

Final Project Exploratory Visualization

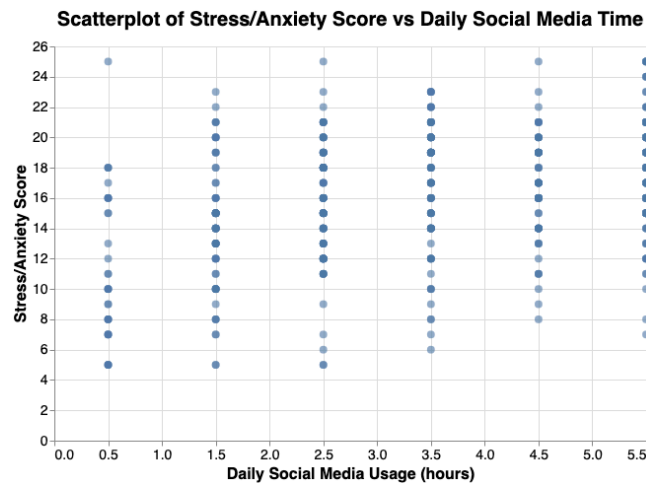
Helin Yilmaz, Maya Alhashem, Nory Arroyo, Oviya Adhan, Vanessa Navarro

W209 - Section 4

Table of Contents

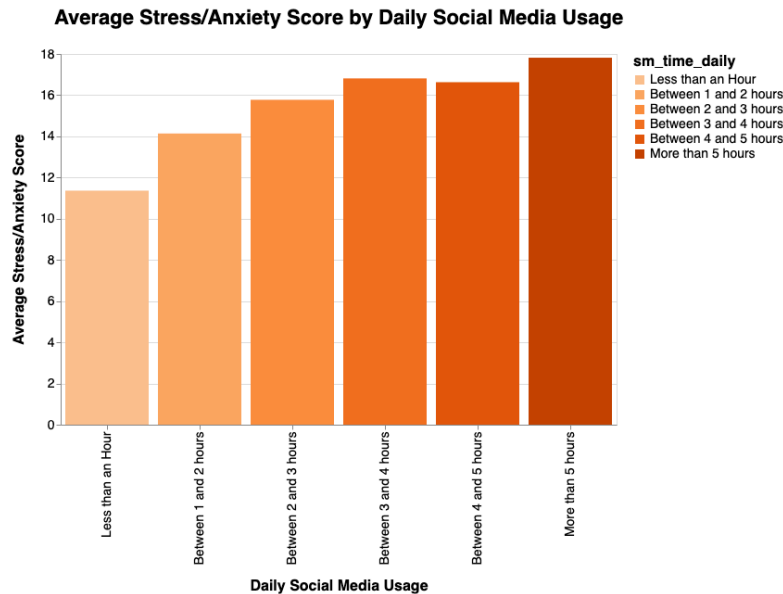
- I. [Hypothesis 1](#)
- II. [Hypothesis 2](#)
- III. [Hypothesis 3](#)
- IV. [Hypothesis 4](#)

HYPOTHESIS 1: Individuals who spend more time on social media report higher levels of stress or anxiety, and this relationship varies depending on factors such as age, gender, and sleep habits.



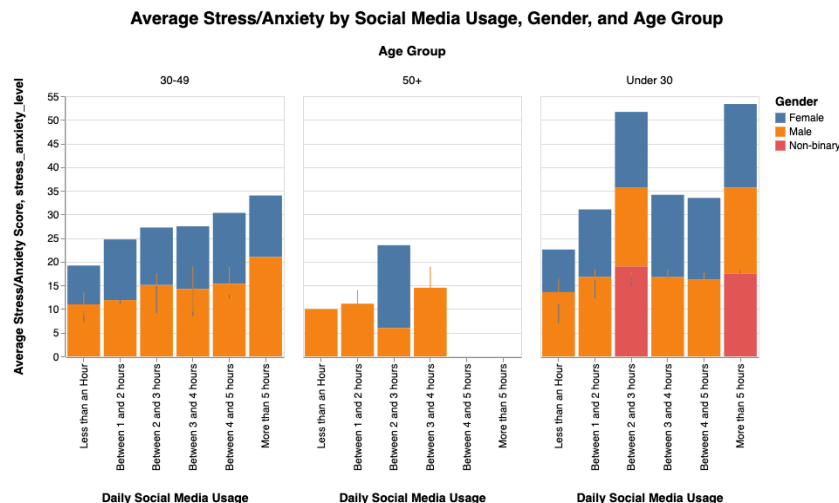
What's informative about this view: The range of the stress/anxiety score is presented in this view of the data, spanning from approximately 4 to 26. This score is a composite of measures including worry, concentration difficulty, social media comparison feelings, frequency of depression, and sleep issues. Similarly, we can view the range of hours for daily social media use, from 0.0 to 5.5 hours. Although it isn't a profound difference, we see that the stress/anxiety score is lower for those who use social media less, specifically in the 0.5 hours category, where scores predominantly range from 4 to 18, with a few outliers slightly higher. As daily social media usage increases, there appears to be a general upward trend in the stress/anxiety scores, with higher scores becoming more frequent in the higher social media usage categories (ex. 4.5 to 5.5 hours), where scores regularly reach 20 and above.

What could be improved about this view: It is difficult to precisely quantify and visually discern the significance of the relationship between these two variables directly from this scatter plot. The overlapping data points make it hard to determine the exact distribution of stress/anxiety scores at each level of social media usage, especially for usage categories with a high density of points. A clear trend or correlation is not immediately apparent due to the spread of the data. Furthermore, this scatter plot alone does not allow us to investigate the influence of age, gender, and sleep habits. It may be more useful to plot this data as a bar chart of averages, rather than a scatter plot.



What's informative about this view: This chart demonstrates a direct relationship in which increasing daily social media usage corresponds with a higher average stress/anxiety score. An exception to the trend is observed for users who spend between 4 and 5 hours on social media, who report lower stress and anxiety levels than those spending 3 to 4 hours daily. The x-axis categories are logically ordered, making the upward trend straightforward to observe. Additionally, presenting the aggregated average score along with respondent counts in the tooltip provides a concise and informative summary of the data.

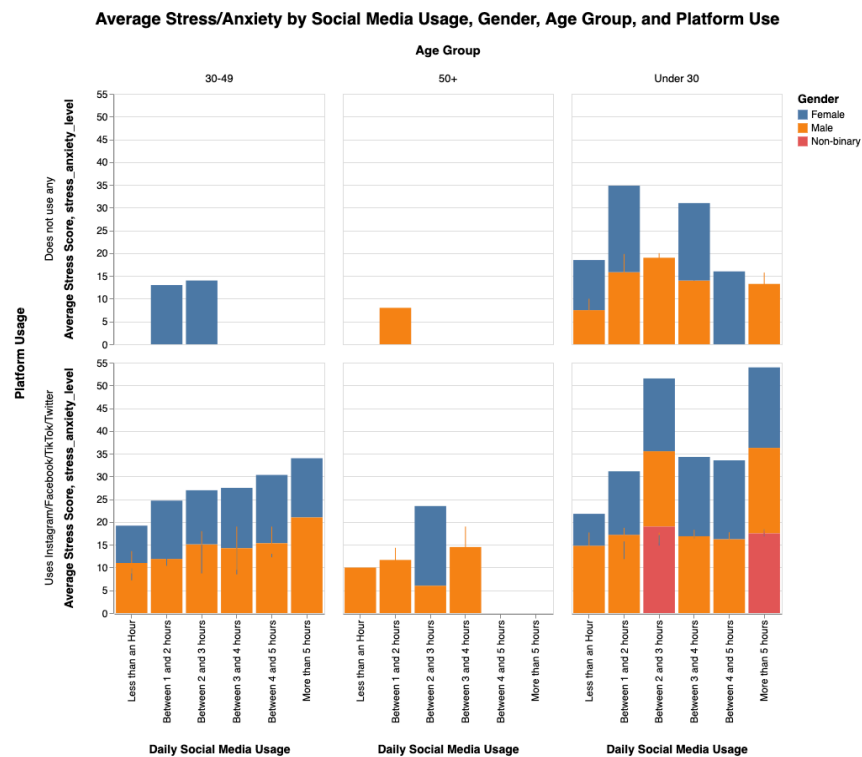
What could be improved about this view: The current visualization does not display variability or uncertainty around the average stress scores. Incorporating error bars would significantly enhance the interpretation. Furthermore, the chart does not reveal how other critical factors from the hypothesis, such as age, sleep habits, or specific social media platforms, influence this relationship. Introducing interactive filtering or faceting based on these variables would allow for a more comprehensive analysis.



What's informative about this view: This visualization presents average stress and anxiety levels by social media usage, further broken down by gender and age group. Individuals under 30 report the highest stress levels, with peaks among those spending 2 to 3 hours or more than 5 hours daily on social media. In the 50 and older group, overall stress is lower, though female respondents primarily clustered in the 2 to 3 hour usage range exhibit the highest stress within that age group. For those aged 30 to 49, a clearer linear trend emerges, showing that increased social media time correlates with

higher stress consistently across males and females. Non-binary respondents are only observed in the under-30 group. The inclusion of error bars effectively communicates variability and confidence in these averages, emphasizing how the relationship between social media use and stress varies across demographic groups.

What could be improved about this view: This visualization could be enhanced by distinguishing between users of specific social media platforms and those who do not engage with platforms where personal information is regularly shared. This approach may provide insights into whether all or only certain types of social media applications contribute to mental health effects.

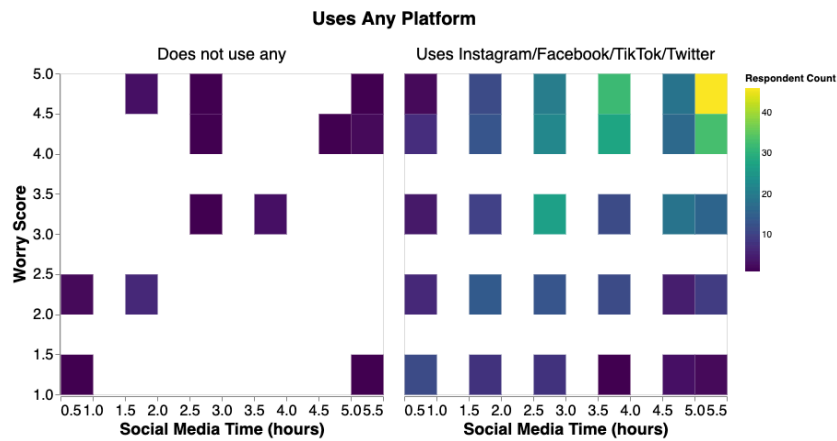


What's informative about this view: Faceting by use of Instagram, Facebook, TikTok, or Twitter versus non-use reveals differences in the relationship between social media and stress/anxiety based on platform engagement. Those aged 40 to 49 and 50 and older who do not use any of these platforms report substantially lower stress and anxiety levels than those who do. Age group segmentation (Under 30, 30-49, 50+) highlights differing stress experiences related to social media usage among younger and older individuals. Younger individuals under 30 consistently report the highest stress levels regardless of platform use. This visualization integrates social media usage time, platform use, gender, and age groups within a clear grid layout and includes error bars to indicate confidence intervals, supporting deeper exploratory analysis.

What could be improved about this view: The combination of row (platform usage) and column (age group) facets creates multiple small charts, which may overwhelm viewers or complicate quick comparisons. This layout requires viewers to spend time to grasp the main insights. Additionally, grouping four common platforms into a simple yes/no category obscures potential differences in how each platform affects stress. Separate analyses for each platform may reveal more nuanced relationships.

Exploring other types of charts:

Heatmap of Social Media Usage vs Worry Score, Faceted by Platform Use



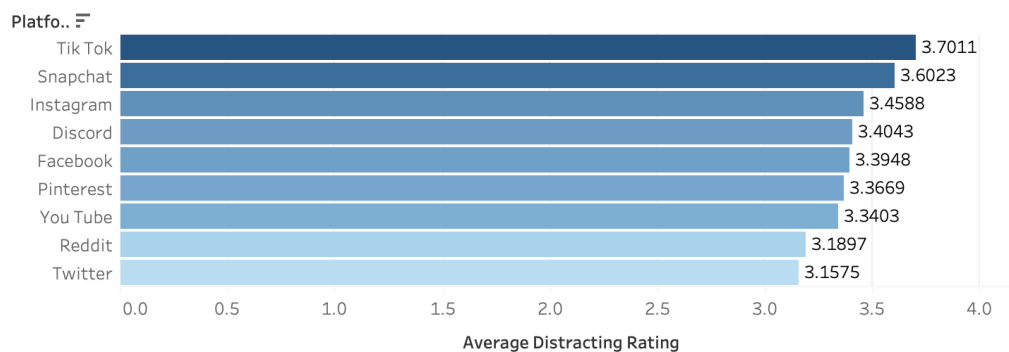
What's informative about this view: This heat map provides a clear visualization of the relationship between worry levels and time spent on social media. It separates respondents into two groups: those who use any of the four platforms (Instagram, Facebook, TikTok, or Twitter) and those who do not. This allows for a straightforward comparison of how platform usage relates to worry scores and social media time.

What could be improved about this view: This visualization does not include breakdowns by age or gender, which are important factors for understanding differences in stress and social media use. Adding these dimensions provides a more detailed picture of how worry levels vary across different demographic groups.

Conclusion: The data generally support the idea that spending more time on social media is linked to higher stress and anxiety levels. This relationship changes depending on age, gender, and whether people use popular platforms like Instagram, Facebook, TikTok, or Twitter. Younger people and those who use these platforms tend to report higher stress levels. However, differences between groups and some exceptions in the data show that factors like the type of platform and sleep habits should be studied more to better understand how social media affects mental health.

HYPOTHESIS 2: Personal media sharing platforms such as Instagram, Facebook, Snapchat, TikTok, and YouTube distract users more than discussion forums and bookmarking platforms such as Reddit, Discord, Twitter, and Pinterest.

Average Distraction Feelings by Social Media Platforms

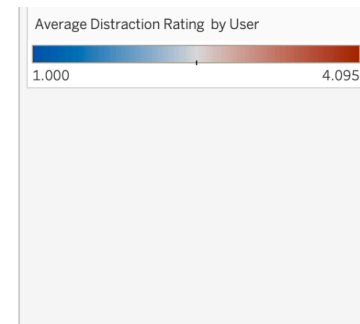


What's informative about this view: The view depicts each platform and the magnitude of the average user rating regarding the feeling of being distracted by social media while doing other tasks. The platform with the highest average value for feelings of distractibility is TikTok, followed closely by Snapchat.

What could be improved about this view: Two things could help improve this visualization. First, it would be beneficial to grant a broader range of colors, making it easier for the user to distinguish between higher and lower values. The second approach would be to implement a second measure to add more context; we could either track time spent on social media or the percentage of users who utilize each platform.

Feelings of Distraction by Daily Time Spent Across Social Media Platforms

Sm Time Daily	Platform								
	Discord	Facebook	Instagram	Pinterest	Reddit	Snapchat	Tik Tok	Twitter	You Tube
Less than an Hour	2.250	1.931	1.933	1.833	1.667	1.857	1.000	1.800	1.783
Between 1 and 2 hours	2.778	2.880	2.846	2.667	2.591	3.294	3.333	2.765	2.717
Between 2 and 3 hours	3.429	3.390	3.366	3.156	3.381	3.400	3.438	3.138	3.304
Between 3 and 4 hours	3.318	3.507	3.414	3.100	3.435	3.528	3.556	3.048	3.416
Between 4 and 5 hours	3.515	3.786	3.811	4.095	3.556	3.821	3.889	3.833	3.719
More than 5 hours	3.800	3.786	3.870	3.816	3.423	3.931	4.033	3.500	3.802

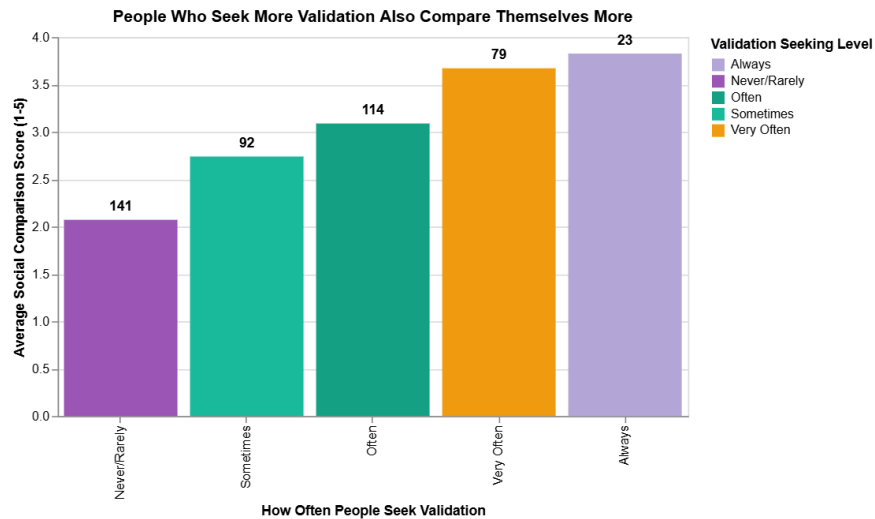


What's informative about this view: This view provides more context for the previous graph we examined. Now we get a wider array of values, and the actual highest average distraction rating is for Pinterest for users who spend between 4 and 5 hours online. Additionally, we can observe that the higher values overall are found in the users who spend more than 5 hours online, across all platforms.

What could be improved about this view: Since we're looking at averages, we could further expand on this visualization by taking into account the number of users that utilize each platform. That way, we can determine the magnitude of the impact on the average calculations. Additionally, we could provide further context on what our numbers represent.

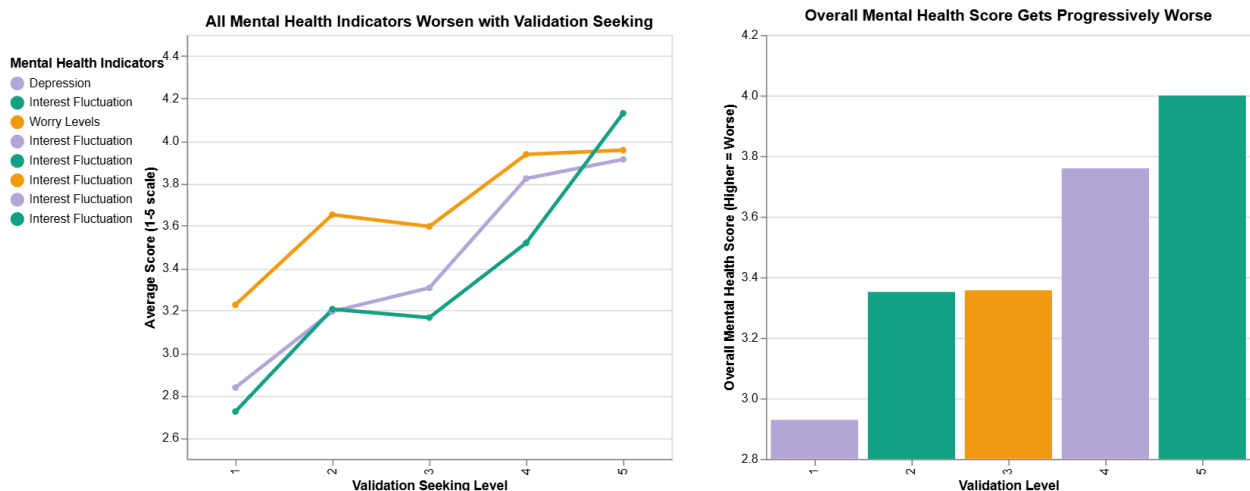
Conclusion: At the beginning of our analysis, we observed that overall, TikTok had the highest average distraction ratings. However, upon further inspection, the highest value across all platforms, taking into account the time spent, was for Pinterest, a bookmarking-type application. However, we can observe that for all personal media sharing platforms, the more time a user spends online, the average distraction rating consistently increases.

HYPOTHESIS 3: Users who frequently seek validation on social media show higher comparison behaviors and report more negative mental health outcomes



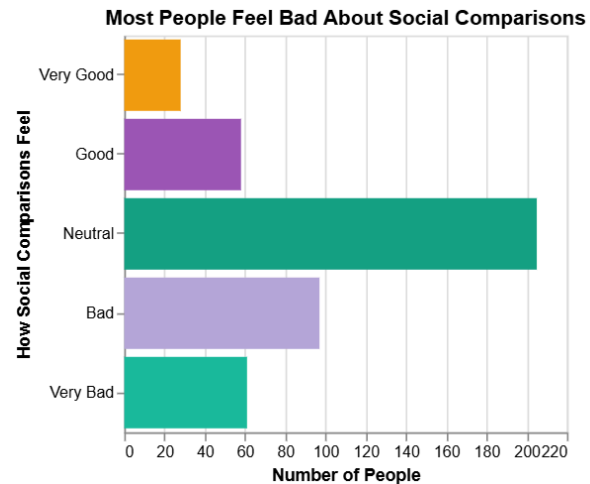
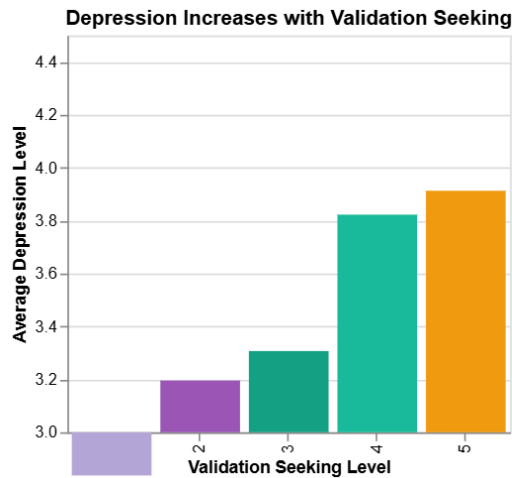
What's informative about this view: Shows a clear linear progression: as validation seeking increases from "Never/Rarely" (around 2.1) to "Always" (around 3.8), comparison behavior consistently increases. It provides solid evidence for the first part of hypothesis 3. Color coding helps distinguish between validation seeking levels.

What could be improved about this view: It could benefit from error bars or confidence intervals to show statistical significance especially that sample sizes differ for each category. Moreover, the scale could be more clearly labeled (it is a 1 to 5 scale so doesn't even start at 0).



What's informative about this view: The line chart effectively shows that ALL mental health indicators (worry, depression, interest fluctuation) worsen with increased validation seeking. The bar chart provides a clear composite mental health score that summarizes the overall trend.

What could be improved about this view: The line chart legend could be clearer - some labels appear duplicated. The legend has mistakes. The bar chart shouldn't vary in color as the color variation does not add any information.



What's informative about this view: Left chart isolates depression specifically, showing a clear upward trend with validation seeking. Right chart reveals that most people feel "Neutral" to "Bad" about social comparisons, supporting the negative impact hypothesis.

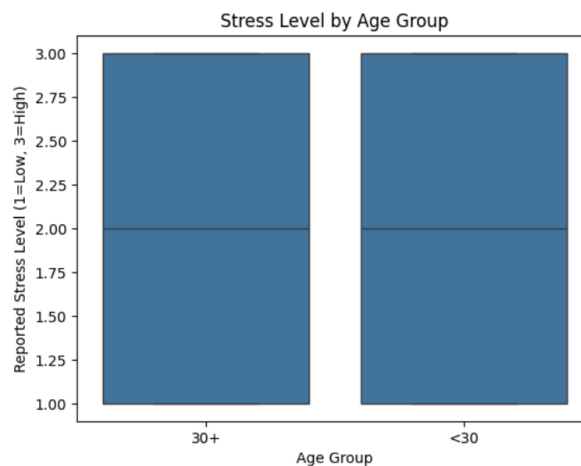
What could be improved about this view: Y-axis range in left chart. Remove meaningless color variation. Missing the connection between how people feel about comparisons and their validation seeking behavior. Could show validation seeking levels within each "feeling" category to strengthen the connection.

Conclusion: The data visualized strongly support hypothesis 3. The visualizations demonstrate:

1. Clear relationship between validation seeking and comparison behavior (1st visualization)
2. Consistent negative mental health outcomes across multiple indicators (2nd visualization)
3. Specific evidence for depression increases and general negative feelings about comparisons (3rd visualization)

Out of the 3 attempts, visualization 2 is the best because it tests both parts of hypothesis 3 - showing that validation seeking leads to both higher comparison behaviors AND worse mental health outcomes across multiple indicators.

HYPOTHESIS 4: The impact of social media use on mental health is stronger for individuals under 30 compared to those 30 and older.

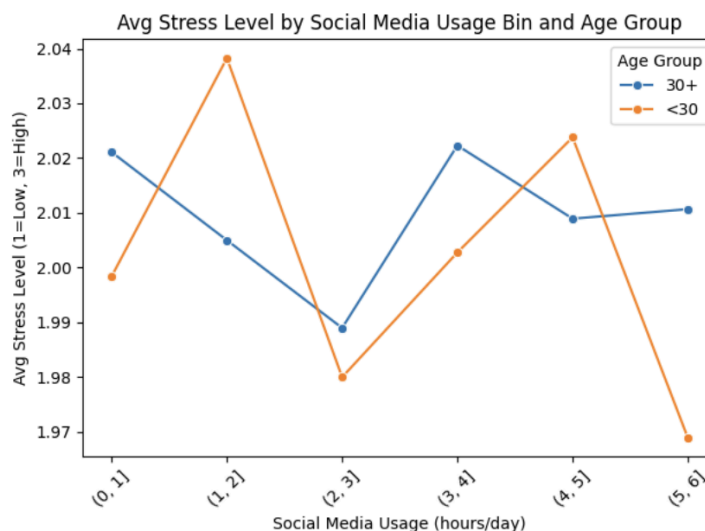


What's informative about this view:

This chart visualizes the distribution of reported stress levels between individuals under 30 and those that are 30 and older. It shows that the median stress levels (converted to a numeric scale) and overall distribution are nearly identical across both groups. This suggests that the overall frequency of reported stress may not significantly differ between age groups.

What could be improved about this view:

As we can see, the plot lacks visual difference because the counts are nearly identical in distribution within each group. So, this view does not completely tell us whether stress levels are more sensitive to social media usage among younger individuals. To investigate this more deeply, I can examine how stress correlates within each age group, that is, whether the *relationship* between social media use and stress differs between younger and older respondents.

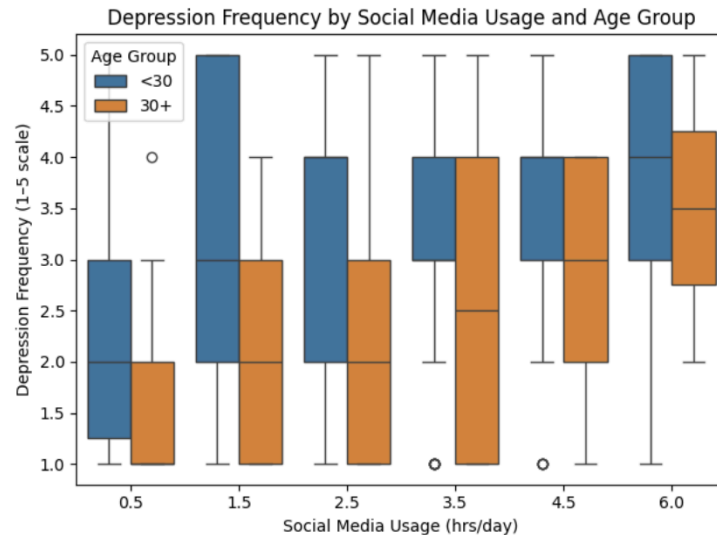


What's informative about this view:

This binned line chart displays the average stress levels reported by individuals across increasing ranges of daily social media use. The orange line represents individuals under 30, while the blue line represents those 30 and above. The plot shows more of a fluctuation in stress levels for those under 30 considering there are sharper drops and rises compared to the 30+ group. This could suggest the <30 group may be more sensitive to changes in screen time than their older counterparts.

What could be improved about this view:

The overall variation in stress levels is relatively small, since it hovers between 1.98 and 2.04 on a 1-3 scale. This limits how much insight can be drawn from visual differences. Also, the smaller sample sizes in higher usage bins, particularly among the <30 group, could be exaggerating the perceived peaks or dips. In a last attempt I could add error bars or incorporate additional mental health indicators like sleep quality or depression frequency to enhance interpretability.



What's informative about this view:

This boxplot demonstrates the distribution of depression frequency scores across different social media usage levels, grouped by the two age categories <30 and 30+. Individuals under 30 have higher medians and greater variability in depression frequency compared to older users. This pattern supports the hypothesis that younger users may experience greater mental health impacts with increased social media exposure.

What could be improved about this view:

While the boxplot provides a clear comparison of central tendency and spread, it does not account for differences in sample size across usage bins. Some bins could be based on smaller groups, which could distort the apparent variability. Also, depression frequency is a self-reported ordinal variable, so treating it like a continuous variable could introduce some imprecision in interpreting averages and variability.

Conclusion: Considering all three attempts, the data provides partial support for the hypothesis. In the second attempt, we observed greater variability in stress levels among the <30 individuals, compared to the older group. The third attempt strengthens this interpretation by showing depression frequency (a mental health indicator) had a consistently higher median among the <30 group across several usage levels. While the trends align with the hypothesis, further analysis accounting for sample size differences and additional variables could strengthen the claim.