2024

Exploring U.S. Natural Disasters and Psychological Distress: From Time Series Trends to Machine Learning Insights on Hurricane Helene

Sarah Jane Fullerton

Claremont McKenna College

Exploring U.S. Natural Disasters and Psychological Distress:
From Time Series Trends to Machine Learning Insights on Hurricane Helene

submitted to
Professor Mark Huber

by
Sarah Jane Fullerton

for
Senior Thesis
Fall 2024
December 1st, 2024

1 Abstract

This research investigates the historical trends of psychological distress in the U.S. in relation to natural disaster occurrences. By analyzing long-term data, we examine how significant natural disasters relate to levels of psychological distress over time. The research employs Exploratory Data Analysis (EDA) and Time Series Analysis to identify patterns and trends between the frequency and intensity of natural disasters and the rise of psychological distress across various periods in U.S. history. Additionally, real-time data from Reddit was collected through a custom-built Reddit web scraper specialized for Hurricane Helene. This dataset was labeled for sentiment and used to train machine learning models for sentiment analysis, providing valuable tools for understanding emotional responses in real-time. Their adaptability makes them applicable for future use in crisis response. The findings of this research offer a dual perspective: understanding the broader historical relationship between natural disasters and psychological distress, and providing insights into emotional reactions to 2024 events.

Contents

2 Acronyms

BERT	Bidirectional Encoder Representations from Transformers. 38
EDA	Exploratory Data Analysis. 1
FEMA	Federal Emergency Management Agency. 9, 10, 25, 32, 33
Fy	Fiscal Years. 25
GBD	Global Burden of Disease. 2, 10, 20, 42
GRU	Gated Recurrent Unit. 3, 38
IHME	Institute For Health Metrics and Evaluation. 2, 10, 11
LSTM	Long Short-Term Memory. 3, 38
NSDUH	National Survey on Drug Use and Health. 2, 9–11, 19, 42
PRAW	Python Reddit API Wrapper. 33, 34, 37
PTSD	Post Traumatic Stress Disorder. 6–8, 23, 25
SAMHSA	Substance Abuse and Mental Health Services Administration. 9
SPD	Serious Psychological Distress. 2, 5, 9, 11, 12, 19, 22–28, 32, 33, 41, 42
YLD	Years Lived with Disability. 2, 5, 10, 11, 20, 22–24, 28–33, 41

3 List of Figures

4 Background

Natural disasters, such as hurricanes, floods, and wildfires have become more frequent and intense due to climate change [7], causing not only physical destruction but also significant psychological effects on those affected. This paper aims to explore the emotional responses to extreme weather events and their relationship to mental health, using comprehensive data analysis methods to examine both short-term and long-term psychological effects over time. Understanding these emotional impacts is crucial for improving mental health support and disaster preparedness strategies, particularly for vulnerable populations.

4.1 Relevant Climate-Mental Health Research

Several studies have linked natural disasters to mental health issues such as Post Traumatic Stress Disorder (PTSD), anxiety, and depression. This research aims to contribute to understanding the role of demographic factors, such as age, gender, and location, in shaping emotional responses. Several studies have explored the impact of natural disasters on mental health, particularly the prevalence of PTSD, anxiety, and depression. For example, Monsour et al. (2021) focused on the effects of tropical cyclones and sea-level rise on mental illness symptoms in Miami-Dade and Broward counties, specifically examining the prevalence of PTSD and depression in response to climate-induced stressors [13]. This study has inspired the current research, which focuses on a similar theme but takes a different approach by analyzing emotional responses to Hurricane Helene in the sentiment analysis section (9.3).

4.1.1 Existing Research and Gaps

This section examines existing research on psychological effects, emotional responses, and their relationship to extreme weather events through themes that frequently arise.

Main Themes in Relevant Research

- **Climate Change as a Public Health Threat due to Mental Health and Emotional Response Impacts**
  - *Cianconi et al.* conduct a systematic review examining the direct and indirect pathways through which climate change impacts mental health. They highlight exposure to extreme weather events, such as floods and droughts, as critical stressors that contribute to PTSD, depression, and anxiety. Their study underscores the need for a multidisciplinary approach, combining environmental data with mental health outcomes to identify vulnerable populations [4].
  - *Berry HL, Bowen K, and Kjellstrom* propose a causal pathways framework for understanding the impacts of climate change on mental health. This conceptual framework identifies both direct effects, such as trauma from natural disasters, and indirect effects, such as the erosion of community resilience and increased social isolation due to climate-related migration. Their work suggests that mental health research should not only focus on

immediate emotional responses but also on long-term consequences, including disruptions in social structures and livelihoods [3].

  - *Lawrance et al.* focus on the policy implications of the mental health effects of climate change. Their research identifies a link between extreme temperatures and mental health outcomes, such as suicides, and emphasizes the need for mental health interventions during climate crises. Their study suggests that existing mental health infrastructure may be insufficient to handle the rising tide of climate-related psychological distress [12].

- **Increase in the Frequency and Toll of Natural Disasters in the U.S.**

  - *Research by Galea et al.* highlighted the profound mental health consequences of Hurricane Katrina, showing high rates of PTSD and depression among survivors, particularly in marginalized communities. The study's focus on long-term psychological effects of such disasters underlines the increasing mental health burden as climate change amplifies extreme weather events [10].
  - *Anderson and Bell*'s study explored the association between heatwaves and increased mortality, as well as the indirect impact on mental health, including anxiety and depression, in urban areas during extreme heat events [2].

- **Demographic Vulnerability**

  - *A study by Enarson* demonstrated that women often experience greater psychological distress in the aftermath of climate-induced disasters due to gender-specific roles and responsibilities, pointing to the need for gender-sensitive disaster preparedness strategies [5].

- **Social Media Sentiment Analysis and Mental Health**

  - *Sentiment Analysis of Twitter Posts During Hurricane Harvey*: Using natural language processing (NLP), Rosenberg et al. (2023) analyzed Twitter data during Hurricane Harvey, classifying tweets as positive, negative, or neutral. The study found that negative sentiment spiked before and after the hurricane made landfall, indicating heightened public anxiety and fear [16].
  - *Prediction and Analysis of Sentiments of Reddit Users towards the Climate Change Crisis*: This study used machine learning algorithms to analyze Reddit posts related to climate change. The researchers found that discussions around climate crises were dominated by negative sentiments, such as fear and helplessness, reflecting the broader public anxiety about the climate crisis [15].

Research Gaps While prior studies have focused on hurricanes and their mental health impacts, particularly in regions like Florida [13], there remains a gap in leveraging sentiment analysis for specific hurricanes, such as Hurricane and Tropical Storm Helene. This research addresses this gap by analyzing sentiment data related to Hurricane Helene, public emotional responses during this event, as detailed in Section 9.3.

Research Contribution The existing body of research highlights substantial progress in understanding the mental health impacts of climate change. My study builds on this foundation by combining time series analysis of climate and mental health data spanning several decades with sentiment analysis of social media data during Hurricane Helene. This approach seeks to address gaps in existing research, providing a comprehensive view of both long-term trends and real-time emotional responses. By examining these patterns, my work contributes new insights into the emotional reactions to extreme weather events, furthering our understanding of how climate change impacts mental health.

5 Research Objectives

5.1 Focus: The Impact of Natural Disasters on Mental Health

The rise in the frequency and intensity of extreme weather events, such as hurricanes, heatwaves, floods, and wildfires, is largely driven by climate change. This poses significant challenges for communities worldwide, with both immediate and long-term consequences. Extreme weather events are defined as events such as hurricanes, floods, wildfires, and heatwaves. Metrics such as intensity and frequency will be explored in this study to evaluate the impact of these events.

5.2 Mental Health, Emotional Responses, and Psychological Disorders

Natural disasters often trigger a range of emotional responses with potential to lead to mental health conditions such as PTSD, anxiety, and depression. When emotional responses to traumatic events that are not treated, they can lead to long-term disorders like depression, substance abuse, acute stress disorder, and anxiety disorders [17]. This research will assess emotional response data over time utilizing public databases, and emotional response in current social media data.

5.3 Research Questions and Hypotheses

5.3.1 Research Question 1

What is the relationship between the frequency/intensity of extreme weather events and the prevalence of mental health disorders, particularly anxiety and depression?

- **Null Hypothesis (H_0):** The frequency and intensity of extreme weather events have no significant effect on the prevalence of mental health disorders, such as anxiety and depression.
- **Alternative Hypothesis (H_1):** The frequency and intensity of extreme weather events significantly affect the prevalence of mental health disorders, such as anxiety and depression.

5.3.2 Research Question 2

How are emotional responses to natural disasters, such as Hurricane Helene, reflected on popular social media platforms, such as Reddit?

- **Null Hypothesis (H_0):** Emotional responses to natural disasters, such as Hurricane Helene, observed on Reddit do not significantly vary in negative patterns.
- **Alternative Hypothesis (H_1):** Emotional responses to natural disasters, such as Hurricane Helene, observed on Reddit significantly vary in negative patterns.

6 Methodology

6.1 Research Design

This research employs a mixed-methods approach, combining exploratory data analysis, sentiment analysis of social media data, and time series analysis of climate and mental health data to explore the emotional responses to extreme weather events.

The public datasets used for the analysis in this thesis were sourced from publicly available online databases and supplemented by data collected via the social media platform Reddit. All datasets, processing, and analysis used for this research are available in my GitHub repositories:

- Time Series Analysis with data sourced from Public Databases: `https://github.com/sjanefullerton/time-series`
- Reddit Sentiment Analysis: `https://github.com/sjanefullerton/Web-Scraper`
- Sentiment Analysis: `https://github.com/sjanefullerton/Sentiment-Analysis`

These repositories allow for the replication and extension of the research presented in this thesis.

6.2 Data Collection from Public Sources

The following datasets were used for the analysis:

- **OpenFederal Emergency Management Agency (FEMA) Disaster Declarations** [6]:
  - Time span: 1953-2023
  - Purpose is to explore climate-related impacts and their correlation with public mental health trends.
- **Substance Abuse and Mental Health Services Administration (SAMHSA) NSDUH Population Data** [14]:
  - Time span: 2004-2022
  - Utilized SAMHSA's Data Analysis System to retrieve SPD X Gender crosstab analysis for each year.
  - **SPD** is an indicator of functioning impairment due to multiple mental health symptoms [11].

- **IHME GBD Study Data** [8]:
  - Time span: 1990-2021
  - Utilized GBD's interactive data visual tool to retrieve yearly YLD counts caused by depression and anxiety.
  - **YLD** is an abbreviation for years lived with disability, which can also be described as years lived in less than ideal health [9].
- **Reddit Sentiment Data**
  - Data was collected from Reddit discussions surrounding extreme weather events, focusing on public emotional responses during Hurricane Helene.
  - A detailed description of the custom web scraper used for data collection can be found in Section 9.3.

6.3 Data Preprocessing

Data preprocessing was carried out to clean, transform, and standardize the datasets for subsequent analysis. Each dataset underwent specific steps tailored to its format and content. The final preprocessed datasets used for analysis are listed below:

6.3.1 Natural Disaster Dataset

- **Tool:** Processed and validated using R.
- **Steps:**
  1. Filtered FEMA disaster declarations data to retain only records related to natural disasters (e.g., tornadoes, floods, hurricanes).
  2. Validated the data by checking for missing values, duplicate rows, and outliers.
  3. Ensured consistent categorical values for key attributes such as `incident_type` and `state`.
- **Outcome:** The filtered data was saved as `natural_disaster_data.csv`, focusing on relevant incidents between 1953 and 2023.

6.3.2 IHME Dataset

- **Steps:** The dataset required minimal cleaning. It was renamed to `IHME.csv` for consistency and ease of access.
- **Outcome:** The dataset retained its original structure, ensuring that the information on global burden of disease metrics was preserved for analysis.

6.3.3 NSDUH Dataset

- **Tool:** Processed using Python (PySpark) for efficient handling of large-scale data.
- **Steps:**

1. Extracted the year from file names and added it as a column.
2. Standardized column names (`SERIOUS PSYCHOLOGICAL DISTRESS` to `SPD`, and gender-related columns to `GENDER`).
3. Filtered rows where the `SPD` column indicated positive cases (`1`), signifying serious psychological distress.
4. Sorted the data by year and retained relevant columns such as percentages, confidence intervals, and counts.

- **Outcome:** The cleaned data was exported as `NSDUH_PreProcessed.csv`, consolidating yearly records from 2004 to 2022 for analysis.

6.3.4 Reddit Dataset

- **Tools:** Processed using Python (pandas library).
- **Steps:**
  1. Combined multiple datasets (e.g., comments, posts, titles) into a single file for unified analysis.
  2. Removed duplicates and filtered out bot-generated comments (e.g., containing "I am a bot").
  3. Applied text cleaning:
     - Converted text to lowercase.
     - Removed digits, punctuation, links, emojis, and non-ASCII characters.
  4. Extracted and retained the cleaned text column for sentiment analysis.
- **Outcome:** The cleaned dataset, `combined_clean.csv`, contains deduplicated, filtered text ready for sentiment analysis.

6.4 Data Validation

Data validation was conducted to ensure the integrity, consistency, and accuracy of the datasets. Each dataset underwent checks for missing values, duplicates, outliers, and format consistency. Below are the validation steps performed for each dataset:

IHME Dataset

- **Missing Values:** Confirmed the absence of missing values across all relevant columns.
- **Data Consistency:** Checked for uniformity in the time range (1990–2021) and verified that all years were accounted for.
- **Outliers:** Inspected key health metrics (e.g., YLD values) to ensure all entries fell within expected ranges.

NSDUH Dataset

- **File Integrity:** Cross-checked the extracted year column against file names to confirm alignment.

- **Missing Values:** Identified and addressed any missing entries in critical fields, such as `SPD` and `GENDER`.
- **Data Consistency:** Verified that all records included standardized column names and followed a uniform format.
- **Outliers:** Filtered out any anomalous percentages or counts that deviated significantly from expected trends.

Natural Disaster Dataset

- **Incident Types:** Ensured the `incident_type` field contained only natural disaster categories (e.g., Tornado, Flood, Hurricane).
- **Missing Values:** Checked for and addressed any gaps in key columns like `state` and `declaration_type`.
- **Duplicate Rows:** Removed duplicate entries identified during processing.
- **Outliers:** Reviewed numeric columns such as `fy_declared` and `disaster_number` for anomalies and inconsistencies.
- **Categorical Values:** Validated that categorical columns, such as `state` and `incident_type`, contained only expected values.

Outcome: After applying these validation procedures, all datasets were deemed ready for analysis, with no significant data quality issues remaining.

6.5 Data Analysis Techniques

The **exploratory data analysis** (see Section 7) focused on the initial data exploration. Detailed **time series analysis** and **statistical analysis** are discussed in Section 8. These analyses were used to examine the relationship between climate change and mental health trends. Statistical methods were employed to validate the results. Lastly, detailed in Section 9 is the **sentiment analysis,** which includes machine learning techniques applied to analyze social media data.

7 Exploratory Data Analysis

7.1 Natural Disasters

7.1.1 Frequency and Temporal Trends

Disaster Types by Frequency

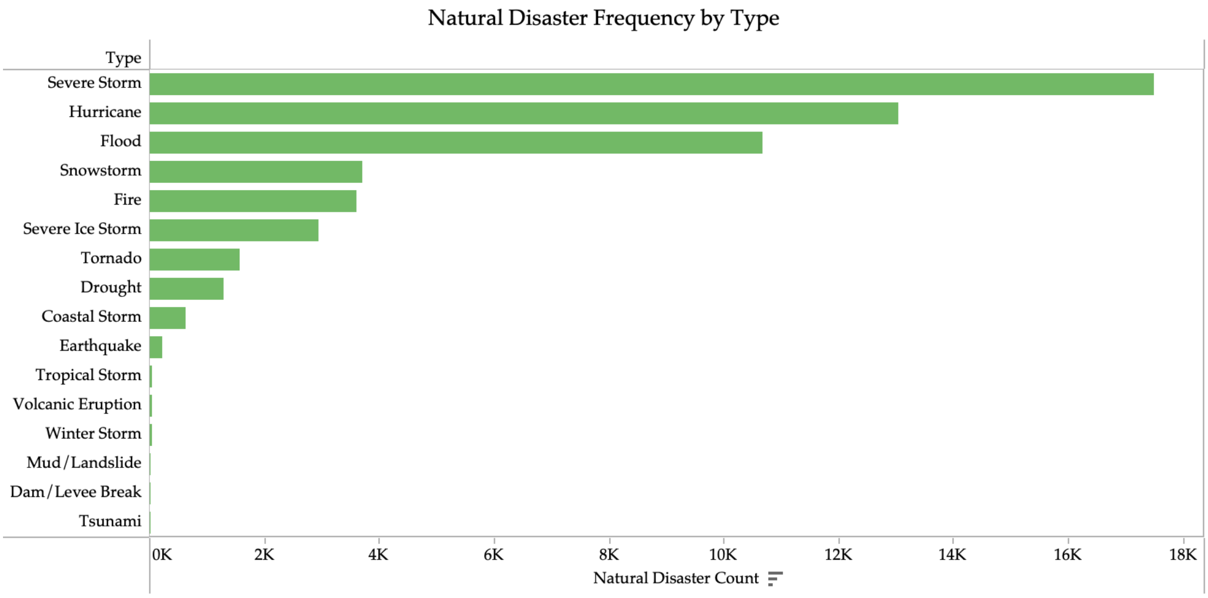

Figure 1: Bar Graph of Natural Disaster Type Frequency.

Figure 1 displays the frequency of natural disasters in the United States, categorized by type. The bar graph provides a clear visualization of how often each disaster type has occurred over the analyzed period. Severe storms, hurricanes, and floods appear most frequently, reflecting their prevalent impact across the country. These disaster types are followed by snowstorms and fires, which also contribute significantly to the total count.

Yearly Trends by Disaster Type

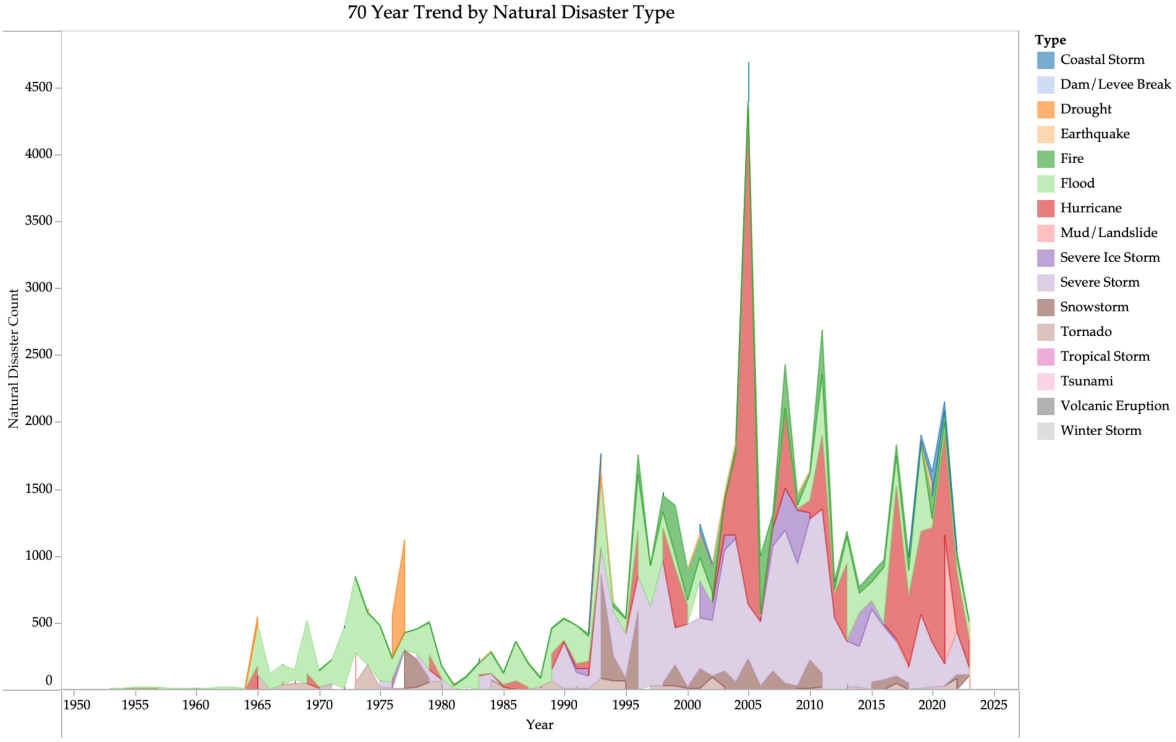

Figure 2: 70 year natural disaster type trends.

Figure 2 illustrates the yearly trends in natural disaster counts over a 70-year period, with different colors representing various disaster types. A prominent peak is observed in 2005, driven by Hurricane Katrina, which significantly increased the total disaster count for that year. Beyond this outlier, the data reveal a steady upward trend in the frequency of natural disasters over time. Storms, floods, and hurricanes, represented by light purple, green, and red, respectively, dominate the dataset, accounting for a substantial proportion of the overall natural disaster counts. This increase over the years may reflect not only heightened weather activity but also improved reporting and data collection methods.

Total Disasters Heatmap by Year colored by frequency

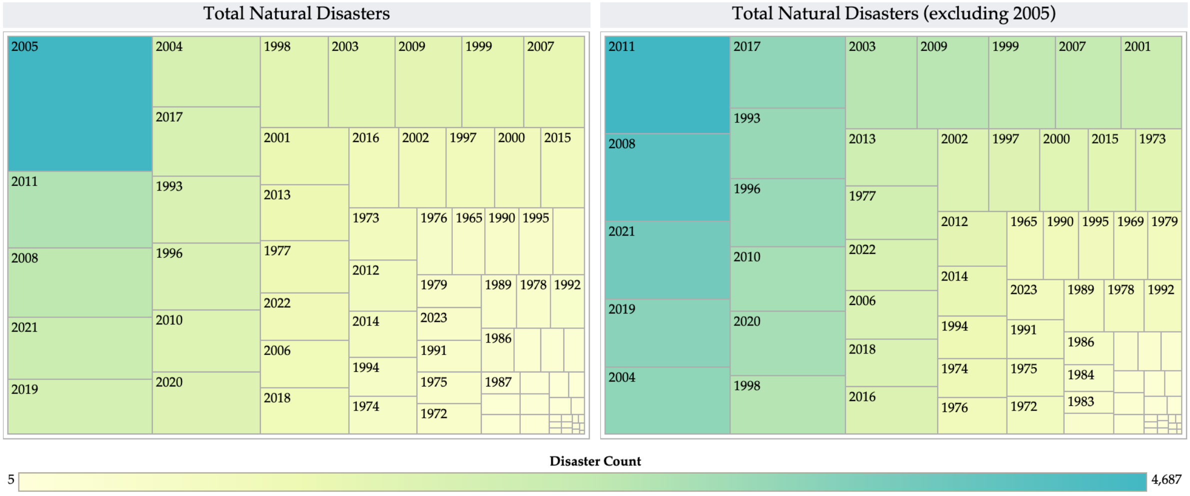

Figure 3: Heatmap of natural disaster counts by year.

Figure 3 presents heatmaps of natural disaster counts by year, highlighting frequency trends. The left heatmap includes all years, with 2005 standing out as the darkest shade, indicating the highest recorded disaster count. This aligns with the spike in Hurricane Katrina-related disasters from the previous figure. To provide a clearer visualization of patterns across other years, the right heatmap excludes 2005 as an outlier. Without this dominant year, the data reveal that the highest counts are concentrated closer to the present, with 2011 having the highest disaster frequency, followed by 2008, 2021, and 2019. This pattern suggests a continuing rise in natural disaster occurrences in recent decades.

Seasonality

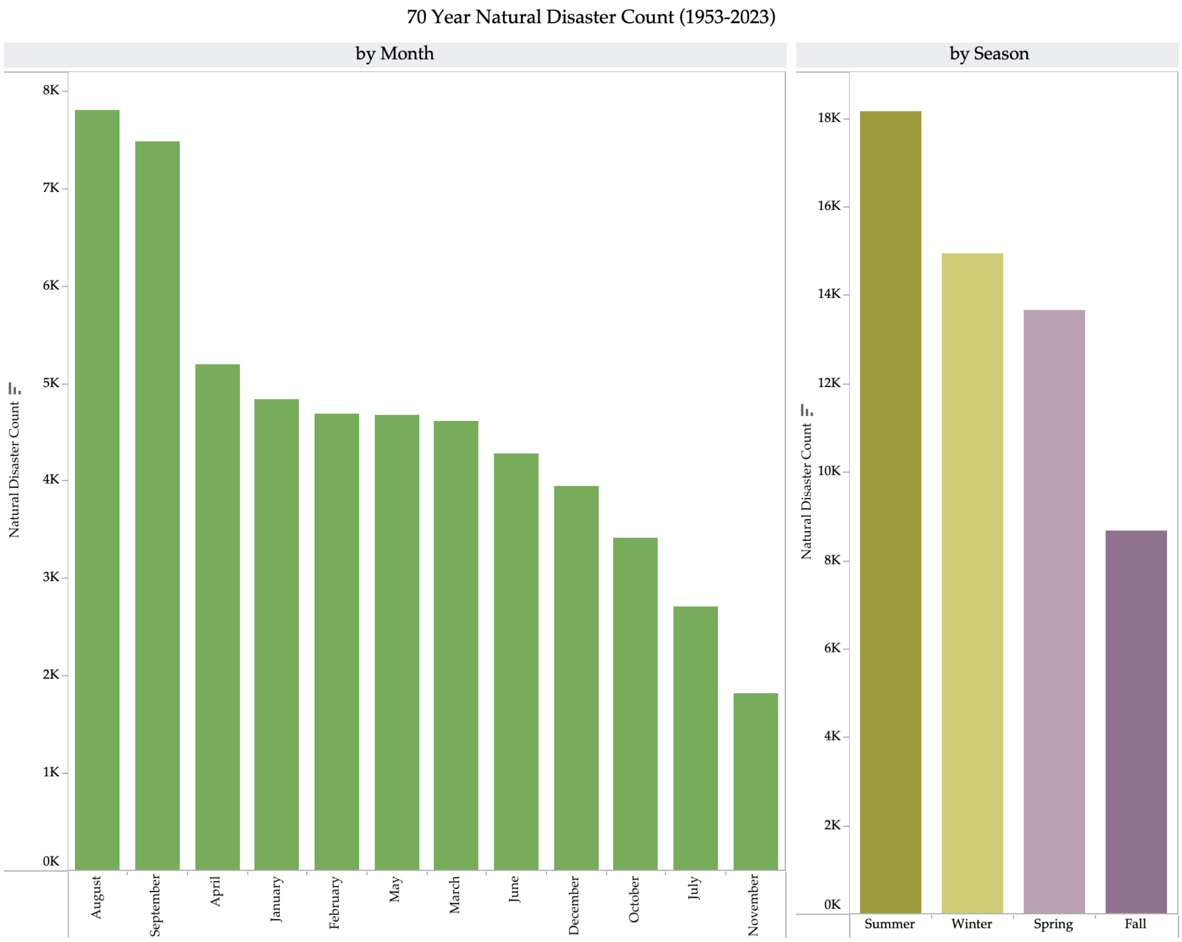

Figure 4: 70 Year Seasonality Bar Graphs.

Figure 4 illustrates the seasonality of natural disasters over a 70-year period, with two bar graphs side by side. The graph on the left displays natural disaster counts by month, while the graph on the right aggregates the data by season.

- From the monthly analysis, we observe that August and September have the highest natural disaster frequencies, followed by April and January. These peaks are likely influenced by seasonal patterns in hurricane and winter storm activity, which are most prominent during late summer and early winter, respectively.

- The seasonal analysis further highlights these trends, with summer showing the highest overall natural disaster counts. Winter follows in second place, likely due to the prevalence of winter storms and other weather-related events. Spring and fall rank third and fourth, respectively, reflecting lower natural disaster activity during these periods compared to summer and winter.

- This seasonality analysis emphasizes the cyclical nature of natural disasters and the importance of preparing for heightened risks during specific times of the year, particularly late summer and winter months.

7.1.2 Geographic Analysis

Disaster Distribution by State

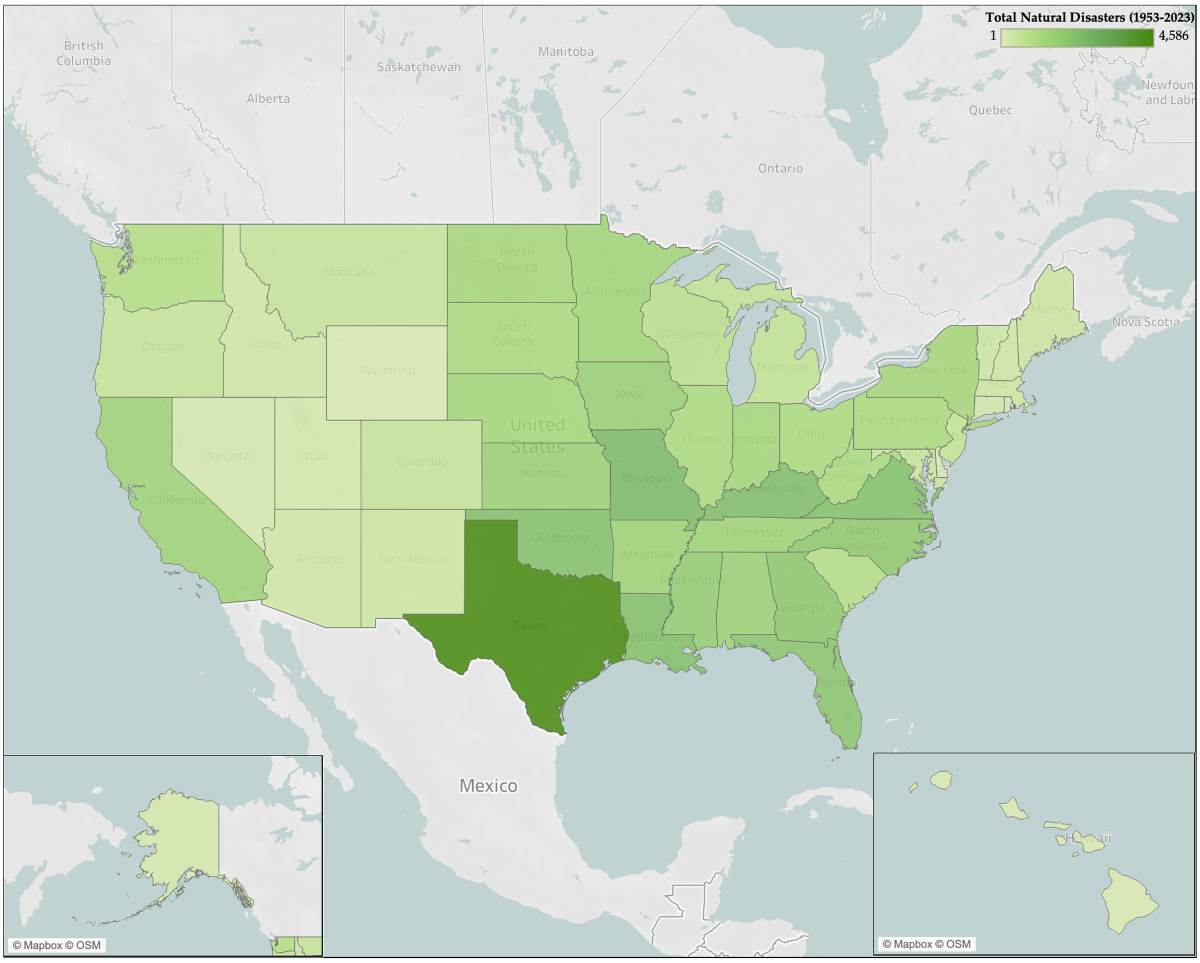

Figure 5: Heatmap of natural disaster types by state.

Figure 5 provides a visual representation of natural disaster frequency across the United States, with states shaded according to their total number of recorded natural disasters. Darker shades indicate higher frequencies, and it is immediately apparent that Texas has the highest count of natural disasters. Other states with significant disaster frequencies include Louisiana, Florida, and North Carolina, reflecting a pattern consistent with regions prone to hurricanes and other extreme weather events. This map sets the stage for the next figure (Figure 6), which breaks down the top 20 states by disaster frequency and identifies the specific types of disasters that occur most often. The heatmap underscores the concentration of disaster events in Texas and other high-risk states, aligning with the detailed breakdown in the following section.

Disaster Types by State: Top 20

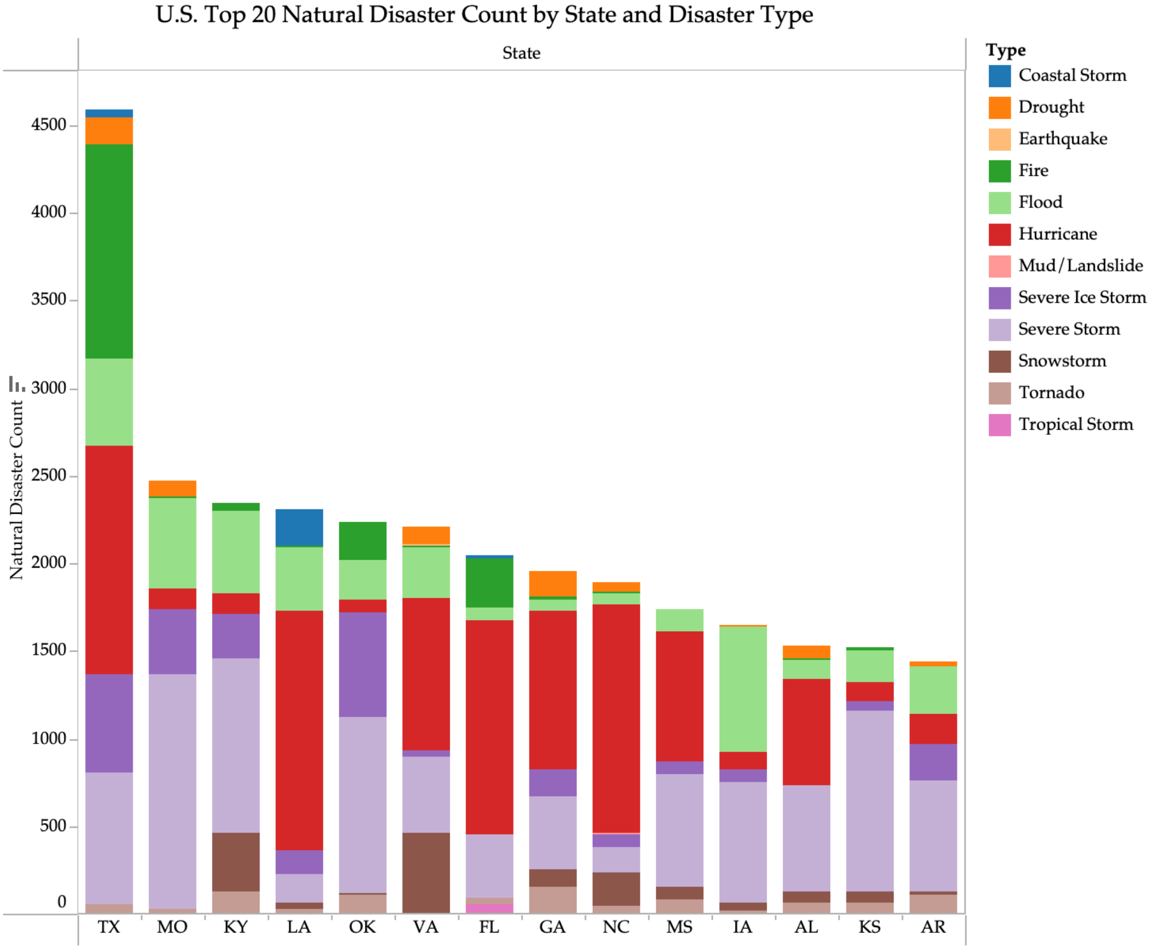

Figure 6: Top 20 Natural Disaster Count by Disaster Type.

As shown in Figure 6, hurricanes are the most frequent disaster type in states like TX, LA, FL, NC, and VA. Texas, ranking highest overall, is frequently affected by hurricanes as well as other disaster types like tornadoes and floods, reflecting its vulnerability to diverse extreme weather conditions. Similarly, Louisiana, Florida, North Carolina, and Virginia are also highly susceptible to hurricanes due to their coastal geography.

7.2 NSDUH Population

SPD Counts Heatmap by Gender

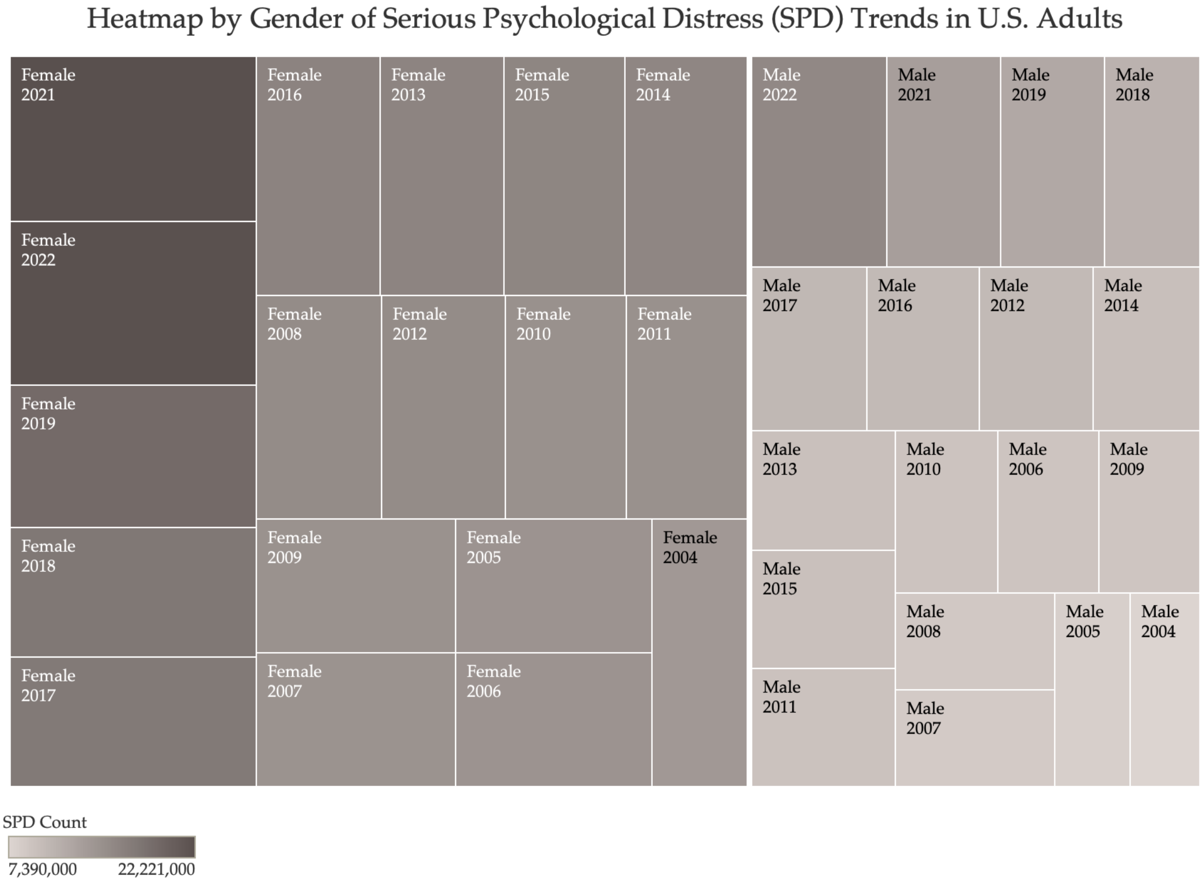

Figure 7: Heatmap showing SPD counts by gender over time.

The intensity of color reflects the total number of SPD cases. Figure 7 provides a gender-wise breakdown of SPD counts over time. Female populations consistently exhibit higher SPD counts, as indicated by darker heatmap shades. This pattern aligns with the earlier trend analysis, where females showed steeper increases in distress levels. This reveals a clear upward trend in SPD cases over the years, with notable differences between male and female populations. Heatmaps further emphasize the temporal and gender-specific concentration of psychological distress, highlighting the need for targeted mental health interventions.

7.3 YLDs (GBD Study)

Anxiety and/or Depression Frequency by State Heatmap

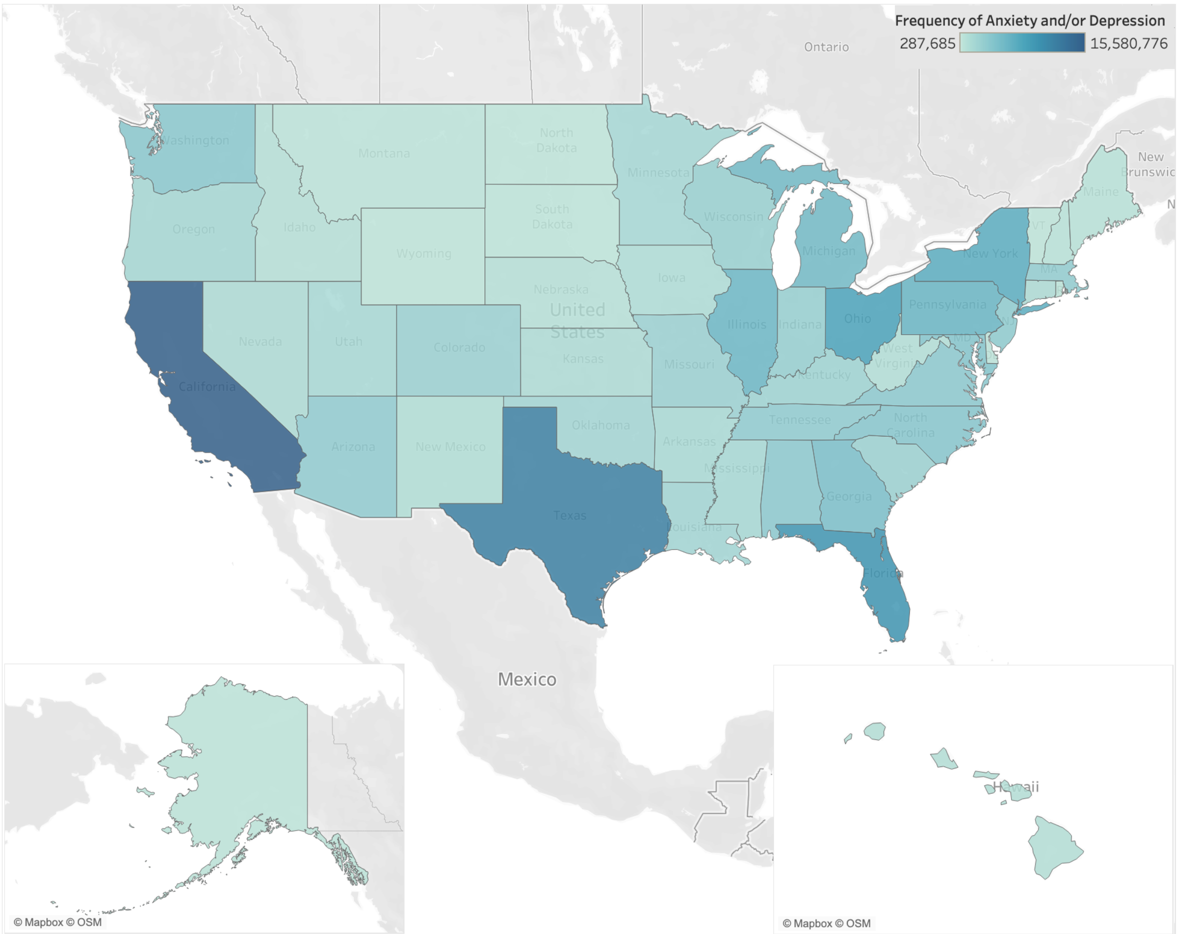

Figure 8: State YLD heatmap.

The heatmap in Figure 8 highlights the distribution of anxiety and/or depression frequencies across various states. The darkest regions, which indicate the highest frequencies, are observed in California, Texas, Florida, and Ohio. These states are notable not only for their large populations but also for their diverse socioeconomic conditions, urbanization levels, and varying exposure to climate-related stressors.

California, for example, has been significantly impacted by wildfires and prolonged droughts, which may contribute to heightened levels of anxiety and depression. Texas and Florida, with their frequent experiences of extreme weather events like hurricanes and heatwaves, also show elevated levels of mental health challenges. Ohio's darker shading might be linked to industrial decline and socioeconomic factors contributing to mental health burdens.

Heatmaps by Gender

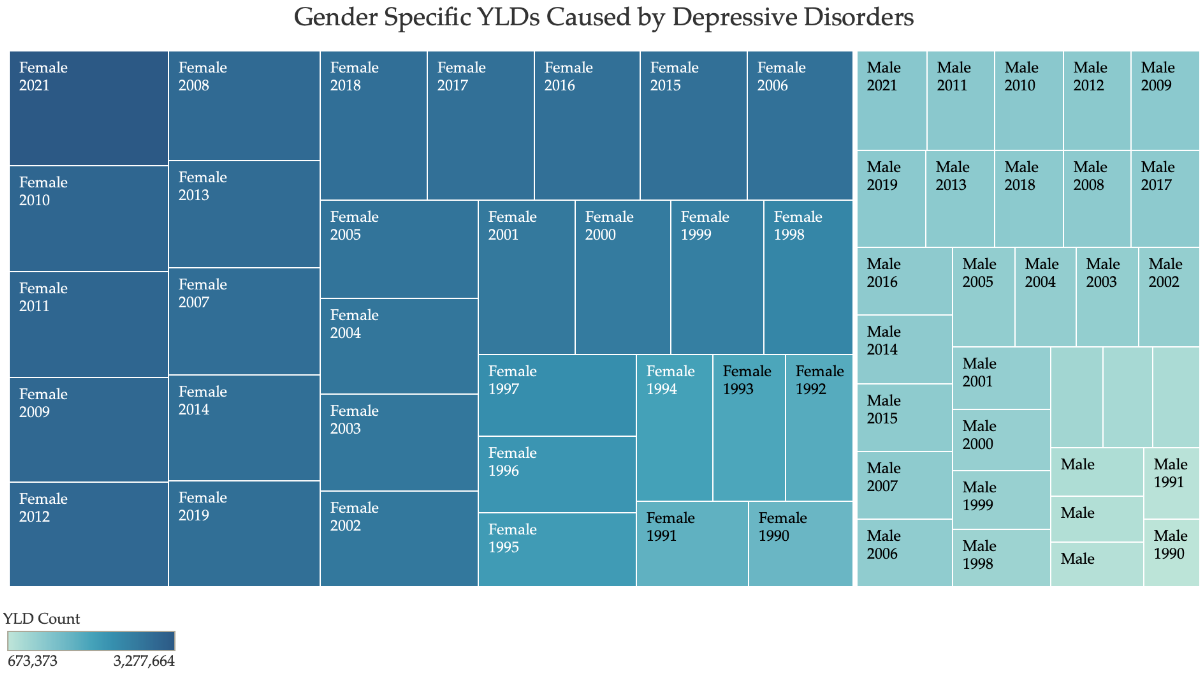

Figure 9: Depression heatmap by gender.

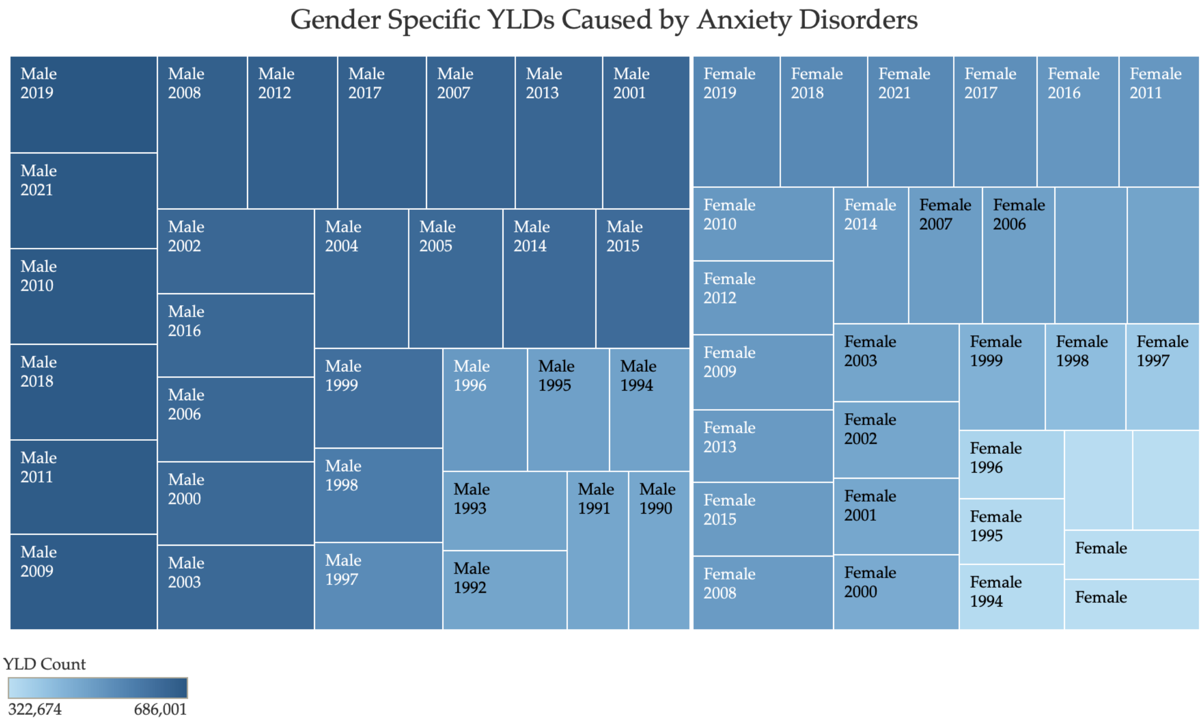

Figure 10: Anxiety heatmap by gender.

The heatmaps in Figure 9 and Figure 10 visually emphasize the gendered trends observed, with darker shades of depression in women and anxiety in men across all years. The darker shades indicate higher prevalence, and shows that females have higher rates of depression across all years.

Insights and Gender-Specific Trends

The exploratory analysis reveals notable gender differences in the prevalence of depression and anxiety. Over time, depression has consistently shown higher rates in women compared to men, a trend that aligns with established research linking hormonal, psychological, and societal pressures to higher depression risks among women. Conversely, anxiety prevalence is marginally higher in men, which challenges common assumptions about anxiety being more prevalent among women. This discrepancy may reflect under-reporting by women or differing expressions of anxiety symptoms across genders. Moreover, the overall trend highlights that depression is a more prevalent condition than anxiety across all years, suggesting it may represent a larger mental health burden. These insights are critical for designing gender-sensitive mental health policies and interventions.

7.4 Relationships

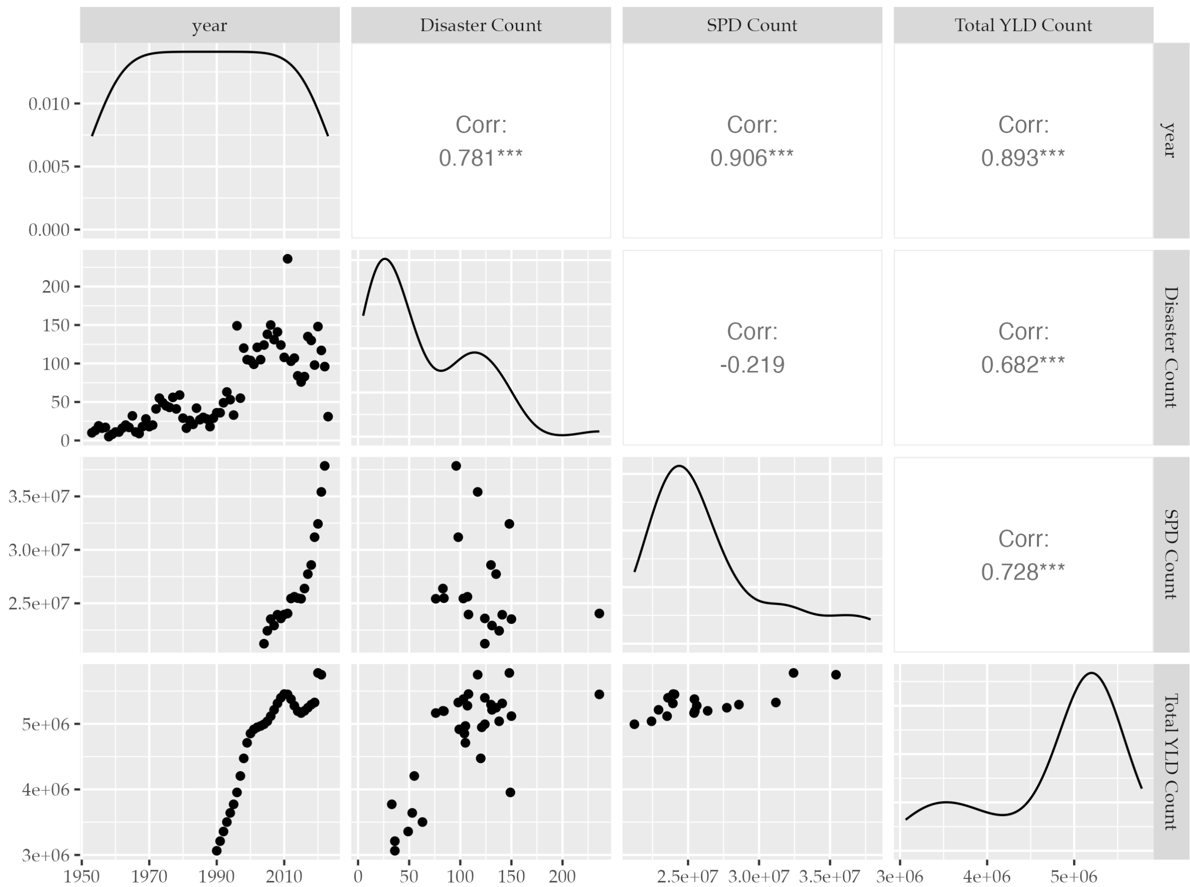

Figure 11: Correlogram scatter plot matrix between year, disaster, SPD, and YLD counts.

The figure presents a correlation scatter plot matrix between the variables' year, disaster, SPD, and YLD counts, showing the relationships between these factors over time. All datasets exhibit an upward trend as years progress, suggesting a temporal increase in each of the measured variables.

- **Disaster-Year (0.781)**: There is a strong positive correlation between disaster and year, indicating that the frequency or intensity of disasters has been increasing over time. This trend suggests that emotional and health impacts related to extreme weather events may also be on the rise.
- **Disaster-SPD (-0.219)**: The weak negative correlation between disaster and SPD counts suggests that as disaster events become more frequent or intense, the specific emotional responses captured by SPD are slightly reduced. This could imply that people's emotional expressions shift or are more focused on other aspects during extreme events.
- **Disaster-YLD (0.682)**: The moderate positive correlation between disaster and YLD counts suggests that more frequent or intense disasters are associated with a higher burden of long-term health impacts, potentially including both physical and emotional health consequences.
- **SPD-Year (0.906)**: There is a very strong positive correlation between SPD and year, showing that emotional distress related to extreme weather events has been increasing over time. This could reflect a growing sense of fear, anxiety, or other negative emotional states as such events become more frequent.
- **SPD-YLD (0.893)**: The strong positive correlation between SPD and YLD indicates that higher emotional distress (captured by SPD) is linked to a higher burden of health impacts, which may include mental health challenges like PTSD or anxiety disorders.
- **YLD-Year (0.893)**: The strong positive correlation between YLD and year suggests that over time, the long-term health impacts (including emotional and psychological tolls) of extreme weather events have been escalating, supporting the broader trend of increasing disaster severity and frequency.

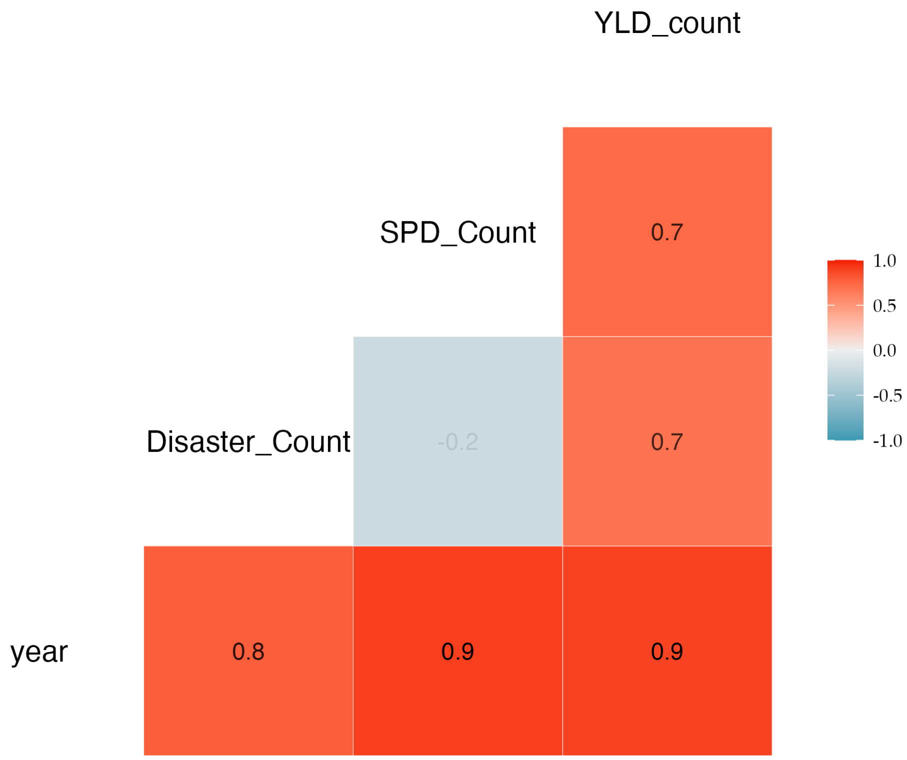

Figure 12: Correlation plot showing the relationships between year, disaster, SPD, and YLD counts.

This plot visually complements the correlation matrix by illustrating the strength of relationships between the variables through color hues. The color palette ranges from red (negative correlation) to blue (positive correlation), with the intensity of the color indicating the strength of the correlation. From the plot, the strong positive correlation between **year** and both **SPD** and **YLD** is clearly visible, represented by deep blue colors. This reflects the increasing emotional and health impacts over time. The moderate positive correlation between **disaster** and **YLD** is also evident, shown with a lighter blue. The slight negative correlation between **disaster** and **SPD** appears as a pale red, confirming the subtle inverse relationship observed in the matrix. The correlation plot helps reinforce the quantitative analysis by providing a more intuitive understanding of how these variables evolve together over time.

8 Time-Series with Statistical Analysis

8.1 U.S. Natural Disaster Temporal Trends

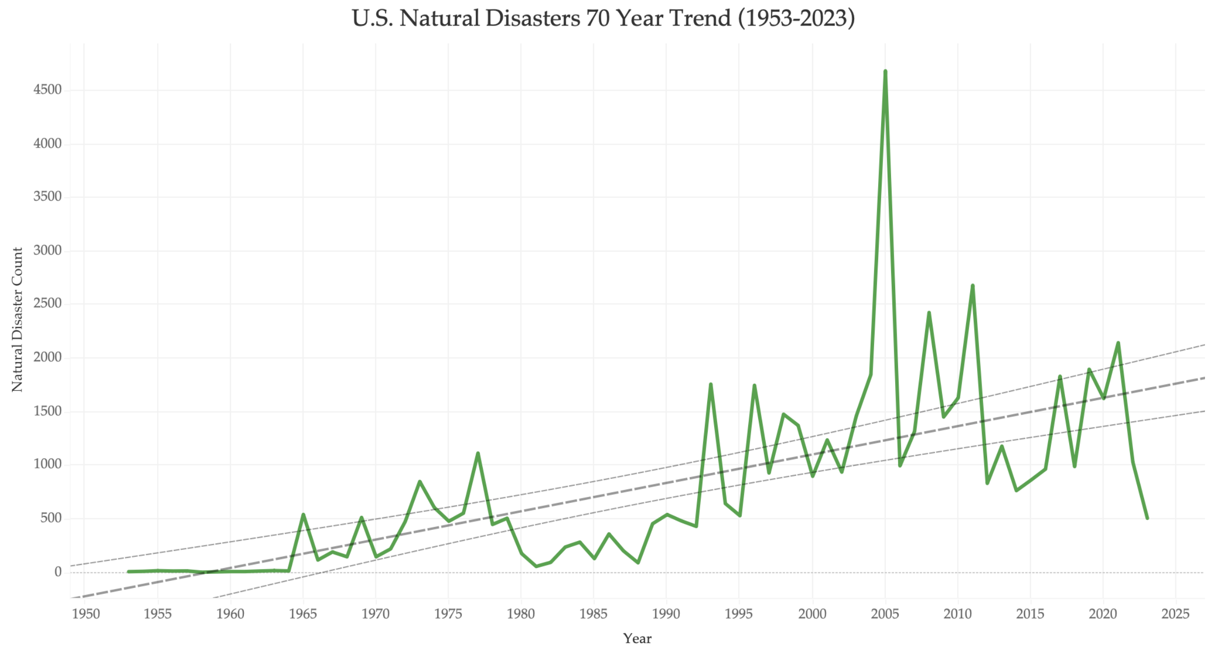

Figure 13: Trend line of natural disaster count by fiscal years declared.

The temporal analysis of FEMA natural disaster declarations reveals a distinct upward trend over the 70-year period. Figure 13 shows the count of disaster declarations by Fiscal Years (Fy). A linear trend model was computed for the data,

$$\text{Number of Disasters} = -51,881.7 + 26.49 \times \text{Year}.$$

This suggests an average increase of approximately 26.49 disaster declarations per year. It is statistically significant with a p-value of less than 0.0001. The R-squared value of 0.45 indicates a moderate correlation between Fy and the count of disasters. This trend likely reflects a combination of environmental changes, increased population density in vulnerable areas, and evolving policies for disaster declaration.

The increasing frequency of natural disasters could directly influence the emotional responses of affected populations, which is the core focus of this thesis. As natural disaster frequency rises, it is essential to consider the psychological impacts, such as anxiety, PTSD, and other emotional responses, which may intensify as people experience more frequent disruptions. This data supports the hypothesis that extreme weather events may have a growing emotional toll, which could be measured using sentiment analysis and time-series analysis of public responses, as discussed in the study.

8.2 SPD Temporal Trends (2004-2022)

The overall trend in SPD U.S. Adults prevalence across the analyzed time period is presented in Figure 14. This visualization reveals significant temporal patterns in the rate of serious psychological distress. Factors such as societal changes, policy shifts, or major external events, including economic downturns and natural disasters, may contribute to these trends, reflecting the interplay between environmental stressors and mental health.

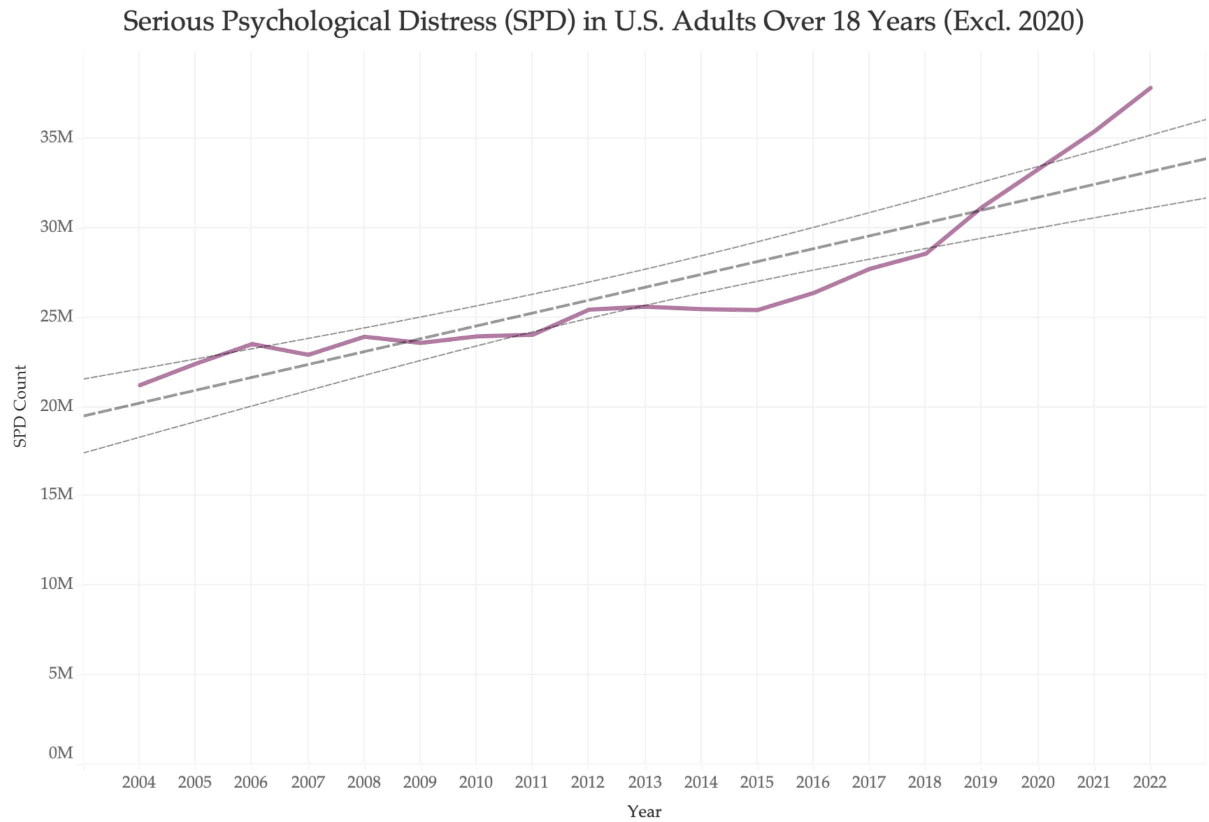

Figure 14: Linear trend and model of SPD counts over time.

A linear trend model highlights a significant increase in SPD counts over the analyzed years. The temporal trend of the sum of SPD counts for each year reveals a steady increase in reported cases over time, as shown in Figure 14. A linear trend model was computed with the following results:

- R-Squared: 0.804, indicating strong correlation.
- p-value: < 0.0001, confirming significance.
- Standard Error: 2.02×10^6.

The formula for the model is:

$$\text{SPD Count} = 719,688 \times \text{Year} - 1.42 \times 10^9$$

This suggests an average yearly increase of approximately 719,688 cases. This increase aligns with the broader hypothesis of this thesis, which investigates how external stressors, such as natural disasters, may exacerbate psychological distress over time. As SPD counts rise, understanding the temporal patterns and correlating them with environmental or societal stressors becomes vital. The results support the importance of contextualizing mental health trends alongside external events to assess their potential cumulative emotional and psychological impact.

SPD Counts by Gender Over Time

To further examine the nuances of SPD trends, Figure 15 depicts SPD prevalence by gender. A linear trend model was computed for the weighted count of SPD cases by year, revealing statistically significant results for both genders. The formula accounts for gender-specific intercepts and slopes, indicating differing trends over time.

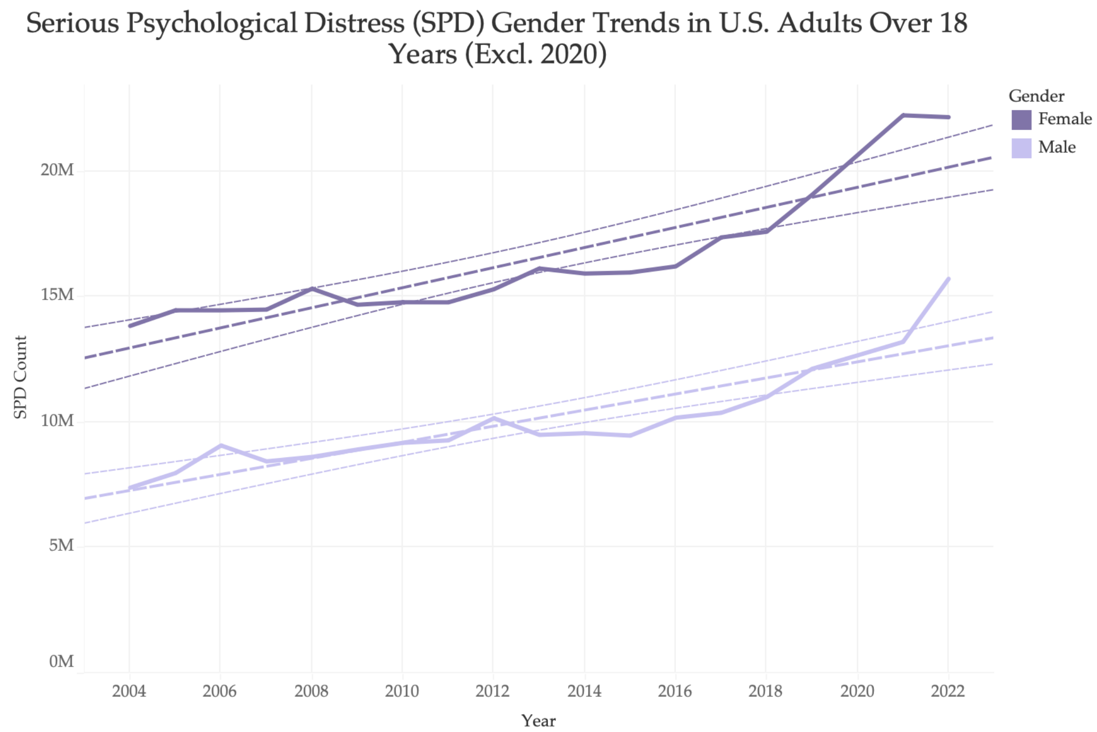

Figure 15: Linear trend and model of SPD counts by gender over time.

- Overall Fit: The R-squared value of 0.931 indicates a very strong correlation between year and SPD counts, accounting for gender-specific trends.
- Statistical Significance: The p-value for the overall model and gender as a factor is < 0.0001, confirming the reliability of the observed trends.
- Coefficients:
  - Female:
  $$\text{Weighted Count} = 399,661 \times \text{Year} - 7.88 \times 10^8$$
  This suggests an average yearly increase of approximately 399,661 cases for females.
  - Male:
  $$\text{Weighted Count} = 320,027 \times \text{Year} - 6.34 \times 10^8$$
  This suggests an average yearly increase of approximately 320,027 cases for males.

The gender-specific trends underscore that while both males and females have seen increases in SPD prevalence over the years, females experience a slightly steeper rise. This disparity may reflect societal and psychological differences, including how external stressors, such as natural disasters and economic instability, affect different genders. These findings are crucial in the context of this thesis, as they highlight the necessity of considering demographic-specific emotional responses when analyzing the psychological impact of extreme weather events and other stressors.

8.3 Depression and Anxiety Induced YLDs Temporal Trends

Overall Trends

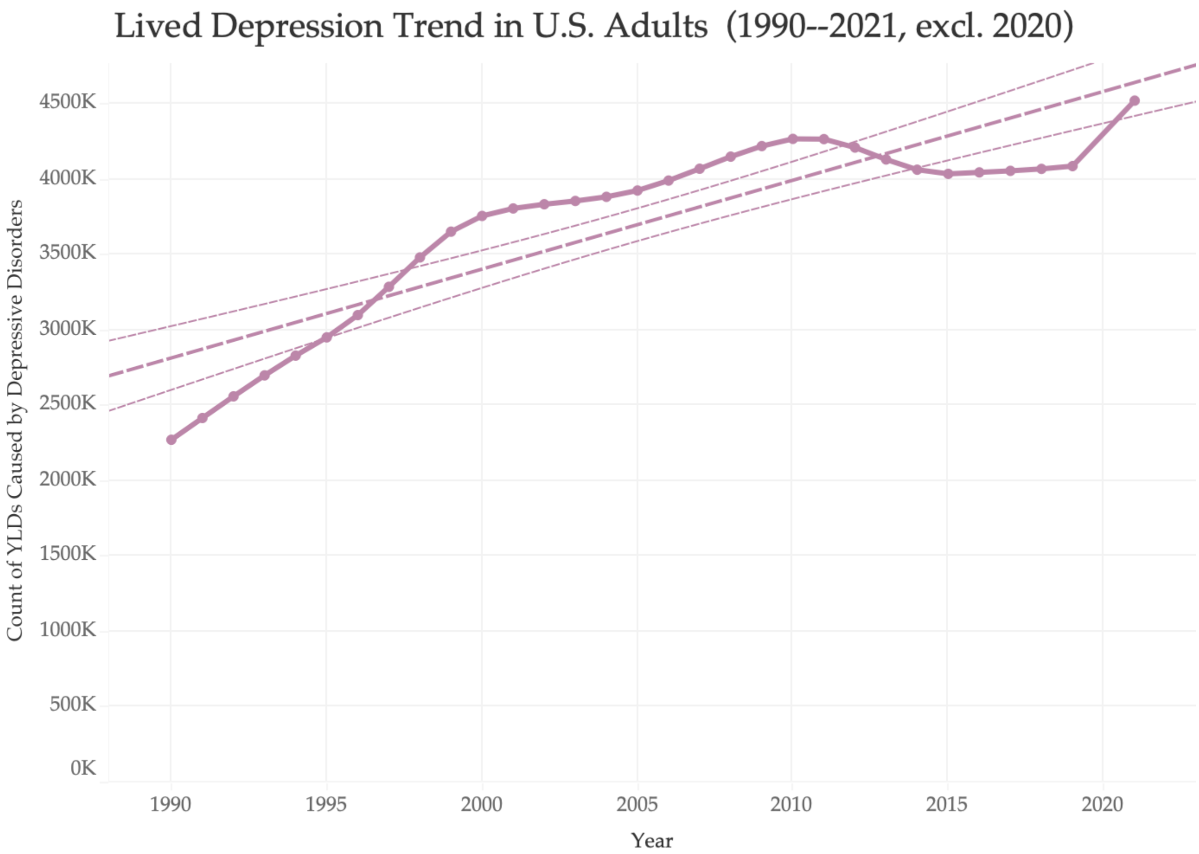

Figure 16: Overall depression trends.

Figure 16 graph shows the overall depression trends over time, highlighting its growing prevalence. To understand the burden of depressive disorders over time, Figure 16 presents the trend of Years Lived with Disability (YLD) caused by depressive disorders in the U.S. A linear trend model was computed for the total YLD count by year, yielding statistically significant results.

- **Overall Fit:** The R-squared value of 0.774 indicates a strong correlation between year and YLD counts.
- **Statistical Significance:** The p-value for the model is < 0.0001, confirming the reliability of the observed trends.

- **Coefficients:**
$$\text{YLD Count} = 58,939.6 \times \text{Year} - 1.14 \times 10^8$$
This suggests an average yearly increase of approximately 58,940 YLD cases caused by depressive disorders.

The steady rise in YLD caused by depressive disorders highlights the growing public health burden posed by mental health issues. These findings are particularly relevant in the context of this thesis, as they underline the interplay between rising emotional distress and increasing extreme weather events. Depressive disorders, exacerbated by environmental stressors such as disasters, could compound the psychological toll on affected populations. Incorporating these trends into sentiment and time-series analyses allows for a deeper understanding of how societal well-being is influenced by environmental and external stressors. Figure 17 illustrates the overall trends in Years Lived with Disability (YLD) caused by anxiety disorders in the U.S. over time. The data reveals a steady increase in YLD counts, reflecting the rising prevalence of anxiety disorders.

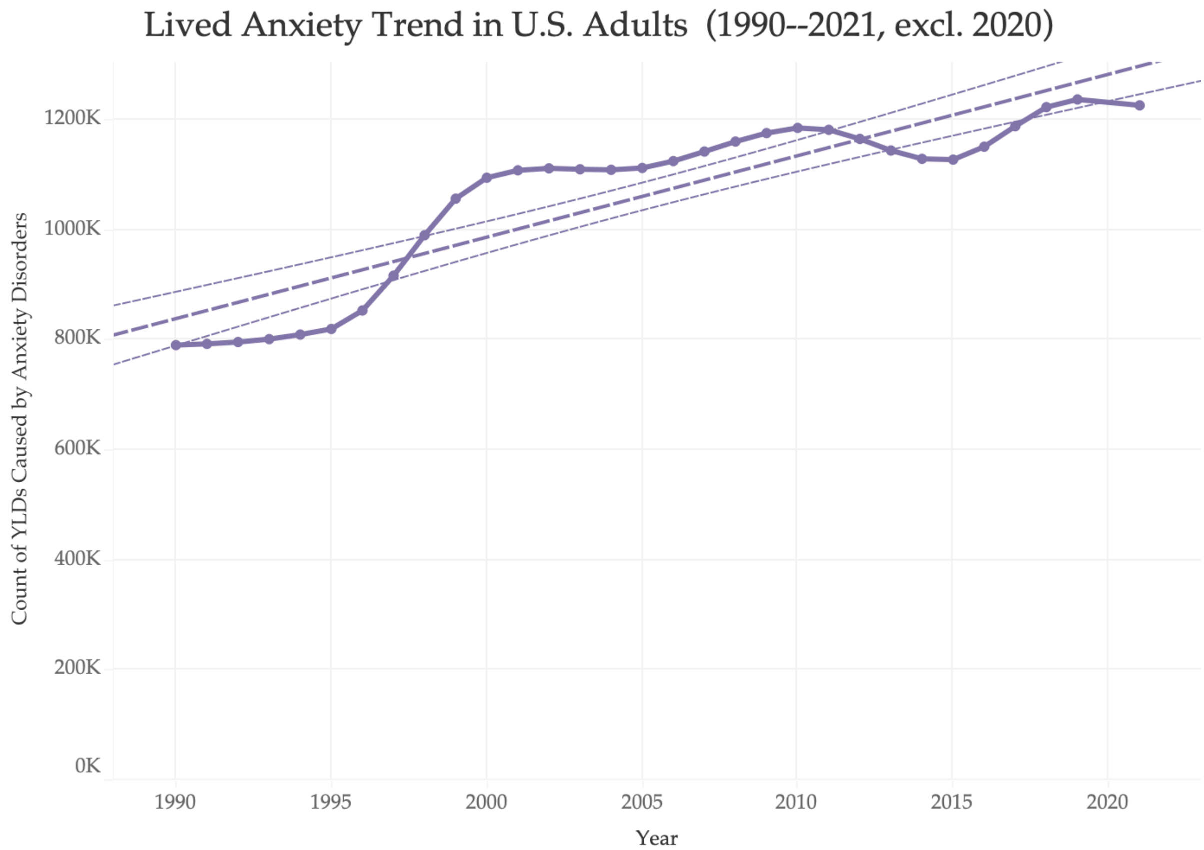

Figure 17: Linear trend and model of YLD caused by anxiety disorders over time.

- **Overall Fit:** The R-squared value of 0.803 indicates a strong correlation between year and YLD counts.

- **Statistical Significance:** The p-value for the model is < 0.0001, confirming the reliability of the observed trends.

- **Coefficients:**

$$\text{YLD Count} = 14,803.4 \times \text{Year} - 2.86 \times 10^7$$

  This suggests an average yearly increase of approximately 14,803 YLD cases caused by anxiety disorders.

The rising trend in YLD caused by anxiety disorders underscores the growing mental health challenges faced by society. These findings are critical in the context of this thesis, as they highlight the compounding emotional toll associated with increasing exposure to extreme weather events. Anxiety disorders are particularly sensitive to external stressors, such as natural disasters, which can exacerbate symptoms and contribute to a broader public health burden. Analyzing these trends provides valuable context for understanding the interplay between mental health and environmental stressors in the study.

As seen in Figure 16 and Figure 17, the overall prevalence of both depression and anxiety has increased over time, emphasizing the growing burden of these mental health conditions. These trends suggest the need for broader mental health interventions and policies. Depression prevalence has consistently exceeded that of anxiety over time, which aligns with its classification as a more pervasive condition globally. The data underscores the need to focus on targeted mental health interventions addressing this disparity.

Trends by Gender

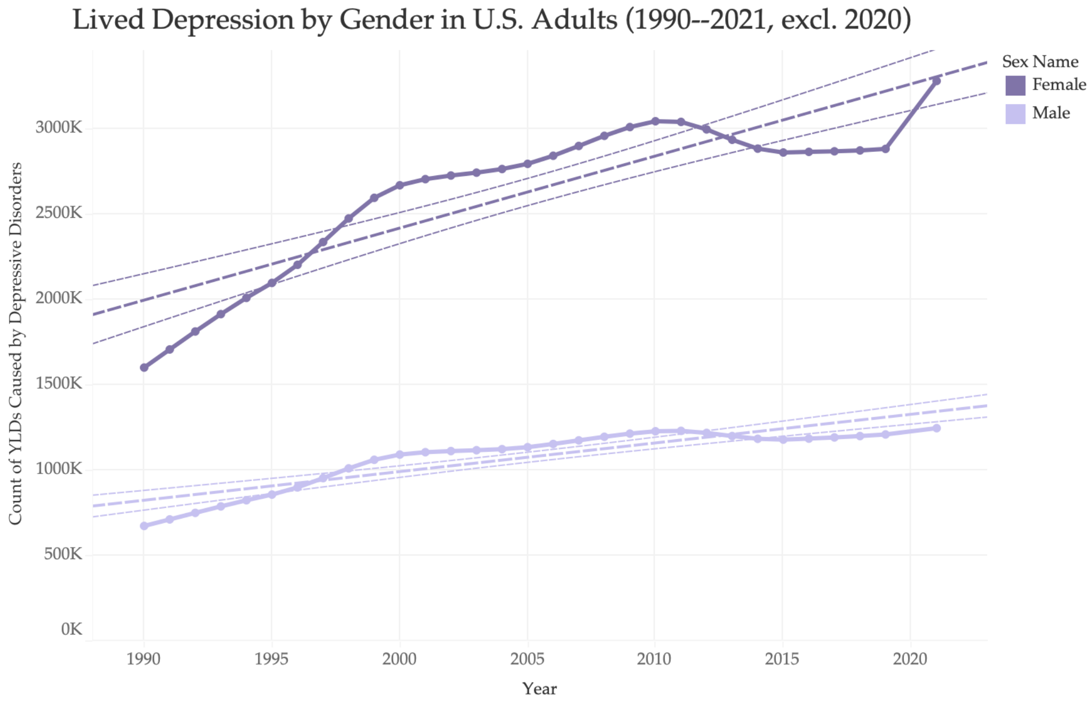

Figure 18: Depression trends by gender.

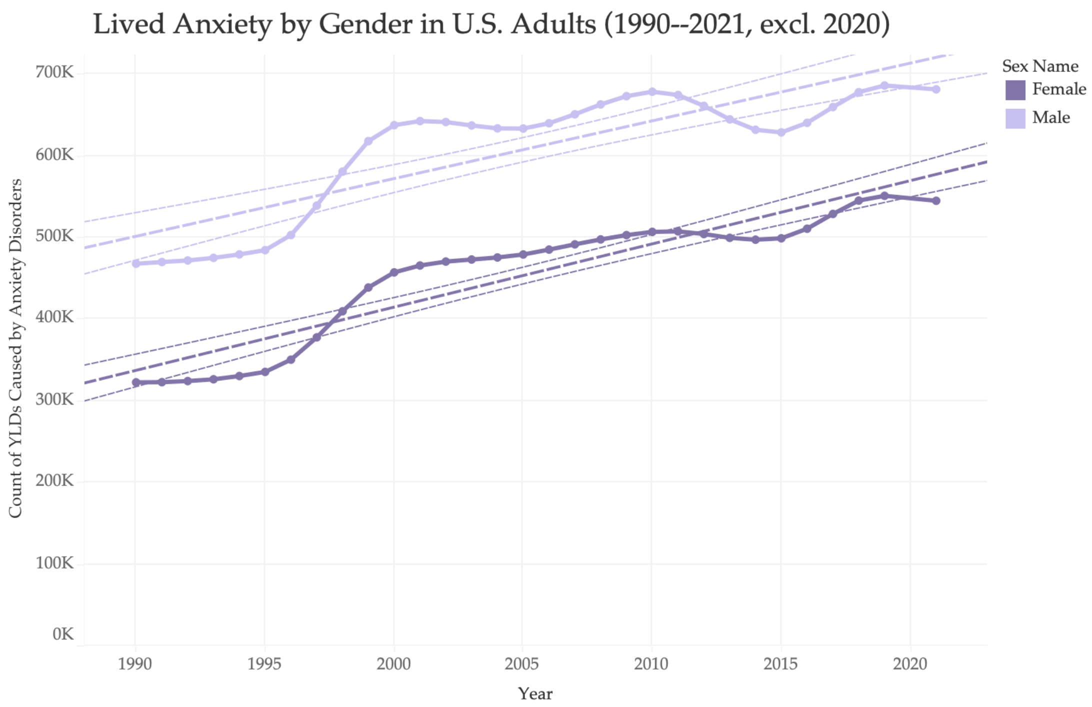

Figure 19: Anxiety trends by gender.

For depression (Figure 18), females consistently show a significantly higher prevalence compared to males. Conversely, anxiety trends (Figure 19) reveal a slightly higher prevalence among males. Both conditions demonstrate a steady upward trend over time, reflecting the broader increase in mental health challenges. This gender-based differentiation aligns with existing literature, which attributes the disparities to a mix of sociocultural and biological factors.

- Overall Fit: The R-squared values for depression (0.965) and anxiety (0.901) indicate strong models that explain a significant portion of the variance in YLD trends by gender over time.
- Statistical Significance: Both models report p-values < 0.0001, confirming the reliability of the observed trends.
- Gender-Specific Trends:
  - Depression (Female):

    $$\text{YLD Count} = 42,171.6 \times \text{Year} - 8.19 \times 10^7$$

    Females show an average yearly increase of 42,172 YLD cases due to depression.
  - Depression (Male):

    $$\text{YLD Count} = 16,768.1 \times \text{Year} - 3.25 \times 10^7$$

    Males show an average yearly increase of 16,768 YLD cases due to depression.

– Anxiety (Female):

$$\text{YLD Count} = 7,739.84 \times \text{Year} - 1.51 \times 10^7$$

Females show an average yearly increase of 7,740 YLD cases due to anxiety.

– Anxiety (Male):

$$\text{YLD Count} = 7,063.59 \times \text{Year} - 1.36 \times 10^7$$

Males show an average yearly increase of 7,064 YLD cases due to anxiety.

The trends underscore notable gender disparities in mental health outcomes, with females disproportionately affected by depression and males reporting slightly higher prevalence of anxiety. These findings highlight the need for targeted mental health interventions that address the unique sociocultural and biological influences on gendered mental health outcomes. Moreover, the steady upward trends in YLD for both depression and anxiety, observed across genders, underscore the growing public health burden posed by these conditions.

8.4 Statistically Significant Findings

8.4.1 U.S. Natural Disaster Temporal Trends

The FEMA natural disasters from 1953 to 2022 reveals a statistically significant upward trend in disaster frequency (Figure 13), with a p-value ≤ 0.0001. This trend supports the broader hypothesis that increased frequency and intensity of extreme weather events are likely to influence demographic responses. We establish a foundational link between environmental changes and potential impacts on emotional and psychological outcomes, aligning with the conditions under which the alternative hypothesis (H_1) may hold.

8.4.2 SPD Temporal Trends (2004–2022)

The significant upward trend in SPD prevalence over time (Figure 14)—with an R-squared value of 0.804 and p-value ≤ 0.0001—indicates that serious psychological distress is increasing in the U.S. The strong correlation between natural disaster frequency and SPD prevalence, as shown in Figure 12, suggests that environmental stressors such as extreme weather events could influence mental health outcomes. These findings challenge the null hypothesis (H_0) by demonstrating a potential link between disaster trends and mental health outcomes.

8.4.3 Depression and Anxiety-Induced YLD Trends (1990–2021)

The upward trends in depression and anxiety-induced YLDs (Figure 16 and Figure 17) highlight the growing mental health burden over time. The significant gender disparities observed (Figure 18 and Figure 19)—with women reporting higher depression prevalence and men slightly higher anxiety prevalence—align with (H_1), suggesting that demographic differences (e.g., gender) influence mental health outcomes in response to broader societal and environmental changes.

8.4.4 Correlations Between Variables

The correlation analyses (Figure 11 and Figure 12) further connect these findings to the hypotheses:

- Natural Disasters and SPD: The strong positive correlation between disaster frequency and SPD prevalence supports the premise of (H_1), indicating that extreme weather events may exacerbate emotional and mental health challenges.
- Temporal Correlations: The strong correlation between time and variables such as YLD and SPD counts reinforces the idea that both environmental and mental health trends are escalating together, indirectly supporting (H_1).
- Dataset Timeline Gaps: The negative correlation between natural disasters and SPD prevalence, likely due to non-overlapping timelines, highlights the complexity of data integration.

9 Sentiment Analysis

The Reddit data related to Hurricane Helene was analyzed for sentiment using machine learning techniques. A web scraper was used to collect the data, and several machine learning models were used to label data sentiments.

9.1 Tropical Storm and Hurricane Helene on Reddit

Recent social media discourse has highlighted the impacts and ongoing recovery efforts related to Hurricane Helene, a significant natural disaster that affected several U.S. states between late September and early October 2024. According to the Federal Emergency Management Agency (FEMA) [6], the timeline of the storm's impact across states is as follows (as of November 21, 2024):

- **Florida**: September 23, 2024 – October 7, 2024 (Hurricane)
- **Georgia**: September 24, 2024 – October 30, 2024 (Hurricane)
- **North Carolina**: Beginning September 25, 2024 (Tropical Storm, ongoing)
- **South Carolina**: September 25, 2024 – October 7, 2024 (Hurricane)
- **Tennessee**: September 26, 2024 – September 30, 2024 (Tropical Storm)
- **Virginia**: September 25, 2024 – October 3, 2024 (Tropical Storm)

This timeline provides critical context for understanding the discussions on Reddit regarding the storm's impacts, community responses, and recovery efforts.

9.2 Building and Deploying a Reddit Web Scraper

The analysis of emotional responses related to Hurricane Helene on Reddit was conducted using a custom web scraper developed with Python and the `Python Reddit API Wrapper (PRAW)` library. The process is outlined as follows:

9.2.1 API Setup

The PRAW library was used to authenticate a Reddit client with a user agent and credentials. This allowed for the extraction of posts and comments while adhering to Reddit's API usage guidelines.

```
[language=Python]
reddit = praw.Reddit(
    client_id='your_client_id',
    client_secret='your_client_secret',
    user_agent='script:my_hurricane_scraper:v1.0'
)
```

9.2.2 Subreddit Selection

To analyze discussions about the emotional and mental health impacts of Hurricane Helene, relevant subreddits were identified. Posts within the month of the event were targeted from subreddits such as depression, NorthCarolina, florida, Tennessee, and Virginia.

```
subreddits = ['mentalhealth', 'MentalHealthSupport', 'depression',
               'anxiety', 'florida', 'Georgia', 'NorthCarolina',
               'Tennessee', 'Virginia']

for subreddit in subreddits:
    print(f"Scraping subreddit: {subreddit}")

    # Fetch posts: mix search, top, and new
    filters = [
        lambda: reddit.subreddit(subreddit).top(time_filter="week", limit=100),
        lambda: reddit.subreddit(subreddit).search(keywords, sort="relevance",
        time_filter="month", limit=100)
    ]

    for filter_fn in filters:
        try:
            # Fetch and process posts here
            pass
        except Exception as e:
            print(f"Error fetching posts: {e}")
```

9.2.3 Defining Keywords

A curated list of keywords was developed to filter posts and comments for relevance. These keywords represent positive, negative, and neutral emotions, as well as terms specific to Hurricane Helene and related natural disasters. The keywords were stored in the variables positive_emotions_keywords, negative_emotions_keywords, and relevant_helene_ and were applied using a filtering function, contains_keyword(text).

```
positive_emotions_keywords = [
```

```
    'happy', 'joyful', 'grateful', 'thankful', 'hopeful', 'relieved', ...
]

negative_emotions_keywords = [
    'sad', 'angry', 'frustrated', 'fearful', 'scared', 'hopeless', ...
]

relevant_helene_keywords = [
    'helene', 'storm helene', 'evacuate', 'evacuation', 'aftermath', 'damage', ...
]

# Function to check if a text contains both sentiment and Helene-related keywords
def contains_keyword(text):
    text = text.lower()

    sentiment_match = any(keyword in text for keyword in
                          positive_emotions_keywords +
                          negative_emotions_keywords +
                          neutral_emotions_keywords)

    helene_match = any(keyword in text for keyword in helene_keywords)

    return sentiment_match and helene_match
```

9.2.4 Keyword Matching and Sentiment Classification

The function `get_sentiment(text)` was developed to classify the sentiment of posts and comments containing both sentiment-related keywords and Hurricane Helene-related terms. Sentiments were categorized as Positive, Negative, or Neutral. Posts were retrieved from the `top`, `new`, and `hot` sorting categories across subreddits to ensure diverse content. Titles, self-text, and comments were analyzed for keyword matches and sentiment classification.

```
# Classify sentiment based on the presence of sentiment keywords
def get_sentiment(text):
    text = text.lower()
    if contains_keyword(text):
        if any(keyword in text for keyword in negative_emotions_keywords):
            return 'Negative'
        elif any(keyword in text for keyword in positive_emotions_keywords):
            return 'Positive'
        else:
            return 'Neutral'
    return 'No relevant keywords found'

# Scraping and sentiment analysis for subreddit posts
for subreddit in subreddits:
    posts = reddit.subreddit(subreddit).search(
        'mental health OR anxiety OR depression',
```

```
        sort='relevance',
        time_filter='month',
        limit=100
    )
    for post in posts:
        if contains_keyword(post.title):
            sentiment = get_sentiment(post.title)
            titles_data.append({'Text': post.title, 'Sentiment': sentiment})
```

9.2.5 Data Storage

The extracted data (titles, post content, and comments) was stored in separate CSV files for further analysis.

```
titles_df = pd.DataFrame(titles_data)
posts_df = pd.DataFrame(posts_data)
comments_df = pd.DataFrame(comments_data)

titles_df.to_csv('titles.csv', index=False)
posts_df.to_csv('posts.csv', index=False)
comments_df.to_csv('comments.csv', index=False)
```

9.2.6 Combining and Cleaning the Data

After data collection, various CSV files were combined into one dataset. The data was then cleaned to ensure consistency and relevance for sentiment analysis. Key steps in the cleaning process include removing duplicates, filtering out irrelevant entries (e.g., bot-related content), and eliminating unnecessary columns.

```
data = ["2hot_no_neutral_helene_comments.csv", ...]
# List of file names to combine
dataframes = []

# Load and combine datasets
for file in data:
    file_path = os.path.join("data", "processed", file)
    if os.path.exists(file_path) and os.path.getsize(file_path) > 0:
        try:
            df = pd.read_csv(file_path)
            dataframes.append(df)
        except Exception as e:
            print(f"Error reading {file_path}: {e}")
    else:
        print(f"Skipping empty or missing file: {file_path}")

if dataframes:
    combined_comments = pd.concat(dataframes, ignore_index=True)
    combined_comments.to_csv("data/processed/combined.csv", index=False)
```

9.2.7 Text Cleaning

A custom text cleaning function was applied to standardize the dataset by converting text to lowercase, removing digits, punctuation, URLs, emojis, and non-ASCII characters. This ensured the text was ready for analysis.

```
def clean_text(text):
    # Convert to lowercase and clean
    text = text.lower()
    text = re.sub(r'\d+', '', text)
    text = re.sub(rf"[{string.punctuation}]", "", text)
    text = re.sub(r'http\S+|www\S+|https\S+', '', text)
    text = emoji_pattern.sub(r'', text)
    text = re.sub(r'[^\x00-\x7F]+', '', text)
    return text
```

9.2.8 Final Data Preparation

After cleaning, unnecessary columns were removed, ensuring that the dataset was formatted correctly for sentiment analysis. The cleaned dataset was then saved for further processing.

```
comments['cleaned_text'] = comments['Text'].apply(clean_text)
df_cleaned = comments.drop(columns=['Text', 'Sentiment', 'Type',
'Post URL', 'Created', 'Comment Author', 'Author'])
df_cleaned.to_csv("combined_clean.csv", index=False)
```

The full Python code and instructions for replicating this process are available on the GitHub repository: `https://github.com/sjanefullerton/Web-Scraper`

9.2.9 Applications and Benefits

This scraper enables researchers to collect large-scale, real-time social media data, offering insights into public discourse surrounding Hurricane Helene. Potential applications include:

- **Sentiment Analysis**: Identifying public reactions to the disaster.
- **Trend Analysis**: Understanding how discussions evolve over time.
- **Disaster Response**: Extracting actionable information for emergency management.

By leveraging the `PRAW` library, the scraper demonstrates a flexible and efficient approach to social media data collection, paving the way for further computational analyses in disaster research.

9.3 Model Testing and Results

Several models were tested for sentiment classification, each selected based on its suitability to handle the textual nature of the data.

9.3.1 Linear Regression

Linear regression was initially tested as a baseline model. It was found to be the most accurate for **smaller datasets**. However, its performance declined as the size of the dataset increased, likely due to its limited ability to capture complex patterns within the data. While linear regression proved effective in identifying simple sentiment, it struggled with the complexities of disaster-related discourse.

9.3.2 Deep Learning (LSTM/GRU)

Subsequently, deep learning models, particularly **LSTM** and **GRU** networks, were tested. These models are designed to capture long-range dependencies in text and were anticipated to outperform linear regression on more complex datasets. Indeed, they demonstrated an improvement in handling the semantic complexity of the posts. However, performance was constrained by the size of the labeled dataset used for training. Despite this limitation, **LSTM and GRU** still outperformed traditional methods like linear regression when applied to **larger, more complex** datasets.

9.3.3 Transformer Models (BERT-based)

The **Transformers** package, particularly models like **Bidirectional Encoder Representations from Transformers (BERT)**, **RoBERTa**, and **DistilBERT**, was found to deliver the **best performance** for sentiment classification in this study. Pretrained on massive corpora, these models excel at capturing **contextual meaning** and understanding nuanced emotional language. After fine-tuning the model on a manually labeled subset of the data, Transformer-based models demonstrated **superior performance**, particularly when identifying sentiments in posts related to **Hurricane Helene**. Transformer models are especially well-suited for handling the semantic complexity of long, conversational text.

9.4 Model Selection and Final Approach

Based on the comparative performance of these models, **Transformer-based models** were selected as the most effective approach for sentiment classification in this study. Their ability to **capture nuanced meanings** and understand **contextual sentiment** made them particularly well-suited for the diverse and complex nature of the data. Furthermore, fine-tuning these models with a **manually labeled dataset** allowed for the achievement of high classification accuracy, which was further supported by post-analysis evaluation metrics.

9.5 Visualizing Sentiments

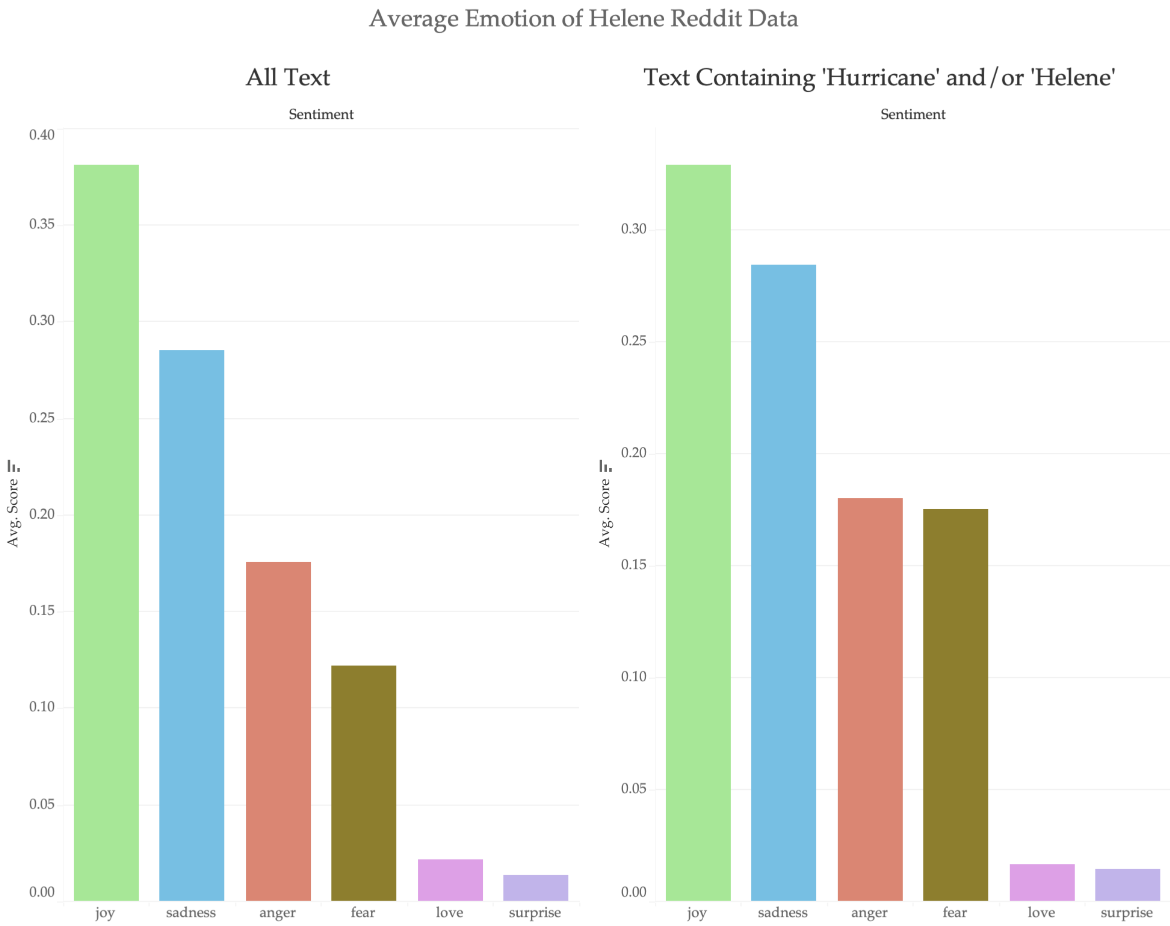

Figure 20: Bar graph of sentiment distribution.

Emotion Distribution Overview
The results from the sentiment analysis of Reddit comments related to Hurricane Helene reveal a complex emotional landscape, with varying levels of sentiment intensity. The following is an interpretation of the findings:

- **Joy**: Despite the destructive nature of hurricanes, a significant portion of the comments express joy. This could reflect moments of relief, hope, or positive emotions in response to the aftermath of the storm. It might indicate community support, personal resilience, or recovery.

- **Sadness**: A strong emotional response to loss, grief, and empathy is expressed through sadness. Given the extensive property damage, injuries, and loss of life caused by the hurricane, this is an expected emotional response in disaster-related discussions.

- **Anger**: Anger often manifests in situations where individuals feel frustrated or powerless. In this case, anger could be directed toward government response, the destruction caused by the hurricane, or the helplessness of affected communities.

- **Fear**: While fear is present, it is the least expressed emotion. This suggests that the immediate fear of the hurricane's impact is lower. The discussions may have focused more on the aftermath and recovery than the anticipation of the disaster.
- **Love** and **Surprise**: These emotions were less prevalent in the dataset. Love might be less common in crisis situations, where the focus is on immediate survival and logistical concerns.
- Both sadness and anger appear to be on average higher when looking at Reddit data that contained the words 'Hurricane' and or 'Helene' specifically.

9.6 Results

The results of these tests highlight the importance of selecting the appropriate model based on the characteristics of the data. For smaller datasets, simpler models like **Linear Regression** can be effective. However, for larger and more complex datasets, especially in the context of **disaster-related sentiment**, advanced models such as **Transformers** provide significantly better accuracy. Future work could involve **fine-tuning** these models with larger datasets, which would likely improve accuracy and yield deeper insights into public sentiment during natural disasters.

With the most accurate model tested, we see Reddit sentiments weighted more in joy, sadness, and anger, than fear, love, or surprise.

9.6.1 Support for Alternative Hypothesis (H_1)

- **Observational Evidence:** Sadness and anger suggest significant negative patterns, while variability in emotions, including joy, rejects the null hypothesis of uniformity.
- **Theoretical Implications:** Emotional responses to disasters reflect diverse experiences and coping mechanisms. This aligns with psychological theories where negativity dominates initial reactions, but resilience and hope emerge over time.

9.6.2 Potential for Future Analysis

- **Temporal Analysis:** Examine how emotional responses evolve during different disaster phases (before, during, after).
- **Fine-Tuning Models:** Use larger labeled datasets to enhance model accuracy, especially for less prevalent emotions like love and surprise.
- **Cross-Platform Analysis:** Explore emotional patterns across other platforms (e.g., Twitter, Facebook) to understand differences by medium and demographic.

10 Discussion

10.1 Research Contribution and Contextual Integration

This study builds on the existing literature by linking long-term time series data on climate and mental health with real-time sentiment analysis during Hurricane Helene. The integration of these methods offers a nuanced view of how natural disasters influence mental health and emotional responses over time. By addressing the posed research questions and testing the hypotheses, this work contributes to understanding the interplay between climate-induced stressors and mental health outcomes.

10.2 Discussion of Research Questions and Hypotheses

10.2.1 Research Question 1: Impact of Extreme Weather on Mental Health

The results provide strong evidence supporting the alternative hypothesis (H_1): *The frequency and intensity of extreme weather events significantly affect the prevalence of mental health disorders, such as anxiety and depression.*

- **Long-Term Trends:** Statistically significant upward trends in natural disaster frequency (Figure 13) and mental health outcomes (Figure 14) indicate a correlation between environmental stressors and increased prevalence of anxiety, depression, and SPD.
- **Demographic Insights:** Gender-specific trends in depression and anxiety (Figure 18 and Figure 19) underscore the differential impacts of environmental stressors, with women experiencing higher rates of depression and men slightly higher rates of anxiety.
- **Correlations Between Variables:** The positive correlations between disaster frequency, SPD prevalence, and mental health burdens such as YLDs further reinforce the relationship between extreme weather events and mental health outcomes.

10.2.2 Research Question 2: Emotional Responses to Hurricane Helene on Social Media

The sentiment analysis of Reddit data supports the alternative hypothesis (H_1): *Emotional responses to natural disasters, such as Hurricane Helene, observed on Reddit significantly vary in negative patterns.*

- **Sentiment Variability:** The presence of sadness, anger, and joy reflects a diverse emotional response. While sadness and anger dominate, joy highlights resilience and recovery themes within affected communities.
- **Negative Patterns:** The high prevalence of sadness and anger, particularly in posts containing keywords like "Hurricane" and "Helene," validates the hypothesis that negative emotional patterns are prominent in disaster-related discussions.

- **Low Fear Levels:** Fear was less pronounced, suggesting that discussions focused more on post-event recovery than pre-disaster anxiety, influenced by the timing of data collection.
- **Model Accuracy:** Advanced models like Transformers outperformed simpler models, underscoring the importance of accurate approaches for capturing nuanced emotional patterns in large datasets.

10.3 Alignment with Overall Study Objectives

The findings align with the study's objective to explore both long-term and real-time emotional and psychological impacts of extreme weather events:

- **Temporal Dynamics:** The significant upward trends in disaster frequency and mental health burdens establish a link between environmental changes and psychological outcomes.
- **Social Media Insights:** Real-time sentiment analysis provides valuable context for understanding community emotional responses during disasters, highlighting resilience amid negativity.
- **Demographic Variability:** Gender and emotional variability suggest that mental health interventions would benefit from being tailored to address specific demographic needs and emotional patterns.

10.4 Limitations

The following limitations are noted:

- **Exclusion of 2020 Data:** Due to Covid-related anomalies, population data from 2020 was excluded as recommended by the 2020 NSDUH release [1].
- **Limited Mental Health Data:** Restricted dataset availability due to sensitivity of mental health data.
- **NSDUH and GBD Coverage:** NSDUH lacked SPD data before 2004, and GBD data was only available from 1990 onward.
- **Climate Data Gaps:** Older datasets potentially had missing records.
- **Social Media Noise:** Sentiment analysis may have been impacted by noisy data, including slang, misspellings, and abbreviations.

10.5 Potential Next Steps

This research demonstrates the potential of sentiment analysis for understanding emotional responses to natural disasters using Reddit data. Future work can enhance these findings through the following steps:

- **Temporal Dynamics:** Analyze how emotional responses evolve across disaster phases (pre-event, during, post-event).

- **Cross-Platform Analysis:** Expand data sources to include Twitter, Facebook, and other platforms to capture broader emotional patterns.
- **Dataset Expansion:** Include data from other natural disasters to provide a comprehensive view of emotional responses across varied events.
- **Model Improvement:** Use larger labeled datasets and advanced techniques, such as hyperparameter optimization and transfer learning, to improve sentiment analysis accuracy.
- **Real-Time Tools:** Develop a web application for real-time sentiment tracking to support disaster response and mental health interventions.
- **Targeted Interventions:** Utilize demographic insights and sentiment variability to design tailored mental health support for affected populations.

10.6 Conclusion

This research reveals the profound influence of natural disasters on long-term mental health trends and immediate emotional responses. By integrating time series analysis with real-time sentiment analysis, it provides a multidimensional understanding of the emotional impacts of extreme weather events. The findings highlight correlations between increasing disaster frequency, rising mental health burdens, and the emotional reactions observed during Hurricane Helene.

From a practical standpoint, the research underscores the potential of computational tools for real-time crisis management and mental health support. Advanced sentiment analysis techniques, coupled with demographic insights, offer valuable resources for designing informed and effective interventions.

Ultimately, this research underscores the need for a collaborative response to address the challenges of climate change and mental health, advancing both scientific understanding and practical solutions to support communities facing these crises.

References

[1] 2020 national survey on drug use and health (nsduh) releases, 2020. Substance Abuse and Mental Health Services Administration (SAMHSA). Accessed: 2024-12-01.

[2] BG Anderson and ML Bell. Weather-related mortality: how heat, cold, and heat waves affect mortality in the united states. *Epidemiology*, 20(2):205–213, 2009.

[3] HL Berry, K Bowen, and T Kjellstrom. Climate change and mental health: a causal pathways framework. *Int J Public Health*, 55(2):123–132, Apr 2010. Epub 2009 Dec 22.

[4] P. Cianconi, S. Betrò, and L. Janiri. The impact of climate change on mental health: A systematic descriptive review. *Front Psychiatry*, 11:74, 2020.

[5] Elaine Enarson. *Gender and natural disasters*. 01 2000. Enarson, E. (2000). Gender and natural disasters. International Development Research Centre.

[6] FEMA.gov. Disaster declarations summaries - v2. `https://www.fema.gov/openfema-data-page/disaster-declarations-summaries-v2`, November 2024. Accessed: 2024-11-23.

[7] Center for Climate and Energy Solutions (C2ES). Extreme weather and climate change, 2024. Accessed: 2024-11-24.

[8] Institute for Health Metrics and Evaluation. Global burden of disease (gbd) 1990-2021. `https://www.healthdata.org/research-analysis/gbd-data`, 2024. Accessed: 2024-11-23.

[9] Institute for Health Metrics and Evaluation. Global burden of disease (gbd). `https://www.healthdata.org/research-analysis/about-gbd`, n.d.

[10] S Galea, CR Brewin, M Gruber, RT Jones, DW King, LA King, RJ McNally, RJ Ursano, M Petukhova, and RC Kessler. Exposure to hurricane-related stressors and mental illness after hurricane katrina. *Arch Gen Psychiatry*, 64(12):1427–1434, 2007.

[11] J Gray, AR Santos-Lozada, G Hard, H Apsley, D O'Sullivan, and AA Jones. Serious psychological distress, substance use disorders, and social issues among men and women in the united states during the covid-19 pandemic. *Am J Health Promot*, 37(7):933–939, Sep 2023. Epub 2023 Jul 3.

[12] EL Lawrance, R Thompson, J Newberry Le Vay, L Page, and N Jennings. The impact of climate change on mental health and emotional wellbeing: A narrative review of current evidence, and its implications. *Int Rev Psychiatry*, 34(5):443–498, Aug 2022. Erratum in: Int Rev Psychiatry. 2022 Aug;34(5):iii. doi: 10.1080/09540261.2022.2161567. PMID: 36165756.

[13] Molly Monsour, Emily Clarke-Rubright, Wil Lieberman-Cribbin, Christopher Timmins, Emanuela Taioli, Rebecca M. Schwartz, Samantha S. Corley, Anna M. Laucis, and Rajendra A. Morey. The impact of climate change on the prevalence of mental illness symptoms. *Journal of Affective Disorders*, 300:430–440, 2022.

[14] U.S. Department of Health, Substance Abuse Human Services, Center for Behavioral Health Statistics Mental Health Services Administration, and Quality. National survey on drug use and health (nsduh), 2004-2022. `https://datatools.samhsa.gov/`, 2024. Accessed: 2024-11-23.

[15] Sujana Ray and Senthil Kumar A.M. Prediction and analysis of sentiments of reddit users towards the climate change crisis. pages 1–17, 04 2023.

[16] Emelie Rosenberg, Carlota Tarazona, Fermín Mallor, Hamidreza Eivazi, David Pastor-Escuredo, Francesco Fuso-Nerini, and Ricardo Vinuesa. Sentiment analysis on twitter data towards climate action. *Results in Engineering*, 19:101287, 2023.

[17] SAMHSA. Crisis counseling assistance and training program trainer's toolkit handout 4 recognizing severe reactions to disaster and common psychiatric disorders. `https://www.samhsa.gov/sites/default/files/handout4-recognizing-severe-reactions.pdf`, 2013.

A Appendix

A.1 GitHub Repositories

The following repositories are open for public use and provide all necessary resources for replicating time series analysis, Reddit web scraping and sentiment analysis conducted in this thesis.

A.1.1 Time Series Analysis

https://github.com/sjanefullerton/time-series

A.1.2 Web Scraper

https://github.com/sjanefullerton/Web-Scraper

A.1.3 Sentiment Analysis

https://github.com/sjanefullerton/Sentiment-Analysis