# ANALYSIS OF MENTAL HEALTH CARE IN USA

**ROHITH SINGH THAKUR**

**STUDENT ID: 11753512**

**INSTRUCTOR: Dr. GAHANGIR HOSSAIN**

**INFO 5709 DATA VISUALIZATION**

**DATA VISUALIZATION – TERM PROJECT**

**05/08/2024**

# INTRODUCTION:

Mental health is a critical aspect of wellbeing despite being a challenging and diverse issue. The mental healthcare system needs to be understood to develop strategies, relations and standards that can bring about positive change. Recognizing the importance of mental health is beyond understanding the normal functions of our brains which lead to the way we feel. Sometimes, believe it or not, our brains fail to function properly like they are supposed to, and then things are hard to accomplish. This is called cognitive impairment (CI). It's the way we think, store information and make decisions. For some people, it's like having a fog in their mind. Remembering names or following directions are examples of very simple things which can become difficult tasks. Advanced age typically entails a decline in cognitive ability, creating an obstacle to doing routine things. This can result in misunderstanding, disappointment and can break their confidence as they cannot do what they previously could. Your confidence will be knocked when you are no longer able to work as well as you were before, this will impact how you relate to others and the way you think of yourself. Even so, with proper help at hand people can adapt and move past their emotional difficulties. It's crucial to mention mental health and being in favor of each other's.

# DATASET:

The dataset is collected from data.gov which is a .cvs file extension. Dataset containing a detailed study of the variation in mental health across the USA from 2020 to 2022 while also diagnosing the impact of the COVID-19 period on the general mental health. This dataset contains 10,404 rows.

The dataset examines across the United States of America, the ways in which memory and cognitive issues impact various demographic groups. It takes Age, Education, Phase, Cognitive Impairment level, Confidence level and Gender by segregating these factors into people who took medication, who did not take medication and counseling in a particular time period because these factors can alter mental health.

**Dataset link:**

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

# ATTRIBUTES:

The following are the attributes used in the “Mental Health Care” dataset.

**INDICATOR:** This explains whether the person needs counseling, received counseling and took prescription or not.

**STATE:** The geographical area or location in the U.S. that provides information about mental health care within the country.

**GROUP:** In the dataset, the group or classification to which the indicator belongs.

**SUBGROUP:** An additional split of the state's population under study.

**PHASE:** The specific process of evaluating mental health care, which illustrates a level of intervention or assessment.

**TIME PERIOD:** The definite timeframe in which the mental health data was contrived or has been examined.

**TIME PERIOD LABEL:** A descriptive label showing what time period the data relates to.

**TIME PERIOD START DATE:** The starting date of the survey that is in survey.

**TIME PERIOD END DATE:** The closing date of the timeframe that is being asked.

**VALUE:** The numerical value or magnitude of the indicator within the specified time period.

**LOW COGNITIVE IMPAIRMENT (CI):** The bottom boundary of the confidence interval that is calculated for the measured value.

**HIGH COGNITIVE IMPAIRMENT (CI):** The upper bound of a confidence interval of the measured value.

**CONFIDENCE INTERVAL:** Interval in which the true value of an indicator is likely to exist with a certain level of confidence.

**QUARTILE RANGE:** Ranging from 25% of the data values lying within it indicating wherein data is distributed within the dataset.

**SUPPRESSION FLAG:** Give a flag to them which shows if some data points are hidden because of confidentiality or reliability challenges.

# TOOLS:

- Python
- Tableau

# DATA CLEANING THROUGH EXPLORATORY DATA ANALYSIS:

In [12]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [13]:
```
# Load the dataset
data = pd.read_csv('Mental_Health_Care_Dataset.csv')
```

In [21]:
```
# Print the column names
print("Column names:")
print(data.columns)
```

```
Column names:
Index(['Indicator', 'Group', 'State', 'Subgroup', 'Phase', 'Time Period',
       'Time Period Label', 'Time Period Start Date', 'Time Period End Date',
       'Value', 'LowCI', 'HighCI', 'Confidence Interval', 'Quartile Range',
       'Suppression Flag'],
      dtype='object')
```

The above are the columns in the dataset.

In [22]:
```
# Print the data types
print("\nData types:")
print(data.dtypes)
```

```
Data types:
Indicator                object
Group                    object
State                    object
Subgroup                 object
Phase                    object
Time Period               int64
Time Period Label        object
Time Period Start Date   object
Time Period End Date     object
Value                   float64
LowCI                   float64
HighCI                  float64
Confidence Interval      object
Quartile Range           object
Suppression Flag        float64
dtype: object
```

The type of data that is in the dataset.

```
In [23]: # Print the shape of the dataset
         print("\nShape of the dataset:")
         print(data.shape)
```

```
Shape of the dataset:
(10404, 15)
```

The dimensions of the dataset. As we can see that there are no null values, every data is important.

```
In [24]: # Print the last few rows
         print("\nLast few rows:")
         print(data.tail())
```

```
Last few rows:
                                              Indicator     Group  \
10399  Needed Counseling or Therapy But Did Not Get I...  By State
10400  Needed Counseling or Therapy But Did Not Get I...  By State
10401  Needed Counseling or Therapy But Did Not Get I...  By State
10402  Needed Counseling or Therapy But Did Not Get I...  By State
10403  Needed Counseling or Therapy But Did Not Get I...  By State

              State       Subgroup Phase  Time Period     Time Period Label  \
10399      Virginia       Virginia   3.4           45  Apr 27 - May 9, 2022
10400    Washington     Washington   3.4           45  Apr 27 - May 9, 2022
10401  West Virginia  West Virginia   3.4           45  Apr 27 - May 9, 2022
10402     Wisconsin      Wisconsin   3.4           45  Apr 27 - May 9, 2022
10403       Wyoming        Wyoming   3.4           45  Apr 27 - May 9, 2022

      Time Period Start Date Time Period End Date  Value  LowCI  HighCI  \
10399             04/27/2022           05/09/2022   10.1    7.1    13.8
10400             04/27/2022           05/09/2022   14.8   12.5    17.4
10401             04/27/2022           05/09/2022    9.8    6.6    13.8
10402             04/27/2022           05/09/2022   11.6    8.5    15.3
10403             04/27/2022           05/09/2022   11.5    8.6    15.1

      Confidence Interval Quartile Range  Suppression Flag
10399         7.1 - 13.8       9.6-11.3               NaN
10400        12.5 - 17.4      13.0-20.8               NaN
10401         6.6 - 13.8       9.6-11.3               NaN
10402         8.5 - 15.3      11.4-12.9               NaN
10403         8.6 - 15.1      11.4-12.9               NaN
```

In [25]:
```
# Print the summary statistics
print("\nSummary statistics:")
print(data.describe())
```

```
Summary statistics:
        Time Period        Value        LowCI       HighCI  Suppression Flag
count  10404.000000  9914.000000  9914.000000  9914.000000              22.0
mean      28.134948    17.450736    14.771565    20.475661               1.0
std       11.040210     8.270565     7.659396     9.052521               0.0
min        1.000000     1.400000     0.800000     2.000000               1.0
25%       20.000000    10.300000     8.000000    12.900000               1.0
50%       29.000000    16.200000    13.900000    19.200000               1.0
75%       37.000000    24.000000    20.800000    27.400000               1.0
max       45.000000    62.900000    53.200000    71.900000               1.0
```

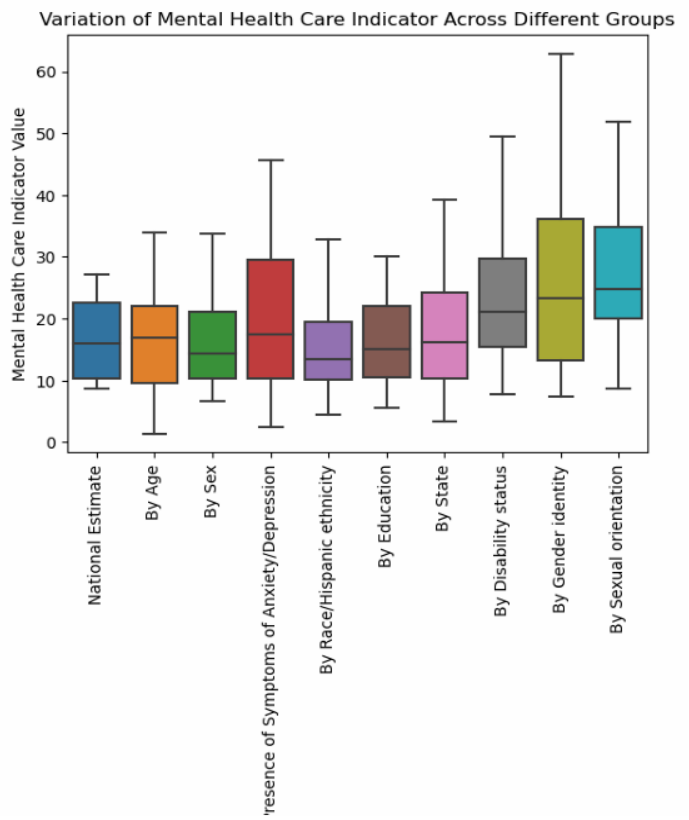
In [26]:
```
# Check for missing values
print("\nMissing values:")
print(data.isnull().sum())
```

```
Missing values:
Indicator                    0
Group                        0
State                        0
Subgroup                     0
Phase                        0
Time Period                  0
Time Period Label            0
Time Period Start Date       0
Time Period End Date         0
Value                      490
LowCI                      490
HighCI                     490
Confidence Interval        490
Quartile Range            3672
Suppression Flag         10382
dtype: int64
```

The dataset's missing values can be examined using the above method. The visualizations used to analyze and determine the relationships between the qualities are shown below.

### 1) How does mental health care indicators vary across different groups?

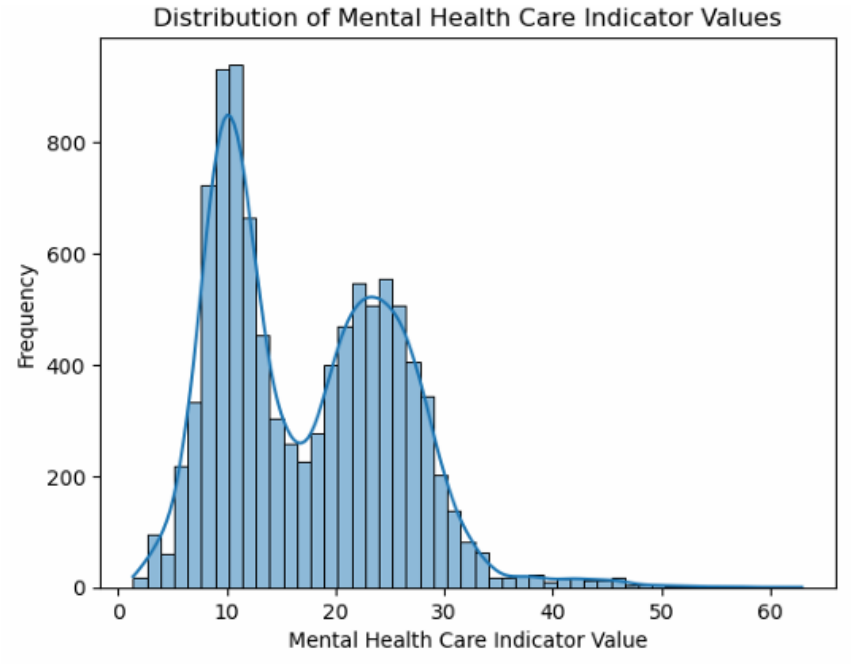
In [18]:
```
# Exploratory Data Analysis
def plot_question(question_num, question, data):
    print(question)
    if question_num == 1:
        sns.boxplot(x='Group', y='Value', data=data)
        plt.xticks(rotation=90)
        plt.title("Variation of Mental Health Care Indicator Across Different Groups")
        plt.xlabel("Group")
        plt.ylabel("Mental Health Care Indicator Value")
        plt.show()
```




This graphic shows how a mental health care indicator varies between various groups. The values of this indicator for the various categories stated on the x-axis are displayed on the y-axis of the bar chart. Clearly, this mental health care indication is more highly valued for groups like "By sexual orientation" and "By disability identity" than it is for other groups like "National Estimate" or "By age." A straightforward comparison of the mental health care indicator levels across various population segments is made possible by this visual representation.

**2) What is the distribution of mental health care indicator values?**

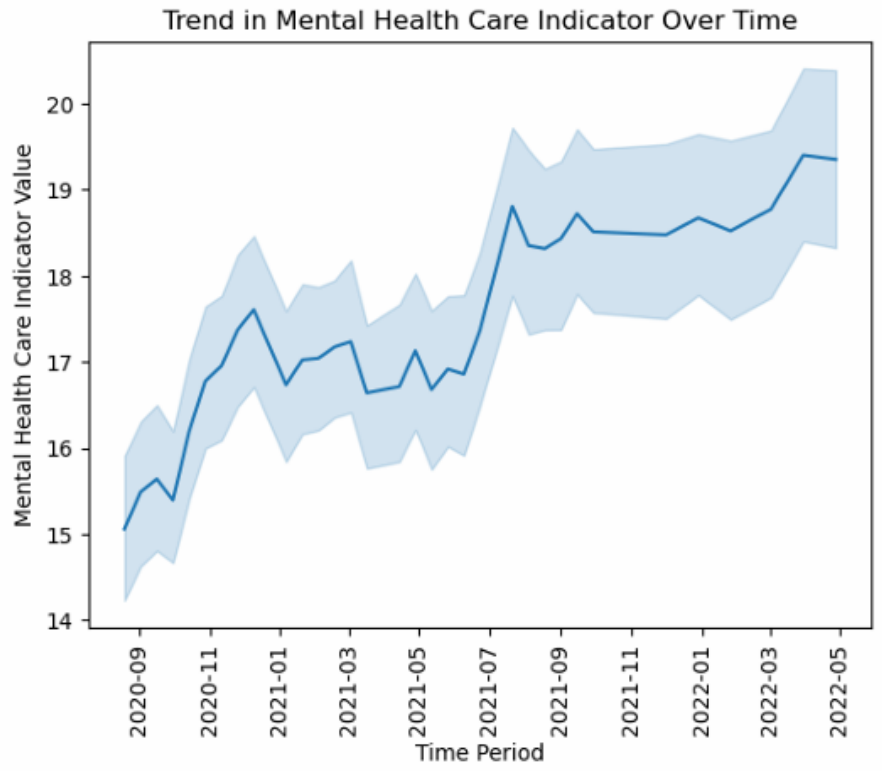
```
elif question_num == 2:
    sns.histplot(data['Value'], kde=True)
    plt.title("Distribution of Mental Health Care Indicator Values")
    plt.xlabel("Mental Health Care Indicator Value")
    plt.ylabel("Frequency")
    plt.show()
```




The histogram's distribution of mental health care indicator values looks to be hybrid, with two separate peaks centered around values of 10 and 30. A longer tail extending towards higher indicator values and a small rightward tilt.

### 3) Is there a trend in mental health care indicator over time?

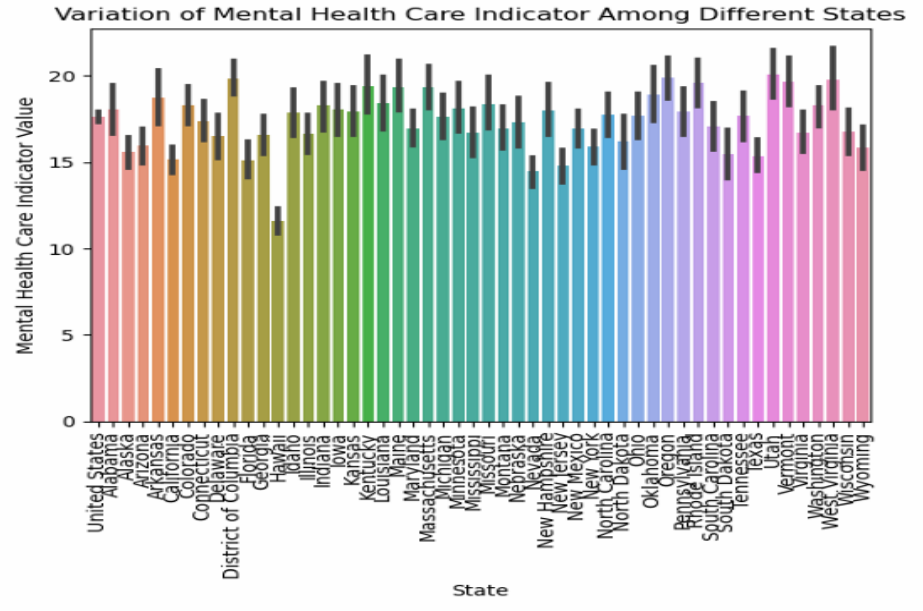
```
elif question_num == 3:
    data['Time Period Start Date'] = pd.to_datetime(data['Time Period Start Date'])
    sns.lineplot(x='Time Period Start Date', y='Value', data=data)
    plt.xticks(rotation=90)
    plt.title("Trend in Mental Health Care Indicator Over Time")
    plt.xlabel("Time Period")
    plt.ylabel("Mental Health Care Indicator Value")
    plt.show()
```




The dataset provides a range of temporal and categorical dimensions for exploring trends, patterns, and disparities in mental health care indicators.

### 4) How does mental health care indicator vary among different states?

```
elif question_num == 4:
    sns.barplot(x='State', y='Value', data=data)
    plt.xticks(rotation=90)
    plt.title("Variation of Mental Health Care Indicator Among Different States")
    plt.xlabel("State")
    plt.ylabel("Mental Health Care Indicator Value")
    plt.show()
```




The presence of cognitive Impairment ('Low Cognitive Impairment', 'High Cognitive Impairment') allows for assessing the uncertainty associated with the reported values.

# HYPOTHESIS:

1) Does the degree of cognitive impairment in mental health and average mental health care value correlate with one another?

2) Is there a statistically significant difference in average mental health values among different gender identities and education levels in the United States, as well as in the high confidence intervals for mental health care data?

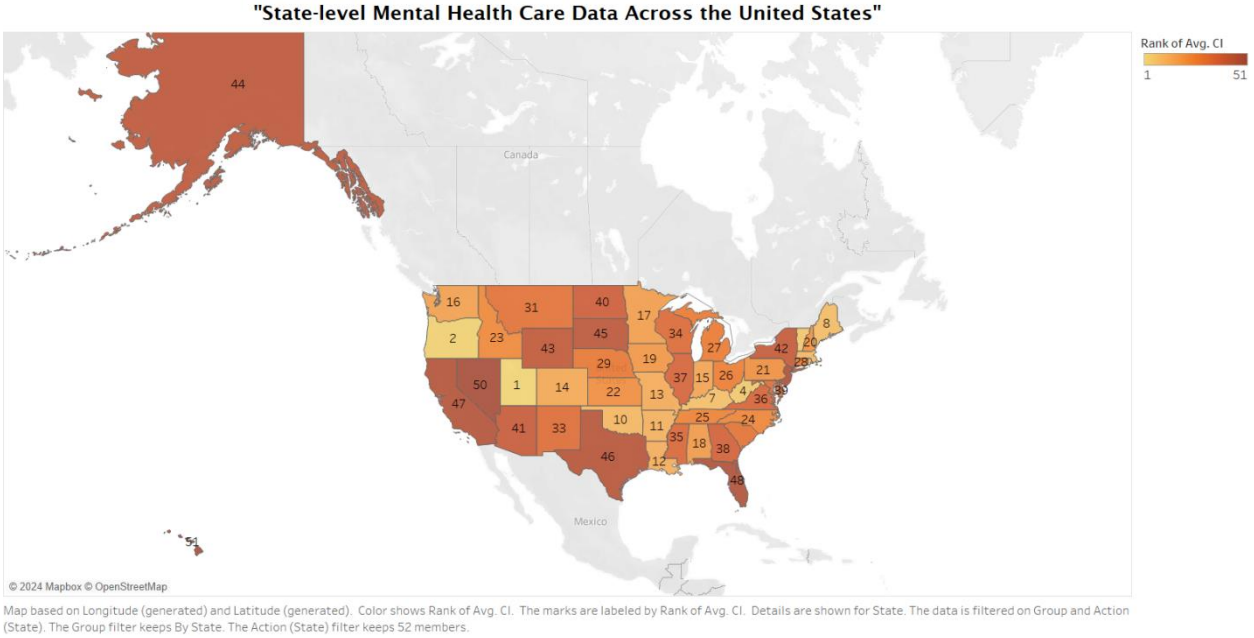
3) Does the average confidence interval for mental health care data vary significantly among different racial subgroups over time?

**1) Does the degree of cognitive impairment in mental health and average mental health care value correlate with one another? Which state ranks the First in CI level?**

To answer this question, I chose Map Bar graph and Treemaps. The visuals are explained below.

## a) Filled Map

The given visual displays state-level mental health care data across the United States using a Geographical Filled Map. States are colour-coded which is correspondingly expressing the value, probably indicating the ranking of Cognitive Impairment (CI) levels during the time span of the covid-19 pandemic from 2020 to 2022.



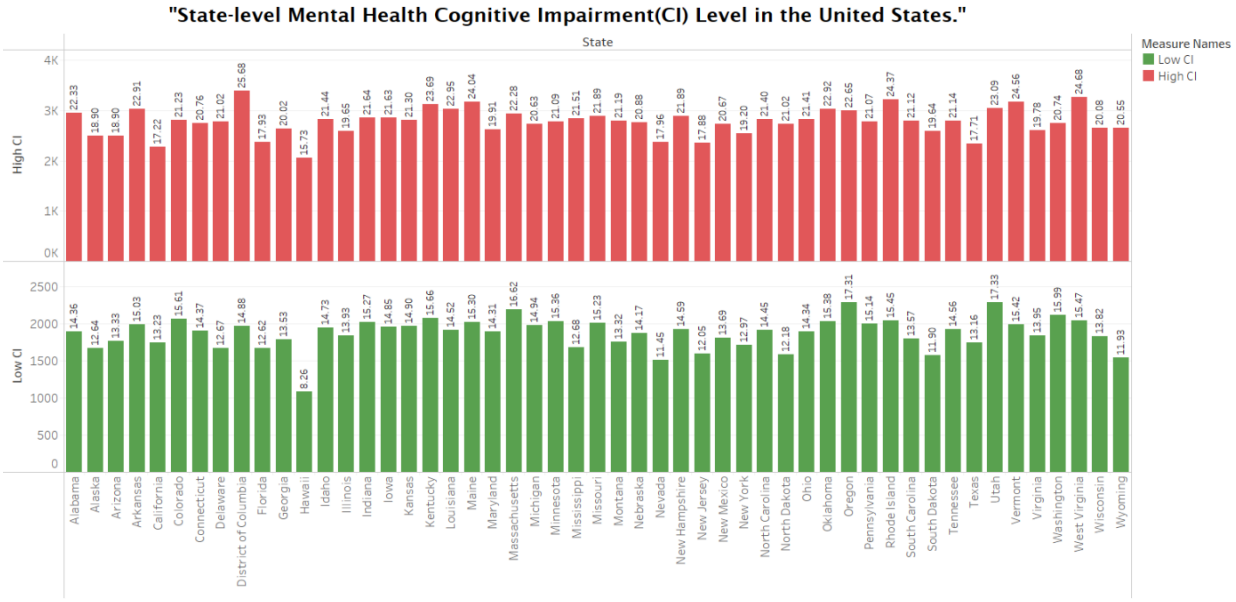

The map highlights the gaps in the ranks of different states -- for instance, some states have lighter shades (that represent lower value) and the other states have darker shades (which may indicate higher value). The legend on the right-hand side provides ranking of the states, allowing for interpretation of the color-coding.

- Utah ranks 1, as it has the highest average number of values of cognitive impairment level represented the gold color.

- Several states in the central and eastern regions, such as Texas, Florida, and New York, have relatively lower ranks of cognitive impairment levels, shown in darker shades of orange.
- States in the western and midwestern regions, like Oregon, Washinton and Minnesota tend to have higher cognitive impairment levels as they rank 2,16 and 17 respectively, represented by lighter shades of gold.

## b) Bar Graph

The given visual displays state-level data on mental health cognitive impairment (CI) levels in the United States during the COVID-19 pandemic period from 2020-2022. It presents two measures: SUM(High CI) and SUM(Low CI) for each state, shown as stacked bar charts.




High CI and Low CI for each State. Color shows details about High CI and Low CI. For pane Sum of High CI: The marks are labeled by average of High CI. For pane Sum of Low CI: The marks are labeled by average of Low CI. The data is filtered on Group and Action (State). The Group filter keeps By State. The Action (State) filter keeps 52 members. The view is filtered on State, which keeps 52 of 52 members.

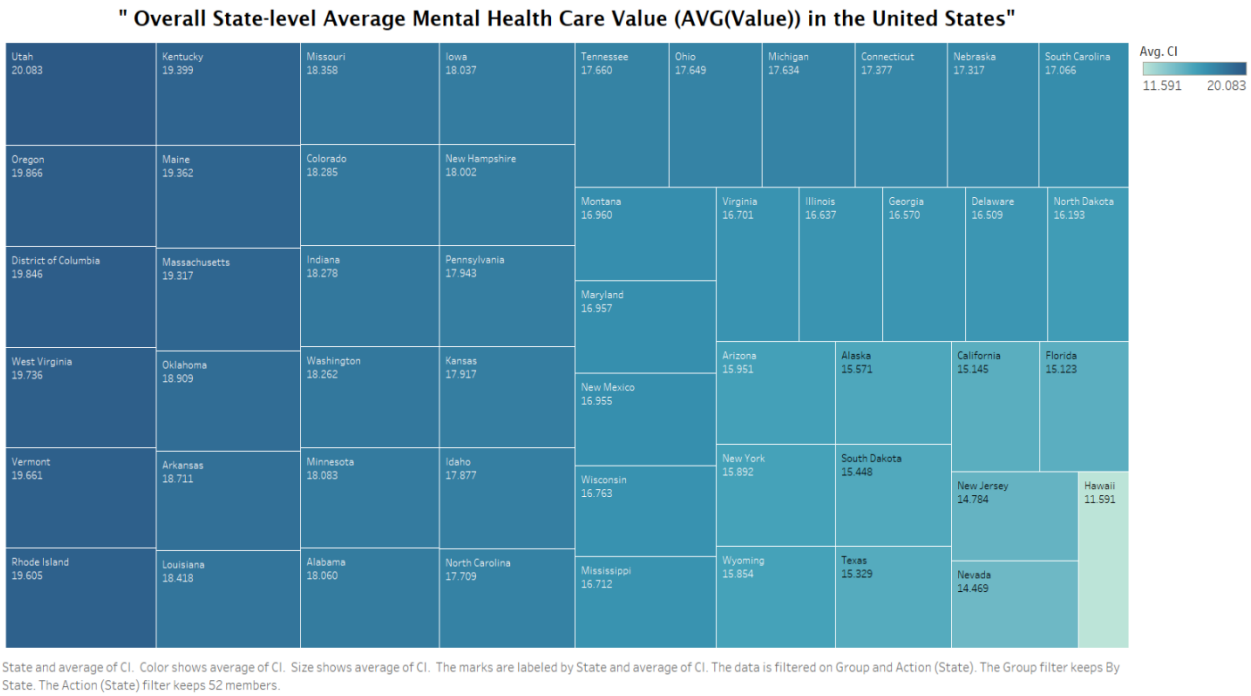
The y-axis represents the values of the two measures, while the x-axis lists all the states in USA. The bars are color-coded, with red bars representing SUM(High CI) and green bars representing SUM(Low CI).

- There is a significant variation in the CI levels across different states, as evident from the varying bar heights.
- States like Alaska, Arizona, Arkansas, and Alabama appear to have higher levels of both High CI and Low CI, as indicated by their taller bars.

- States like Vermont, West Virginia, and Wyoming seem to have relatively lower levels of CI, with shorter bars.
- The highest SUM(High CI) value is observed for District of Columbia (around 3,390) and Avg CI value is 25.68, followed by West Virginia (around 3,258).
- The lowest SUM(High CI) value is observed for Hawaii (around 2060) and Avg CI value is 15.73.
- The highest SUM(Low CI) value is observed for Utah (around 2,287)and Avg CI value is 17.33, followed by Oregon (around 2,284).
- The lowest SUM(Low CI) value is observed for Hawaii (around 1082) and Avg CI value is 8.26.

## c) Tree Map

The image displays the "Overall State-level Average Mental Health Care Value (AVG(Value)) in the United States" as a treemap chart. Every block in the treemap represents a US state, the size of the block indicates the average value for mental health care in the state, and the color indicates the range of values, lighter shades imply lower values and darker shades greater values.

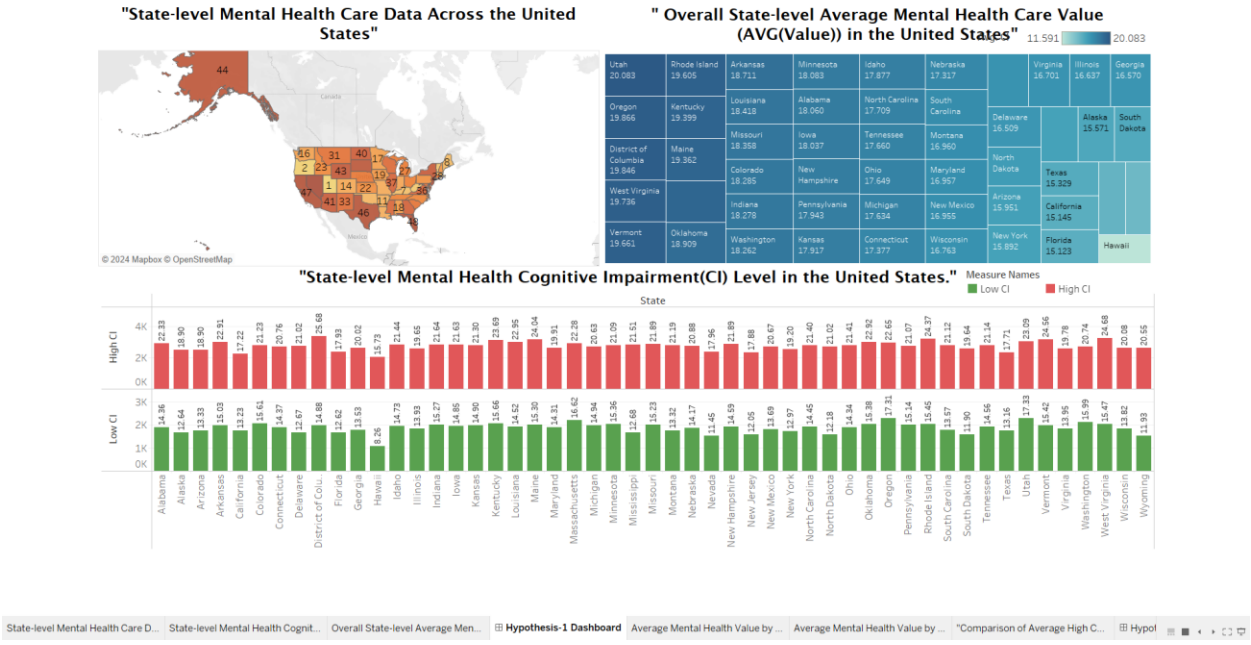



- The average values (AVG(Value)) on the chart vary from roughly 11.591 to 20.083, as indicated by the range indicator in the legend. Although there is no set order for the states, the treemap style makes it easy to compare the averages between them visually.

- For instance, Utah stands out with one of the darkest shades indicating one of the highest average values of approximately 20.083, while Hawaii is at the opposite end with one of the lightest shades and the lowest average value of approximately 11.591. This suggests significant disparities in mental health care value across the states.

The dashboard given below is a combination of the above visualizations. It is an interactive dashboard which gives a clear overview of the ranking of the states, cognitive impairment levels and average of both HIGH CI and LOW CI respectively and Average CI value.

To answer the first hypothesis. Indeed, there is often a relationship between degree of cognitive impairment in mental health and average mental health care value correlate with one another. “Utah” ranks first in CI level.

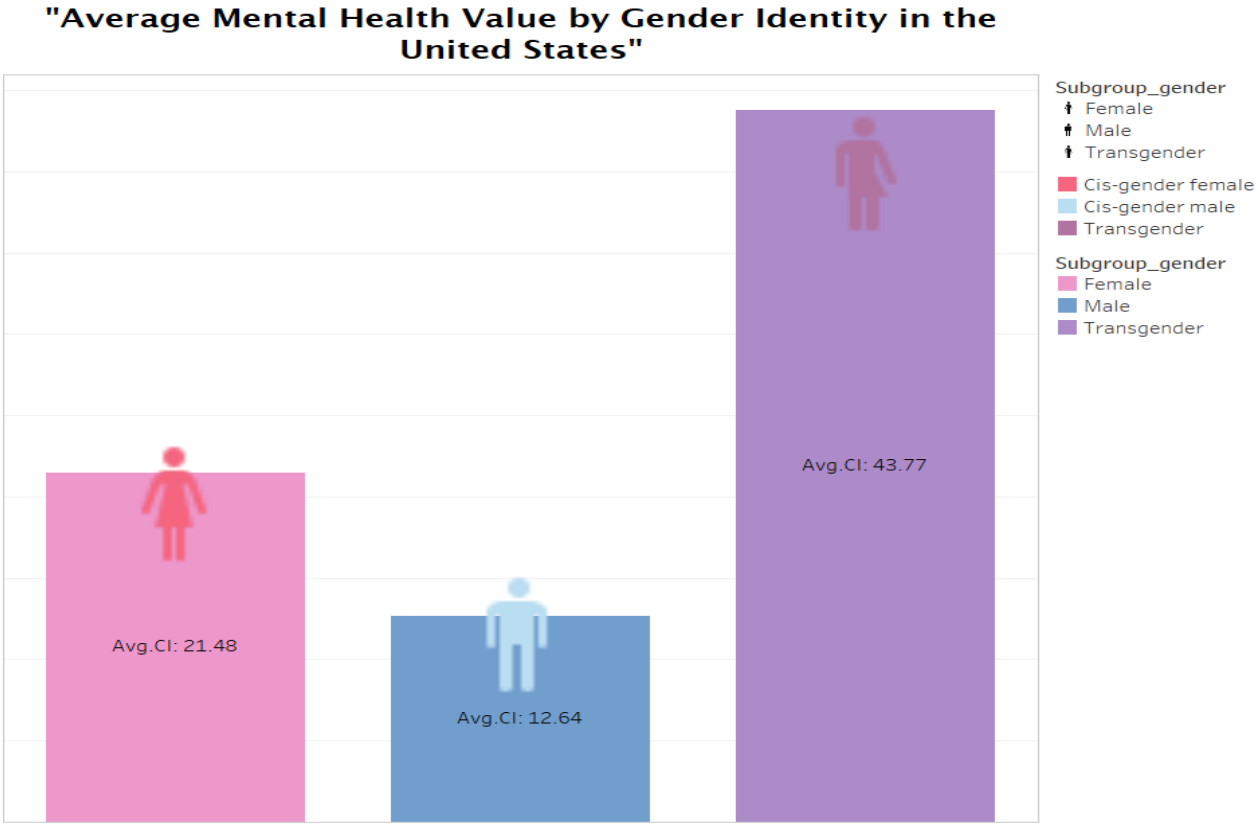



**2) Is there a statistically significant difference in average mental health values among different gender identities and education levels in the United States, as well as in the high confidence intervals for mental health care data?**

To analyze this hypothesis, I chose bar graph, packed bubble bar chart and bullet graphs. Clear explanation is given below.

## a) Bar Graph

The bar chart visual shows the "Average Mental Health Value by Gender Identity in the United States". Three vertical bars are displayed, each of which stands for a distinct gender identity: female, male and transgender. The bars are colored according to gender: pink for female, blue for men and purple for transgender. The data display to provide a statistical measure with a cognitive impairment level for each gender identity's mental health score, as indicated by the "Avg CI" value displayed on each bar.

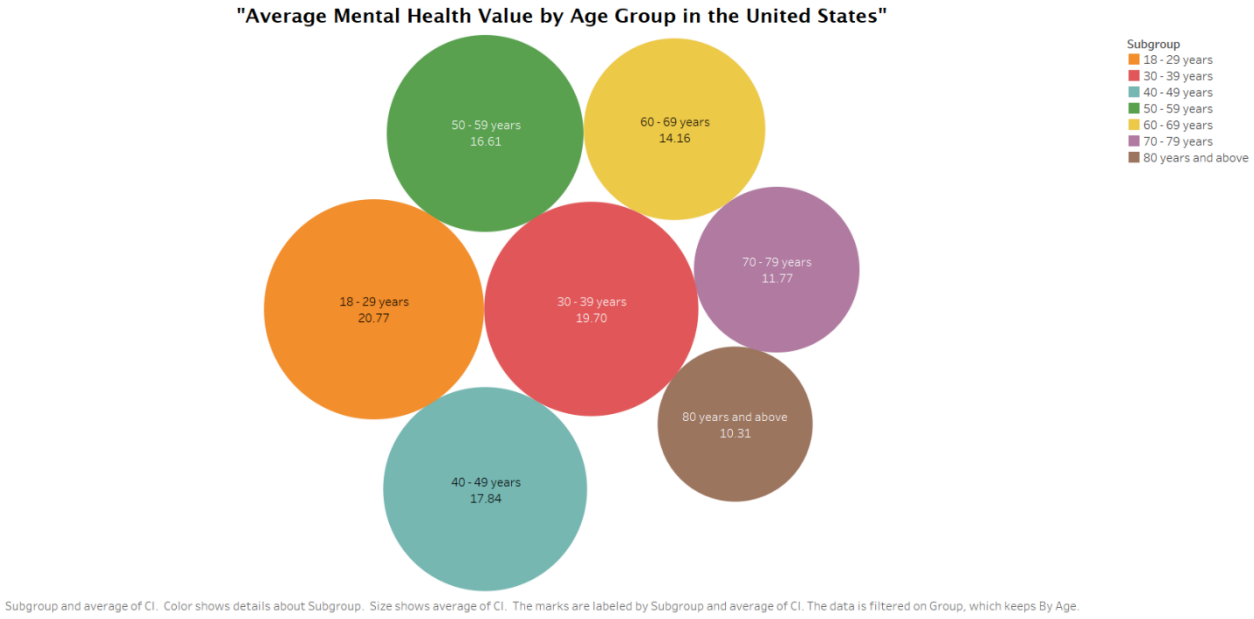



Average of CI and average of CI for each Subgroup. For pane Average of CI: Color shows details about Subgroup_gender. The marks are labeled by average of CI. For pane Average of CI (2): Color shows details about Subgroup (copy). Shape shows details about Subgroup_gender. The data is filtered on Group, which keeps By Gender identity.

The bar chart shows that the average mental health value is highest for the Transgender identity group, with an Avg CI of 43.77, followed by the Female group at 21.48, and the Male group at 12.64.

## b) Packed Bubble chart

A packed bubble chart labeled "Average Mental Health Value by Age Group in the United States" is shown below. The bubble chart shows a number of circles, each of which represents an age group. The size of the circle represents the average mental health value for that particular group. The age range that is included is "18-29 years" to "80 years and above".




The graph shows a trend whereby the average mental health score declines as age increases. Age groups that are younger, such as "18-29 years," have the greatest average mental health values. As one gets older, the values gradually decline, with the "80 years and above" group having the lowest values. According to it, younger people's concerns about mental health might have been more during the COVID-19 period from 2020 to 2022 compared to other age groups whereas older people showed less reported incidences of mental health issues during this period.

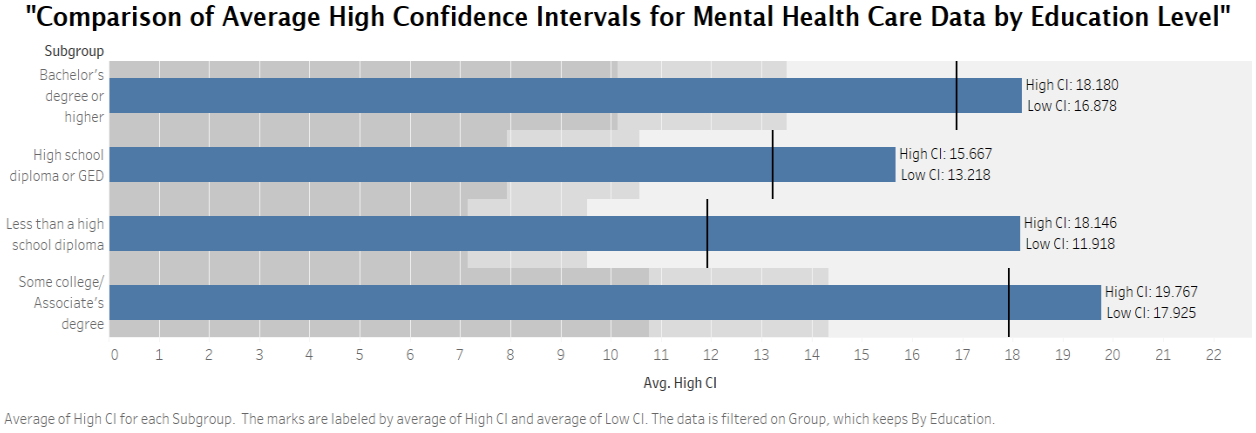
- The "18-29 years" group has the highest average mental health value at 20.77.
- The "80 years and above" group has the lowest average mental health value at 10.31.
- Each age group is represented by a different color, and the values are displayed within the circles.

## c) Bullet Bar Graph

The visualization presents a bullet bar chart titled "Comparison of Average High Confidence Intervals for Mental Health Care Data by Education Level". The chart compares the average high confidence intervals (CI) of mental health data across different education levels. The education levels are listed as:

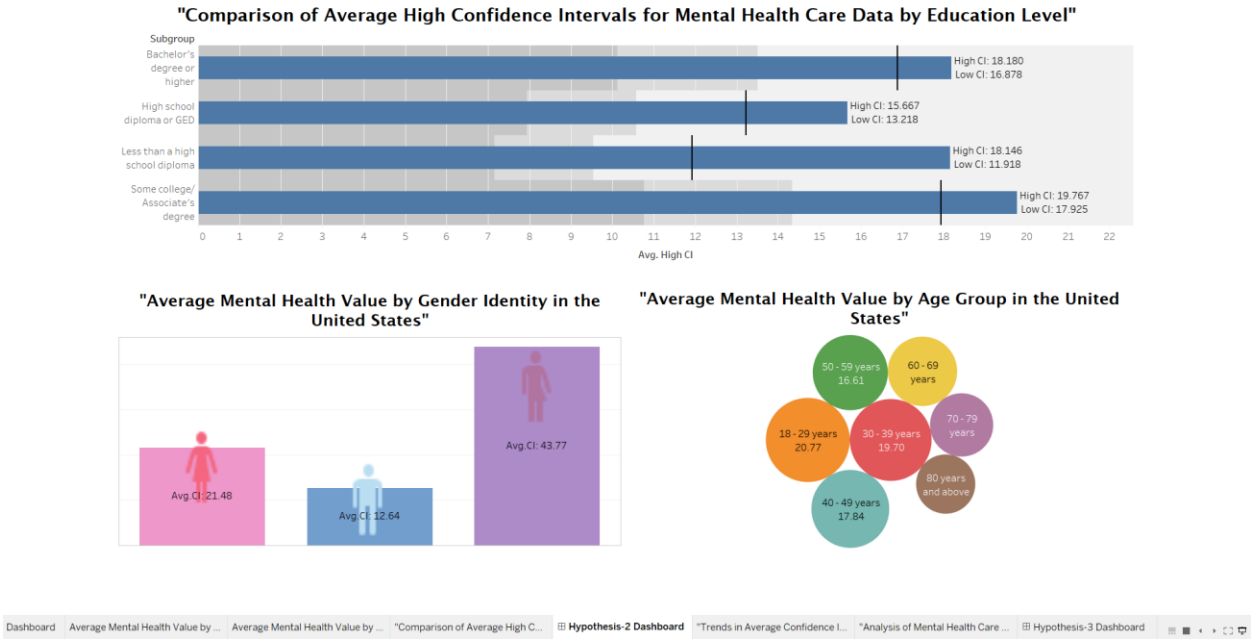
- Bachelor's degree or higher
- Some college/Associate's degree
- High school diploma or GED
- Less than a high school diploma

Each education level has a corresponding horizontal bar representing the average high confidence interval for that subgroup. The bars include error bars, indicating the range from the average low confidence interval to the average high confidence interval. Values for the high and low confidence intervals are noted at the end of each bar.




- According to the bar chart, those who identify as having a "Bachelor's degree or higher" have the highest average high CI value and a wider confidence interval range, indicating a bigger variation and maybe more demands or concerns for mental health care in this category.
- On the other hand, the group with "Less than a high school diploma" had the narrowest confidence interval and the lowest average high CI value, which may indicate that they have less concerns or needs for mental health care.

- The trend that can be seen in this case is that the range of the confidence intervals and the average high CI value both shrink with decreasing educational attainment.
- This may suggest that there is a relationship between the reported level of mental health care needs and educational attainment, with higher levels of education perhaps being linked to higher levels of self-reported mental health care needs over the COVID-19 time period from 2020 to 2022.




Yes, there is a statistically significant difference. The above is a dashboard that analyzes difference in average mental health values among different gender identities and education levels in the United States, as well as in the high confidence intervals for mental health care data.

**3) Does the average confidence interval for mental health care data vary significantly among different racial subgroups over time?**

To analyze the question, I chose Line Graph and Scatter plot. Below is the detailed explanation.

The below condition is used in the calculated field to rename the data.

```
IF [Subgroup] = "Non-Hispanic Asian, single race" then
   "Asian"
ELSEIF [Subgroup] = "Non-Hispanic Black, single race" then
   "Black"
ELSEIF [Subgroup] = "Non-Hispanic White, single race" then
   "White"
ELSEIF [Subgroup] = "Non-Hispanic, other races and multiple races" then
   "Other races"
ELSE
   [Subgroup]
END
```

## a) Line Graph

The image demonstrates a line graph titled, "Patterns in Average Margin of Error for Mental Health Care Data by Race/Ethnicity Over Time". This graph follows the average confidence interval (AVG CI) that serves as a statistical measure to quantify the uncertainty in the estimated value during the period of July 2020 to April 2022.

Five different racial/ethnic subgroups are represented by different colored lines:

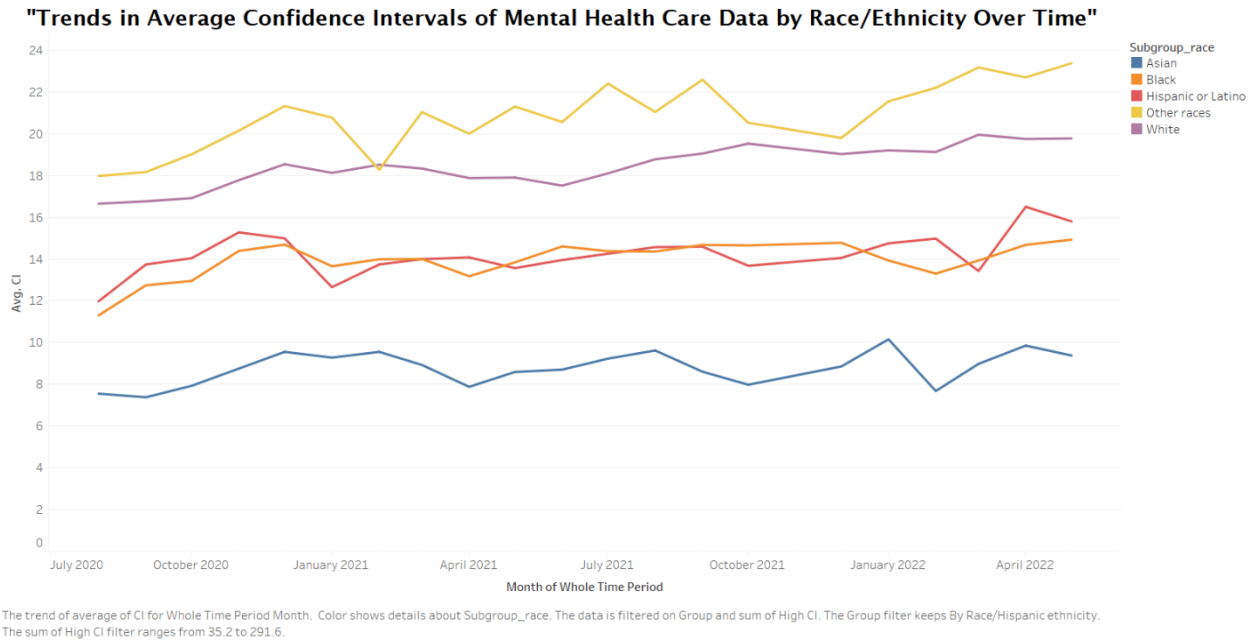
-Asian (blue)

- Black (Orange)

- Hispanic or Latino (Red)

- Other races (Yellow)

- White (purple)



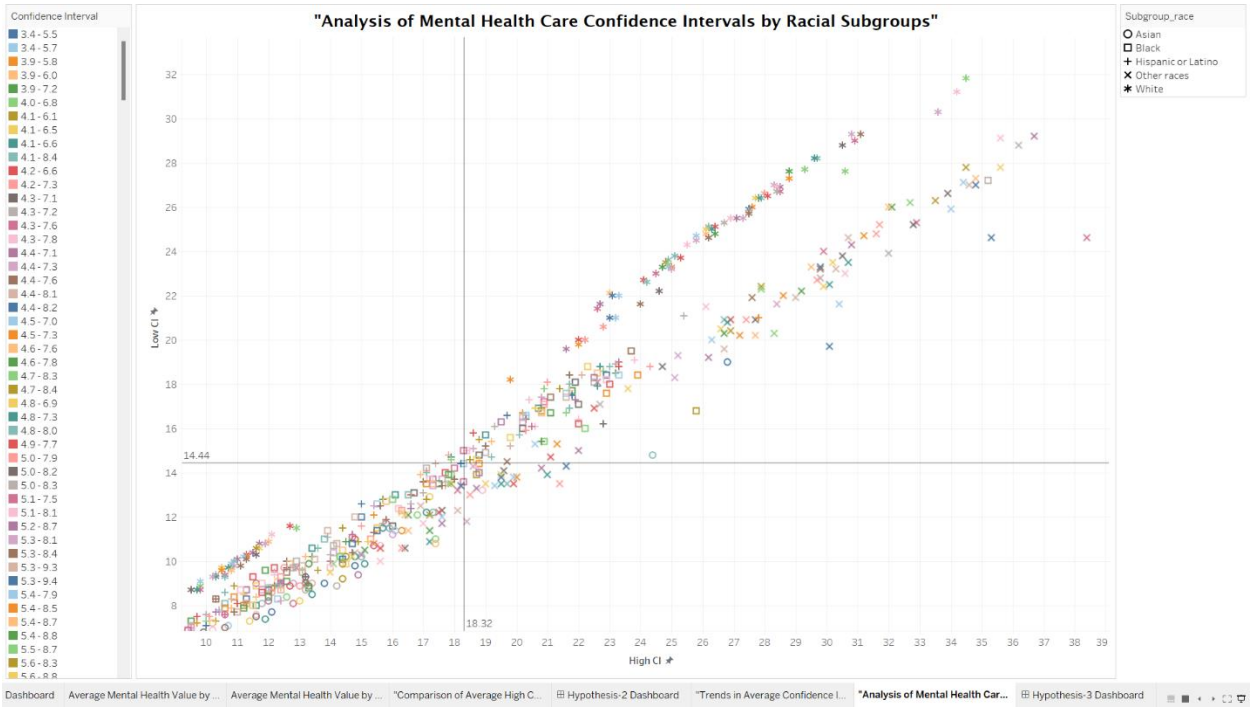

The x-axis will indicate the time period such as months and years while the y-axis will represent the confidence interval values. The diagram depicts deviations in the AVG Confidence Interval for each subset group during the given time period.

- The Asain subgroup (blue line) is characterized by the lowest AVG CI values which means data about mental health care for this group is the most stable and precise.
- One noteworthy summit area can be seen for the Asian subgroup at the end of October, 2020 which then gradually goes down before rising up once more, starting from January, 2021.
- The Black subgroup goes up the AVG CI in January 2021 but it is also relatively high and explicitly decreasing towards April 2022.
- The population from Hispanic or Latino group rises sharply in April 2021, and then it diminishes but rises towards the end of the period.
- The Other races subgroup shows a large peak in January 2022.
- The WHITE subgroup's AVG CI has a downfall from July 2020 to October 2020, onward the line fluctuates and drop a bit around January 2021 and picks the highest point in January 2022.

## b) Scatter plot

The image visualizes the scatter plot of “Mental Health Care Confidence Intervals by Racial Subgroups”. In the plot, the points data is displayed along with x-axis representing the low cognitive Impairment (Low CI) while the y-axis signifying the high cognitive impairment (High CI). Both intervals presumably share the same units of measurement.

- Asian (Circle)
- Black (Square)
- The Hispanic or Latin (Plus)
- Other races (Cross)
- White (Asterisk)

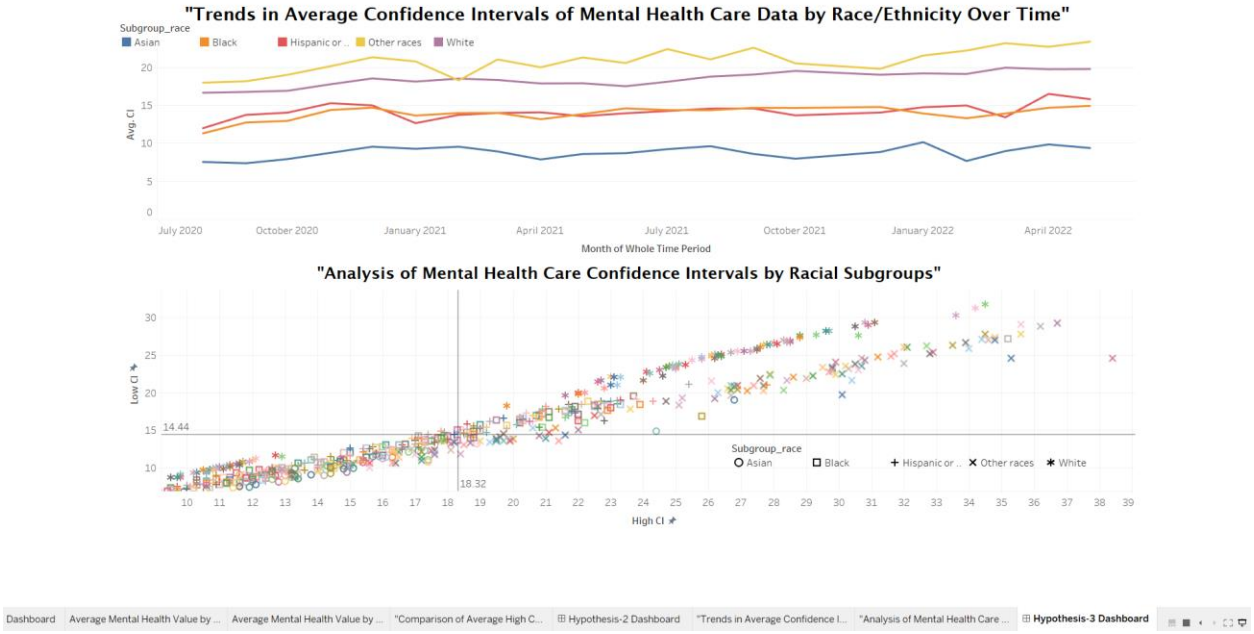



- The data points within Scatter plot are lined up the diagonal line which suggests positive relationship between low CI and high CI values. In other

words, the right end of the confidence interval grows along with the left bound.

- For the confidence interval, there are almost the same distribution of the racial subgroups but there is some variation. The graph does not show any single racial category concentrated in a certain area, thus indicating that the confidence intervals for mental health services data are not significantly different for the represented racial/ethnic groups.
- The Average of High CI 18.32 and Low CI is 14.44 respectively.
- Among the White data, there are an exception or two, where some data points have much higher Low CI and High CI comparatively to the majority of data points in all the groups.
- Substituting the majority of data points with about 10 to 30 for both the Low CI and the High CI, the uncertainty is delineated as moderate in the mental health care data estimates.

The dashboard given below is used to display the Average Confidence Intervals of Mental Health Care Data by Race/Ethnicity Over Time and confidence Intervals of Racial Subgroups.




By analysis of the dataset the answer of third hypothesis, yes, the average confidence interval for mental health care data do not vary significantly among different racial subgroups over time.

## DISCUSSIONS:

From Hypothesis 1, the following discussions can be made:

- The rates of cognitive impairment in the various states of the United States of America show geographic disparities. Utah is the state with the highest levels of cognitive impairment, but the rates are somewhat lower in Texas, Florida, and New York.
- As people age, there is a tendency for average mental health scores to decline; older people tend to have the lowest scores, while younger people tend to have the highest scores.

From Hypothesis 2, the following discussions can be made:

- The indicator of reported mental health care needs is not correlated with educational attainment since individuals with higher educational degrees tend to have larger error margins and more likely seek medical assistance when suffering from mental health illnesses.
- The gender concept is related to the mean health values, where transgender individuals rank highest, followed by females in second place and males in third, suggesting that the mental health of transgender individuals differs from that of females and males.

From Hypothesis 3, the following discussions can be made:

- It was also observed that there were temporal changes in the confidence intervals with racial framework divergences. While some groups exhibit varying degrees of oscillations, others have a constant confidence interval level.
- Racial groupings differ in how their intervals of confidence in the upcoming period evolve. This reflects how data reporting has changed over time and how accurate sample estimation procedures have become during the study period.

## CONCLUSION:

The study shows that during the COVID-19 pandemic, there were notable differences in the mental health care data in the US among various age groups, racial subgroups, gender identities, and educational attainment levels. These results emphasize the necessity for equal access to care and support services, as well as the significance of specialized therapies and assets to address a range of mental health issues.

## REFERENCES:

[1]. Reinert, M. (2019-2020). SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health. Retrieved from https://www.samhsa.gov/data/nsduh/state-reports-NSDUH-2020

[2]. Guo, Z., Zhang, Y., & Liu, Q. (2022). Bibliometric and visualization analysis of research trend in mental health problems of children and adolescents during the COVID-19 pandemic. Frontiers in Public Health, 10, 1040676. Published online January 6,2023.doi: 10.3389/fpubh.2022.1040676 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9853402/

[3]. National Alliance on Mental Illness (NAMI), US Mental Health Statistics by 2024, https://worldpopulationreview.com/state-rankings/mental-health-statistics-by-state