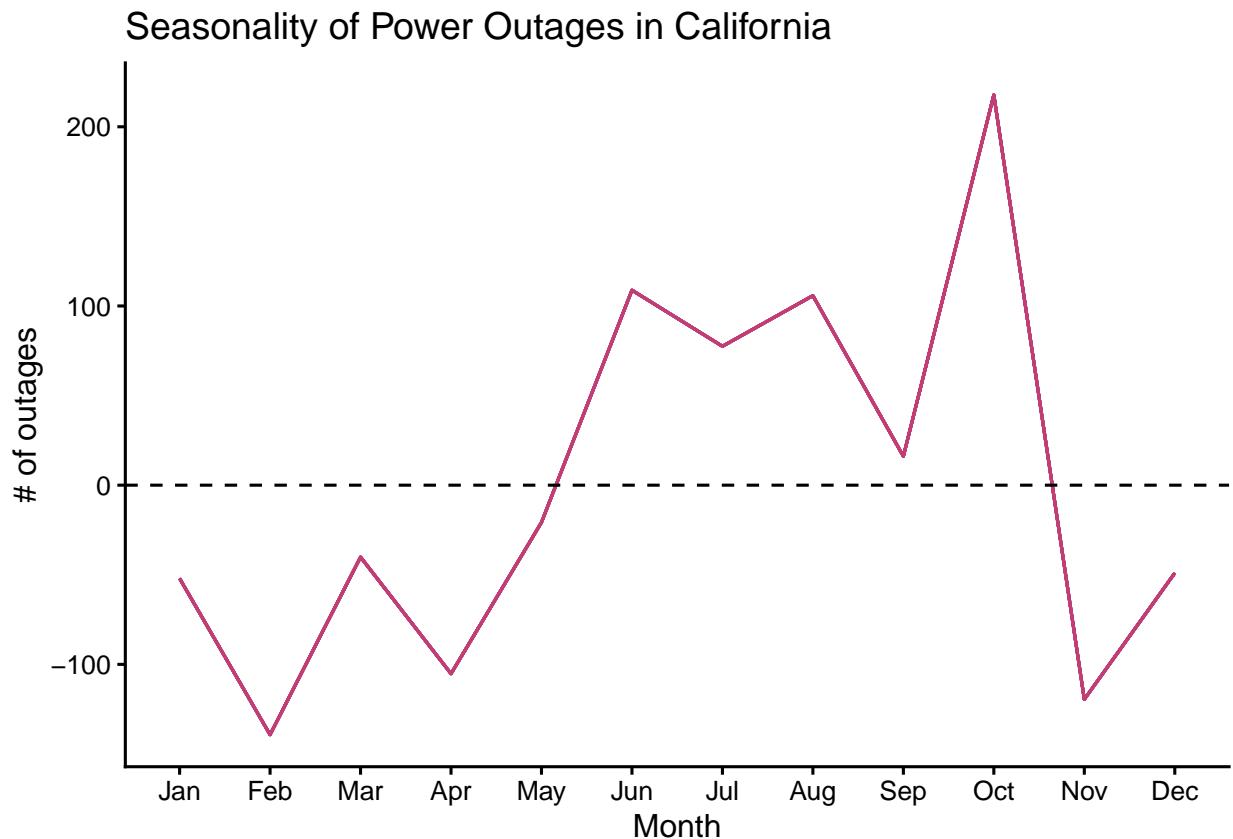


Figure 5: Seasonality component of the California power outage count time series yearly data group by month.

#### 4.2.2 2.2 Florida

#### 4.2.3 2.3 Pennsylvania

#### 4.2.4 2.4 Texas

#### 4.2.5 2.5 Comparision between States

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

#### 4.3.1 3.1 California

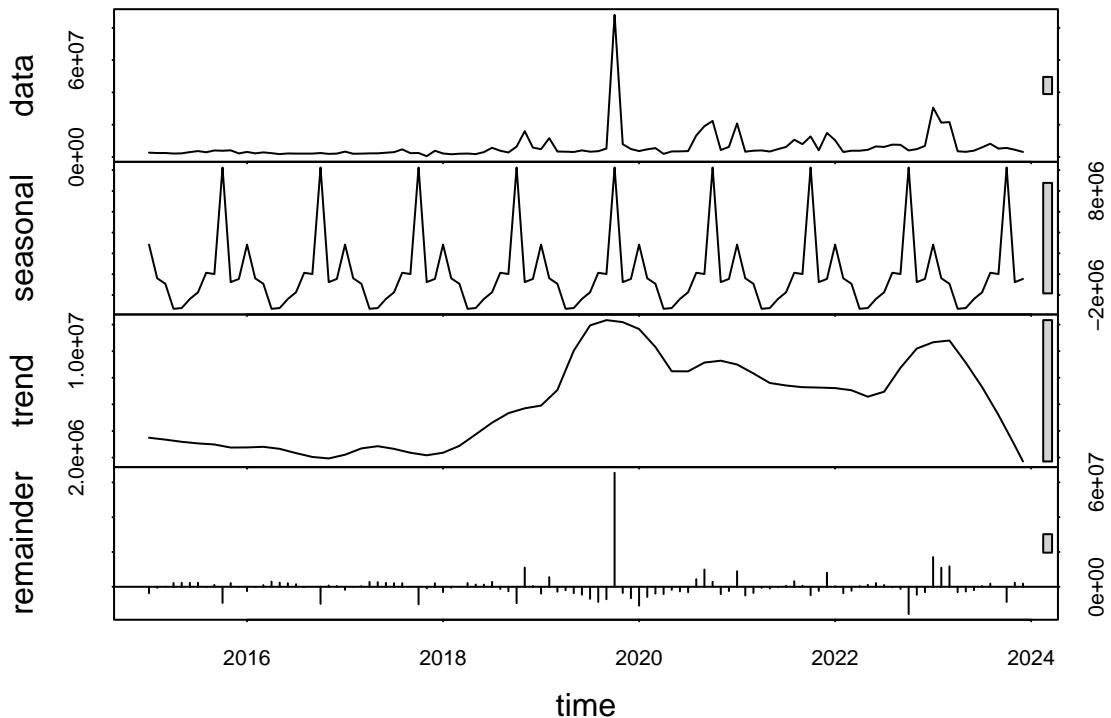


Figure 6: Decomposed components of the California customer weighted outage time series.

```
##  
## Seasonal Mann-Kendall trend test (Hirsch-Slack test)  
##  
## data: CA_CWA_ts  
## z = 6.1698, p-value = 6.839e-10  
## alternative hypothesis: true S is not equal to 0  
## sample estimates:  
##      S varS  
## 206 1104
```

```

## Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data: CA_CWA_ts
## alternative hypothesis: two.sided
##
## Statistics for individual seasons
##
## H0
##          S  varS   tau     z Pr(>|z|)
## Season 1: S = 0  26  92 0.722 2.606 0.0091491 **
## Season 2: S = 0  12  92 0.333 1.147 0.2514522
## Season 3: S = 0  20  92 0.556 1.981 0.0476040 *
## Season 4: S = 0  12  92 0.333 1.147 0.2514522
## Season 5: S = 0  20  92 0.556 1.981 0.0476040 *
## Season 6: S = 0  24  92 0.667 2.398 0.0164887 *
## Season 7: S = 0  20  92 0.556 1.981 0.0476040 *
## Season 8: S = 0  18  92 0.500 1.772 0.0763327 .
## Season 9: S = 0  18  92 0.500 1.772 0.0763327 .
## Season 10: S = 0 10  92 0.278 0.938 0.3480826
## Season 11: S = 0  8  92 0.222 0.730 0.4655122
## Season 12: S = 0 18  92 0.500 1.772 0.0763327 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

### 4.3.2 3.2 Florida

### 4.3.3 3.3 Pennsylvania

### 4.3.4 3.4 Texas

### 4.3.5 3.5 Comparision 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.

## **6 References**

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