HIT-750 (DATA ANALYTICS)

Team-3 : Prof: Dr. Isaac Mativo

01) Mounika Atmakuri

02) Sridhar Reddy Maram

03) Everet Thokozani Kalongonda

04) Shriya Kandikonda Venkatadri

Global Trends in Mental Health Disorders

Overview of the topic :

- **Prevalence of Mental Disorders**
    - Approximately **1 in every 8 people** worldwide live with a mental disorder, highlighting the widespread impact of these conditions [2].
- **Mental Health and Society**
    - Mental health is a **crucial aspect of overall well-being**, influencing people's ability to work, maintain relationships, and engage with their communities [1].
- **Burden of Untreated Mental Illnesses**
    - Despite the existence of **effective prevention and treatment options**, a **large proportion of affected individuals do not receive proper care** due to barriers such as stigma, lack of resources, and poor-quality treatment [1,2]
- **Nature of Mental Disorders**
    - Mental disorders involve **significant disturbances in thinking, emotional regulation, and behavior**, affecting an individual's daily functioning [2].
- **The Need for Reliable Data**
    - **Comprehensive data collection** is crucial to understanding mental health trends, identifying risk factors, and improving access to effective treatments [1].

- **Variety of Mental Health Conditions**
    - There are **numerous types of mental disorders**, including depression, anxiety, schizophrenia, and bipolar disorder, each with varying degrees of severity and impact [2].
- **Gender Disparities in Mental Health**
    - Studies estimate that **1 in 3 women and 1 in 5 men** will experience major depression at some point in their lives [1].
- **Public Perception and Stigma**
    - Many individuals feel **uncomfortable discussing their mental health symptoms**, which can lead to underreporting and misrepresentation of the true prevalence of mental illnesses [1].

Need for the Comprehensive Data :

Mental health disorders are a significant public health concern worldwide, affecting individuals' well-being, productivity, and societal engagement. Understanding global trends in these disorders is crucial for developing effective prevention and treatment strategies, allocating resources appropriately, and reducing the stigma associated with mental health issues. The increasing prevalence of mental health conditions necessitates a comprehensive analysis to inform policy-making and healthcare practices.

Research Questions :

This project aims to explore global trends in the prevalence of mental health disorders using a dataset titled **"Global Trends in Mental Health Disorders"** available on Kaggle. The focus is on identifying patterns of mental disorders such as depression, anxiety, schizophrenia, and substance abuse across various regions and over time.

1. How have the prevalence rates of major mental health disorders (e.g., depression, anxiety, schizophrenia) changed globally over the past decade?
2. Are there notable differences in the prevalence of mental health disorders across geographic regions?
3. What demographic factors (e.g., gender, age) contribute to the variation in mental health trends?

Dataset Overview :

- **Origin and Authors**

- The dataset titled "Global Trends in Mental Health Disorder" is available on Kaggle
- Saloni Dattani, Lucas Rodés-Guirao, Hannah Ritchie and Max Roser (2023) - "Mental Health" Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/mental-health' [Original Online Resource]

- **Contents of the Dataset [05]**

The dataset provides global statistics on various mental health disorders across different countries and years. It includes the following columns:

- Entity – The country or region being analyzed.
- Code – The standardized country code.
- Year – The year of data collection.
- Schizophrenia (%) – The percentage of the population diagnosed with schizophrenia.
- Bipolar Disorder (%) – The percentage of individuals with bipolar disorder.
- Eating Disorders (%) – The prevalence of eating disorders.
- Anxiety Disorders (%) – The percentage of people experiencing anxiety disorders.
- Drug Use Disorders (%) – The proportion of the population with substance abuse disorders.
- Depression (%) – The prevalence of depression in the population.
- Alcohol Use Disorders (%) – The percentage of individuals affected by alcohol-related disorders.

- **Data Preparation**

The following data preparation steps were undertaken:

1. **Data Cleaning**: Removed missing values and handled inconsistencies in country names and codes.
2. **Transformation**: Standardized date formats and normalized prevalence percentages.

3. **Aggregation**: Grouped data by region and year for trend analysis.

Plots from the RStudio :

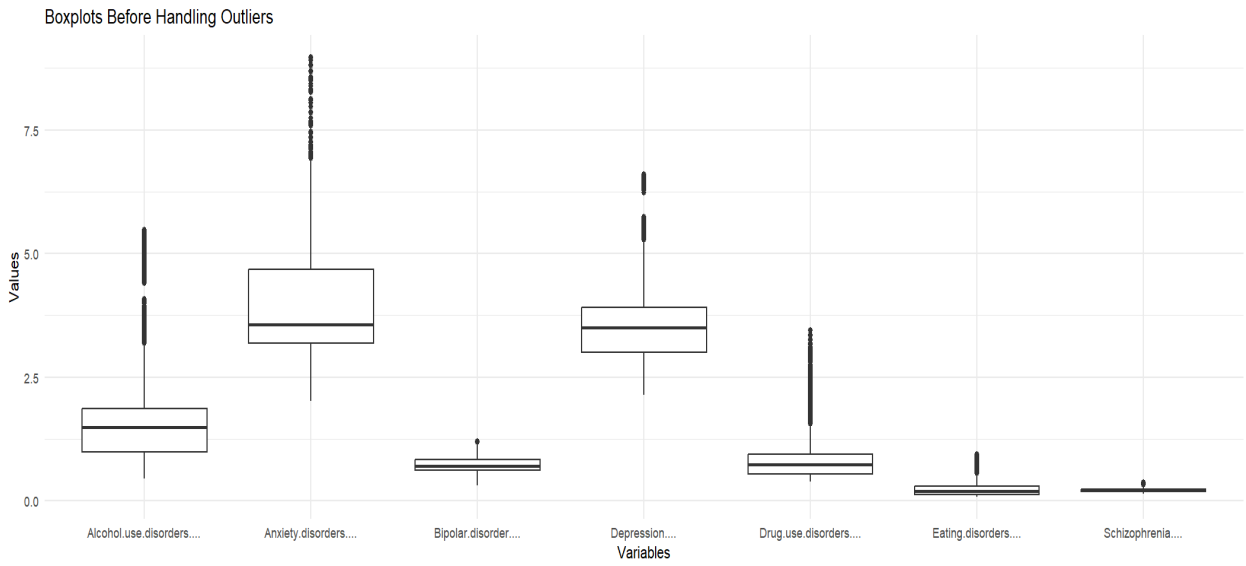

Fig-1 : Boxplots before handling outliers

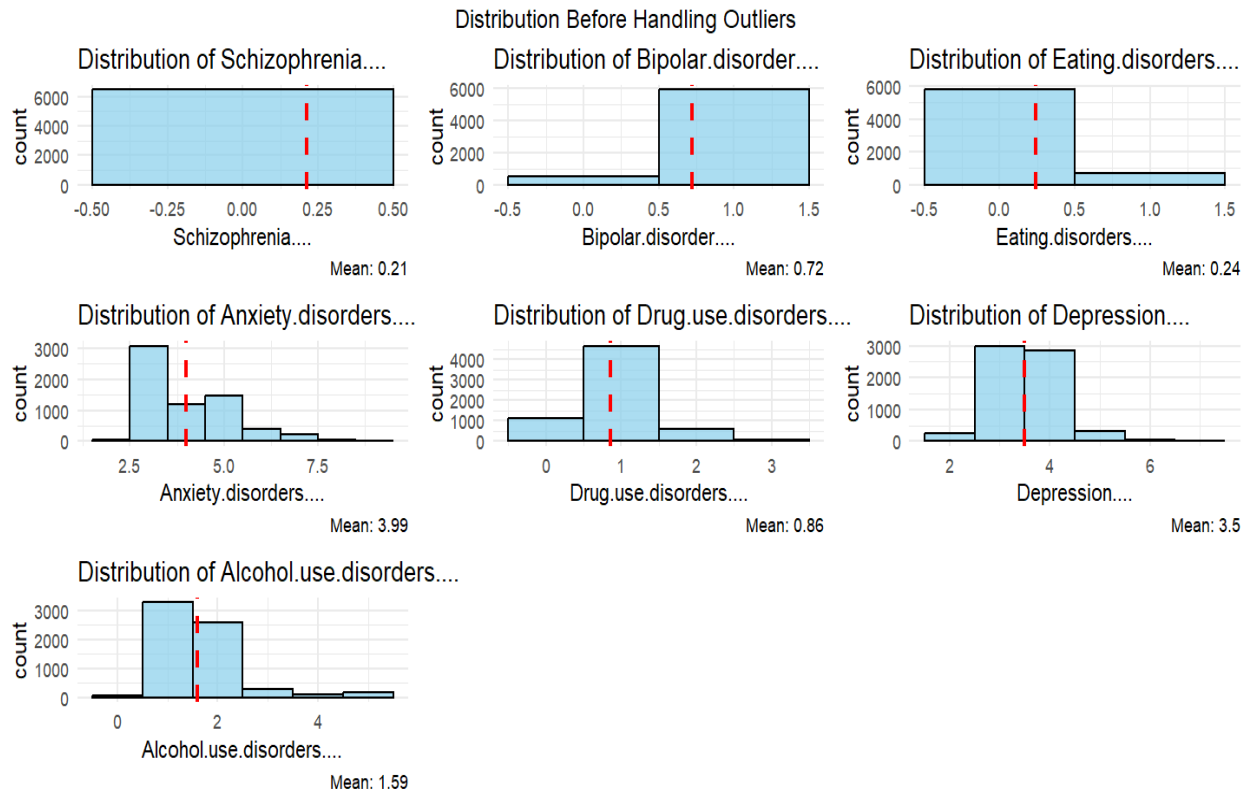

Fig-2 : Histographs before handling outliers

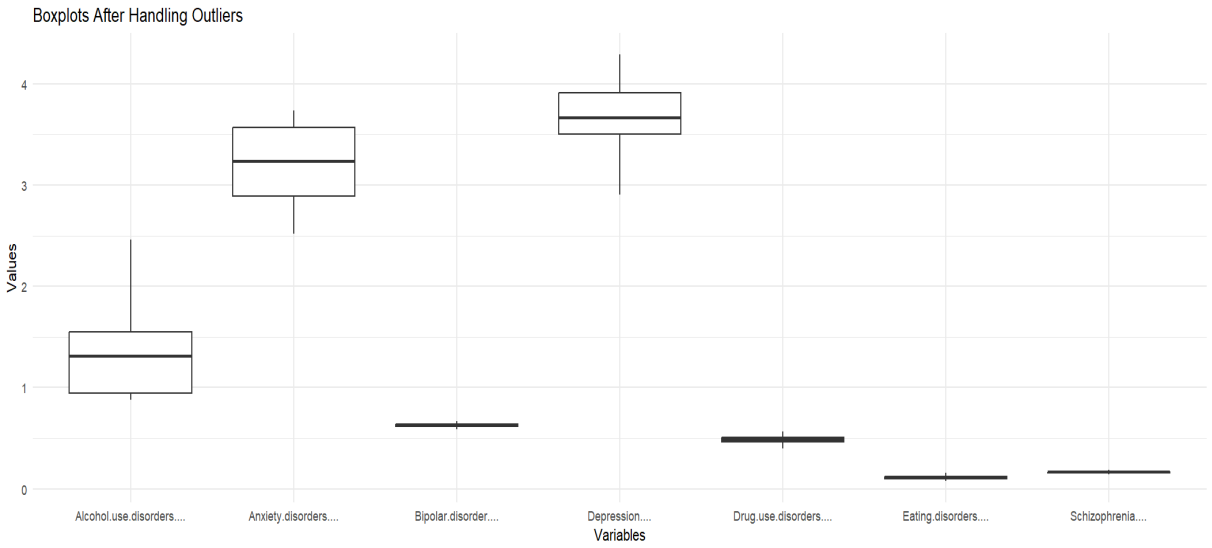

Fig-3 : Boxplots after handling outliers

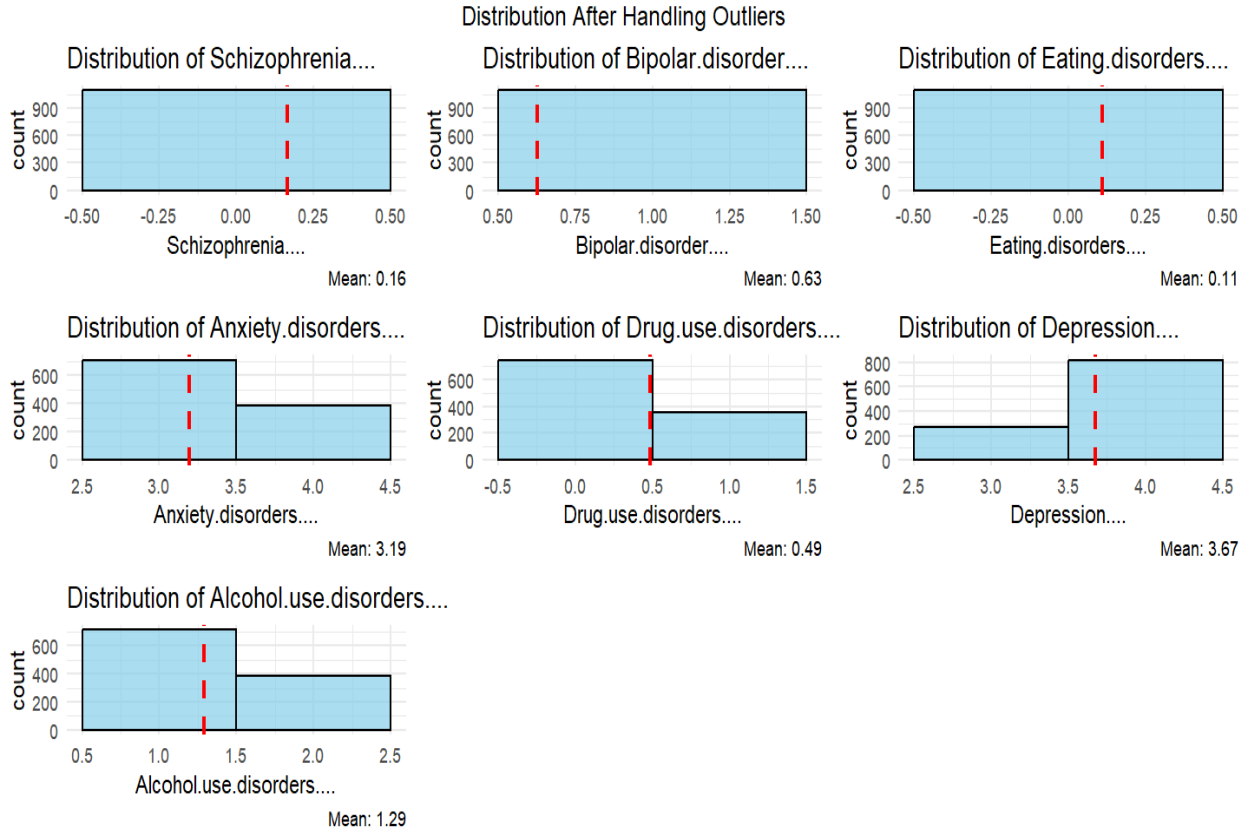

Fig-4 : Histographs after handling outliers

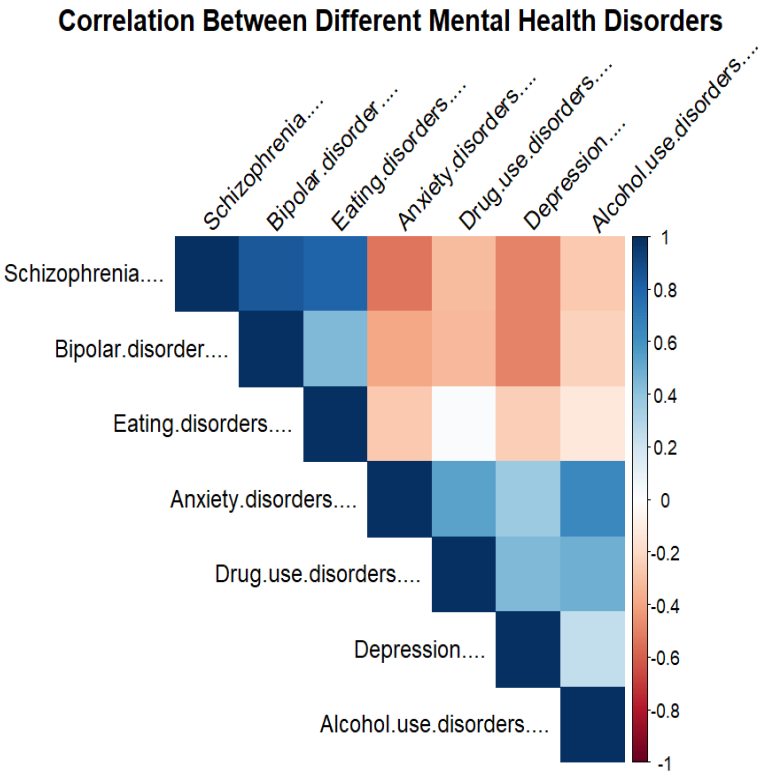

Fig-5 : Correlation Matrix for different Mental Health Disorders

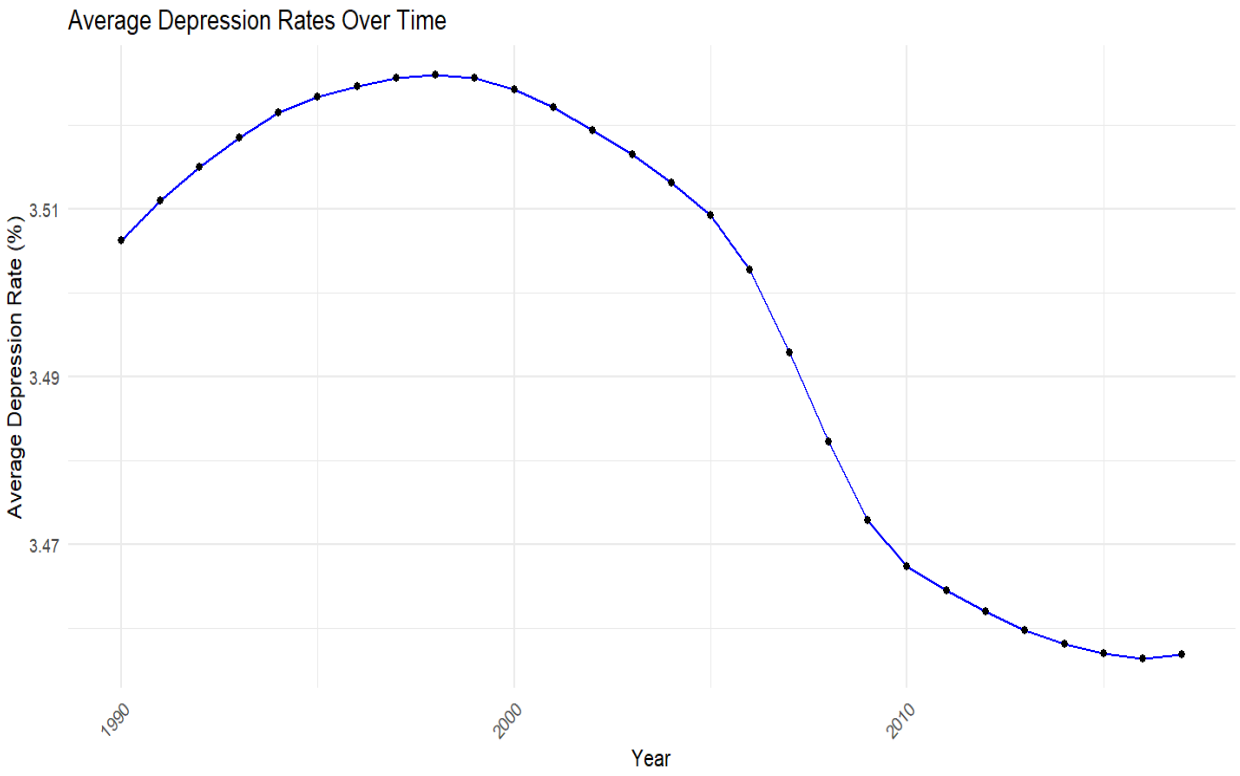

Fig-6 : Avg depression Rate over the time period

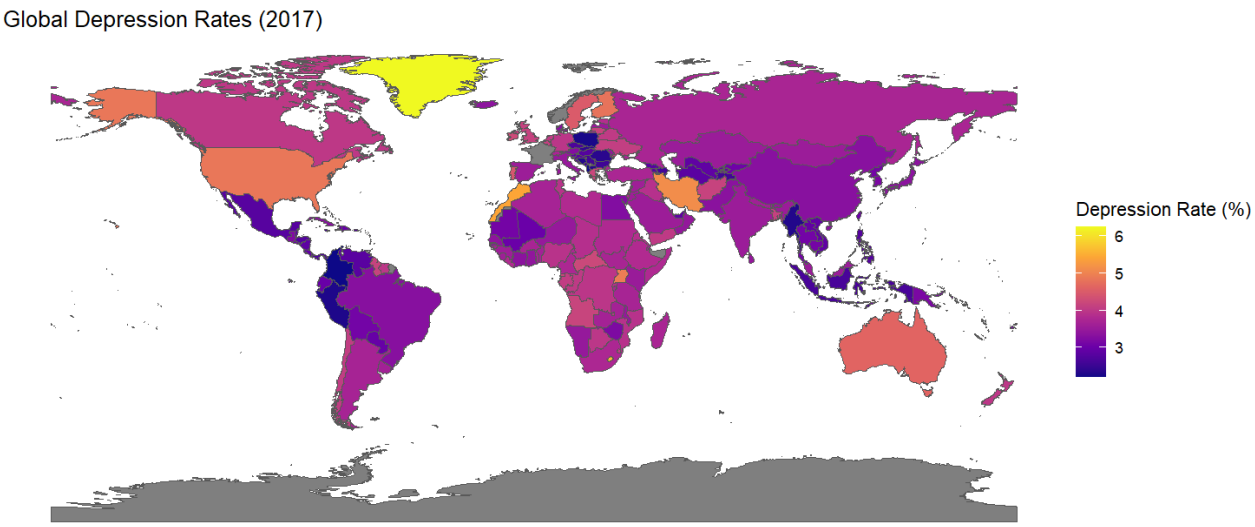

Fig-5 : Global distribution of Depression rate(%) for 2017

Appendix :

```
# ============================

# Install and Load Required Packages

# ============================


# Install required packages (if not already installed)

install.packages(c("dplyr", "ggplot2", "tidyr", "data.table", "readr", "gridExtra", "rnaturalearth",
"rnaturalearthdata"))

install.packages("sf")  # Required for handling spatial data


# Load required libraries

library(dplyr)

library(ggplot2)

library(tidyr)

library(readr)

library(data.table)
```

```
library(gridExtra)
library(grid)
library(rnaturalearth)
library(rnaturalearthdata)
library(sf)

# Suppress warnings
options(warn = -1)

# ============================
# Load and Explore the Dataset
# ============================

# Read the CSV file (update file path if necessary)
csv_data <- read.csv("Mental_health_Depression_disorder_Data.csv", stringsAsFactors = FALSE)

# Display the first 5 rows of the data
head(csv_data)

# Display column names
colnames(csv_data)

# Display the structure of the dataset
str(csv_data)

# ============================
# Data Type Conversion
# ============================
```

```
# Specify the columns to convert to numeric
columns_to_convert <- c("Year", "Schizophrenia....", "Bipolar.disorder....", "Eating.disorders....")

# Convert specified columns to numeric
csv_data[columns_to_convert] <- lapply(csv_data[columns_to_convert], as.numeric)

# Check the structure again after conversion
str(csv_data)

# ============================
# Check for Duplicates
# ============================

# Count and display total duplicates
total_duplicates <- sum(duplicated(csv_data))
cat("Total duplicate rows:", total_duplicates, "\n")

# Remove duplicates (uncomment to apply)
# csv_data <- distinct(csv_data)

# Display dimensions of the dataset after removing duplicates
dim(csv_data)

# ============================
# Handle Missing Values
# ============================

# Count missing values per column
missing_values <- colSums(is.na(csv_data))
```

```
cat("Missing values per column:\n")
print(missing_values)

# Drop rows with missing values in selected columns
selected_columns <- c("Code", "Year",
                      "Schizophrenia....", "Bipolar.disorder....",
                      "Eating.disorders....", "Anxiety.disorders....",
                      "Drug.use.disorders....", "Depression....",
                      "Alcohol.use.disorders....")

# Drop rows where data is missing in these columns
csv_data <- csv_data[complete.cases(csv_data[, selected_columns]), ]

# Display missing values after handling
missing_values_after_handling <- colSums(is.na(csv_data))
cat("Missing values after handling:\n")
print(missing_values_after_handling)

# Display dimensions of the cleaned dataset
dim(csv_data)

# ============================
# Remove Unnecessary Columns
# ============================

# Remove 'index' column if it exists
if ("index" %in% colnames(csv_data)) {
  csv_data$index <- NULL
}
```

```
# Display first 5 rows after column removal
head(csv_data)

# ============================
# Boxplots Before Handling Outliers
# ============================

# Function to create boxplots for numeric columns
plot_boxplots <- function(df, title) {
 # Identify numeric columns (excluding 'Year')
 numeric_columns <- sapply(df, is.numeric)
 numeric_columns <- names(df)[numeric_columns]
 numeric_columns <- setdiff(numeric_columns, "Year")

 # Convert data into long format for plotting
 df_long <- df[, numeric_columns] %>%
  pivot_longer(cols = everything(), names_to = "Variable", values_to = "Value")

 # Create and display boxplot
 p <- ggplot(df_long, aes(x = Variable, y = Value)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = title, x = "Variables", y = "Values") +
  theme_minimal()

 print(p)
}
```

```
# Plot boxplots before handling outliers
plot_boxplots(csv_data, "Boxplots Before Handling Outliers")

# ============================
# Visualizing Trends & Insights
# ============================

# Trend of depression over time
depression_trend <- csv_data %>%
 group_by(Year) %>%
 summarise(avg_depression = mean(Depression...., na.rm = TRUE))

ggplot(depression_trend, aes(x = Year, y = avg_depression, group = 1)) +
 geom_line(color = "blue") +
 geom_point() +
 labs(title = "Average Depression Rates Over Time",
    x = "Year",
    y = "Average Depression Rate (%)") +
 theme_minimal() +
 theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Choropleth Map: Global Depression Rates (2017)
world <- ne_countries(scale = "medium", returnclass = "sf")
map_data <- csv_data %>%
 filter(Year == 2017) %>%
 right_join(world, by = c("Code" = "iso_a3")) %>%
 st_as_sf()

ggplot(map_data) +
```

```
 geom_sf(aes(fill = Depression....)) +
 scale_fill_viridis_c(option = "plasma", name = "Depression Rate (%)") +
 labs(title = "Global Depression Rates (2017)") +
 theme_void()

# Animated Bubble Chart
dynamic_plot <- ggplot(csv_data,
                aes(x = Depression...., y = Anxiety.disorders....,
                  size = Bipolar.disorder...., color = Entity)) +
 geom_point(alpha = 0.7, show.legend = FALSE) +
 scale_size(range = c(2, 12)) +
 labs(title = "Year: {frame_time}",
    x = "Depression Rate",
    y = "Anxiety Rate") +
 transition_time(as.integer(Year)) +
 ease_aes("linear")

animate(dynamic_plot, fps = 10, width = 800, height = 600)

# ============================
# Handle Outliers Using IQR
# ============================

# Function to remove outliers using IQR
remove_outliers <- function(df) {
 df_clean <- df
 numeric_columns <- sapply(df_clean, is.numeric)
 numeric_columns <- names(df_clean)[numeric_columns]
 numeric_columns <- setdiff(numeric_columns, "Year")
```

```
 # Loop through numeric columns to remove outliers
 for (col in numeric_columns) {
   Q1 <- quantile(df_clean[[col]], 0.25, na.rm = TRUE)
   Q3 <- quantile(df_clean[[col]], 0.75, na.rm = TRUE)
   IQR <- Q3 - Q1
   lower_bound <- Q1 - 1.5 * IQR
   upper_bound <- Q3 + 1.5 * IQR

   # Filter outliers
   df_clean <- df_clean %>%
     filter(df_clean[[col]] >= lower_bound & df_clean[[col]] <= upper_bound)
 }
 return(df_clean)
}

# Iteratively remove outliers until no more are detected
data_clean <- csv_data  # Copy of the original data
repeat {
 previous_rows <- nrow(data_clean)  # Rows before removing outliers
 data_clean <- remove_outliers(data_clean)

 # Break loop if no more outliers are removed
 if (nrow(data_clean) == previous_rows) {
   break
 }
}

# Plot boxplots after handling outliers
```

```
plot_boxplots(data_clean, "Boxplots After Handling Outliers")

# Display dimensions after handling outliers
dim(data_clean)

# ============================
# Plot Histograms Before & After Handling Outliers
# ============================

# Load required packages
library(ggplot2)
library(gridExtra)

# Function to create histograms with mean lines
plot_histograms_grid <- function(df, title) {
 numeric_columns <- sapply(df, is.numeric)
 numeric_columns <- names(df)[numeric_columns]
 numeric_columns <- setdiff(numeric_columns, "Year")

 # List to store plots
 plot_list <- list()

 for (col in numeric_columns) {
  # Create the histogram
  p <- ggplot(df, aes_string(x = col)) +
   geom_histogram(binwidth = 1, fill = "skyblue", color = "black", alpha = 0.7) +
   geom_vline(aes_string(xintercept = paste0("mean(", col, ", na.rm = TRUE)")),
          color = "red", linetype = "dashed", size = 1) +
   ggtitle(paste("Distribution of", col)) +
```

```
    theme_minimal() +
    labs(caption = paste("Mean:", round(mean(df[[col]], na.rm = TRUE), 2)))

  # Append the plot to the list
  plot_list[[col]] <- p
 }

 # Arrange all plots in a grid
 grid.arrange(grobs = plot_list, ncol = 3, top = title)
}

# Plot histograms before handling outliers
plot_histograms_grid(csv_data, "Distribution Before Handling Outliers")

# Plot histograms after handling outliers
plot_histograms_grid(data_clean, "Distribution After Handling Outliers")

# ============================
# Trend Analysis & Correlation
# ============================
library(corrplot)
library(dplyr)
# Summary statistics for disorders by year
disorder_summary <- data_clean %>%
 group_by(Year) %>%
 summarise(
  avg_depression = mean(Depression...., na.rm = TRUE),
  avg_anxiety = mean(Anxiety.disorders...., na.rm = TRUE),
  avg_bipolar = mean(Bipolar.disorder...., na.rm = TRUE),
```

```
  avg_schizophrenia = mean(Schizophrenia...., na.rm = TRUE),
  avg_eating = mean(Eating.disorders...., na.rm = TRUE),
  avg_drug = mean(Drug.use.disorders...., na.rm = TRUE),
  avg_alcohol = mean(Alcohol.use.disorders...., na.rm = TRUE)
 )

# Correlation matrix
cor_matrix <- cor(data_clean %>%
            select(Schizophrenia...., Bipolar.disorder...., Eating.disorders....,
                Anxiety.disorders...., Drug.use.disorders...., Depression....,
                Alcohol.use.disorders....), use = "complete.obs")

# Plot correlation heatmap
corrplot(cor_matrix, method = "color", type = "upper",
      tl.col = "black", tl.srt = 45,
      title = "Correlation Between Different Mental Health Disorders",
      mar = c(0,0,1,0))
```

References :

01. Dattani, S., Rodés-Guirao, L., Ritchie, H., & Roser, M. (2023, June 20). *Mental health*. Our World in Data. https://ourworldindata.org/mental-health
02. World Health Organization: WHO. (2022, June 8). *Mental disorders*. https://www.who.int/news-room/fact-sheets/detail/mental-disorders
03. COVID-19 Mental Disorders Collaborators, & Santomauro, D. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. In The Lancet (Vol. 398). https://doi.org/10.1016/S0140-6736(21)02143-7
04. Wu, Y., Wang, L., Tao, M., Cao, H., Yuan, H., Ye, M., Chen, X., Wang, K., & Zhu, C. (2023). Changing trends in the global burden of mental disorders from 1990 to 2019 and

predicted levels in 25 years. *Epidemiology and Psychiatric Sciences*, *32*. https://doi.org/10.1017/s2045796023000756

05. *Global Trends in Mental health Disorder*. (2022, December 14). Kaggle. https://www.kaggle.com/datasets/thedevastator/uncover-global-trends-in-mental-health-disorder