The Rise of Mental Illness Mental: Data Science to the Rescue

Said Moussadeq

2024-02-09

Introduction:

Mental health is a fundamental human right affecting our thoughts, feelings, and acts. Maintaining excellent mental well-being is related to improved learning, creativity, higher levels of productivity, better social relationships, good physical health, increased life expectancy, a positive self-image, and so on. In 2019-2020, according to Mental Health America (MHA), 20,78% of adults were experiencing a mental illness, which is the equivalent of over 50 million Americans. Poor mental is related to genetic factors (nature) and environmental influences (nurture). In search of answers, data science, as an interdisciplinary field, will help make the numbers more transparent regarding the frequent causes of such dilemmas. This paper will dive deep into possible environmental contributors to mental health, such as sleep, diet, the use of substance, social media, and isolation.

Research questions:

- Is there a link between sleep quality, overall well-being, and mental health?
- To what degree can diet influence mental health?
- Could the use of substances be a significant factor in the deterioration of mental health?
- Is there a connection between social media addiction and poor mental health? To What extent does Social media hurt our mental health?
- How does social isolation affect mental health?

Approach:

Based on the collected datasets, we can use covariance and correlation statistical measures to assess relationship and dependency between multiple variables. We, also, can use covariance and correlation to calculate the strength and direction of the variables.

How your approach addresses (fully or partially) the problem.

Environmental behaviors or nurture such as sleep, diet, the use of substances, social media, and isolation do have an affect on our well-being. Knowing how much of an affect does all this factors have on our well-being will guide us towards a strong mental health.

Data (Initial data)

social media and well-being:

- This data reseach for study of the digital behavior on social media and how is affecting mental health
- year:2022
- Location: Kisii University College Faculty of Health Sciences

- Likert scale question to measure intensity of well-being
- link: https://data.mendeley.com/datasets/jxkcm7s638/1

Meaning in life: a major predictive factor for loneliness comparable to health status and social connectedness:

- The dataset includes scores for predictors of loneliness, including summary scores of sociodemographic, lifestyles, self-rated health status questionnaires and cognitive constructs about meaning in life .
- year:2020
- Location: Universitat de Barcelona Facultat de Medicina, Institut Guttmann
- link: https://data.mendeley.com/datasets/zy39mdzxpg/2

Getting More Sleep Linked to Higher Well-Being

- The dataset was meant to evaluate the relationship between sleep and well-being, using a margin of error for well-being index score for different ages.
- year:2014
- Location: telephone interviews/ all over the U.S. states
- link: https://news.gallup.com/poll/181583/getting-sleep-linked-higher.aspx

US prevalence of mental health disorders and theirassociation with co-occurring substance use disorder

- mental health & substance use disorders
- National Survey on Drug Use and Health conducted by the Substance Abuse and Mental Health Services Administration
- year:2017
- Link: https://cdn.mdedge.com/files/s3fs-public/JFP06809400.PDF

Food and mental health: relationship between food and perceived stress and depressive symptoms among university students in the United Kingdom.

- The associations between consuming 'unhealthy' foods and higher depressive symptoms and perceived stress among male and female students as well as the associations between Eating healthy foods and lower depressive symptoms
- year:2014
- The current study assessed the association between nutritional behaviour (twelve independent variables), and stress and depressive symptoms (dependent variables) in a sample from three UK countries.
- link: https://www.semanticscholar.org/paper/Food-and-mental-health%3A-relationship-between-food-Ansari-Adetunji/7089d0982e871d4024e26d04a9e00d478c418e81
- Becky Depression Inventory: https://www.researchgate.net/figure/BDI-II-cut-off-scores_tbl4_239591738

Libraries

- library(readxl)
- library(tidyverse)
- library(ppcor)
- library(readr)
- library(dplyr)
- library(GGally)
- library(ggplot2)
- library(reshape2)

Plots and Table Needs:

- Scatter plot
- Bar Chart
- Correlation Matrix/heatmap

Questions for future steps:

Mental health is a complex subject. Correlation isn't a proof of causation. Other unseen factors or data might be contributing to good or poor mental health. Further and up-to-date, independent and honest studies will be needed to deal with one of the biggest issues facing humanity. Humans, as one of the most developed machines, are at risk of losing the full functionality of their hardware and software (brain, mind and body). In ideal world, free of mental health complications, I see Data science and artificial intelligence playing a huge role in changing the way we tackle such dilemmas. I see wearable for sleep and blood sugar and diet in general, monitors in your toilets, and monitors all over the house. We are going to be well-measured and aware of our data day by day. For sure, we will also challenge ourselves and see what can be done about genetic factors.

Data Cleaning, vision and more (Final Project 2 and 3)

- Data was imported using either using: 'readxl' or 'data.frame'. Also, image to text online converter was used on the last piece of data
- Data was cleaned by filtering rows, selecting specific columns, converting data types, and renaming columns. Data was cleaned manually on excel or by using Rstudio environment.
- Image to text will be very helpful, but most of them fail to convert data types.
- The way to uncover new information in data is through exploratory data analysis (EDA). The goal is to discover patterns, relationships, anomalies, or insights that are not immediately obvious.
- Scatter plot, Bar Chart and Correlation Matrix are great match to the project.

1. Loneliness effects on Mental and Physical Health

Through the use of correlation matrix using the data below, we can better visualize and see what is the data is saying about loneliness, mental and physical health.

```r
# The dataset was loaded into RStudio using the "Import Dataset" feature available in the IDE.
df1 <- read_excel("C:/Users/TheArchitect/Desktop/data science project/loneliness_MiL_dataset.xlsx")
summary(df1)
```

```
##       age           gender          education_level    income_corrected
##  Min.   :40.00   Length:2388        Length:2388        Min.   :0.1912
##  1st Qu.:48.00   Class :character   Class :character   1st Qu.:1.9124
##  Median :54.00   Mode  :character   Mode  :character   Median :2.3550
##  Mean   :54.31                                         Mean   :2.5980
##  3rd Qu.:60.00                                         3rd Qu.:3.1873
##  Max.   :68.00                                         Max.   :7.6496
##    nutrition         cog_act        exercise             sleep
##  Min.   : 2.000   Min.   : 1.00   Length:2388        Min.   : 4.00
##  1st Qu.: 7.000   1st Qu.:11.00   Class :character   1st Qu.: 5.00
##  Median : 8.000   Median :14.00   Mode  :character   Median : 7.00
##  Mean   : 8.203   Mean   :13.53                      Mean   : 8.17
##  3rd Qu.:10.000   3rd Qu.:16.00                      3rd Qu.:10.00
##  Max.   :14.000   Max.   :23.00                      Max.   :24.00
##  household_arrang   social_interaction physical_health  mental_health
```

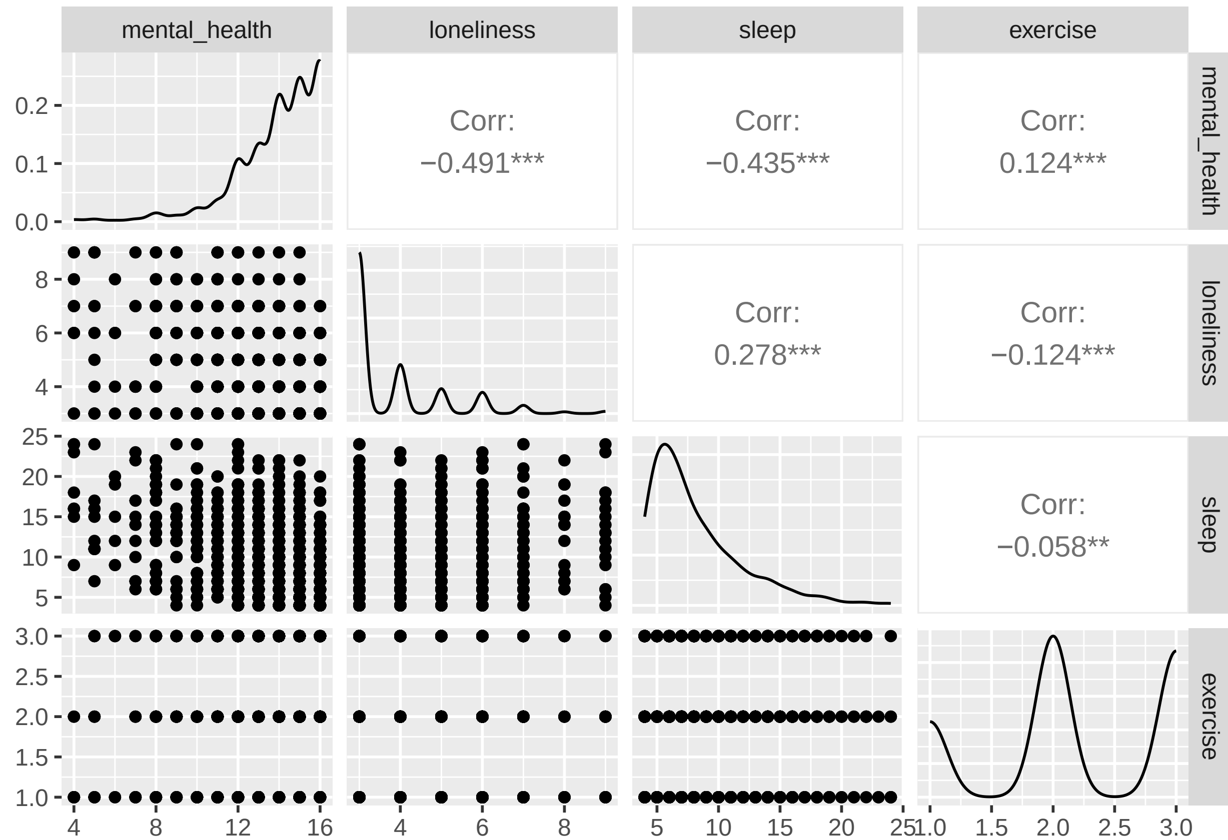
```
##  Length:2388        Min.   : 0.00      Min.   : 2.000   Min.   : 4.00
##  Class :character   1st Qu.:12.00      1st Qu.: 8.000   1st Qu.:13.00
##  Mode  :character   Median :14.00      Median : 9.000   Median :14.00
##                     Mean   :13.92      Mean   : 8.438   Mean   :13.99
##                     3rd Qu.:17.00      3rd Qu.: 9.000   3rd Qu.:16.00
##                     Max.   :20.00      Max.   :10.000   Max.   :16.00
##  cognitive_health  engagmt_life    purpose_life    sense_coherence
##  Min.   :12.00    Min.   :16.00   Min.   : 6.00   Min.   :17.00
##  1st Qu.:47.00    1st Qu.:56.00   1st Qu.:25.00   1st Qu.:59.00
##  Median :53.00    Median :61.50   Median :30.00   Median :68.00
##  Mean   :50.79    Mean   :61.37   Mean   :28.81   Mean   :66.79
##  3rd Qu.:57.00    3rd Qu.:68.00   3rd Qu.:33.00   3rd Qu.:75.00
##  Max.   :60.00    Max.   :80.00   Max.   :36.00   Max.   :91.00
##      ucla_1          ucla2           ucla3          loneliness
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :3.000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:3.000
##  Median :1.000   Median :1.000   Median :1.000   Median :3.000
##  Mean   :1.343   Mean   :1.222   Mean   :1.242   Mean   :3.807
##  3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:4.000
##  Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :9.000
##    exclusion
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.06198
##  3rd Qu.:0.00000
##  Max.   :1.00000
```

Given that variables like exclusion, ucla3, ucla2, ucla_1, sense_coherence, purpose_life, engagmt_life, household_arrang, cog_act, income_corrected, education_level, gender, and age are peripherally not connected to our study, it's crucial to retain only to those variables that are directly related to our research. For instance, factors such as age and gender, among others, do not hold significance for our investigation. This approach allows us to concentrate on variables that are more closely aligned with our research objectives.

```r
# Exclude unecessary variables
columns_to_exclude <- c("exclusion", "ucla3", "ucla2", "ucla_1",
                        "sense_coherence", "purpose_life", "engagmt_life",
                        "household_arrang", "cog_act", "income_corrected",
                        "education_level", "gender", "age")

# Drop the columns from df1
df1_filtered <- df1[, !(names(df1) %in% columns_to_exclude)]

# Display the structure of the filtered dataframe
str(df1_filtered)
```

```
## tibble [2,388 x 8] (S3: tbl_df/tbl/data.frame)
##  $ nutrition         : num [1:2388] 14 13 13 13 13 13 13 13 13 13 ...
##  $ exercise          : chr [1:2388] "Low" "High" "High" "High" ...
##  $ sleep             : num [1:2388] 12 4 6 5 8 5 4 6 7 11 ...
##  $ social_interaction: num [1:2388] 13 12 15 14 14 15 17 18 8 19 ...
##  $ physical_health   : num [1:2388] 9 9 10 9 9 8 9 9 9 10 ...
##  $ mental_health     : num [1:2388] 14 16 16 16 16 16 16 15 15 15 ...
##  $ cognitive_health  : num [1:2388] 39 60 59 59 56 58 58 46 59 59 ...
##  $ loneliness        : num [1:2388] 6 3 4 3 3 3 3 3 6 3 ...
```

```r
# Scatter Matrix plot
df1_filtered$exercise <- as.numeric(factor(df1_filtered$exercise, levels = c("Low", "Moderate", "High")))
selected_vars <- df1_filtered[c("mental_health", "loneliness", "sleep", "exercise")]
library(GGally)

scatter_plot_matrix <- ggpairs(selected_vars)

# Print the scatter plot matrix
print(scatter_plot_matrix)
```



The scatter plot matrix shows relationships between mental health, loneliness, sleep, and exercise. Here's an explanation of each correlation:

- Mental Health vs. Loneliness: The correlation coefficient is -0.491, indicating a moderate negative correlation. This suggests that as mental health scores increase (improve), loneliness scores tend to decrease.
- Mental Health vs. Sleep: The correlation coefficient is -0.435, indicating a moderate negative correlation. This implies that higher mental health scores are associated with lower sleep scores.
- Mental Health vs. Exercise: The correlation coefficient is 0.124, which is a weak positive correlation. This suggests that there is a slight tendency for individuals with higher mental health scores to have higher levels of exercise.
- Loneliness vs. Sleep: The correlation coefficient is 0.278, indicating a weak positive correlation. This suggests that individuals with higher loneliness scores may have higher sleep scores.
- Loneliness vs. Exercise: The correlation coefficient is -0.124, indicating a weak negative correlation. This suggests that individuals who report higher levels of exercise tend to have lower loneliness scores.

- Sleep vs. Exercise: The correlation coefficient is -0.058, which is a very weak negative correlation. This indicates a negligible relationship between sleep scores and exercise levels. Also, the asterisks next to the correlation coefficients denote the level of statistical significance, with more asterisks indicating a higher level of significance. In this case:
- "***" typically denotes $p < 0.001$, which is considered highly statistically significant.
- "**" might denote $p < 0.01$, which is considered statistically significant but less so than the three asterisks.

In a way or another, loneliness, sleep, and exercise are statistically related to mental health. They are all interconnected. Movement, social interactions, and good quality of sleep will all definitely affect the quality of our well-being.

2. Social media and mental health

For this piece of dataset, the focus will be only on the frequency of social media interaction and its association or effects on mental health. Through the use of histogram graph, using these data, we can better understand social media interaction and its association or effects on mental health

```r
# Moving to df2, Gender and Age variables are not of value to our research.
df2 <- read_excel("C:/Users/TheArchitect/Desktop/data science project/Digital Behavior and Mental Healt
summary(df2)
```

```
##    Respondent          Age          Gender
##  Min.   :  1.00   Min.   :21.0   Length:300
##  1st Qu.: 75.75   1st Qu.:27.0   Class :character
##  Median :150.50   Median :32.0   Mode  :character
##  Mean   :150.50   Mean   :34.2
##  3rd Qu.:225.25   3rd Qu.:42.0
##  Max.   :300.00   Max.   :55.0
##  Frequency of Social Media Interaction Self-reported Mental Health Status
##  Length:300                            Length:300
##  Class :character                      Class :character
##  Mode  :character                      Mode  :character
##
##
##
##  Impact on Mental Health (Score)
##  Min.   :2.000
##  1st Qu.:3.100
##  Median :3.800
##  Mean   :3.621
##  3rd Qu.:4.300
##  Max.   :4.800
```

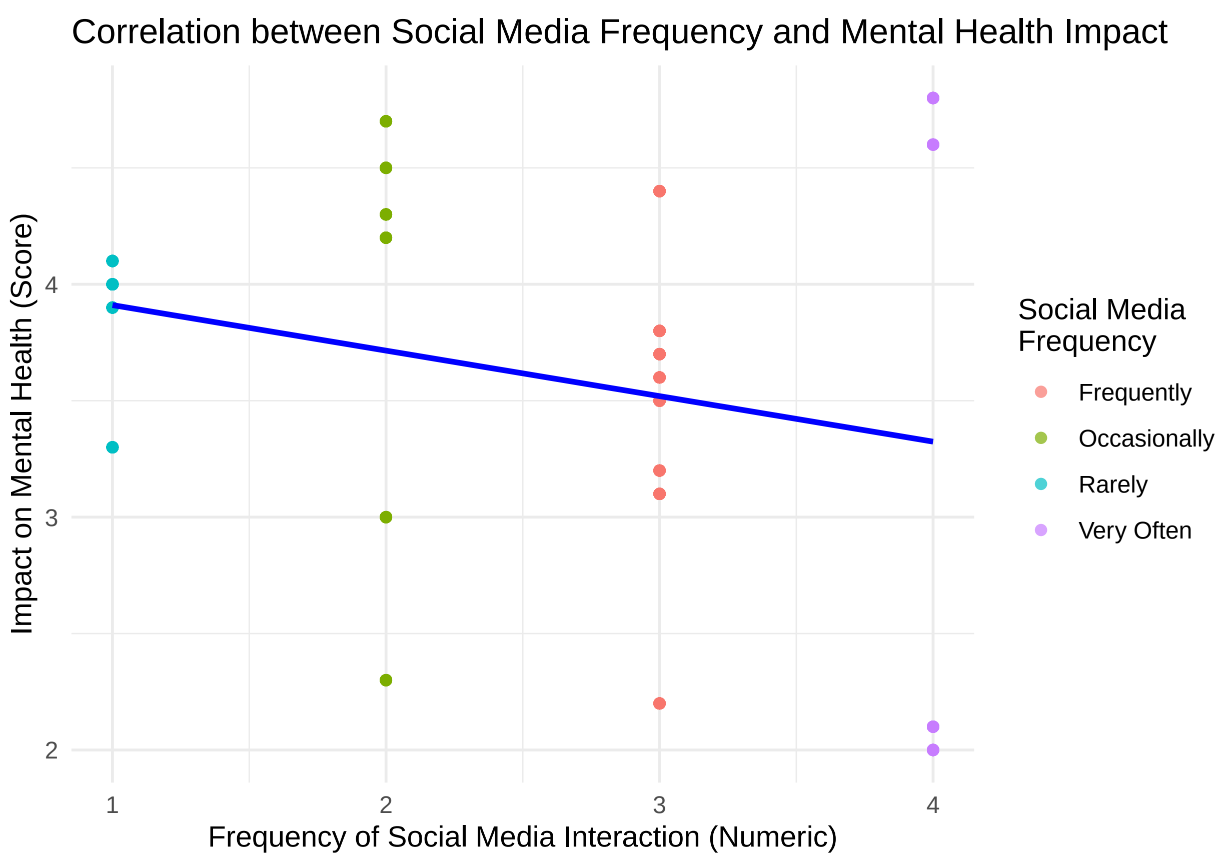
```r
# Convert "Frequency of Social Media Interaction" to numeric
# Define a mapping from categories to numeric values
# Example: "Never" = 0, "Rarely" = 1, "Occasionally" = 2, "Frequently" = 3, "Very Often" = 4
# Adjust the mapping based on the actual categories in your dataset
frequency_map <- setNames(c(0, 1, 2, 3, 4), c("Never", "Rarely", "Occasionally", "Frequently", "Very Of

# Apply the mapping to create a new numeric column
df2$Frequency_Numeric <- as.numeric(frequency_map[df2$`Frequency of Social Media Interaction`])

# Calculate the correlation
```

```r
# Ensure "Impact on Mental Health (Score)" is numeric, which it should be based on the file structure
correlation_result <- cor(df2$Frequency_Numeric, df2$`Impact on Mental Health (Score)`, use = "complete

# Print the correlation result
print(correlation_result)
```

```
## [1] -0.238329
```

```r
# Create a scatter plot with a regression line
ggplot(df2, aes(x = Frequency_Numeric, y = `Impact on Mental Health (Score)`)) +
  geom_point(aes(color = `Frequency of Social Media Interaction`), alpha = 0.7) + # Add points, color-c
  geom_smooth(method = "lm", color = "blue", se = FALSE) + # Add linear regression line
  labs(title = "Correlation between Social Media Frequency and Mental Health Impact",
       x = "Frequency of Social Media Interaction (Numeric)",
       y = "Impact on Mental Health (Score)") +
  theme_minimal() +
  scale_color_discrete(name = "Social Media\nFrequency")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Correlation, Variability and Outliers:

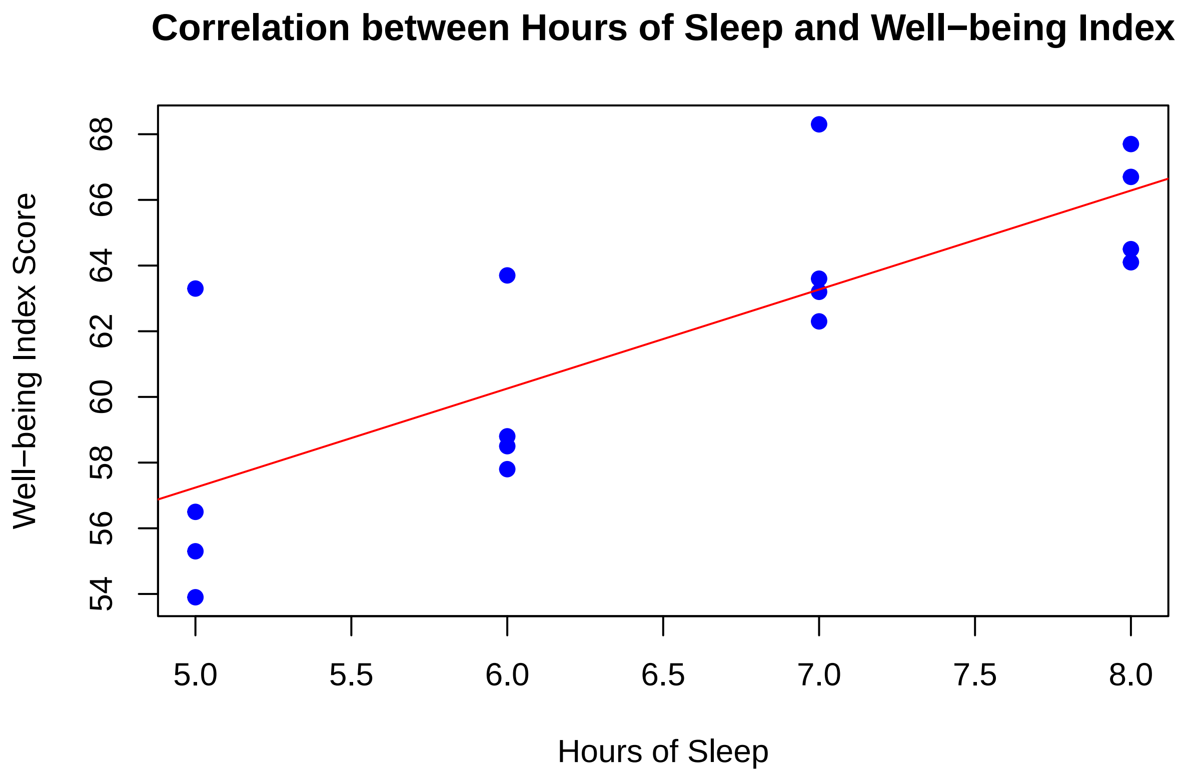
The negative slope of the trend line indicates a negative correlation between the frequency of social media use and mental health impact score. This suggests that as people use social media more frequently, their mental health impact score tends to decrease. The spread of data points around the trend line indicates variability in the impact of social media on mental health. For example, even among individuals who use social media 'Very Often', there are varying levels of reported impact on mental health. There appear to be some potential outliers, particularly at the ‘Rarely’ and ‘Frequently’ levels, where some individuals have much higher mental health impact scores compared to others in the same category.

3. Getting More Sleep Linked to Higher Well-Being

The following data was imported manually, using the ‘data.frame’ function, as the data is not too long and image to excel option, using tesseract library (optical character recognition) didn’t go as desired. In order to reveal any correlation, a scatter plot or a bubble plot to compare hours of sleep and well-being index scores.

```
# Creating the data frame with well-being index scores and hours of sleep
df3 <- data.frame(
  Age = c(18, 30, 45, 65),
  Five_Hours = c(56.5, 53.9, 55.3, 63.3),
  Six_Hours = c(58.5, 57.8, 58.8, 63.7),
  Seven_Hours = c(62.3, 63.2, 63.6, 68.3),
  Eight_Hours = c(66.7, 64.1, 64.5, 67.7)
)
```

```
# Calculate the average sleep across all hours columns
df3$Average_Sleep <- rowMeans(df3[, 2:5])
# Convert the wide format data to long format for easier plotting
df3_long <- melt(df3, id.vars = "Age", variable.name = "Hours_of_Sleep", value.name = "Wellbeing_Index")

# Convert the 'Hours_of_Sleep' from factor to numeric for plotting purposes
# We need to extract the numeric part of the factor levels, assuming they are like "Five_Hours", "Six_H
hours_mapping <- c(Five_Hours = 5, Six_Hours = 6, Seven_Hours = 7, Eight_Hours = 8)
df3_long$Hours_of_Sleep <- hours_mapping[as.character(df3_long$Hours_of_Sleep)]

# Ensure all data is finite
df3_long <- df3_long[is.finite(df3_long$Hours_of_Sleep) & is.finite(df3_long$Wellbeing_Index), ]

# Plotting the scatter plot
plot(df3_long$Hours_of_Sleep, df3_long$Wellbeing_Index, main = "Correlation between Hours of Sleep and W

# Adding the regression line
abline(lm(Wellbeing_Index ~ Hours_of_Sleep, data = df3_long), col = "red")
```



Strong positive correlation: The upward slope suggests a positive correlation between hours of sleep and well-being index scores. This implies that as the number of hours of sleep increases, the well-being index score tends to increase as well. The closeness of the points to the regression line indicates the strength of the correlation. This means that sleep duration is a good predictor of well-being index.

4. US prevalence of mental health disorders and their association with co-occurring substance use disorder

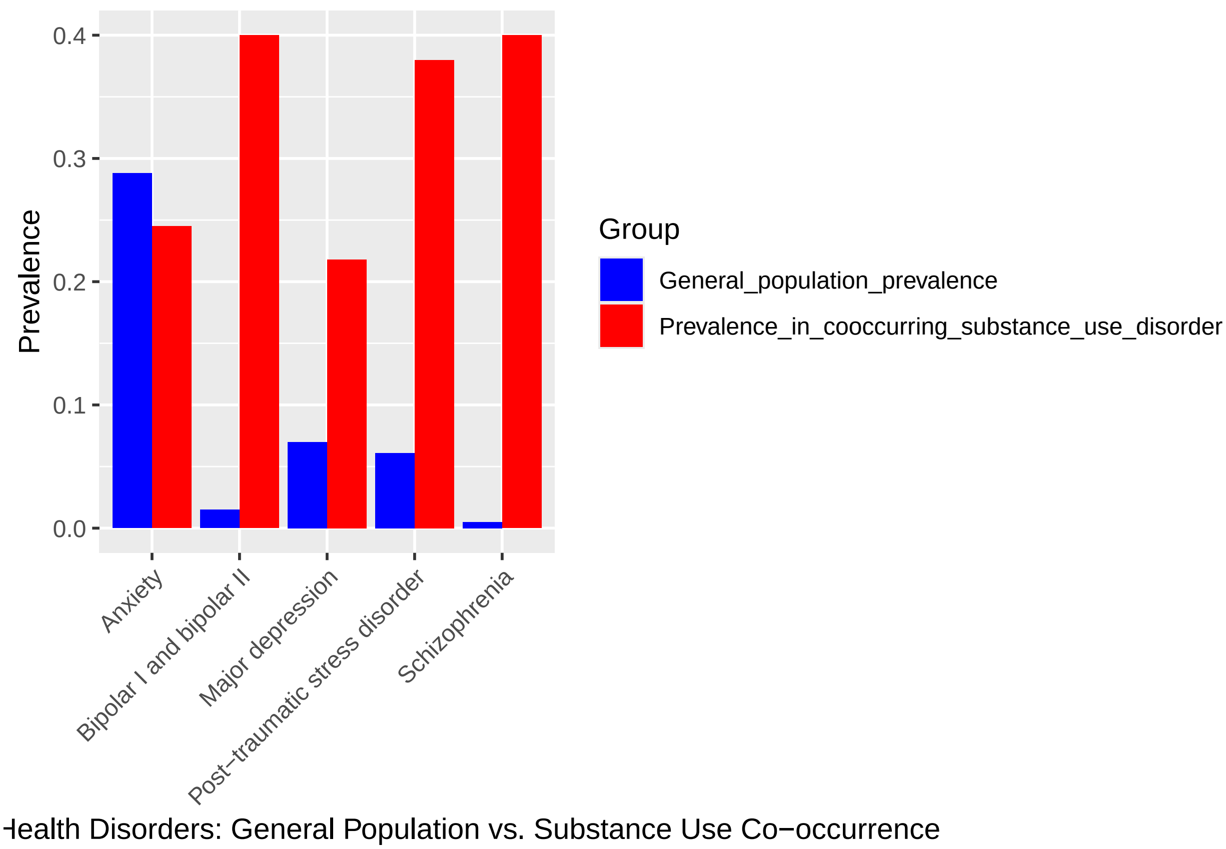
The next data is intended to explore the relationship between mental health disorders and substance use disorder. To detect the prevalence of mental health disorders and substance use disorder and compare them, we can create a chart that visualizes both the general population prevalence and the prevalence in individuals with co-occurring substance use disorder for each mental health disorder listed. A good choice for this comparison would be a bar chart, where we can have side-by-side bars for each disorder showing the two different prevalence rates.

```
# load the data
df4 <- read_excel("C:/Users/TheArchitect/Desktop/data science project/data.xlsx")
head(df4)
```

```
## # A tibble: 5 x 3
##   `Mental health disorder`       General population pre~1 Prevalence in co-occ~2
##   <chr>                                             <dbl>                  <dbl>
## 1 Anxiety                                           0.288                  0.245
## 2 Major depression                                  0.07                   0.218
## 3 Schizophrenia                                     0.005                  0.4
## 4 Bipolar I and bipolar II                          0.015                  0.4
## 5 Post-traumatic stress disorder                    0.061                  0.38
```

```
## # i abbreviated names: 1: `General population prevalence`,
## #   2: `Prevalence in co-occurring\r\nsubstance use disorder`
```

```r
# Create the dataframe
df5 <- data.frame(
  Mental_health_disorder = c("Anxiety", "Major depression", "Schizophrenia", "Bipolar I and bipolar II"
  General_population_prevalence = c(0.288, 0.070, 0.005, 0.015, 0.061),
  Prevalence_in_cooccurring_substance_use_disorder = c(0.245, 0.218, 0.400, 0.400, 0.380)
)

# Convert dataframe to long format to use with ggplot
df5_long <- reshape2::melt(df5, id.vars = "Mental_health_disorder")

# Create the bar chart
ggplot(df5_long, aes(x = Mental_health_disorder, y = value, fill = variable)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_manual(values = c("General_population_prevalence" = "blue", "Prevalence_in_cooccurring_subs
  labs(x = "Prevalence of Mental Health Disorders: General Population vs. Substance Use Co-occurrence",
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x labels for readability
```



Observations:

- Anxiety: The prevalence of anxiety is high in the general population but slightly lower in the substance use disorder group.
- Bipolar Disorder and Major Depression: The prevalence of bipolar disorder and major depression is significantly higher in the substance use disorder group compared to the general population.
- Post-traumatic Stress Disorder (PTSD): PTSD also has a higher prevalence among those with co-

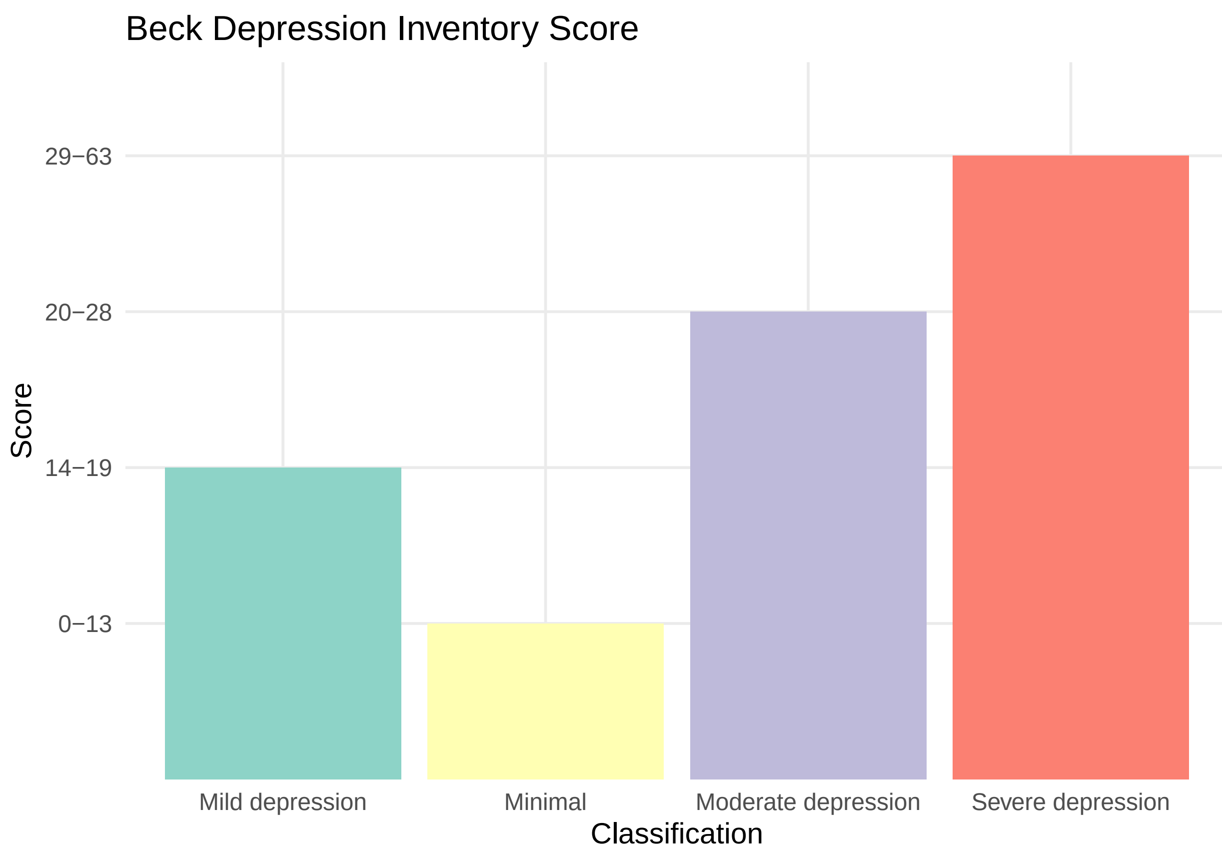
occurring substance use disorders.
- Schizophrenia: This disorder has the highest relative increase in prevalence when comparing the general population to the substance use disorder group.

These observations suggest that substance use disorders are associated with an increased prevalence of certain mental health conditions.

5. Food and mental health: relationship between food and perceived stress and depressive symptoms

Through the use of bar chart, and based on Beck Depression Inventory score, the goal is to explore The associations between consuming 'unhealthy' foods and higher depressive symptoms and perceived stress among male and female students as well as the associations between Eating healthy foods and lower depressive symptoms suggest that interventions to reduce depressive Symptoms and stress among students could also result in the consumption of healthier foods and/or vice versa.

```r
# Data from the provided classification and scores
scores <- c('0-13', '14-19', '20-28', '29-63')
classifications <- c('Minimal', 'Mild depression', 'Moderate depression', 'Severe depression')

# Create a data frame
depression_df <- data.frame(Classification = classifications, Score = scores)

# Plotting a colorful bar chart
ggplot(depression_df, aes(x = Classification, y = Score, fill = Classification)) +
  geom_bar(stat = 'identity') +
  scale_fill_brewer(palette = "Set3") +  # This will use a predefined color palette for distinct colors
  labs(title = 'Beck Depression Inventory Score', x = 'Classification', y = 'Score') +
  theme_minimal() +
  theme(legend.position = "none")  # Remove the legend
```



The Beck Depression Inventory (BDI) is a self-report questionnaire that is widely used to assess the severity of depression in individuals. It was created by Dr. Aaron T. Beck and consists of multiple statements related to symptoms of depression such as mood, pessimism, sense of failure, self-dissatisfaction, guilt, punishment, self-dislike, self-accusations, suicidal ideas, crying, irritability, social withdrawal, indecisiveness, body image change, work difficulty, insomnia, fatigue, appetite, weight loss, and somatic preoccupation. Respondents are asked to reflect on each statement and select the one that best describes their experience over the past two weeks. The BDI scores can be classified into ranges(as the chart shows) that correspond to minimal(0-13), mild(14-19), moderate(20-28), and severe levels of depression(29-63).

Bar Chart of Depressive Symptoms Score by Food Group

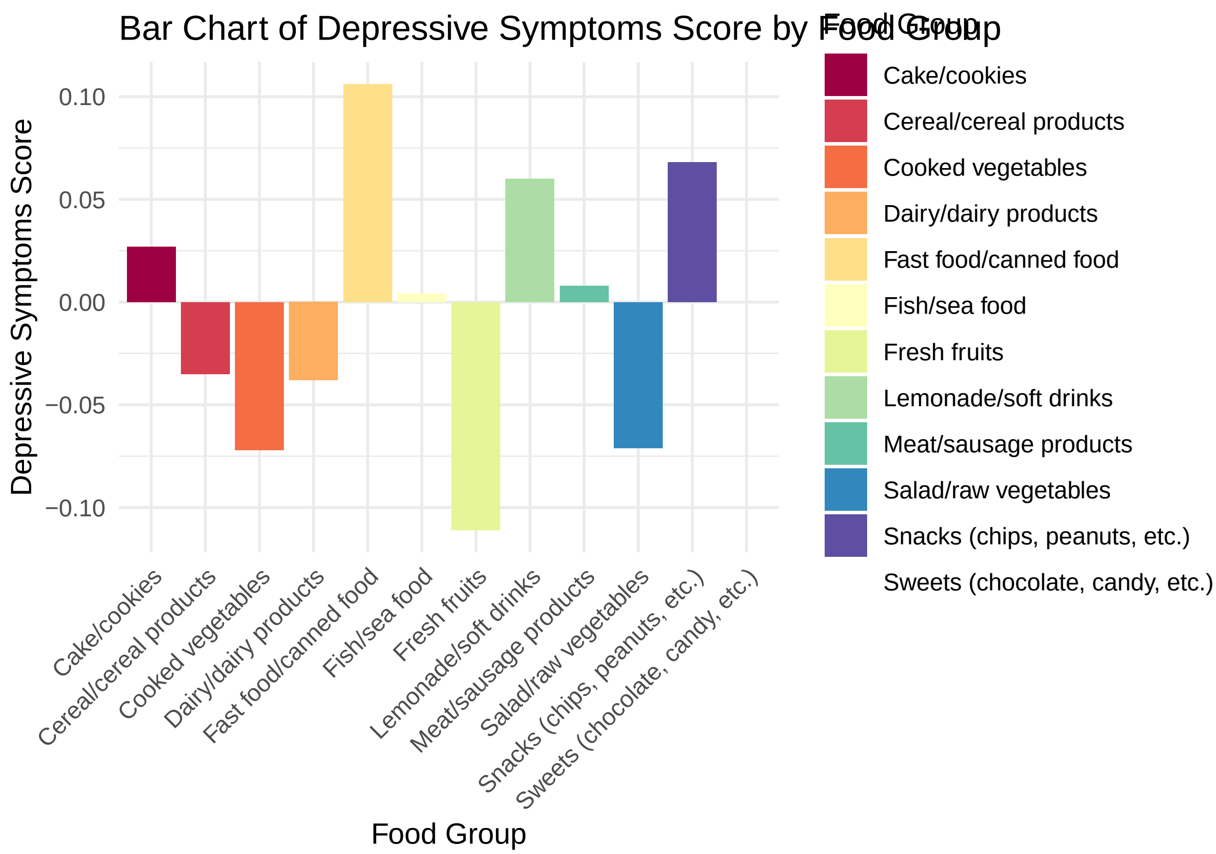
```r
# after loading the data using image to excel converter(nanonets.com)and manually cleaning and skipping
df5 <- read_excel("C:/Users/TheArchitect/Desktop/data science project/df666.xlsx")
summary(df5)
```

```
##   Food Group         Perceived Stress Depressive Symptoms Score
##  Length:12           Min.   :0.0010   Min.   :-0.11100
##  Class :character    1st Qu.:0.0010   1st Qu.:-0.04625
##  Mode  :character    Median :0.0020   Median : 0.00600
##                      Mean   :0.1033   Mean   : 0.00225
##                      3rd Qu.:0.1422   3rd Qu.: 0.06200
##                      Max.   :0.6650   Max.   : 0.10600
```

```r
# Create a long format of the data for ggplot2
df5_long <- reshape2::melt(df5, id.vars = 'Food Group')

# Create a bar chart
```

```r
ggplot(df5, aes(x = `Food Group`, y = `Depressive Symptoms Score`, fill = `Food Group`)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_brewer(palette = "Spectral") + # This adds color to the bars
  theme_minimal() +
  labs(title = "Bar Chart of Depressive Symptoms Score by Food Group",
       x = "Food Group",
       y = "Depressive Symptoms Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # This rotates the x labels for better reada
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Spectral is 11
## Returning the palette you asked for with that many colors
```



Higher Depressive Symptoms Scores vs Lower Depressive Symptoms Scores:

- Higher Depressive Symptoms Scores: Food groups that have bars extending above the horizontal axis are associated with higher depressive symptoms scores. For instance, "Cake/cookies" and "Fast food/canned food" show positive associations, meaning that higher consumption of these foods may correlate with higher levels of reported depressive symptoms.
- Lower Depressive Symptoms Scores: Conversely, food groups with bars extending below the horizontal axis, such as "Fresh fruits" and "Salad/raw vegetables", show negative scores, which could suggest that their consumption is associated with lower levels of depressive symptoms.

Final Thoughts:

Analysis and Insights

Our findings indicate:

- A moderate negative correlation between mental health and loneliness, suggesting that better mental health is associated with less loneliness and vice versa.
- A moderate negative correlation between mental health and sleep quality, implying that improved mental health accompanies better sleep.
- A weak positive correlation between mental health and exercise, hinting at a slight benefit of exercise on mental health.
- Individuals with substance use disorders experience significantly higher rates of various mental health conditions, including major depression, bipolar disorder, PTSD, and especially schizophrenia, compared to the general population.
- Consumption of foods like cake, cookies, and fast food is correlated with higher depressive symptom scores, while eating fresh fruits and salads is associated with lower depressive symptom scores.

Implications for the Audience

While some correlations between lifestyle factors and mental health may seem weak initially, their consistent impact over time shouldn't be underestimated. Focusing on small, sustainable lifestyle changes can have a compounding positive effect on our mental wellbeing. Small, consistent improvements in diet, sleep, social behaviors, or exercise can lead to major health gains.

Limitations and Future Work

Correlation does not imply causation, and other unmeasured factors may influence mental health. Future studies should include longitudinal data to establish causality and explore genetic factors. The evolution of data science and AI, including wearables and home monitors, may revolutionize our understanding and management of mental health.

Conclusion

Data science can be a powerful tool to fight the mental health crisis. By effectively analyzing massive amounts of data on opioids(which was not covered, but there are robust findings indicating its effects on worsening mental health status), sleep, exercise, loneliness, substance use, and diet, we can find hidden patterns and the actual impact causes of mental well-being. This knowledge will empower us to build better, targeted solutions for individuals and change how we approach mental health as a human beings.