A Comprehensive Study of Sleep, Age, and Mental Health Dynamics

Introduction

Mental health is a crucial aspect of every human well-being, influencing various dimensions of an individual's life, including emotions, thoughts, behaviors, and overall quality of life. In recent years, there has been a growing recognition of the significance of mental health in society, with an increasing number of individuals experiencing mental health challenges worldwide.

This survey focuses on conducting a comprehensive survey to delve into the multifaceted aspects of mental health, aiming to deepen the understanding of psychological well-being and its implications. The primary purpose of this survey is to shed light on the current state of mental health among individuals across diverse demographic backgrounds, encompassing various age groups, genders, mental health problems, and sleep patterns. By examining the prevalence and patterns of mental health issues, this study intends to contribute valuable insights into the factors that influence mental health disparities and identify vulnerable populations.

Furthermore, this survey aims to examine the impact of psychological factors, such as stress, depression, anxiety, and panic attacks, on mental health outcomes. Understanding these intricate dynamics will enable us to develop targeted strategies that promote positive mental health and prevent the onset of mental disorders. By implementing this comprehensive mental health survey, I anticipate generating a dataset that can serve as a valuable resource for researchers, policymakers, and mental health professionals. By disseminating the findings of this study, I aspire to raise awareness about the importance of mental health, reduce the associated stigma, and advocate for equitable access to mental health services.

Research that is grounded in previous research was written by Faravelli, C. et al. in 2012 titled "Associations between Perceived Stress, Depression, Anxiety, and Panic Attacks: A Population-Based Study." This research study aimed to examine the associations between perceived stress, depression, anxiety, and panic attacks among a population of medical students. The researchers explored the relationships and correlations between these variables to understand better the interplay among stress, depression, anxiety, and panic attacks in this specific population.

The findings of the study indicated that higher levels of perceived stress are linked to increased rates of depression, anxiety, and panic attacks among the medical student population. It also highlighted the interconnectedness of these mental health conditions, suggesting that high levels of stress can contribute to the development or exacerbation of depressive and anxiety symptoms, as well as an increased risk of experiencing panic attacks. It emphasizes the importance of addressing stress levels as a potential risk factor for the development of these mental health conditions.

The significance of this survey is that it identifies that mental health affects individuals' well-being, quality of life, and daily functioning. By surveying this area, there's a deeper understanding of individuals' specific challenges when dealing with these disorders. The findings can inform the development of interventions tailored to address the unique needs of affected individuals, promoting better mental health outcomes, and improving their overall quality of life.

Problem Statement

This survey is designed by a cross-sectional study. My survey focuses on three key variables: sleep duration, age, and the type of mental health disorder. These variables play crucial roles in shaping an individual’s mental health trajectory, and examining their intersections can provide detailed insights into the complex interplay between sleep, age, and mental well-being.

The first aspect under consideration is the relationship between sleep duration and individuals’ self-rated mental health. Sleep is a fundamental biological process that influences cognitive functions, emotional regulation, and overall psychological resilience. Investigating how variations in sleep duration correlate with self-reported mental health on a scale of 1-10, where 10 indicates good mental health, and 1 signifies poor mental well-being, which also offers valuable information for preventative strategies and therapeutic interventions.

The second point of my study explores the connection between age and mental health scale. Age is a factor that inherently brings about life transitions, exposing individuals to different stressors and challenges. By examining how age influences mental health perceptions, I aim to identify potential age-specific risk factors or protective factors contributing to variations in mental health outcomes.

Finally, I turn the attention to the interrelation between the type of mental health disorder and age. Certain mental health disorders may manifest at different stages of life, and understanding these patterns can aid in tailoring targeted interventions based on the age of onset and progression of specific disorders.

My survey seeks to unravel the intricate connections between sleep, age, and mental health, contributing to a more comprehensive understanding of the factors that influence mental well-being. Through this exploration, I aspire to inform mental health practices and policies, fostering a more detailed and practical approach to promoting mental health across diverse populations.

Research Method

Data Collection

This survey was administered through an online questionnaire platform to enhance accessibility and reach. Participants are provided with a link to the survey, where they complete the questionnaire at their convenience. The responses are collected anonymously to encourage honest and open sharing. The total number of participants is 60 after removing the incomplete responses.

Visual representations and demographic interpretations of the findings highlight key trends, patterns, or relationships from the data. Graphical representations were created through bars and charts where they represent the survey results with percentages and scales based on the nature of the data and the survey questions, such as:

Age? Numerical Value

Gender? Male and Female

Have you ever experienced a mental health problem? *

- ○ Yes
- ○ No
- ○ Maybe

What type of a mental disorder you have experienced before? *

- ○ Depression
- ○ Anxiety
- ○ Panic Attacks
- ○ Stress
- ○ Schizophrenia
- ○ Post - Traumatic Stress Disorder (PTSD)
- ○ Bipolar Disorder
- ○ Eating Disorder

How many hours do you sleep per day? *

- ○ Less than 5
- ○ 5-7
- ○ 7-9
- ○ 9+

Are you taking any medication? *

- ○ Yes
- ○ No

When was the last time you were really happy? *

- ○ Few days ago
- ○ Few weeks ago
- ○ Few months ago
- ○ Few years ago
- ○ Don't remember

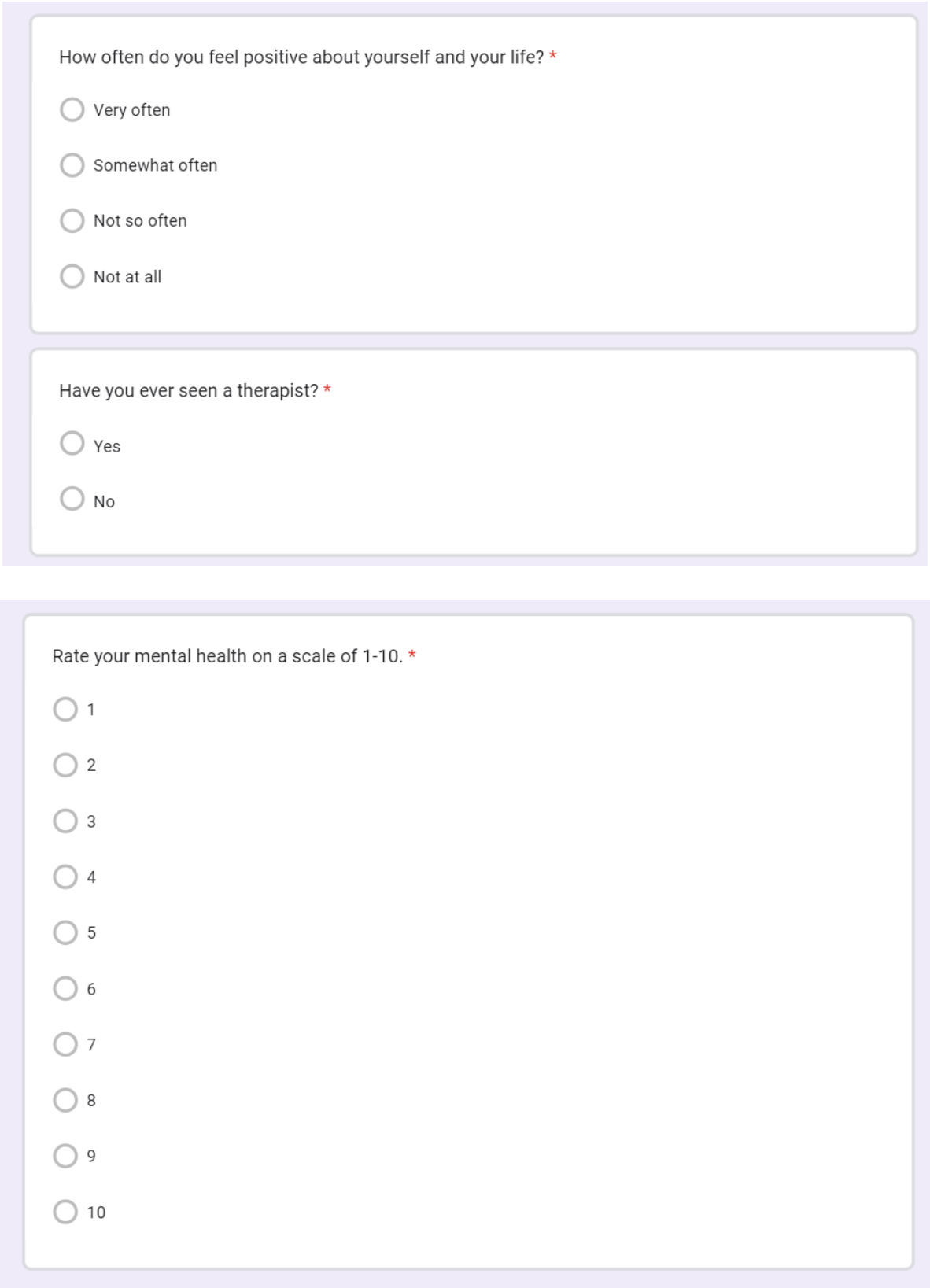

Figure 38: Survey Questions

The survey context provides a framework and background information for conducting a mental health survey. This survey context includes the following elements: Survey Questions guide the survey design and data collection process. It mainly focuses on the prevalence of anxiety and depression, factors contributing to stress among different demographic groups.

A literature review helps identify knowledge gaps, establish the theoretical framework, and ensure that the survey builds upon existing surveys. Target Population considers age, gender, and other relevant demographic characteristics. Survey Design developed a comprehensive survey instrument that captures the necessary data to answer survey questions. Data Collection determines the most appropriate method for data collection - online surveys.

Data Pre-Processing

The data is initially in XLSX format. All survey questions are transformed into corresponding variables, and the date and time information from the survey responses is excluded. The data is then imported into R. A row contains a missing age value, as one participant chose not to disclose their age. This response is subsequently omitted. Categorical variables are converted into factors with their respective levels to facilitate statistical analysis.

```
### Importing Data
library(xlsx)
sd <- read.xlsx("Mental Health Survey (Responses) (1).xlsx",1)

#### Removing rows with na values
sd <- sd[-35,]

#converting Character variables to Factors for statistical Analysis
for (i in 1:length(sd)) {
  if(is.character(sd[[i]]) == TRUE){
    sd[[i]] <- factor(sd[[i]])
  }
}
```

```
> str(sd)
'data.frame':   60 obs. of  10 variables:
 $ Age                : num  25 23 24 31 24 24 23 29 23 24 ...
 $ Gender             : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 2 1 1 2 2 ...
 $ Mental_Health_Issue: Factor w/ 3 levels "Maybe","No","Yes": 2 2 3 2 3 3 3 3 2 1 2 ...
 $ Disorder_type      : Factor w/ 6 levels "Anxiety","Bipolar Disorder",..: 6 6 6 6 3 6 5 6 3 6
...
 $ Sleep              : Factor w/ 4 levels "<5","5-7","7-9",..: 3 3 3 3 2 3 2 2 3 2 ...
 $ Medication         : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Happy              : Factor w/ 5 levels "Don't remember",..: 2 1 2 2 2 3 4 2 1 1 ...
 $ Positive           : Factor w/ 4 levels "Not at all","Not so often",..: 3 3 4 4 4 4 2 3 3 2
...
 $ Therapist          : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ Scale              : num  10 7 9 10 9 9 3 8 6 7 ...
```

Figure 39: Data Pre-Processing Code and Results

Before Processing, this data is directly from the survey with unnecessary variables like timestamp, survey questions, names, etc., and I removed them in the next step.

	A	B	C	D	E	F	G	H	I	J	K
1	Timestamp	Age	Gender	r experienced a mental hental disorder you have exany hours do you sleep pe	you taking any medicatic	the last time you were reu	feel positive about yourse	you ever seen a therapr	mental health on a scale of 1-10.		
2	7/2/2023 18:58:32	25	Female	No	Stress	7-9	No	Few days ago	Somewhat often	No	10
3	7/2/2023 18:58:46	23	Male	No	Stress	7-9	No	Don't remember	Somewhat often	No	7
4	7/2/2023 18:59:49	24	Male	Yes	Stress	7-9	No	Few days ago	Very often	No	9
5	7/2/2023 19:02:19	25	Male	No	Stress	7-9	No	Few days ago	Very often	No	10
6	7/2/2023 19:04:50	25	Female	Yes	Depression	5-7	No	Few days ago	Very often	No	9
7	7/2/2023 19:20:30	25	Male	Yes	Stress	7-9	No	Few months ago	Very often	No	9
8	7/2/2023 21:53:16	25	Female	Yes	Panic Attacks	5-7	No	Few weeks ago	Not so often	No	3
9	7/2/2023 23:27:50	25	Female	No	Stress	5-7	No	Few days ago	Somewhat often	No	8
10	7/2/2023 23:50:39	25	Male	Maybe	Depression	7-9	No	Don't remember	Somewhat often	Yes	6
11	7/3/2023 0:10:45	25	Male	No	Stress	5-7	No	Don't remember	Not so often	No	7
12	7/3/2023 21:47:12	25	Female	Yes	Panic Attacks	5-7	No	Few months ago	Somewhat often	Yes	8
13	7/3/2023 21:49:53	25	Male	Maybe	Stress	7-9	No	Few weeks ago	Not so often	No	7
14	7/3/2023 21:50:55	25	Male	Maybe	Stress	5-7	No	Few days ago	Very often	No	9
15	7/3/2023 21:52:04	25	Female	Yes	Stress	Less than 5	No	Don't remember	Somewhat often	No	7
16	7/3/2023 21:54:49	25	Female	Yes	Anxiety	5-7	No	Few weeks ago	Very often	No	7
17	7/4/2023 4:05:32	25	Female	Yes	Panic Attacks	5-7	No	Few days ago	Very often	No	7
18	7/4/2023 4:20:58	25	Female	No	Stress	9+	No	Few days ago	Very often	No	9
19	11/11/2023 18:43:30	25	Male	Yes	Depression	Less than 5	No	Few months ago	Not so often	No	5
20	11/11/2023 18:43:54	25	Female	Maybe	Stress	5-7	No	Few months ago	Somewhat often	No	7
21	11/11/2023 18:49:04	25	Male	Yes	Panic Attacks	5-7	No	Few months ago	Somewhat often	No	6
22	11/11/2023 18:49:46	25	Male	Maybe	Anxiety	Less than 5	No	Few weeks ago	Somewhat often	No	7
23	11/11/2023 18:50:27	25	Female	Maybe	Depression	Less than 5	Yes	Few years ago	Not at all	Yes	3
24	11/13/2023 9:00:53	25	Male	Maybe	Anxiety	9+	No	Few days ago	Very often	No	9
25	11/13/2023 9:02:00	25	Male	Yes	Depression	Less than 5	Yes	Few years ago	Not so often	Yes	3
26	11/13/2023 9:03:57	25	Male	Maybe	Panic Attacks	5-7	No	Few months ago	Somewhat often	No	7
27	11/13/2023 9:07:21	25	Female	No	Stress	7-9	No	Few days ago	Very often	No	10
28	11/13/2023 9:08:57	25	Female	Maybe	Panic Attacks	7-9	No	Few weeks ago	Somewhat often	No	8
29	11/13/2023 9:10:52	25	Female	Yes	Depression	Less than 5	Yes	Few years ago	Not so often	Yes	4
30	11/13/2023 9:15:55	27	Male	No	Stress	7-9	No	Few days ago	Somewhat often	No	9
31	11/13/2023 17:07:47	32	Male	Yes	Panic Attacks	Less than 5	Yes	Few years ago	Not so often	Yes	5

Figure 40: Original Survey Data Before Pre-Processing

After processing, the data looks like this with the required variables and relevant column names.

Age	Gender	Mental_Health_Issue	Disorder_type	Sleep	Medication	Happy	Positive	Therapist	Scale
25	Female	No	Stress	7-9	No	Few days ago	Somewhat often	No	10
23	Male	No	Stress	7-9	No	Don't remember	Somewhat often	No	7
24	Male	Yes	Stress	7-9	No	Few days ago	Very often	No	9
31	Male	No	Stress	7-9	No	Few days ago	Very often	No	10
24	Female	Yes	Depression	5-7	No	Few days ago	Very often	No	9
24	Male	Yes	Stress	7-9	No	Few months ago	Very often	No	9
23	Female	Yes	Panic Attacks	5-7	No	Few weeks ago	Not so often	No	3
29	Female	No	Stress	5-7	No	Few days ago	Somewhat often	No	8
23	Male	Maybe	Depression	7-9	No	Don't remember	Somewhat often	Yes	6
24	Male	No	Stress	5-7	No	Don't remember	Not so often	No	7
26	Female	Yes	Panic Attacks	5-7	No	Few months ago	Somewhat often	Yes	8
25	Male	Maybe	Stress	7-9	No	Few weeks ago	Not so often	No	7

Figure 41: Data After Pre-Processing

Dashboard

Tableau is a potent data visualization and business intelligence tool that empowers users to craft interactive, shareable dashboards. Dashboards comprise visualizations like charts, graphs, maps, and other elements, carefully curated to convey insights from data in a coherent and accessible manner Jones (2014).

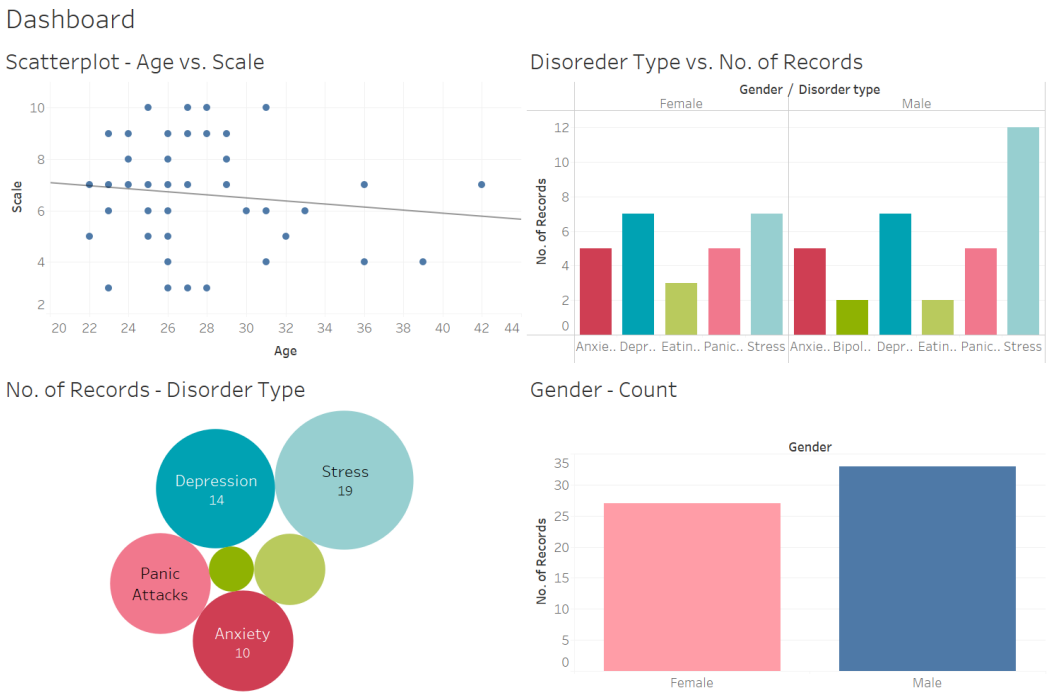

Figure 42: Survey Dashboard

To create a dashboard, I used Tableau and designed a comprehensive visualization to interpret the dataset to extract meaningful insights visually.

Research Question 1: Sleep vs. Mental Health Scale

The first research question concerns the association between sleep and mental health scale. The difference in the scale's means between different sleep duration groups is explored.

Descriptive Statistics

Boxplot

The boxplot in R is a powerful tool for summarizing and visualizing data distribution. Its importance lies in its ability to quickly convey critical aspects of a dataset, identify outliers, and facilitate comparisons between groups, making it an essential component of exploratory data analysis.

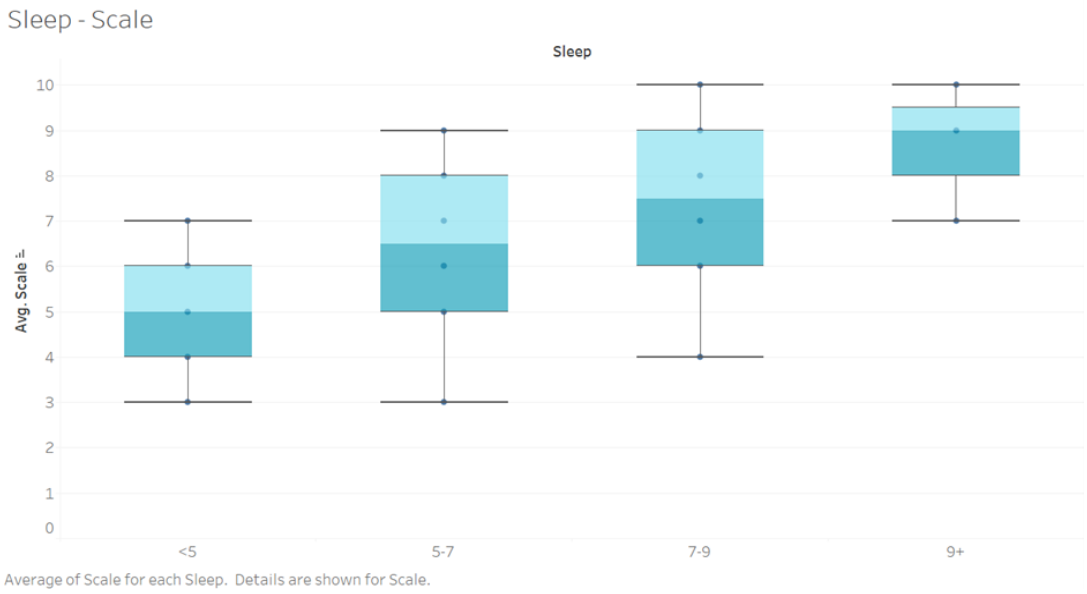

Figure 43: Sleep vs. Scale Boxplot

Based on the survey depicted in the above plot, the following observations are made:

The mental health scale distribution for the "Less than 5 hours sleep" group ranges from 3 to 7. Approximately 50% of participants in this group fall within the scale of 3 to 5, while the remaining 50% fall between the scale of 7 and 5.

For the "5-7 hours' sleep" group, the mental health scale distribution extends from 3 to 9. About 50% of participants in this group have a scale below 7, and the remaining 50% have a scale between 7 and 9.

In the "7-9 hours' sleep" group, the mental health scale distribution spans from 4 to 10. Here, 75% of participants have a scale above 6, while the remaining 25% fall between the scale of 4 and 6.

The "9+ hours sleep" group exhibits a mental health scale distribution from 7 to 10. Approximately 50% of participants in this group are distributed between 9 and 10, and the other 50% are distributed between 7 and 8.

It is important to note that as the duration of sleep increases among participants in this survey, the lower whisker of the boxplots ascends.

Histogram

Histograms in R play a crucial role in exploratory data analysis by providing an intuitive and informative representation of the distribution of continuous variables. Their importance lies in their ability to reveal critical features of the data, guide further analysis, and support decision-making in data exploration and interpretation.

```
hist(sd$Scale, probability = T, col = "lightblue", ylim = c(0,0.24),
     main = "Histogram of Mental Health Rating",xlab = "Rating")
curve(dnorm(x,mean = mean(sd$Scale), sd = sd(sd$Scale)),
      lwd= 2, col = "red", add = TRUE)
```

Figure 44: Histogram Code for Scale

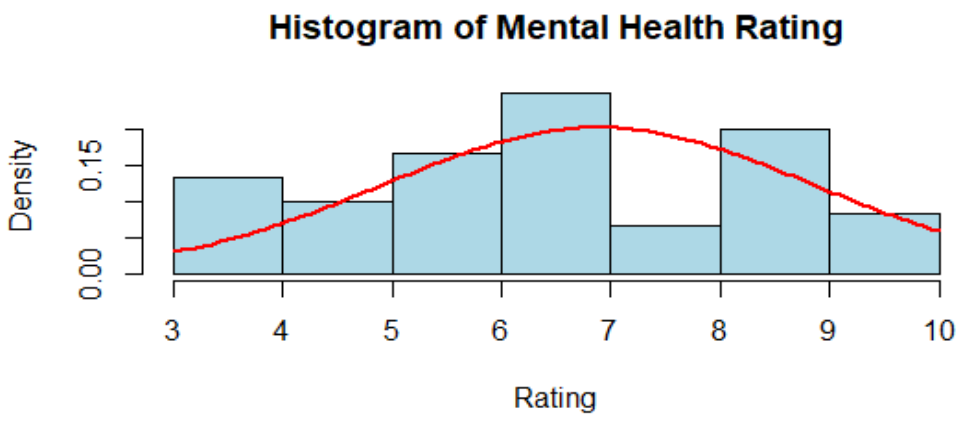

Figure 45: Histogram of Scale

From the above histogram, the distribution is close to a normal distribution. Also, we can see the bell-shaped curve with only one peak, which is symmetric around the mean. This normal distribution of the dependent variable may increase the validity of the ANOVA test.

Analysis

ANOVA

Analysis of Variance (ANOVA) is a statistical method used to assess whether there are any statistically significant differences between the means of three or more independent groups Soetewey (2022).

- If the p-value is less than the chosen significance level (e.g., 0.05), null hypothesis is rejected.
- Rejecting the null hypothesis implies that at least one group is different from the others.
- Reject H0: Conclude a significant difference in at least one group means.
- Fail to Reject H0: Conclude that insufficient evidence suggests a significant difference in group means.

ANOVA allows for the comparison of multiple groups simultaneously and is useful when there are more than two groups to compare. The F-statistic assesses whether the variability between group means is more significant than within groups.

Here, ANOVA analyzes the association between sleep and mental health scale. The code and the results are below. ANOVA Formulas are also added below.

1. Total Sum of Squares (SST):

$$SST = \Sigma_{i=1}^{k}\Sigma_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$$

k: $Number\ of\ groups$

n_i: $Number\ of$ observations in group i

X_{ij}: $j - th\ observation\ in\ group\ i$

$\bar{X}$: $Overall\ mean$

2. Between-Groups Sum of Squares (SSB):

$$SSB = \Sigma_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

$\bar{X}_i$: $Mean\ of\ group\ i$

3. Within-Groups Sum of Squares (SSW):

$$SSW = \Sigma_{i=1}^{k}\Sigma_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2$$

4. Degrees of Freedom:

$df_{Total} = N - 1, where\ N\ is\ the\ total\ number\ of\ observations.$

$df_{Between} = k - 1, where\ k\ is\ the\ number\ of\ groups.$

$df_{within} = N - k$

5. Mean Squares:

$$MS_{Between} = \frac{SSB}{df_{Between}}$$

$$MS_{within} = \frac{SSW}{df_{within}}$$

6. F-statistic:

$$F = \frac{MS_{Between}}{MS_{Within}}$$

7. The p-value is obtained from the F-distribution with degrees of freedom and df_{within}.

Figure 46: ANOVA Formulas

```
m <- aov(Scale ~ Sleep , data = sd)
```

Figure 47: ANOVA Test

```
> summary(m)
             Df Sum Sq Mean Sq F value  Pr(>F)
Sleep         3  92.77  30.925    12.6 2.1e-06 ***
Residuals    56 137.41   2.454
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure 48: ANOVA Test Results

It indicates a statistically significant difference in the mean mental health scale among the different sleep duration groups. The p-value associated with the F statistic is very small (2.1e-06), falling below conventional significance levels (e.g., 0.05). These results suggest that at least one pair of sleep duration groups has a significantly different mean mental health scale. As the ANOVA says, there is at least one pair of sleep groups that have other mean mental health; I wanted to extend my statistical analysis to the group level to see which groups have statistically significant differences in mean mental health scale by doing the post hoc analysis using Tukey post hoc method.

Tukey

Tukey's Honestly Significant Difference (HSD) is a valuable tool in data analytics, particularly when handling multiple groups. It assists researchers in identifying and comprehending specific differences among groups after detecting a significant overall effect in ANOVA. This contributes to a more detailed interpretation of experimental results Soetewey (2022).

```
> TukeyHSD(m)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Scale ~ Sleep, data = sd)

$Sleep
             diff        lwr      upr     p adj
5-7-<5  1.6956522  0.2896498 3.101655 0.0120234
7-9-<5  3.1052632  1.6443335 4.566193 0.0000036
9+-<5   3.7500000  1.3984437 6.101556 0.0005076
7-9-5-7 1.4096110  0.1237410 2.695481 0.0263718
9+-5-7  2.0543478 -0.1926420 4.301338 0.0847373
9+-7-9  0.6447368 -1.6370247 2.926498 0.8770523
```

Figure 49: TukeyHSD Analysis

The Tukey HSD post-hoc test output provides information on the pairwise comparisons between sleep duration groups regarding their mean mental health scale.

- The “diff” column shows the mean mental health scale differences between the respective sleep duration groups.

- The “lower” and “up” columns represent the lower and upper bounds of the 95% confidence interval for each difference.

- The “p adj” column provides the adjusted p-values after correction for multiple comparisons.

- The mean mental health scale for the group with sleep duration 5-7 significantly differs from the group with sleep duration <5 ($p = 0.012$).

- The mean mental health scale for the group with sleep duration 7-9 is significantly different from the group with sleep duration <5 ($p = 0.0000036$) and the group with sleep duration 5-7 ($p = 0.026$).

- The mean mental health scale for the group with sleep duration 9+ is significantly different from the group with sleep duration <5 ($p = 0.0005076$) and the group with sleep duration 5-7 ($p = 0.0847$).

The groups 5-7 and < 5, 7-9 and <5, 9+ and <5, 7-9 and 5-7, 9+ and 5-7 have statistically significant mean mental health scale differences. The 9+ and 7-9 groups have no statistically significant difference in mean mental health; the p-value is 0.877. This means the participants with sleep duration of 9+ hours and 7-9 hours do not have statistically different mean mental health scale.

Overall Analysis:

- The ANOVA results indicate a statistically significant difference in the mean mental health scale among the sleep duration groups ($F(3, 56) = 12.6$, $p < 0.001$).

- Degrees of Freedom (Df):

– The Sleep factor has 3 degrees of freedom (four sleep duration groups minus one), and the Residuals (error term) have 56 degrees of freedom.

- F Value:

– The F value of 12.6 suggests that the variation in the mean mental health scale among sleep duration groups is more significant than what would be expected by chance alone.

- P-value:

– The p-value (2.1e-06) is much smaller than the conventional significance level of 0.05, indicating a solid indication that at least one pair of sleep duration groups has a significantly different mean mental health scale.

Pairwise Comparisons:

- The Tukey HSD post-hoc test is used to identify which specific sleep duration groups have significantly different mean mental health scale.

Research Question 2: Age vs. Mental Health Scale

The second research question analyzes the correlation between age and mental health scale. Regression Analysis and Spearman's rank correlation test are used for this research question.

Descriptive Statistics

Age

Histogram is performed for age and the code is shown below.

```
hist(sd$Age, probability = T, col = "lightblue", ylim = c(0,0.09),
     main = "Histogram of Age",xlab = "Age")
curve(dnorm(x,mean = mean(sd$Age), sd = sd(sd$Age)),
      lwd= 2, col = "red", add = TRUE)
```

Figure 50: Histogram Code for Age

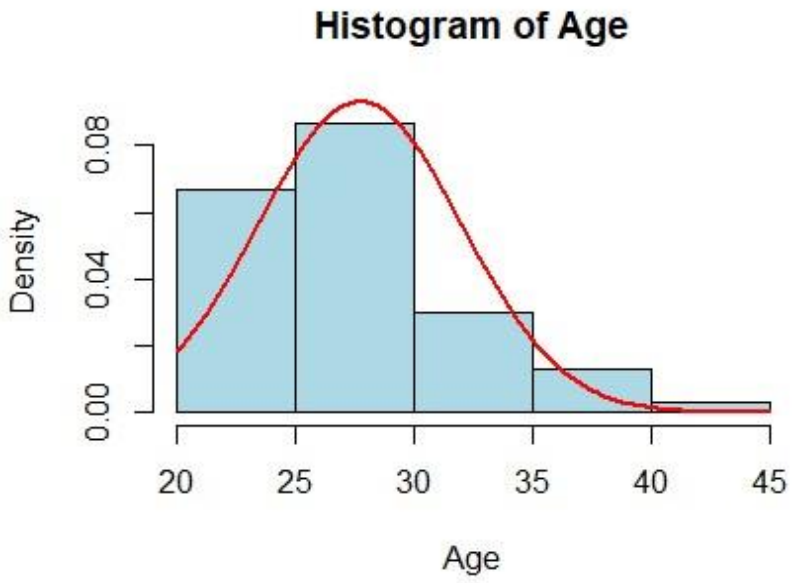

Figure 51: Histogram of Age

The above histogram shows that age in this survey data is a slightly skewed distribution. The data points are not evenly distributed around the mean. The Pearson correlation test cannot be used here as the age distribution is not normally distributed, so Spearman's rank correlation test, which does not assume normal distribution and linearity, is considered in this situation.

Spearman's Rank Correlation Equation:

Spearman's Rank Correlation Coefficient

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Figure 52: Spearman's Rank Correlation Formula

Where d_i is the difference between the ranks of corresponding values in the two variables.

- The value of ρ ranges from -1 to 1.
- ρ =1: Perfect positive monotonic correlation (as one variable increases, the other also increases monotonically).
- ρ = −1: Perfect negative monotonic correlation (as one variable increases, the other decreases monotonically).
- ρ = 0: No monotonic correlation

P-value:

- The p-value associated with Spearman's rank correlation is calculated based on the assumption that ρ follows a t-distribution with n − 2 degrees of freedom.
- The p-value indicates the probability of observing a correlation as extreme as the one computed from your sample if the accurate correlation in the population is zero.

If the p-value is less than the chosen significance level (e.g., 0.05), null hypothesis is rejected and concludes that there is evidence of a significant monotonic correlation.

Spearman's rank correlation is strong to outliers and does not assume linearity. It is beneficial when dealing with ordinal or non-normally distributed data Zach Bobbitt (n.d.).

Regression Analysis

A regression analysis analyzes the relationship between age and mental health scale. The `lm` function in R is used to perform regression analysis. The code and the results are below. The

regression analysis equation represents the relationship between the dependent variable (Y) and one or more independent variables (X1, X2,…., Xk) in a linear regression model Berman (n.d.).

Regression Analysis

$Y = \beta_0 + \beta_1 X + \varepsilon$

$Y: Dependent\ variable$

$X: Independent\ variable$

$\beta_0: Intercept(the\ value\ of\ Y\ when\ X\ is\ 0).$

$\beta_1: Coefficient\ for\ the\ independent\ variable\ X\ (the\ change\ in\ Y\ for\ a\ one\ unit\ change\ in\ X).$

$\varepsilon: Error\ term\ (captures\ unobserved\ facotrs\ influencing\ Y\ not\ included\ in\ the\ model).$

Figure 53: Regression Analysis Formula

```
asm <- lm(Scale~ Age, data = sd)
plot(sd$Age,sd$Scale, pch = 20, col = "red",
     xlab = "Age", ylab = "Mental Health Rating")
abline(asm, col = "blue")
summary(asm)
```

Figure 54: Regression Analysis Code for Age vs. Mental Health Rating

```
Call:
lm(formula = Scale ~ Age, data = sd)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3426 -1.3426 -0.0992  1.6817  3.4364

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.58212    1.66293   5.762 3.35e-07 ***
Age         -0.09737    0.05931  -1.642    0.106
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.947 on 58 degrees of freedom
Multiple R-squared:  0.04441,   Adjusted R-squared:  0.02793
F-statistic: 2.695 on 1 and 58 DF,  p-value: 0.106
```

Figure 55: Regression Analysis Results

- Intercept: The estimated intercept (β0) is 9.58212. This is the predicted "Scale" value when "Age" is 0.
- Age: The estimated coefficient for "Age" (β1) is -0.09737. This represents the estimated change in the "Scale" for a one-unit change in "Age." For the intercept: p-value < 0.001.
- For Age: p-value = 0.106 (not statistically significant at the conventional 0.05 level).
- Multiple R-squared: The proportion of the variance in the dependent variable explained by the independent variable(s). In this case, it's 0.04441, suggesting a weak relationship.
- F-statistic: A test statistic for the overall significance of the model. It tests whether at least one predictor variable has a non-zero coefficient. The p-value is 0.106, indicating that the model's overall significance is insignificant.

Analysis

- The intercept is statistically significant, suggesting that the model intercept differs significantly from zero.
- The coefficient for “Age” is not statistically significant at the 0.05 level (p-value = 0.106), indicating that there is not enough evidence to conclude that “Age” has a meaningful linear relationship with “Scale” in this model.
- The model does not fit the data well, as indicated by the low R-squared values.

- In summary, based on the current model, insufficient evidence suggests a significant linear relationship between “Age” and “Scale.” The model’s predictive power is also limited, as indicated by the R-squared values.

Scatter Plot

Scatterplots are very important in data analytics by providing a visual and intuitive method for exploring relationships, identifying patterns, and validating assumptions. Their versatility establishes them as fundamental tools for both the initial stages of data exploration and the later stages of detailed analysis and model validation Winston Chang (2018).

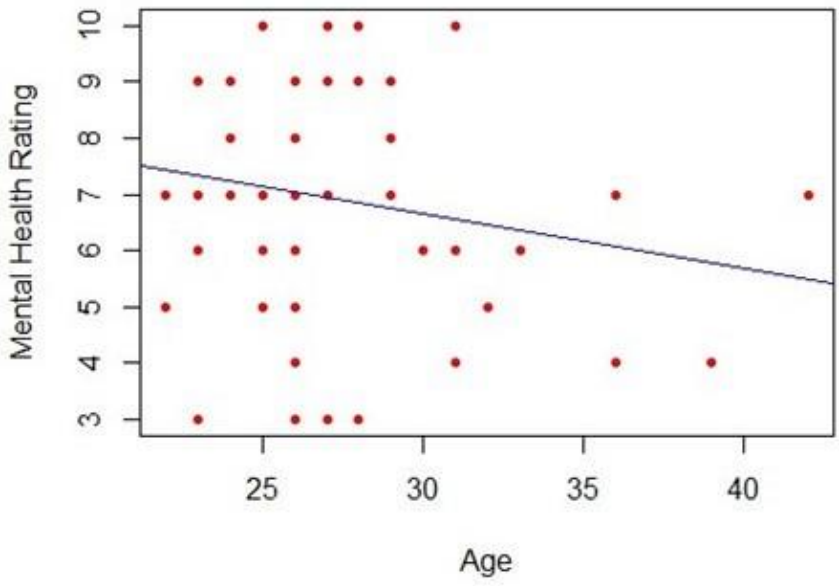

Figure 56: Age vs. Mental Health Rating Scatterplot

The above scatter plot shows a negative relationship because as age increases, the mental health scale decreases. However, this model is not statistically significant as the p-value exceeds 0.05.

Statistical Analysis

Below is the code for Spearman’s rank correlation in R.

```
> cor.test(sd$Age, sd$Scale, method = "spearman")

        Spearman's rank correlation rho

data:  sd$Age and sd$Scale
S = 42807, p-value = 0.1472
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.1894229
```

Figure 57: Spearman's rank correlation

Analysis

Spearman's Rank Correlation Coefficient (rho):

- The estimated Spearman's rank correlation coefficient (rho) is approximately -0.189. This value indicates the strength and direction of the monotonic relationship between age and the mental health scale.

Test Statistic (S):

- The test statistic (S) is 42807. It measures the strength of the monotonic relationship between the variables.

P-value (p-value):

- The p-value associated with the test is 0.1472. This p-value is more significant than the standard significance level of 0.05. Therefore, based on this p-value, there is no enough evidence to reject the null hypothesis that the accurate Spearman's rank correlation equals 0.

Based on the above result, there is insufficient evidence to reject the null hypothesis of no Spearman's rank correlation between age and the mental health scale. In other words, there is no association between participants' age and their mental health scale, according to this survey data. The Spearman's rank correlation coefficient is negative, but the p-value is more significant than 0.05, suggesting that the observed correlation may occur by chance.

Research Question 3: Disorder Type vs. Age

The third research question analyzes the correlation between disorder type and age.

Descriptive Statistics

Boxplot

Boxplot is made for Age vs. Disorder Type.

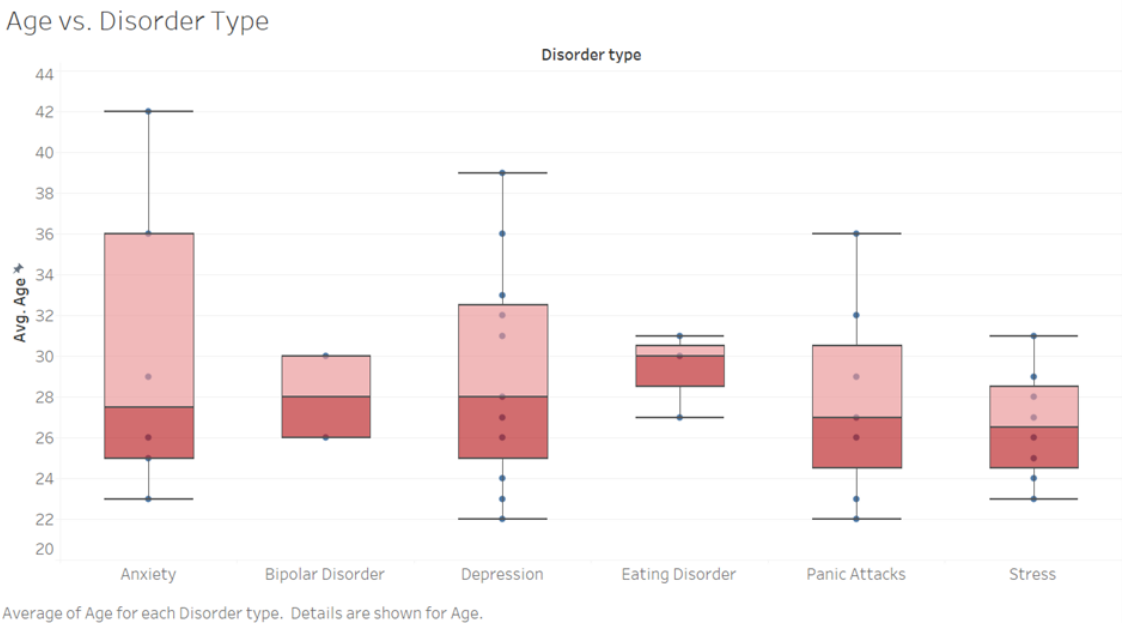

Figure 58: Age vs. Disorder Type Boxplot

Based on the findings from the presented plot and survey data:

Participants in the “Anxiety” disorder group exhibit an age distribution ranging from 23 to 42. Approximately 50% of the participants are above the age of 27, with only 25% falling within the ages of 23 and 24.

Notably, the “Bipolar Disorder” group comprises only 2 participants, aged 26 and 30, respectively.

Participants in the “Depression” disorder group show an age distribution spanning from 22 to 39. About 75% of the participants are below the age of 31, while the remaining 25% fall between the ages of 33 and 39.

The “Eating Disorder” group participants are distributed between the ages of 27 and 31. Half of the participants fall within the age range of 27 to 30, and the remaining 50% are above 31.

The “Panic Attacks” disorder group participants have an age range from 22 to 36. Approximately 75% of the participants are below the age of 31, and the remaining 25% fall between the ages of 31 and 36.

Participants in the “Stress” disorder group are distributed within the age range of 23 to 31. Fifty percent of the participants are below the age of 27, while the remaining 50% are above the age of 27.

From the survey data and the accompanying box-plot graph, it can be concluded that anxiety disorder is prevalent across various age groups in the survey. It is noteworthy that stress is observed exclusively in age groups under 31.

Analysis

ANOVA

I did ANOVA to analyze the correlation between disorder type and age. The code and the results are below.

The formulas are the same as were used for research question #1, but the variables are different: age and disorder type.

```
> d <- aov(Age ~ Disorder_type , data = sd)
> summary(d)
              Df Sum Sq Mean Sq F value Pr(>F)
Disorder_type  5  112.9   22.57   1.263  0.293
Residuals     54  965.3   17.88
```

Figure 59: ANOVA Test

- Degrees of Freedom (df): 5
- Sum of Squares (SumSq): 112.9
- Mean Square (MeanSq): 22.57

- F value (F): 1.263
- p-value (Pr(>F)): 0.293

Residuals:

- Degrees of Freedom (df): 54
- Sum of Squares (SumSq): 965.3
- Mean Square (MeanSq): 17.88

Analysis:

- The ANOVA test compares the variation between the means of different levels of the "Disorder_type" variable to the variation within each group (Residuals).
- The null hypothesis (H0) is that there is no significant difference in the means of "Age" among different levels of "Disorder_type."

The p-value for "Disorder_type" is 0.293 and failed to reject the null hypothesis since the p-value is greater than 0.05 (typical significance level).

There is not enough evidence to suggest a significant difference in the mean "Age" among different levels of "Disorder_type."

Based on this ANOVA test, there is no sufficient evidence to conclude that the mean "Age" significantly differs across different levels of "Disorder_type."

Limitations

The analysis relies on self-reported data, introducing potential biases and inaccuracies. The sample size and diversity may impact the generalization of the findings. Additional variables and longitudinal data could provide a more comprehensive understanding of mental health factors. In conclusion, the project contributes insights into the complex relationships between sleep, age, and mental health but underscores the need for further research to unravel the detailed interplay of these factors.

Conclusion

In conclusion, the project aimed to investigate three research questions related to sleep duration, mental health scale, age, and types of disorders. Here's a summary of the key findings:

Research Question 1: Sleep vs. Mental Health Scale The analysis revealed a significant association between sleep duration and mental health scale. Different sleep duration groups exhibited distinct patterns on the mental health scale. The ANOVA test indicated a statistically significant difference in the mean mental health scale among the sleep duration groups. Post-hoc analysis using Tukey's method identified specific pairwise differences between groups, highlighting substantial variations.

The findings suggest that varying sleep duration may influence mental health outcomes, emphasizing the importance of considering sleep patterns in mental health research and interventions.

Research Question 2: Age vs. Mental Health Scale The investigation into the correlation between age and mental health scale did not yield statistically significant results. The regression analysis did not support a meaningful linear relationship between age and mental health scale. Additionally, Spearman's rank correlation coefficient and associated tests did not provide evidence of a significant monotonic relationship.

While the analysis did not find a clear correlation, it underscores the complexity of factors influencing mental health beyond age alone.

Research Question 3: Disorder Type vs. Age The examination of disorder types and age revealed age distribution within each disorder group. However, the ANOVA test did not detect a significant difference in mean age across disorder types. The p-value exceeded the significance level, leading to the retention of the null hypothesis.

The lack of a significant difference in mean age across disorder types suggests that age alone may not be a decisive factor in predicting the kind of disorder.

Overall Summary: Sleep duration is significantly associated with mental health scale, indicating potential implications for mental health outcomes. Age did not show a significant linear or monotonic correlation with mental health scale, highlighting the multifaceted nature of mental health determinants. No significant difference in mean age across disorder types was observed, suggesting that age alone may not be a distinguishing factor in the prevalence of specific disorders.