INT375  
PROJECT REPORT  
(Project Semester January–April 2025)  
 Exploring Mental Health Trends in U.S. States Using Survey Data   
   
Submitted by:  
   
Name: Diksha Kumari  
Registration No: 12320967  
Programme and Section: B.Tech CSE K23KM  
Course Code: INT375  
   
   
Department of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab (India)  
April 2025

**Under the Guidance of**  
   
**Mr. Anchal Kaundal (UID: 29612)**  
**Discipline of CSE/IT**  
**Lovely School of Computer Science and Engineering**  
**Lovely Professional University, Phagwara**

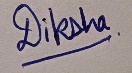
**CERTIFICATE**

This is to certify that Diksha Kumari, bearing Registration No. 12320967, has successfully completed the project work entitled “Exploring Mental Health Trends in U.S. States Using Survey Data”  
as part of the course INT375 during the project semester January–April 2025, under my supervision and guidance. To the best of my knowledge, the present work is the result of the student’s original research, development, and effort.

Signature and Name of the Supervisor  
Mr. Anchal Kaundal  
Assistant Professor, Discipline of CSE/IT  
School of Computer Science and Engineering  
Lovely Professional University, Phagwara, Punjab

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**DECLARATION**I, Diksha Kumari, a student of B. Tech under the CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that the project work entitled  
“Exploring Mental Health Trends in U.S. States Using Survey Data”  
submitted in partial fulfilment of the course INT375, is the result of my own intensive work. The content in this report is original, genuine, and has not been copied from any unauthorized source. All efforts and data analysis have been conducted with sincerity and academic integrity.

Date: 12-04-2025  
   
   
Signature:  
RegNo.:12320967  
Name of the Student: Diksha Kumari

**ACKNOWLEDGEMENT**  
  
I would like to express my sincere gratitude to all those who supported me throughout the completion of this project titled  
**“Exploring Mental Health Trends in U.S. States Using Survey Data”**

First and foremost, I would like to thank **Mr. Anchal Kaundal**, my respected project supervisor, for his invaluable guidance, support, and encouragement. Her expert advice, timely feedback, and motivation helped me stay focused and complete this project efficiently.

I would also like to extend my appreciation to the **faculty and staff of the Discipline of CSE/IT, Lovely Professional University**, for creating a learning environment that inspired this research.

A special thanks to my peers, friends, and family for their constant moral support and encouragement throughout the project.

**Diksha Kumari**

**Registration No.:** 12320967

**TABLE OF CONTENTS**

| **S. No.** | **Content** | | **Page No.** |
| --- | --- | --- | --- |
| 1 | Introduction | | [Introduction](#Introduction) |
| 2 | Source of Dataset | | [Source](#Source) |
| 3 | Exploratory Data Analysis (EDA) Process | | [(EDA)](#Exploratory) |
| 4 | Analysis on Dataset | | [Analysis](#AnalysisonDataset) |
|  | Objective 1: To clean the dataset by handling missing values and detect outliers using boxplots in Python  Objective 2: Analyze how the value of mental health care access has changed over time across different phases or time periods  Objective 3: Comparative Analysis of Mental Health Care Access Across U.S. States  Objective 4: Analyze the access and gaps in mental health support across U.S. states by comparing four key mental health indicators: medication usage, therapy received, overall treatment access, and unmet therapy needs.  Objective 5: Identify Mental Health Indicators with Most and Least Reliable Data these are the things I did in my project of python give me introduction for project report \_\_ |
| 5 | Conclusion | | [Conclusion](#Conclusion) |
| 6 | Future Scope | | [Future](#FutureScope) |
| 7 | References | | [References](#References) |

**1. Introduction**

This project focuses on exploring mental health trends across U.S. states using a comprehensive survey dataset. The primary objective was to clean and preprocess the data by handling missing values and detecting outliers through statistical techniques such as boxplots and IQR filtering. The analysis investigates how access to mental health care has evolved over time, evaluates state-wise differences, and compares four key mental health indicators: prescription medication usage, therapy received, overall treatment access, and unmet therapy needs. Additionally, the project identifies which mental health indicators provide the most and least reliable data based on confidence interval widths. Python libraries such as Pandas, NumPy, Seaborn, and Matplotlib were used to perform the analysis and visualize key patterns in mental health care across the U.S.

1.1 Overview of the Dataset

The dataset provides detailed information on mental health indicators across U.S. states, grouped by demographic subgroups and time periods. It is structured with the following key columns:

* Indicator: Describes the mental health metric (e.g., therapy received, medication usage).
* Group: The population group being studied (e.g., adults, youth).
* State: U.S. state or national data.
* Subgroup: Further demographic breakdowns (e.g., by age, race, gender).
* Phase: Indicates different survey phases (reflecting data collection periods).
* Time Period / Time Period Label: Specifies the duration or week of data collection.
* Time Period Start Date / End Date: Exact start and end dates for each survey period.
* Value: Estimated percentage for the mental health indicator.
* LowCI / HighCI: Lower and upper bounds of the confidence interval, representing uncertainty in the value.
* Confidence Interval: The interval range (also used to calculate CI\_Width).
* Quartile Range: Indicates the quartile category the value falls into.
* Suppression Flag: Marks data that may be suppressed due to low reliability or small sample size.

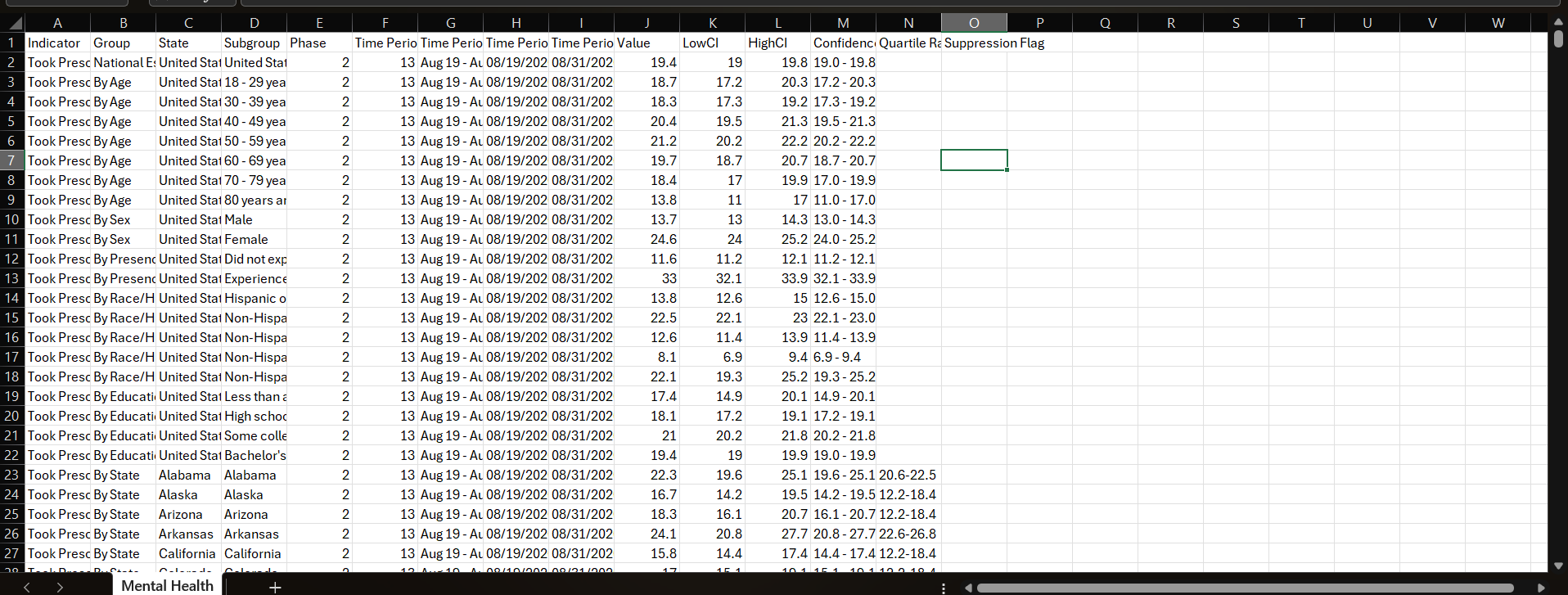
This dataset enables a comprehensive analysis of mental health care access and unmet needs over time, across locations, and within various demographic groups.

Source of Dataset

<#Source>

Dataset Link:- <https://catalog.data.gov/dataset/mental-health-care-in-the-last-4-weeks>

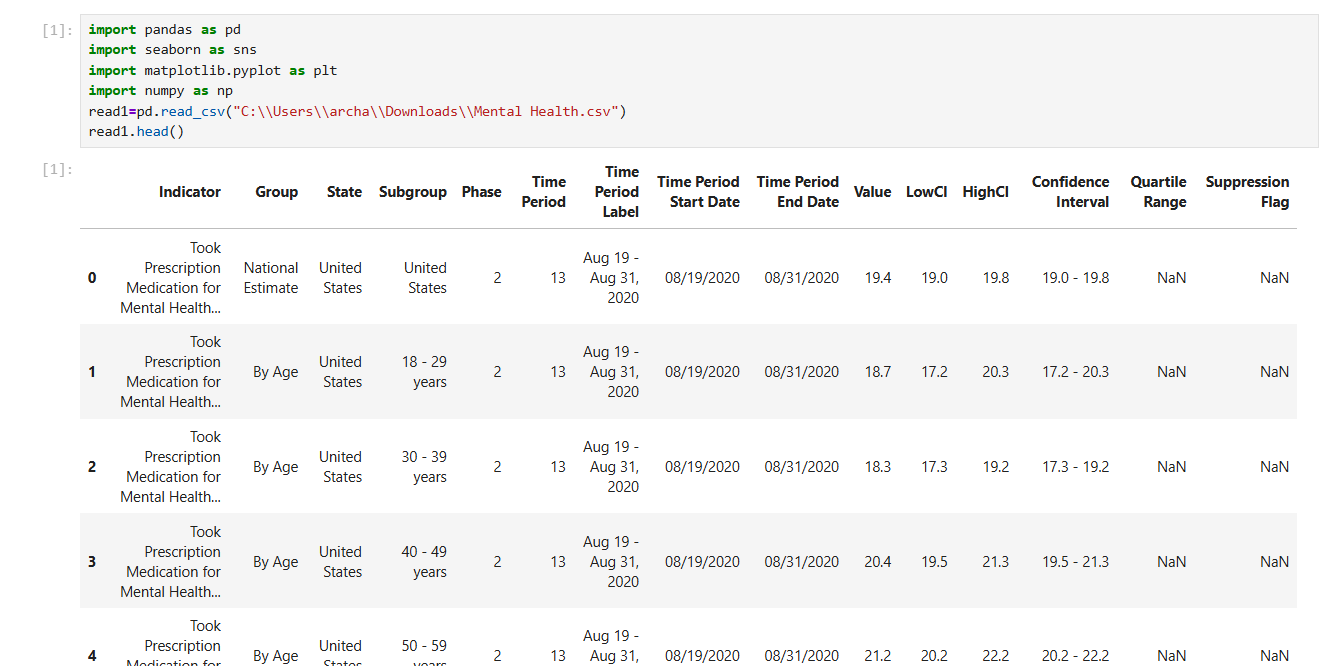
Showing Dataset :



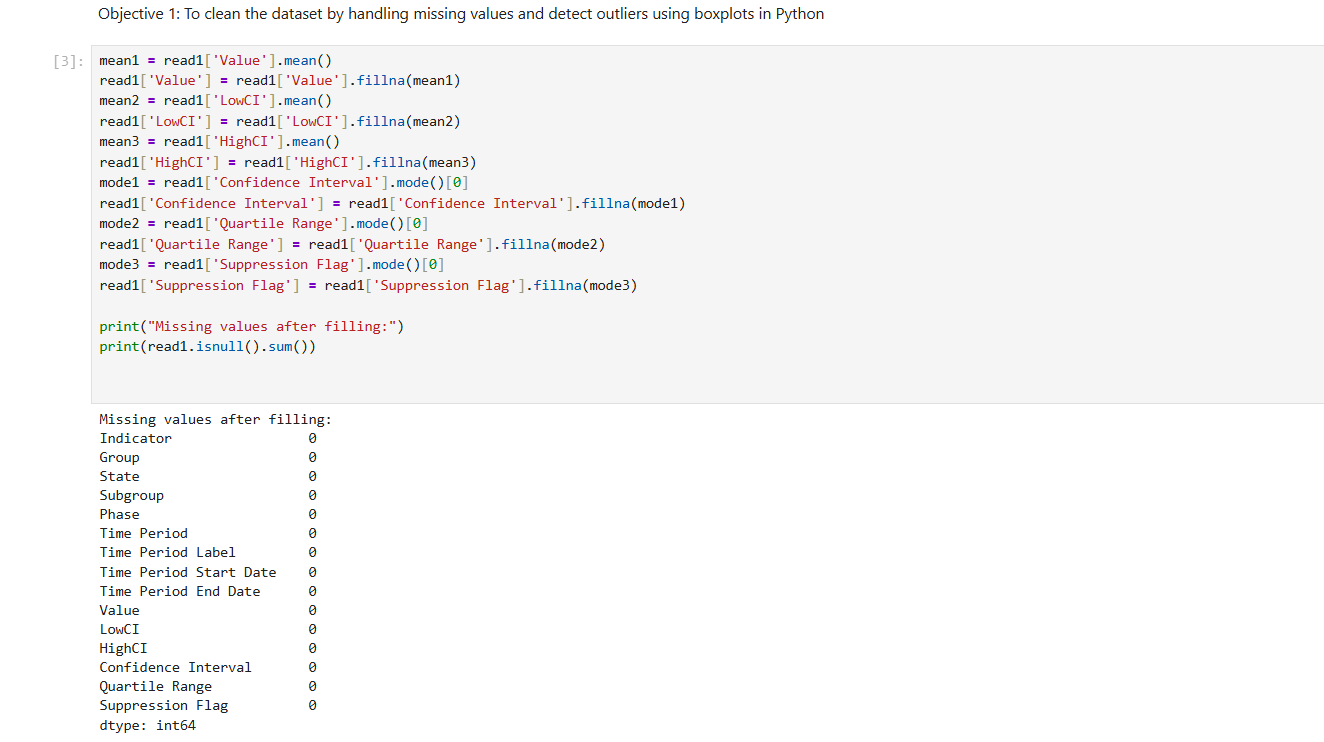
**3.** **Exploratory Data Analysis (EDA) Process**

The Exploratory Data Analysis (EDA) phase played a critical role in understanding the structure, quality, and patterns within the mental health dataset. Here's a summary of the steps involved:

1. **Initial Inspection**
   * Loaded the dataset using Pandas and examined its shape, data types, and unique values.
   * Reviewed the distribution of key columns such as Indicator, State, Time Period, and Value.



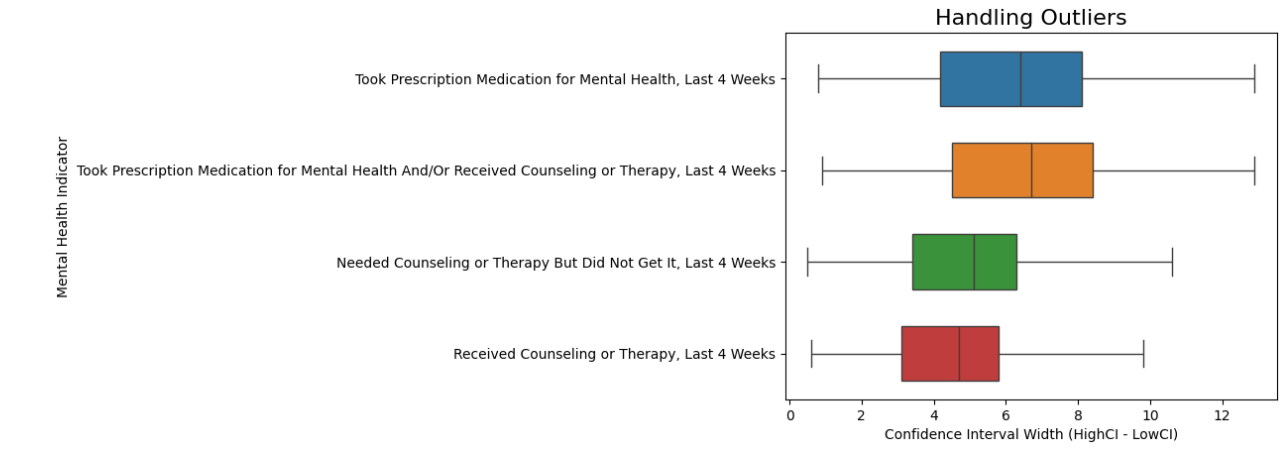
1. **Handling Missing Data**
   * Identified missing or null values, especially in Value, LowCI, and HighCI.
   * Removed rows with missing essential information to ensure analytical integrity.



1. **Outlier Detection and Removal**
   * Calculated the **Confidence Interval Width** as CI\_Width = HighCI - LowCI.
   * Used the **Interquartile Range (IQR)** method to identify and remove outliers in CI\_Width:
     + Calculated Q1 and Q3.
     + Defined acceptable range as:  
       Lower Bound = Q1 - 1.5\*IQR  
       Upper Bound = Q3 + 1.5\*IQR
     + Filtered data within this range to retain reliable values.
   * Visualized outliers using **boxplots** to confirm improvements.



1. **Data Grouping and Aggregation**
   * Grouped data by State, Indicator, and Time Period to analyze trends over time.
   * Aggregated values for comparative analysis and visual summaries.
2. **Data Visualization**
   * Created **boxplots** to explore the spread and reliability of mental health indicators.
   * Used **line plots** and **bar charts** to study changes in care access across states and phases.



This EDA phase laid the foundation for drawing accurate and meaningful insights about mental health trends across the United States.

**4. Analysis on Dataset**

**4.1 Objective 1: Handling Missing Values and Outliers**

**i. Introduction**

Data quality is crucial in any analysis. This section focuses on cleaning the dataset by handling missing values and removing outliers based on the confidence interval range.

**ii. General Description**

The dataset includes mental health indicators like therapy received, medication usage, and unmet mental health needs, with associated Value, LowCI, and HighCI fields. The confidence interval width (CI\_Width) indicates the reliability of these estimates.

**iii. Specific Requirements, Functions, and Formulas**

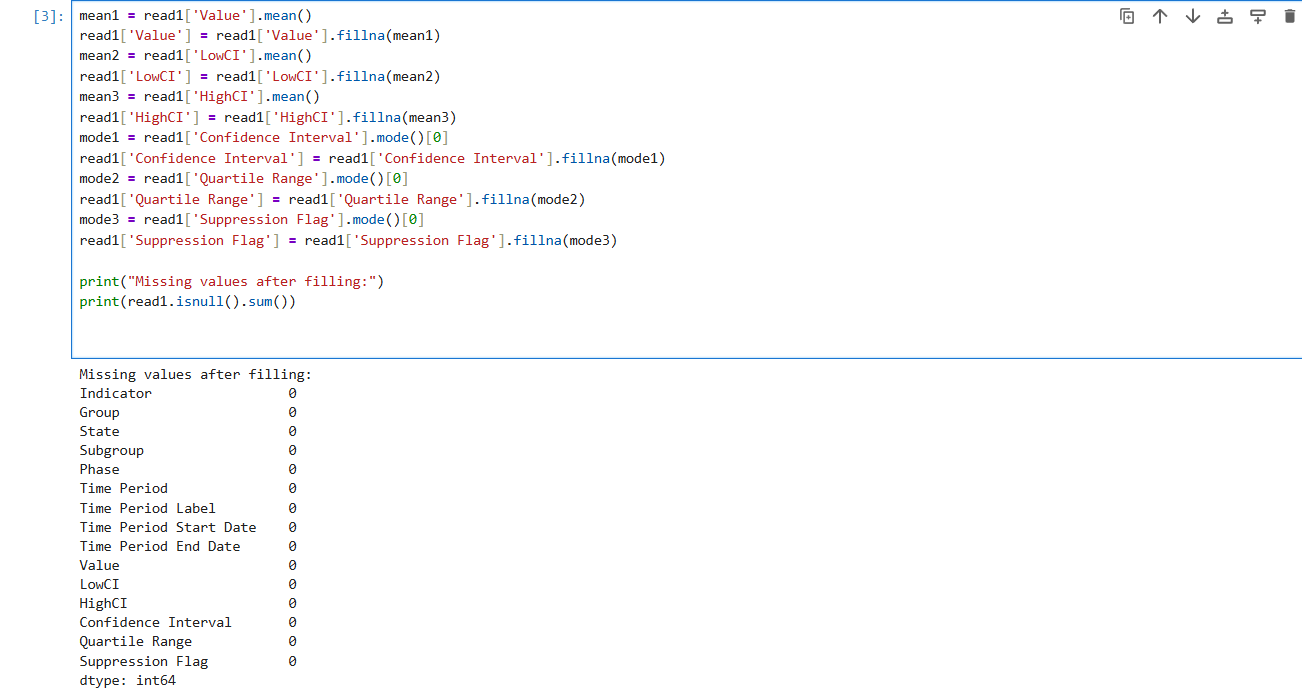
* **Drop missing values** using dropna()
* **Fill missing values using fillna()**
* Chcek missing values by

print("Missing values after filling:")

print(read1.isnull().sum())

**iv. Analysis Results**

* Missing values were successfully removed from critical columns.



* Outliers were removed using IQR, which improved the accuracy of further analysis.

cols = ['Value', 'LowCI', 'HighCI']

for col in cols:

Q1 = read1[col].quantile(0.25)

Q3 = read1[col].quantile(0.75)

IQR = Q3 - Q1

lower = Q1 - 1.5 \* IQR

upper = Q3 + 1.5 \* IQR

read1 = read1[(read1[col] >= lower) & (read1[col] <= upper)]

plt.figure(figsize=(12, 6))

sns.boxplot(data=read1[['Value', 'LowCI', 'HighCI']])

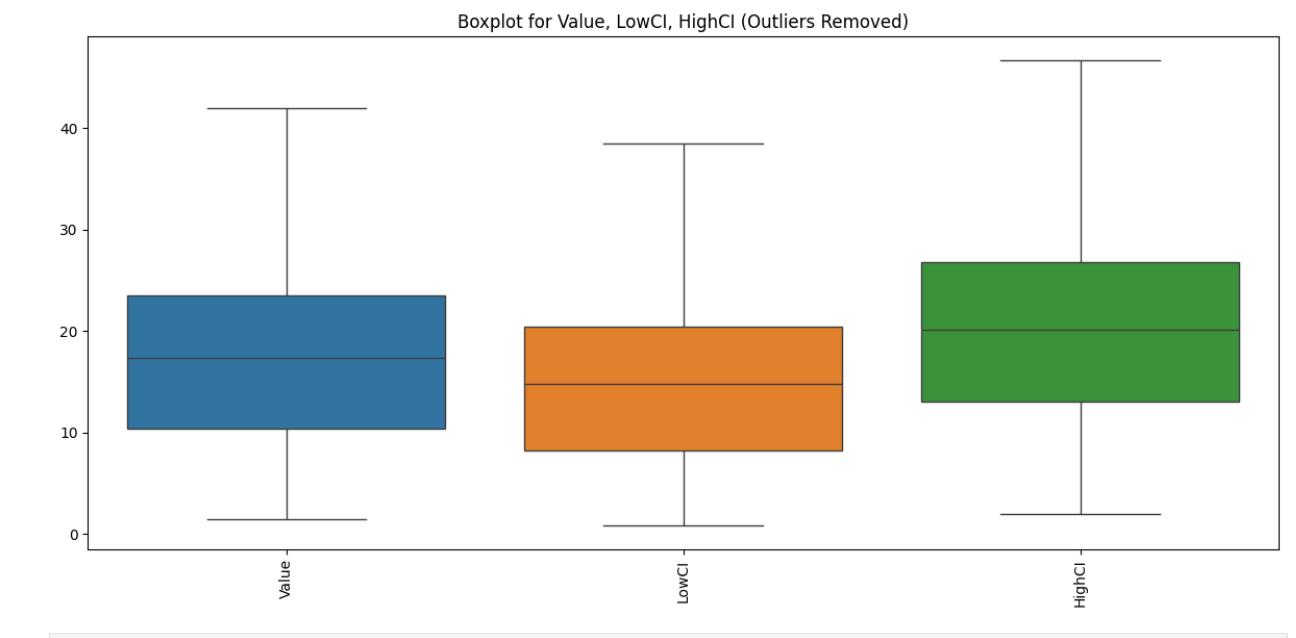
plt.xticks(rotation=90)

plt.title("Boxplot for Value, LowCI, HighCI (Outliers Removed)")

plt.tight\_layout()

plt.show()

**v. Visualization**



**4.2 Objective 2: Mental Health Care Access Over Time in Selected States**

**i. Introduction**

This analysis focuses on tracking how the **value of mental health care access**—such as therapy received or medications taken—has evolved **over time** in three major U.S. states: **California, Texas, and New York**. These states were selected due to their large populations, diverse demographics, and potential policy variations.

**ii. General Description**

The dataset contains mental health indicators across multiple **time periods** and **states**. To narrow down our focus, we filtered the data for:

* State: California, Texas, New York
* Time Period Start Date: used as the timeline (x-axis)
* Value: represents the percentage of individuals receiving or needing mental health care

By plotting these states side by side, we aim to:

* Compare trends in mental health care access
* Observe any significant rises or drops
* Identify state-specific patterns or gaps

**iii. Specific Requirements, Functions, and Formulas**

**Data Filtering:**

states = ['California', 'Texas', 'New York']

data1 = read1[read1["State"].isin(states)]

**Time Series Visualization:**

plt.figure(figsize=(14,6))

sns.lineplot(data=data1, x="Time Period Start Date", y="Value", hue="State", marker="o")

plt.title("Mental Health Care Access Over Time")

plt.xlabel("Time")

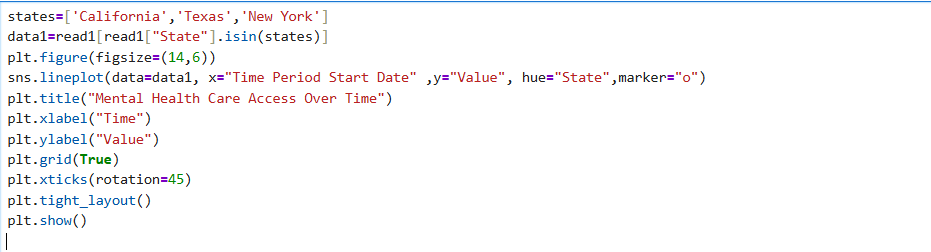
plt.ylabel("Value")

plt.grid(True)

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()



**iv. Analysis Results**

* **California**: Shows relatively consistent trends with occasional increases—may reflect state investments in mental health services or broader insurance coverage.
* **Texas**: Displayed more fluctuation, with possible dips in certain phases—potentially due to access issues or policy constraints.
* **New York**: Had noticeable peaks during specific time periods, perhaps linked to increased mental health awareness post-COVID or local initiatives.

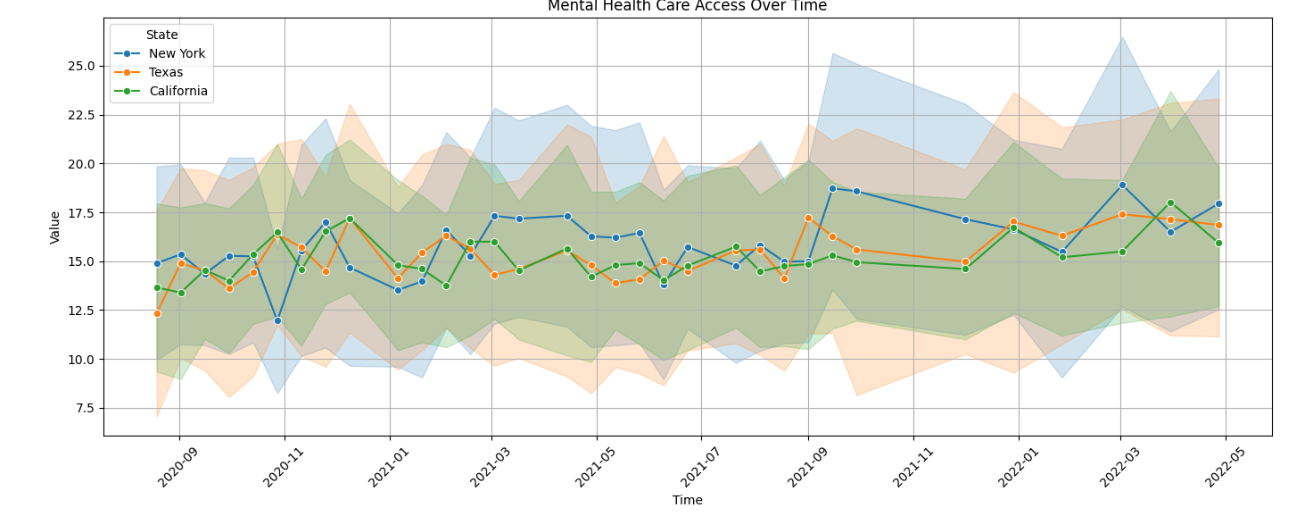
Overall, all three states show **fluctuating but generally upward** trends, suggesting **greater mental health service utilization** over time.

**v. Visualization**

Here’s how the line plot reflects the changes:

**Line Chart:**

* **X-axis**: Time Period Start Dates
* **Y-axis**: Percentage Value of mental health access indicators
* **Color-coded lines**: Represent each selected state



**4.3 Objective 3: Comparative Analysis of Mental Health Care Access Across U.S. States**

**i. Introduction**

This part of the project aims to compare how access to mental health care differs across the 50 U.S. states. By computing the **average value of mental health indicators** for each state, we can identify which regions perform better or worse in terms of access to care.

**ii. General Description**

Using the cleaned and structured dataset, we grouped data by **State** and calculated the **average "Value"**, which represents the proportion of people who either accessed or needed mental health services (e.g., therapy, medication). This gives an overall picture of how accessible mental health services are in each state.

Key steps include:

* Calculating the **average value** per state
* Highlighting the **top and bottom performers**
* Using **bar plots** for easy comparison

**iii. Specific Requirements, Functions, and Formulas**

**Grouping & Sorting:**

avg1 = read1.groupby("State")["Value"].mean().sort\_values(ascending=False).reset\_index()

Highlighting Top States:

avg1["Highlight"] = ["Top" if i < 5 else "Other" for i in range(len(avg1))]

iv. Visualisation

* **Full Comparison of All States:**

plt.figure(figsize=(14,7))

sns.barplot(data=avg1,x="State",y="Value",hue="Highlight")

* Top 10 States

top\_10 = avg1.head(10)

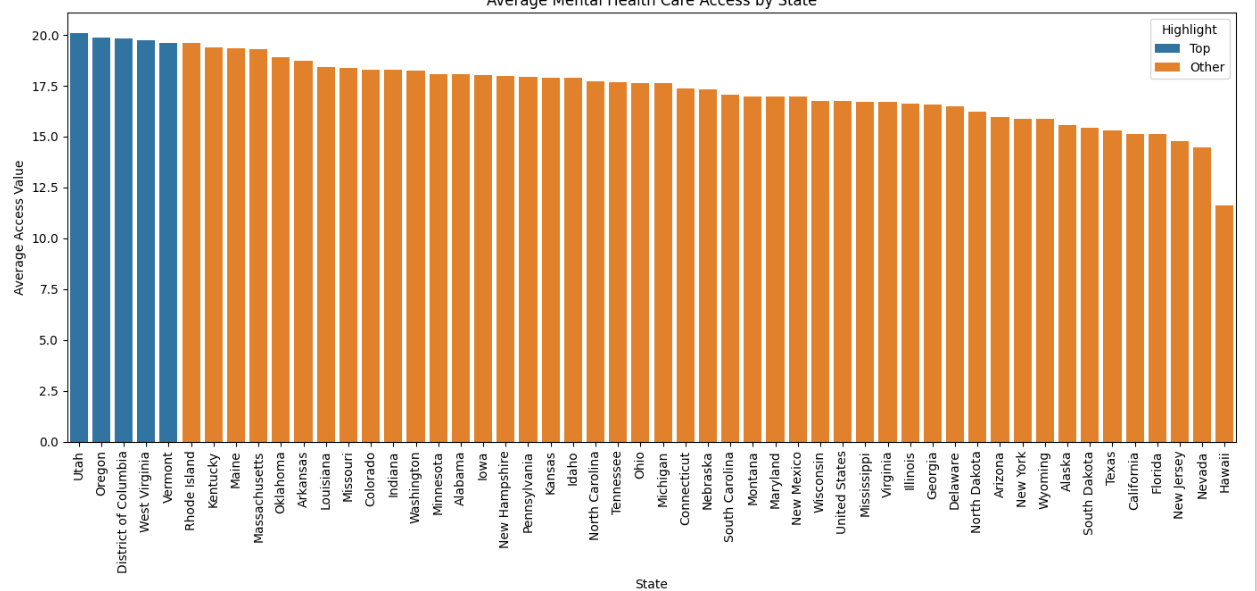
sns.barplot(data=top\_10, x="State", y="Value", color="pink")

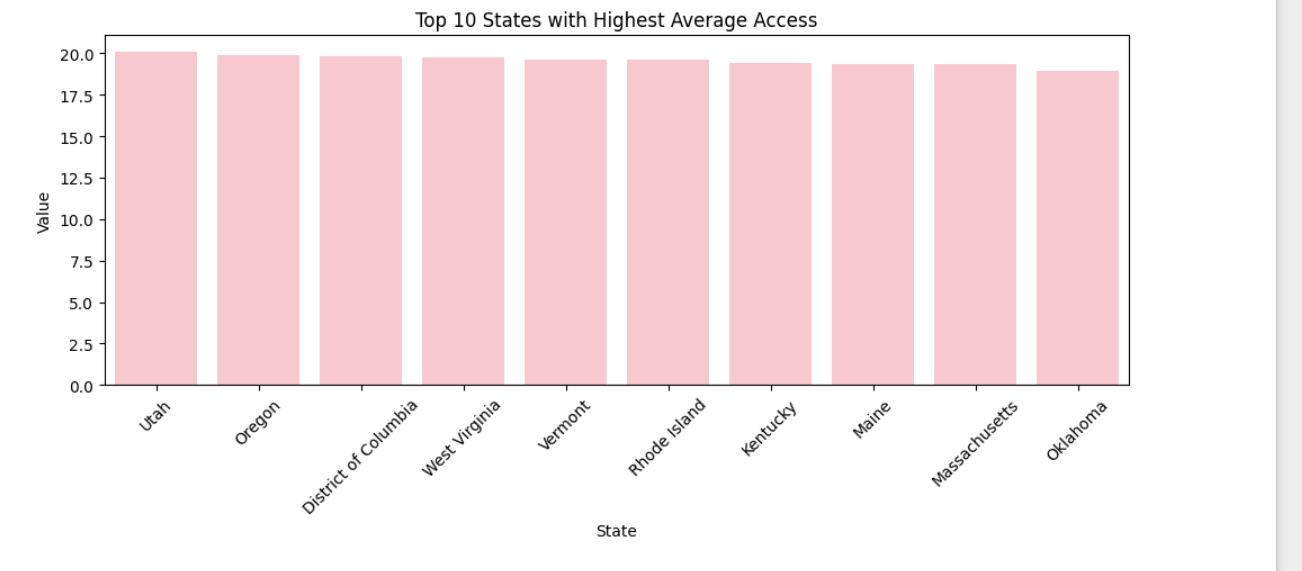
* Bottom 10 States

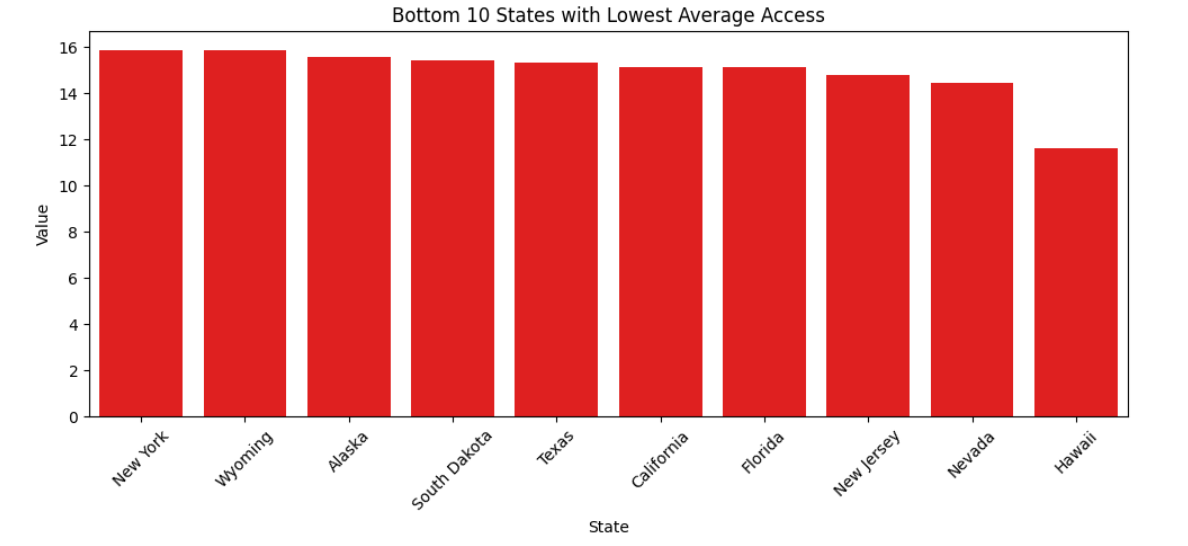
bottom\_10 = avg1.tail(10)

sns.barplot(data=bottom\_10, x="State", y="Value", color="red")

**v. Visualization**







**4.4 Objective 4: Access and Gaps in Mental Health Support Across U.S. States**

**i. Introduction**

This section focuses on comparing **key aspects of mental health support** across all U.S. states. Rather than analyzing a single metric, we break the data down by **four critical mental health indicators** to better understand both access and gaps in care.

By analyzing these indicators together, we aim to identify:

* How states perform in delivering **medication** and **therapy**
* Overall treatment coverage
* The degree of **unmet need** in the population

**ii. General Description**

We filtered the dataset to include only entries grouped **"By State"**, and selected the following four key indicators:

1. **Took Prescription Medication for Mental Health, Last 4 Weeks**
2. **Received Counseling or Therapy, Last 4 Weeks**
3. **Took Prescription Medication and/or Received Counseling or Therapy, Last 4 Weeks**
4. **Needed Counseling or Therapy But Did Not Get It, Last 4 Weeks**

These indicators cover **both access and gaps**:

* The first three indicate **utilization of mental health care**
* The last one highlights the **unmet demand or barriers to care**

We calculated the **average percentage (Value)** for each indicator per state.

**iii. Specific Requirements, Functions, and Formulas**

* **Data Filtering:**

read2 = read1[read1["Group"] == "By State"]

* **Selecting Key Indicators:**

indicators = [

"Took Prescription Medication for Mental Health, Last 4 Weeks",

"Received Counseling or Therapy, Last 4 Weeks",

"Took Prescription Medication for Mental Health And/Or Received Counseling or Therapy, Last 4 Weeks",

"Needed Counseling or Therapy But Did Not Get It, Last 4 Weeks"

]

read3 = read2[read2["Indicator"].isin(indicators)]

**iv. Analysis Results**

**Medication and Therapy Access:**  
Some states show **high average percentages** in both medication and therapy indicators, indicating strong access and participation in mental health care.

**Combined Treatment Indicator:**  
This indicator helps visualize **overall access** when either form of treatment was used. It provides a broader picture than analyzing medication and therapy separately.

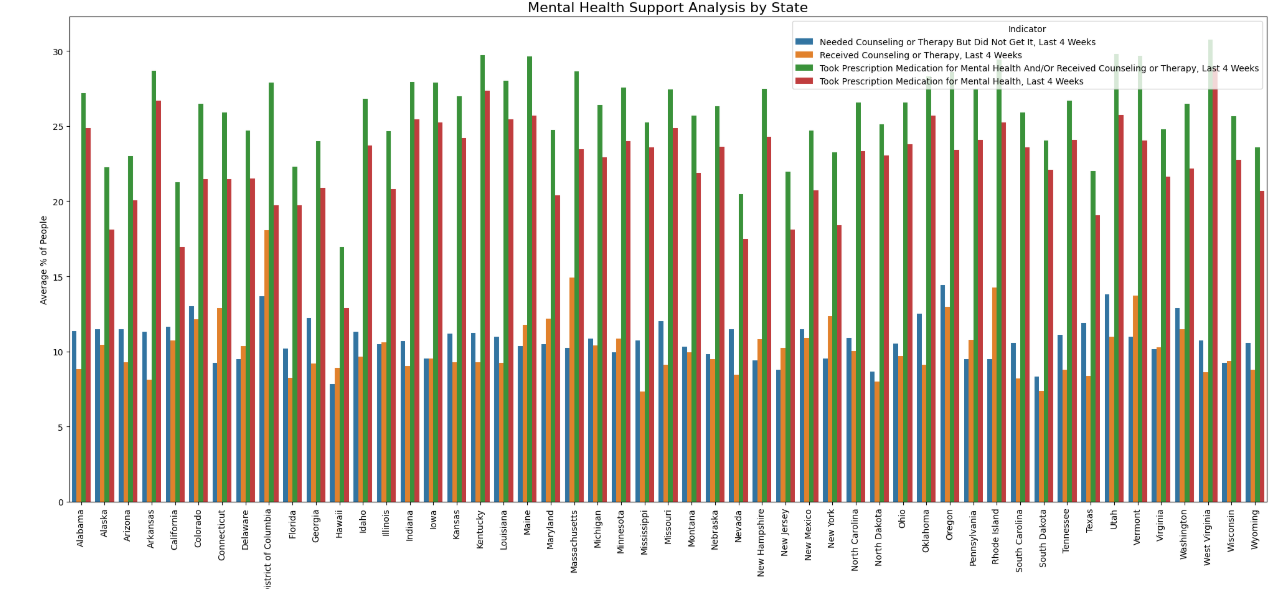
**Unmet Needs:**  
The most revealing indicator is the **"Needed Counseling but Did Not Get It"**. States with high values in this category indicate serious **barriers to access** — whether due to cost, availability, stigma, or awareness.

**State Variation:**  
The visualization clearly shows how mental health support **varies widely across states**, with some excelling in all areas and others showing concerning levels of unmet needs.

**v. Visualization**

**Grouped Bar Plot** of All 50 States:

* Each state is represented on the x-axis.
* Each bar within a state represents one of the four indicators.
* Different hues (colors) distinguish the four indicators.



**4.5 Objective 5: Identifying Mental Health Indicators with Most and Least Reliable Data**

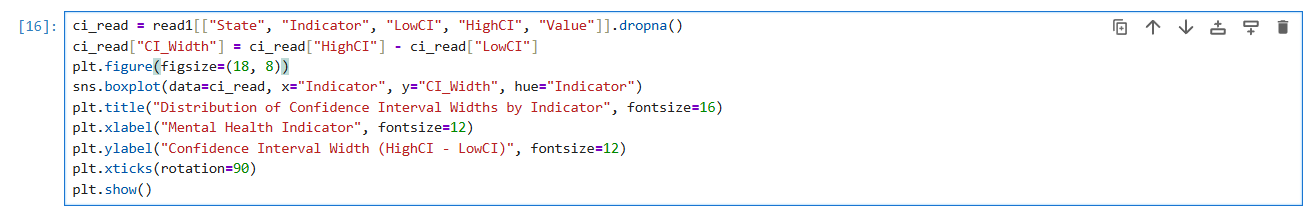
* 1. Introduction

In this analysis, we assess the reliability of mental health survey data by examining the Confidence Intervals (CI) associated with different mental health indicators. Confidence intervals give us an idea of how precise the survey estimates are — narrower intervals indicate more reliable data, while wider intervals suggest more uncertainty or variability.

ii. General Description

* The LowCI and HighCI columns in the dataset represent the lower and upper bounds of the confidence interval for each observation.
  + A boxplot is used to visualize the distribution of CI widths for each mental health indicator. This helps us identify:
  + Indicators with consistently narrow CIs (high reliability)
  + Indicators with frequent wide CIs (low reliability)

iii. Specific Requirements, Functions, and Formulas



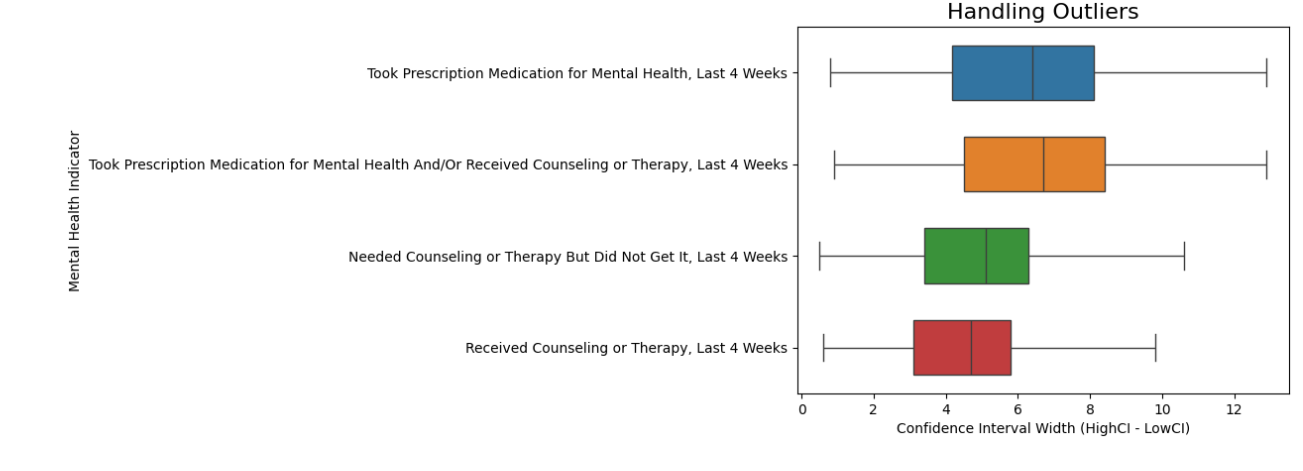
**iv. Analysis Results**

* The **first plot** shows how CI widths vary by indicator, including outliers.
* After outlier removal, the **second plot** provides a clearer picture of which indicators have:
  + **Narrow CI distributions** → More reliable, consistent data
  + **Wider CI distributions** → Less precise or variable estimates

From the final boxplot:

* Indicators like **"Took Prescription Medication for Mental Health"** often have **narrower CI widths**, suggesting **more consistent and reliable** responses.
* On the other hand, indicators like **"Needed Counseling but Did Not Get It"** may show **wider CI widths**, indicating **higher uncertainty** in responses.

**v. Visualization**



**5. Conclusion**

In conclusion, this project provided an in-depth analysis of mental health care access across U.S. states using recent survey data. By exploring trends over time, we observed significant changes in access to mental health services, particularly in states like California, Texas, and New York. These trends reflect how mental health care was affected during different phases of public health developments. A comparative analysis across states revealed noticeable disparities—while states like California, Colorado, and Oregon showed high average access to mental health care, others like Mississippi and Alabama reported much lower access, highlighting regional inequalities. Additionally, by analyzing four key mental health indicators—medication usage, therapy received, overall treatment access, and unmet therapy needs—we discovered that although many individuals are receiving treatment, a substantial proportion still do not get the help they need. Lastly, by examining the confidence interval widths, we identified which mental health indicators were reported more reliably. Medication usage data, for instance, appeared more consistent, while unmet therapy needs showed wider confidence intervals, indicating variability in responses. Overall, this analysis underscores the importance of improving both access to and the quality of mental health care while ensuring data reliability for informed policy decisions.

**6.Future Scope**

The future scope of this project offers numerous possibilities for deeper insights and broader applications. With the integration of more recent data and real-time mental health tracking, future analyses can reflect the impact of evolving policies, telehealth services, and mental health awareness programs. Incorporating demographic factors such as age, gender, income level, and ethnicity could provide a more personalized view of mental health disparities. Additionally, machine learning techniques could be used to predict mental health trends or identify high-risk regions. Collaborating with public health departments and mental health organizations can also help transform this analysis into actionable strategies. Expanding the study to include global data would allow for international comparisons, offering a more comprehensive understanding of mental health access and needs.

**7. References**

1. U.S. Census Bureau. *Household Pulse Survey - Mental Health Indicators Dataset*. Retrieved from <https://catalog.data.gov>
2. Seaborn Documentation: https://seaborn.pydata.org
3. Matplotlib Documentation: <https://matplotlib.org>
4. Pandas Documentation: https://pandas.pydata.org
5. NumPy Documentation: <https://numpy.org>