Stressed, Scrolling, and Sleep-Deprived: What are Teens Telling Us?

by Howard Lin, Yasthil Singh and Luca Cannis

University of Michigan | Milestone I Project

Background & Motivation

In today's digital age, teens are more immersed in social media than ever before, especially post-COVID. Recent research suggests that more teens are view social media as having a harmful effects on people their age, with concerns rising about overuse and negative impacts on confidence, sleep, and overall well-being.

It is therefore critical to understand how social media use effects the mental health of teenagers while also taking into account the role that behavioral factors have on mental health outcomes.

Objectives

This project explores the relationship between social media frequency, sleep hours, and exercise hours in relation to reports of mental suffering in teenagers aged between 13 to 17. We examine trends across the same age, gender and sleep hour groups to explore how each variable can effect each other and ultimately their effect on reports of mental suffering.

While we don't aim to prove causation, our goal is to highlight clear, actionable patterns through data and visual storytelling.

Outcomes

Insights can support school counselors, parents, and policymakers in understanding how social media use and behavioral factors can influence mental health in teenagers. These insights can inform targeted programs that encourage improve mental health outcomes

We focus on key behavioral indicators:

- Social media usage frequency
- Sleep duration
- Stress levels
- Physical activity

Important Observations:

- As teenagers use social media more than one a day - reports of mental suffering significantly increases compared to less than daily social media use.
- As teenagers sleep less reports of mental suffering increase.

Key Questions:

- Do we see age-gender groups with higher physical activity report lower rates of sadness?
- Is there a relationship between social media use and exercise?
- How does social media usage affect sleep quality?
- How average sleep hours effect reported feelings of mental suffering across age-gender-sleep groups?
- Does frequent social media use significantly increase the likelihood of teenagers reporting feelings of sadness or hopelessness?

The result can support the design of future awareness campaigns or digital wellness initiatives, and be integrated into predictive models to better understand which behaviors most significantly impact mental health and sleep outcomes in adolescents.

Executive Summary

Findings

- **Physical Activity & Mental Health**

Teens with higher physical activity levels, report slightly fewer symptoms of sadness or depression.
→ Encouraging regular exercise may support emotional well-being.

- **Sleep & Social media**

Groups with lower average sleep hours tend to show more frequent use of social media and mental health concerns.
→ Reinforces the need to promote healthier sleep routines.
→ Calls students to reduce screen time and support student mental health.

Limitations

- **Self-Reported Data Bias**

Both datasets rely on self-reported responses, which may be affected by memory errors, social desirability bias, or misinterpretation of questions.

- **Data Editing in YRBS**

YRBS applies internal rules to handle conflicting responses, some data is marked as missing, reducing completeness for certain variables.

- **Limited Access to Supplementary Data**

Public datasets on teens (e.g., school or government data) often require formal access, limiting opportunities to cross-reference or enrich the analysis.

- **Feature Engineering Required**

Some variables were derived to match the Kaggle dataset's structure, as not all metrics were directly available in the YRBS dataset.

Recommendations

- **Expand Survey Variables**

Add questions on school performance, family environment, substance use, abuse, bullying, peer dynamics, and access to mental health services for deeper, more holistic insights.

- **Incorporate Additional Groupings**

Use more survey dimensions (e.g., region, household size, academic stress) to explore subgroup patterns.

- **Build Interactive Dashboards**

Share findings visually with educators and mental health professionals to support intervention design.

- **Compare Pre- and Post-Pandemic Trends**

Analyze changes in teen behavior and well-being over time to better understand long-term impacts of COVID-19.

Data Sources: What are the datasets? What are the variables we are using?

Youth Risk Behavior Surveillance System (YRBSS) - 2023 Survey

The 2023 Youth Risk Behavior Survey (YRBS), conducted by the CDC, is designed to monitor health risk behaviours and experiences among high school students in the United States. This dataset contains **self-reported information** about the daily habits and mental health of **20,103 students** surveys. It covers a wide range of topics including physical activity, substance use, diet, and violence. The survey ensures a representative sample of U.S. high school students by randomly selecting schools and students from both public and private schools across all regions in the country. The data collected helps inform public health policy and monitor trends in youth health behaviours over time.

From this survey, 3 key variables are relevant to our project:

1. **Social_media_daily(Q80)**: Captured through questions about the amount of time students spend using electronic devices for non-school purposes.
2. **Sleep_Hours(Q85)**: Assessed through questions on how many hours of sleep students usually get on school nights.
3. **Mental_suffering(Q26)**: Reflected in questions related to mental health, such as feelings of sadness or hopelessness, and experiences with bullying, violence, and social support.

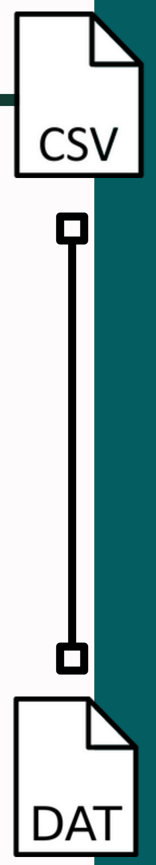
Mental Heath Analysis Among Teenagers dataset - 2025

This dataset includes information from 5,000 teenagers and looks at different parts of their daily lives, such as social media use, physical activity, sleep habits, stress levels, and academic performance. The data comes from surveys, wearable devices, and online activity.

The goal is to find patterns between screen time and teen well-being, especially how it relates to stress, sleep, and school performance.

From this dataset, here are the key variables from the dataset that correspond to the analysis goals and align with the 2023 YRBS survey variables:

- **Social_Media_Hours**
  - Self-reported total hours spent on screens (e.g., phones, computers, TVs) per day.
- **Sleep_Hours**
  - Self-reported average sleep duration per night.
- **Exercise_Hours**
  - Self-reported exercise hour.



🧠 Mental Health Dataset

A synthetic dataset designed to simulate mental health patterns in teenagers, focusing on stress levels and their potential links to lifestyle behaviors such as sleep, social media use, screen time, and exercise. The data was generated for research and educational purposes.

- 🔗 Source: Kaggle - Mental Health Analysis Among Teenagers
- Accessed on: 2024-05-26
- No. of Records: 5,000
- No. of Attributes: 11
- Format: CSV (synthetic, structured tabular)

🧠 XXH2023_YRBS_Data.dat

A nationally representative dataset collected by the CDC, the YRBS tracks adolescent health behaviors across the U.S., including topics such as mental health (e.g. prolonged sadness), social media use, sleep duration, exercise frequency, and demographic factors. This dataset is used as a benchmark for public health analysis.

- 🔗 Source: CDC YRBS Official Data
- Accessed on: 2024-05-26
- No. of Records: ~19,000+
- No. of Attributes Used in Project: ~5 (Age, Sex, Sleep Hours, Q26, Q80)
- Format: Fixed Width File (.dat) + Codebook PDF
- Parsing Method: Python (pandas, read_fwf), SAS script for column mapping

Data Manipulation

Step	Mental Health Dataset	CDC YRBS Dataset
CREATE	Initial cleaning and standardization steps to make the dataset easier to work with. • Renamed all column names to a standard format (lowercase, no spaces). • Renamed gender column to “sex”, and converted values: • 'Male' → 'M', 'Female' → 'F' for consistency. • Ensured screen_time_hours was always greater than or equal to social_media_hours (sanity check for logical consistency)	Raw parsing and initial column transformations: • Renamed Q2 to sex, then mapped values: 1 → F, 2 → M • Used CDC codebook mapping to convert Q1 codes to actual Age values (e.g., 2.0 → 13) • Renamed Q1 to age and converted to integer format • Renamed key survey question columns for clarity: ◦ Q85 → sleep_hours ◦ Q26 → mental_suffering? ◦ Q80 → social_media_daily?
UPDATE	Transformation and formatting for analysis-readiness: • Renamed key columns to snake_case format (age, sleep_hours, etc.) to match the CDC dataset schema. • Rounded sleep_hours to the nearest integer (with Int64 dtype to preserve NaNs if any). • Validated value ranges: clipped or removed rows where screen_time_hours, sleep_hours, or exercise_hours exceeded 24 hours.	Grouped values, binarized responses, and standardized formats: • Defines a dictionary to translate CDC's coded responses into actual sleep hours. (Q85) • Grouped social media frequency (Q80) into binary flags: 1 (numeric) (responses 5–8) vs. 0(numeric) (responses 1–4) • Mapped mental health answers to binary numeric format 1/0
DELETE	Dropped incomplete or invalid records: • Dropped rows with missing values in critical columns like sleep_Hours and survey_stress_score. • Filtered dataset to only include teens aged 13–17. • Removed exact duplicates to avoid bias in aggregation.	Filtered dataset to ensure valid, focused data: • Filtered rows to only include teens aged 13–17 • (Assumed) dropped any rows with missing/invalid Sleep_Hours or Age during aggregation
Final Shape	Summary of your cleaned and structured dataset: • ~5,000 records with cleaned features (e.g., age, sex, sleep_hours, screen_time_hours, exercise_hours) • Dataset now ready to be grouped by age/sex/sleep_hours and merged with the CDC dataset for comparative analysis.	Prepared dataset for merging/grouping: • Original ~19,000 raw records • Aggregated into grouped structure by age, sex, and sleep_hours for merge and comparison with the mental health dataset

Exploratory Insights - Analysis

Sleep data observed in hours was combined from both the Mental Health Dataset from Kaggle (sourced through wearable devices) and the CDC dataset (collected via survey responses). While these datasets were joined from different sources, they did not produce conflicting results. Instead, the relationship between sleep and other variables like exercise and social media use remained consistent with our hypothesis, reinforcing the validity of the observed trends.

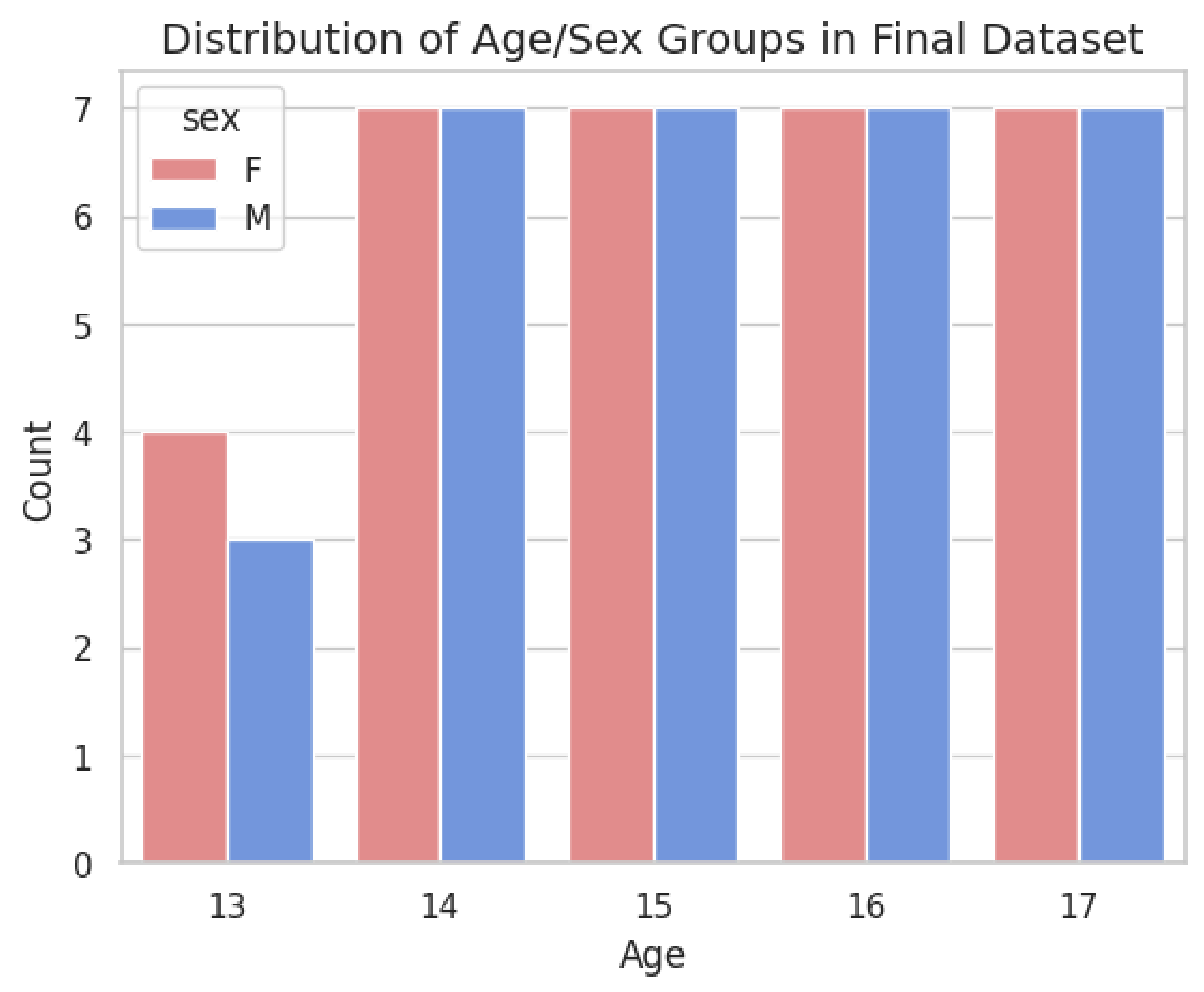
Key Variables Observed:

- Age
- Sex
- Sleep Hours
    - Exercise Time (avg hours)
    - Social Media Use (avg hours + daily count)
    - Screen Time
    - Stress Score
    - Mental Suffering Count

# of Groups

# of Groups Male - (13-17)	32
# of Groups Female - (13-17)	31

The age 13 group has fewer entries (F: 4, M: 3) compared to other age groups (F/M: 7 each), indicating a potential gap in data coverage for the youngest participants.



Interpretation of Heatmap:

Important Positive Correlations:

1. social_media_daily_count & mental_suffering_count: **0.85**
2. avg_exercise_hours & sleep_hours: **0.25**
3. age & mental_suffering_count: **0.41**

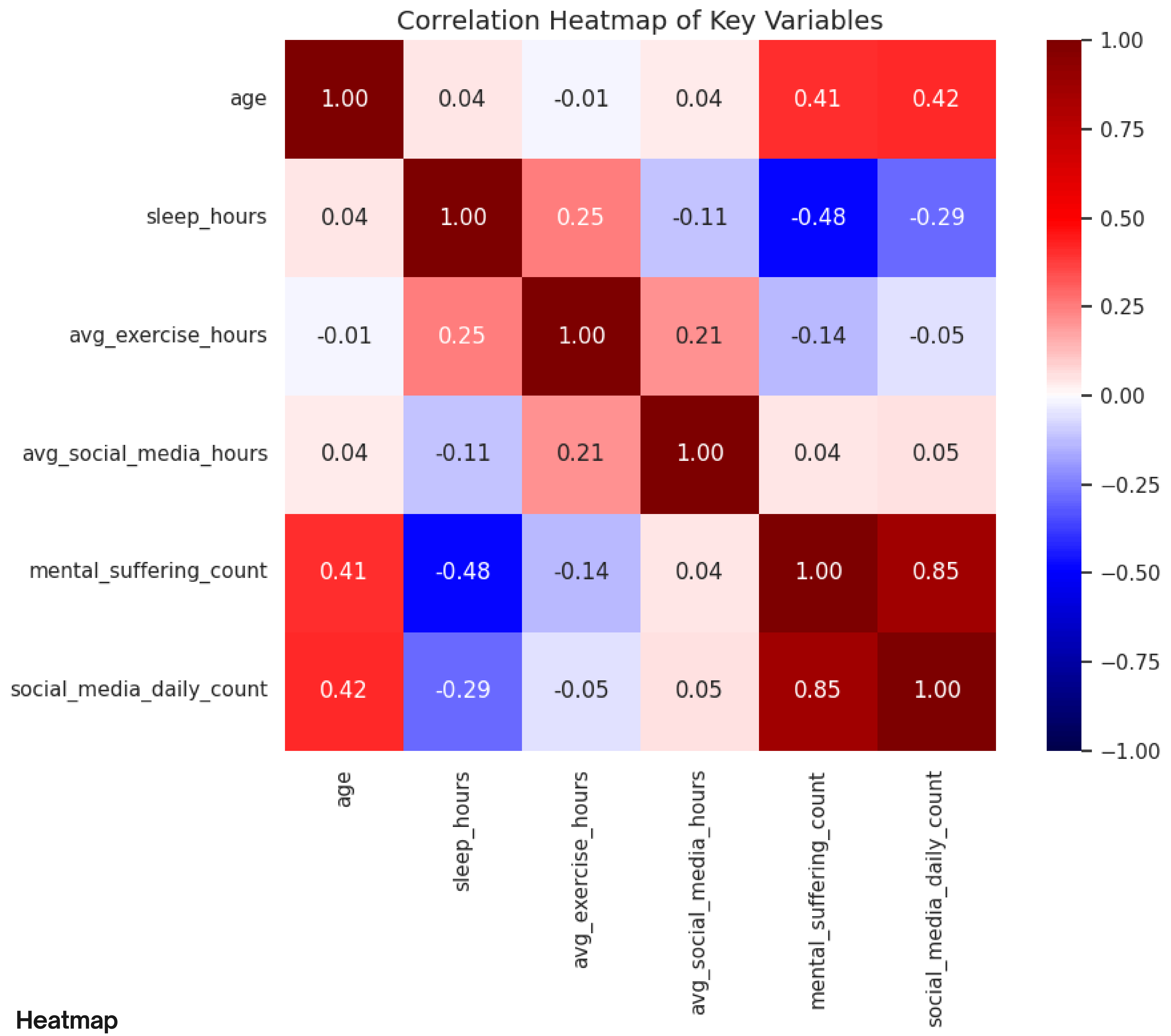
Important Negative Correlations:

1. sleep_hours & social_media_daily_count count: **-0.29**
2. sleep_hours & mental_suffering_count: **-0.48**

Next Steps:

Exploration of the important positive and negative relationships will be show on the following slides:

- Slide 7: Analysis of Social Media
- Slide 8: Analysis of Sleep
- Slide 9: Analysis of Exercise



Heatmap

Analysis: Social Media

What is the relationship between social media use and changes in reports of mental suffering ?

Visual 1: Key Insights

The positive slope of the red line indicates a strong positive linear relationship between higher daily social media usage and increased reports of mental suffering.

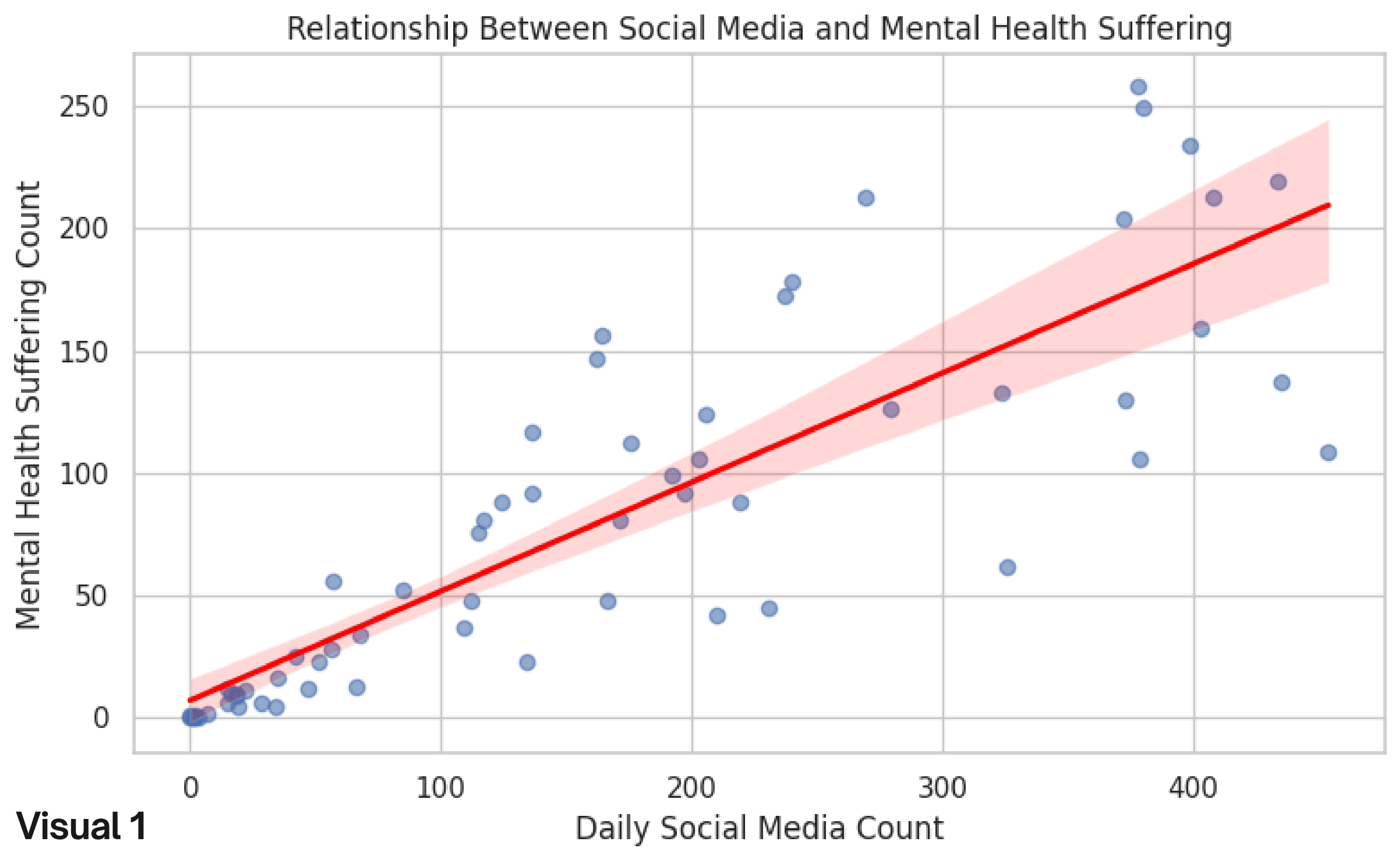
This suggests that as daily social media usage rises, so do mental health struggles.

The tight confidence interval around the regression line shows the relationship is likely consistent and statistically meaningful.

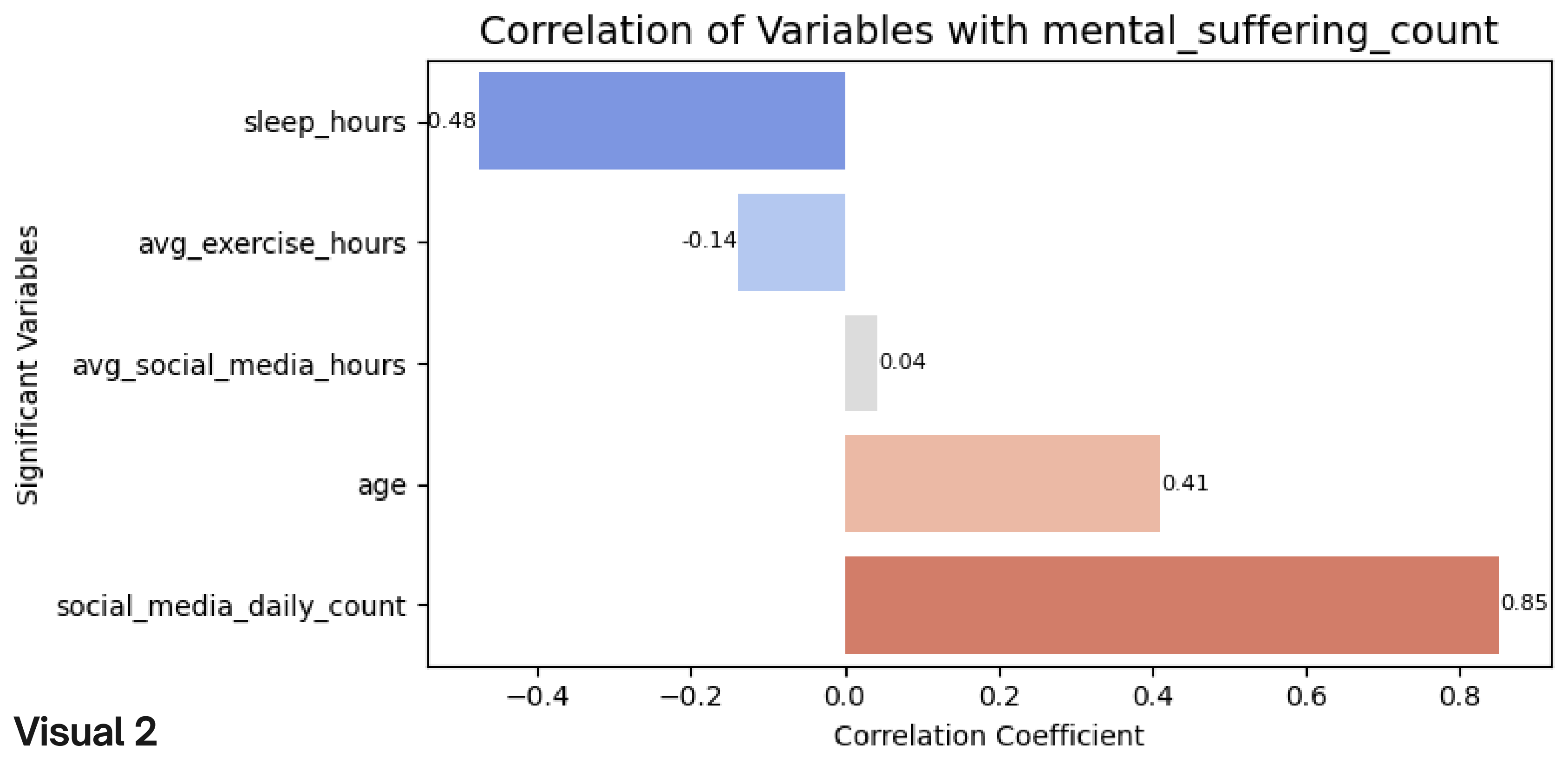
Visual 2: Key Insights

- Social media daily count is the strongest predictor of mental suffering.
- Sleep hours have a meaningful negative relationship — less sleep = more mental suffering.
- Exercise has a small protective effect.

Average hours on social media alone are not as telling — perhaps due to self-reporting inaccuracies.



Visual 1



Visual 2

Analysis: Sleep

Sleep & Mental Health

- When teens sleep only 4 to 7 hours a night, they report higher levels of mental suffering.
- As sleep increases to 8, 9, or 10 hours, reports of mental suffering go down a lot.
- Visual 3 also shows that girls report more mental suffering than boys at every sleep level.

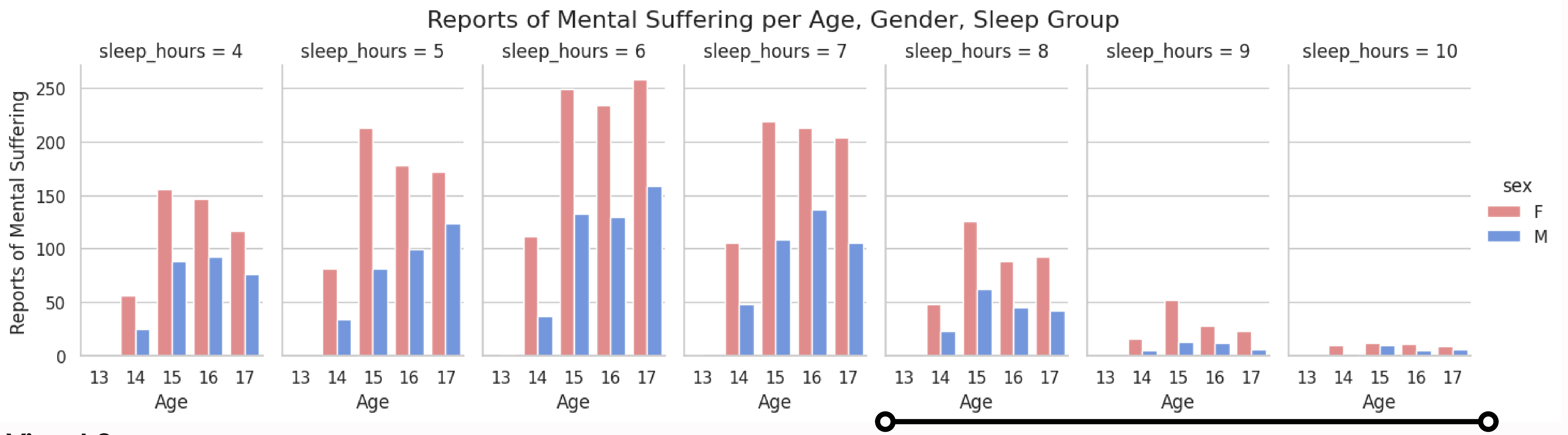
This suggests that getting more sleep helps reduce symptoms of mental health problems.

Sleep & Social Media Use:

- Teens who sleep around 6 to 7 hours report the highest social media use, especially those aged 15 to 17.
- As sleep increases to 8, 9, or 10 hours, social media use goes down a lot. Teens who sleep 9 or 10 hours use social media the least.
- Girls and boys have similar patterns, but girls often use social media a little more.

Overall, the graph shows that teens who get more sleep tend to spend less time on social media.

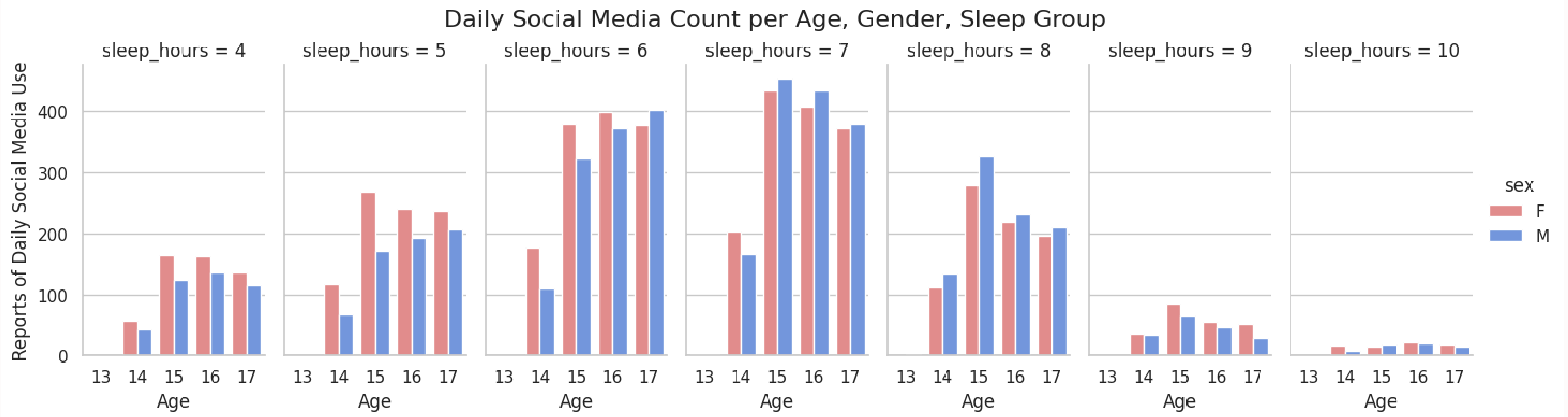
How average sleep hours effect reported feelings of mental suffering?



Visual 3

8 hours or more shows a big difference in teen mental health.

How does social media usage affect sleep quality?



Visual 4

Analysis: Exercise

Exercise and Mental Health:

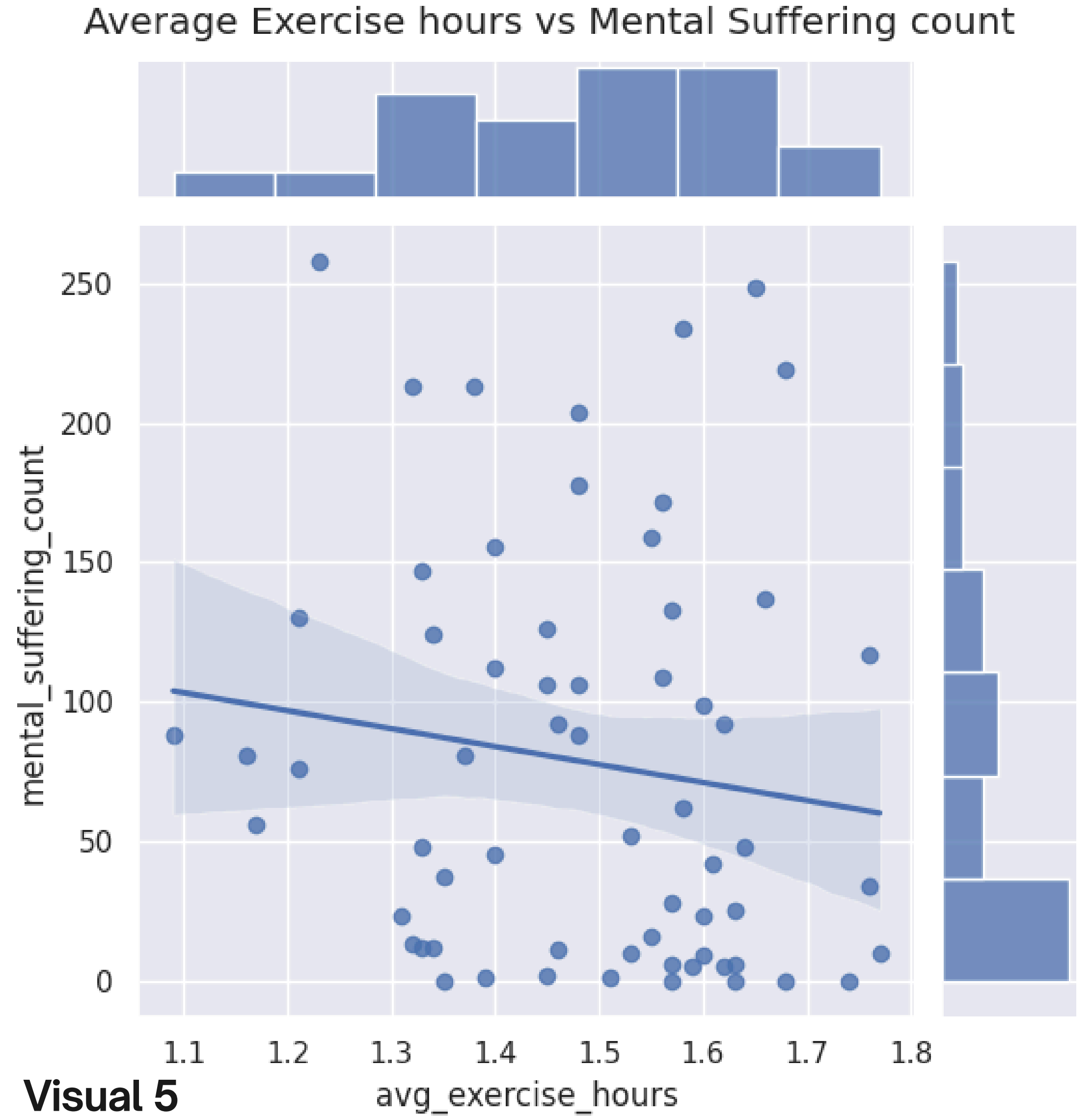
Shows a stronger negative trend between exercise and mental suffering count, suggesting that individuals who exercise more tend to report fewer instances of mental suffering.

The regression line in this case has a more noticeable downward slope, and while there is still variability, the relationship appears more consistent compared to the first plot.

Exercise and Social Media:

There appears to be a very weak negative correlation between exercise and social media daily count, as indicated by the slightly downward sloping regression line, though the data is widely scattered and suggests only a minimal association.

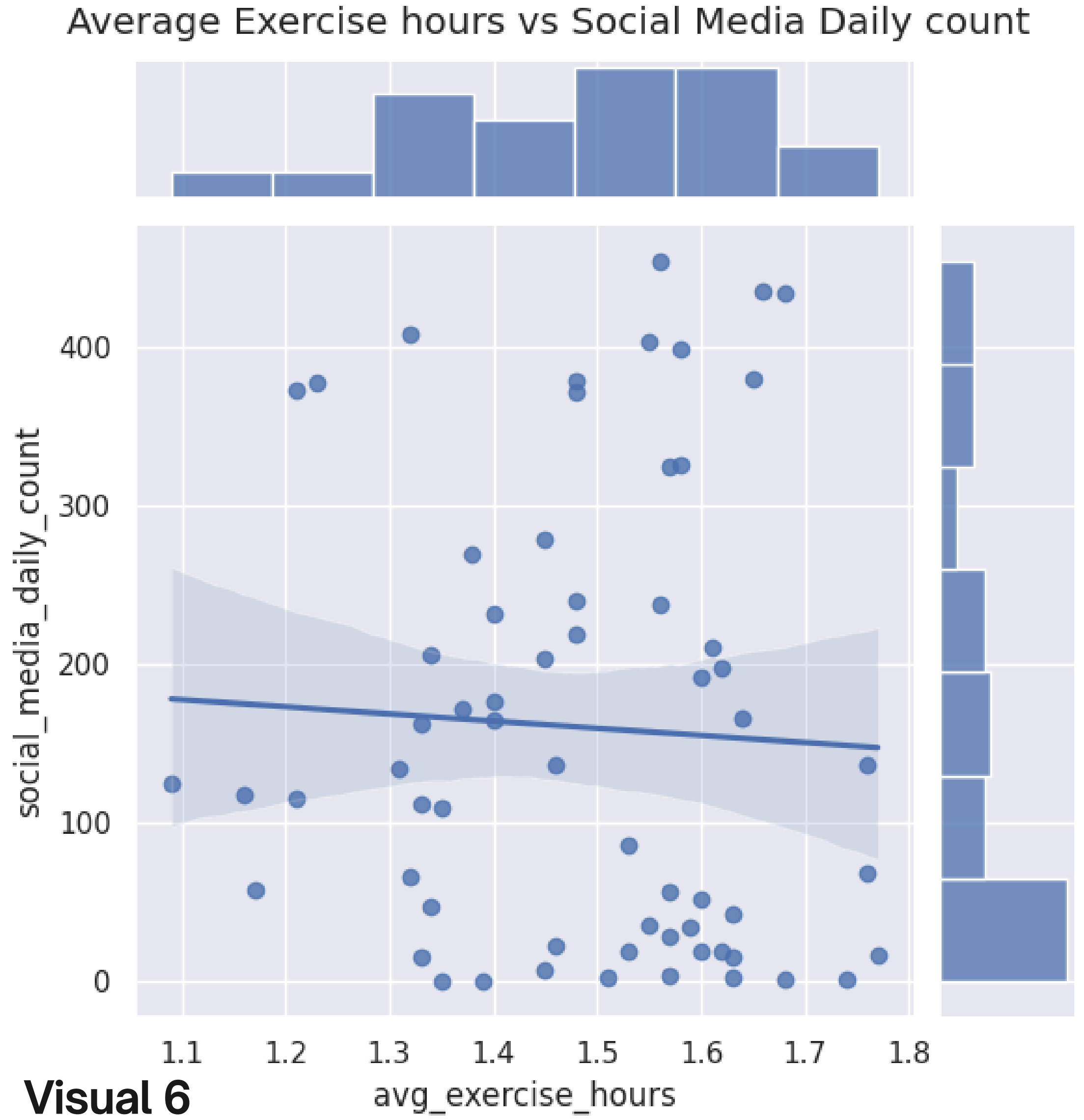
Relationship between exercise hours and reports of mental suffering



Visual 5

It appears that more the exercise hours the less people report mental health suffering - indicates a negative correlation.

Relationship between daily social media use and Exercise Hours



Visual 6

No clear trend was observed between exercise and and social media.

Analysis Conclusion

Interesting Relationships & Insights

From our analysis, **4** key insights emerged.

- **High social media use is strongly associated with mental suffering**
  - Teens who spend more time on social media report significantly higher levels of mental distress. This was the strongest correlation in the data.
- **Sleep is a major protective factor**
  - Teens who get 8 or more hours of sleep report far fewer mental health issues. Sleep is also negatively correlated with social media usage.
- **Exercise offers some mental health benefits**
  - While the effect is smaller, teens, especially boys, who exercise more tend to report lower levels of mental suffering.
- **Age has limited impact**
  - Although older teens report more mental suffering, the trend is not as strong or consistent as lifestyle factors like sleep and screen time.

Promoting healthy sleep habits, limiting social media, and encouraging physical activity could meaningfully reduce mental suffering among teens. Among all factors, sleep and social media usage show the most actionable impact.

Statement of Work

Howard Lin

Howard Lin led the exploratory data analysis (EDA) phase of the project, focusing on identifying and assessing relevant datasets on teenage screen time use. He evaluated each dataset's quality, structure, and completeness to determine its suitability for our research goals. Howard also contributed to planning the data cleaning pipeline and is responsible for implementing preprocessing steps such as handling missing values and correcting anomalies to ensure the data is ready for modeling.
In addition, Howard lead the draft project proposal to ensure alignment between the research questions and available data, offering revisions to sharpen focus and clarity.

Collaboration with Howard has been effective, particularly in early-stage decision-making and dataset selection. In future projects, establishing clearer timelines and task ownership earlier on could further enhance our workflow and coordination.

Yasthil Singh

Yasthil played a key role in maintaining alignment between the project's analytical processes and its central research question. He ensured that all analyses, data usage, and methodological decisions remained focused on addressing the project's core objectives with clarity and precision.
Yasthil also supported multiple stages of the project pipeline, contributing to data cleaning and transformation efforts across both datasets. He assisted in merging the datasets into a final cohesive structure, enabling accurate and efficient downstream analysis.

Throughout the project, Yasthil has provided thoughtful input to ensure analytical consistency and focus on the projects objectives. For future collaboration, improved task delegation and more frequent team check-ins could help streamline communication and reduce overlap in efforts.

Luca Cannis

Luca is responsible for developing the project's exploratory and results-driven visualizations. His work focuses on uncovering patterns and communicating insights related to key variables such as age, gender, sleep duration, and screen time among teenagers. He selected appropriate chart types to clearly illustrate trends and relationships, such as how sleep patterns vary by gender or how screen time may correlate with reported stress levels.
In addition to designing the visualizations, Luca ensures that each graph is accompanied by clear and concise explanations to make the findings accessible to a broad audience, including those without a technical background.

Collaboration with Luca has been productive, especially in translating data insights into clear narratives. For future projects, allocating more time for iteration and peer feedback on visual materials could further strengthen the storytelling aspect of the analysis.

References

Faverio, Michelle, Monica Anderson, and Eugenie Park. Teens, Social Media and Mental Health. Pew Research Center, 22 Apr. 2025, https://www.pewresearch.org/internet/2025/04/22/teens-social-media-and-mental-health/

American Psychological Association. Trends to Watch in 2023: Improving Youth Mental Health. Monitor on Psychology, Jan. 2023, https://www.apa.org/monitor/2023/01/trends-improving-youth-mental-health.

Centers for Disease Control and Prevention. Sleep Duration and Risk of Depression Among U.S. Adults, 2020. National Center for Health Statistics, Data Brief No. 513, May 2024, https://www.cdc.gov/nchs/data/databriefs/db513.pdf. Accessed 10 June 2025.

Data References

Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance System (YRBSS) Data Summary & Trends. CDC, Cell phones, Teens and Mental Health . Accessed 10 June 2025.

Wankhede, Aniruddha. Mental Health Analysis Among Teenagers. Kaggle, https://www.kaggle.com/datasets/aniruddhawankhede/mental-heath-analysis-among-teenagers. Accessed 10 June 2025.