Visualizing the Effects of Social Media Use on Mental Health

Anmolpreet Kaur

Major: Computer Science

Advisor: Professor Holly Rushmeier

May 4, 2023

Abstract

In the past decade, social media has become a central part of many people's lives. Studies show that almost 60% of the world's population is on social media and 92% of people with internet are on social media[1]. While social media allows people to connect with others across the globe and form healthy connections, several recent studies have linked social media use to mental health issues, such as depression and anxiety[2]. Addictiveness, feelings of loneliness, and self-doubt are only some of the most common negative effects of social media, particularly among teens[3]. This project uses D3.js, HTML, CSS, and Python to visualize data about how social media use affects mental health.

I looked at datasets that allow us to make observations about a wide range of users, rather than a subset of social media users. This includes people of different ages, genders, and geographic locations. I processed the data from these datasets using the pandas Python library. I created interactive visualizations using the D3.js library, including a map, bar chart, and scatterplots. Interacting with these visualizations allows us to notice that teens ages 16-19 tend to struggle the most with mental health. It is also seen that social media usage decreases as age increases, but the impact of social media does not have a clear correlation to age. When looking at data geographically, states that have reported better general health on average do not always have better mental health, and vice versa. Overall, these visualizations serve as a starting point for observers to understand how social media tends to affect mental health.

Viewing the Project

To view the webpages that I created with the visualizations, make sure you have Python3 installed. Unzip the code folder. Then, in your terminal, `cd` to the top-level folder for the unzipped code. Run a server using the command `python3 -m http.server`. Then, open index.html in your browser. Add `localhost:8000` to the beginning of the URL (this should be the default host for most devices). You can now click on the link from there to view the visualizations. Go through each visualization and click on the link at the bottom to go to the next one.

Datasets

My goal was to use datasets that cover different demographics, such as age, gender, and geographic location. After searching through a large number of datasets, I chose the following three datasets. The first dataset is called "Social Media and Mental Health survey conducted in a University for research purposes".[4] This dataset includes data on mental and general health by state. The second dataset is called "Data on language analyses in Instagram users, who had engaged in non-suicidal self-injury".[5] This dataset includes data on statistics of Instagram users and whether they have attempted suicide. The third dataset is called "Effects of social media on mental health during Covid-19 lockdowns".[6] This dataset includes data on what social media platforms are used the most across different age groups and the effect they have had on users.

Before using these datasets in my code, I converted them all to CSV format. These datasets had many columns with a lot of entries. I went through all of the variables to find ones appropriate enough to make visualizations. I also made sure that the variables I chose were relevant enough across all datasets. I cleaned up variable names to make them easier to use in my code. Because there were so many entries, I often had to summarize data for certain demographics like age or state (I talk more about this is the next section).

Visualizations

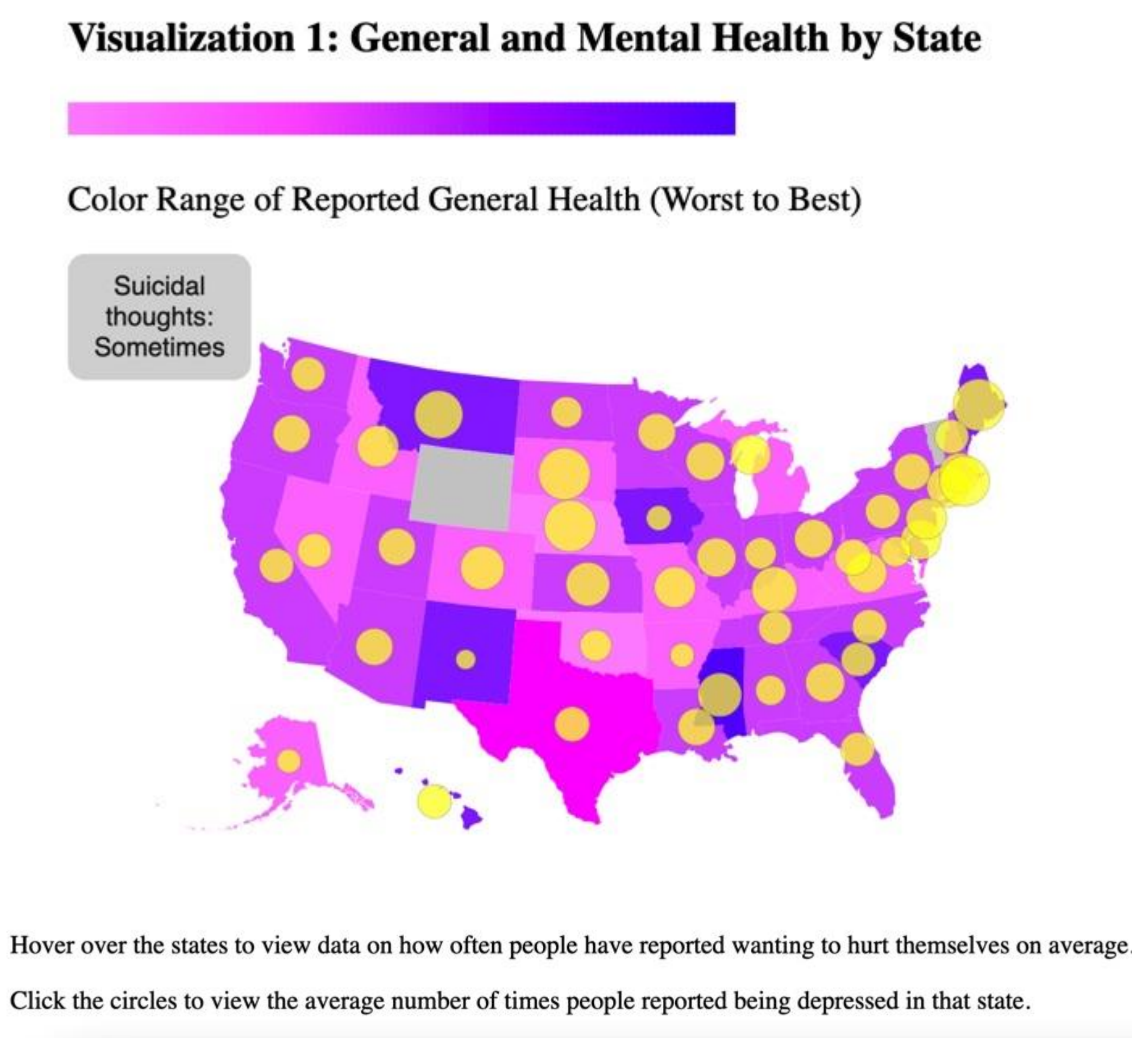
I created four visualizations with the three datasets. Below, I describe what specific data was used from each dataset to create each visualization and how it was processed. I explain what different parts of the visualizations show and how to interact with them. I have included pictures of the visualizations. I have also summarized observations we can make for each visualization by interacting with it.

Visualization 1

The first visualization is a geographic visualization that uses the first dataset, the mental health survey that was conducted in a university in Michigan. Specifically, for each state, the color (a shade of purple) is a representation of the average general health reported. This is the "general_health" column from the dataset. This column had the general health reported on a scale of "Poor" to "Excellent". I used the "pandas" Python library to first get a total count of each of the five possible answers – "Poor", "Fair", "Good", "Very Good", and "Excellent". I then found the average answer for each state by assigning weights numerically to each answer.

In the visualization, the yellow circles represent the average number of times people have felt depressed for each state. This is using the "acha_12months_times_5" column from the dataset, which answers the question, "Within the last 12 months how many times have you – Felt so depressed that it was difficult to function?" I processed the data for this variable similarly to "general_health." The size of the circle corresponds to how many times people reported being depressed on average for the state. The smallest circle size corresponds to the average answer being close to "Never" while the biggest circle corresponds to "11 or more times." Clicking each circle shows a text with the average number of times people reported being depressed in that state.

Hovering over each state in the visualization gives the average amount of times people have reported wanting to hurt themselves. This is using the "phq9_9" column from the dataset, which answers the question, "Over the last two weeks, how often have you been bothered by the following problems? – Thoughts that you would be better off dead, or thoughts of hurting yourself in some way?" This was processed similar to the last two variables.



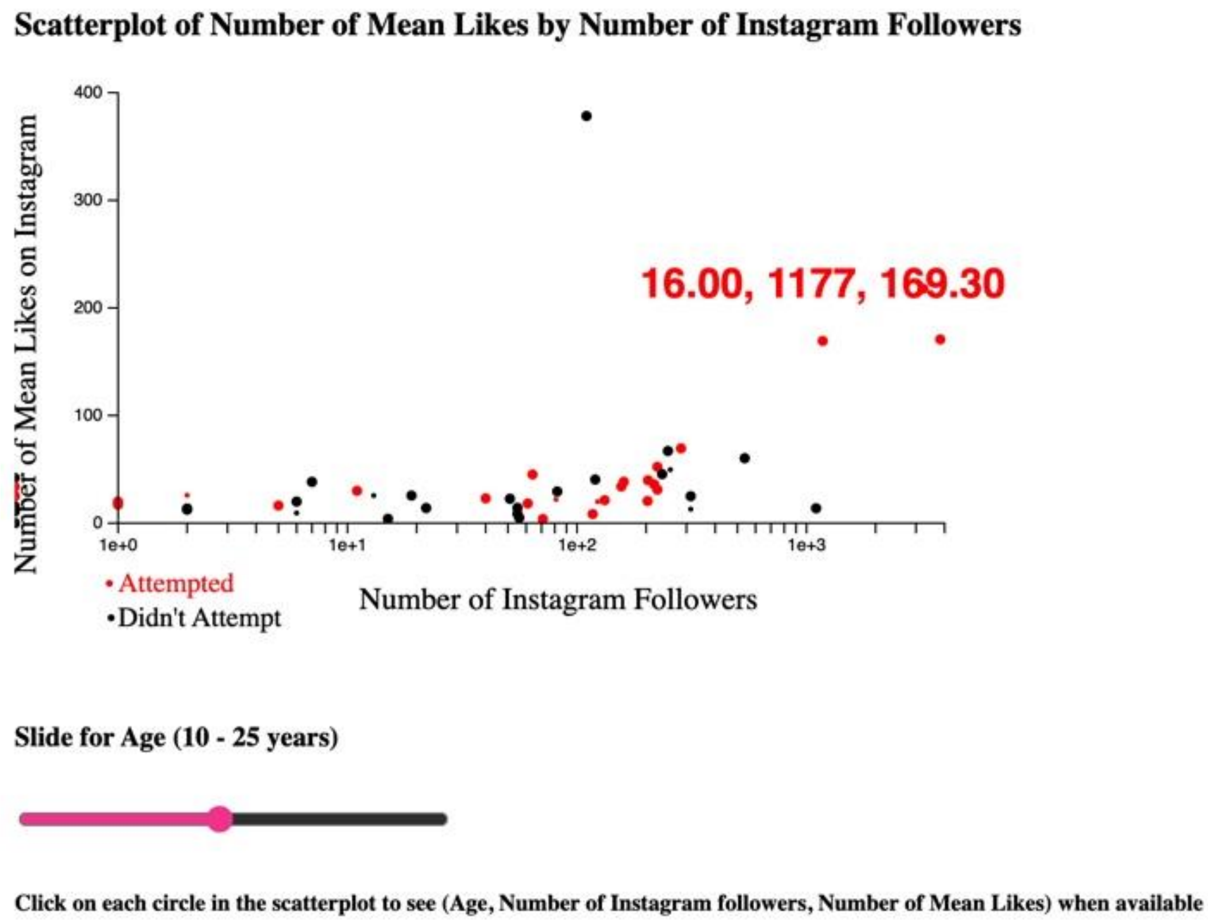
As a side note, some of the circles are clustered on the east coast in the map. I had attempted to include a zoomed in version of that section of the map to make it easier to see.

However, because of the complex coordinates in the "us-states.json" file, it was difficult to recreate that section in a larger view due to limited time. This could be a useful next step.

Looking at and interacting with this visualization, we can see that states that have better general health on average do not always have better mental health, and vice versa. States that have reported feeling depressed a greater number of times do not always correspond to a greater number of times they have wanted to hurt themselves on average. Therefore, there are no clear trends geographically. It would be useful to see if other factors, such as age or gender, show clearer trends in the other visualizations.

Visualization 2

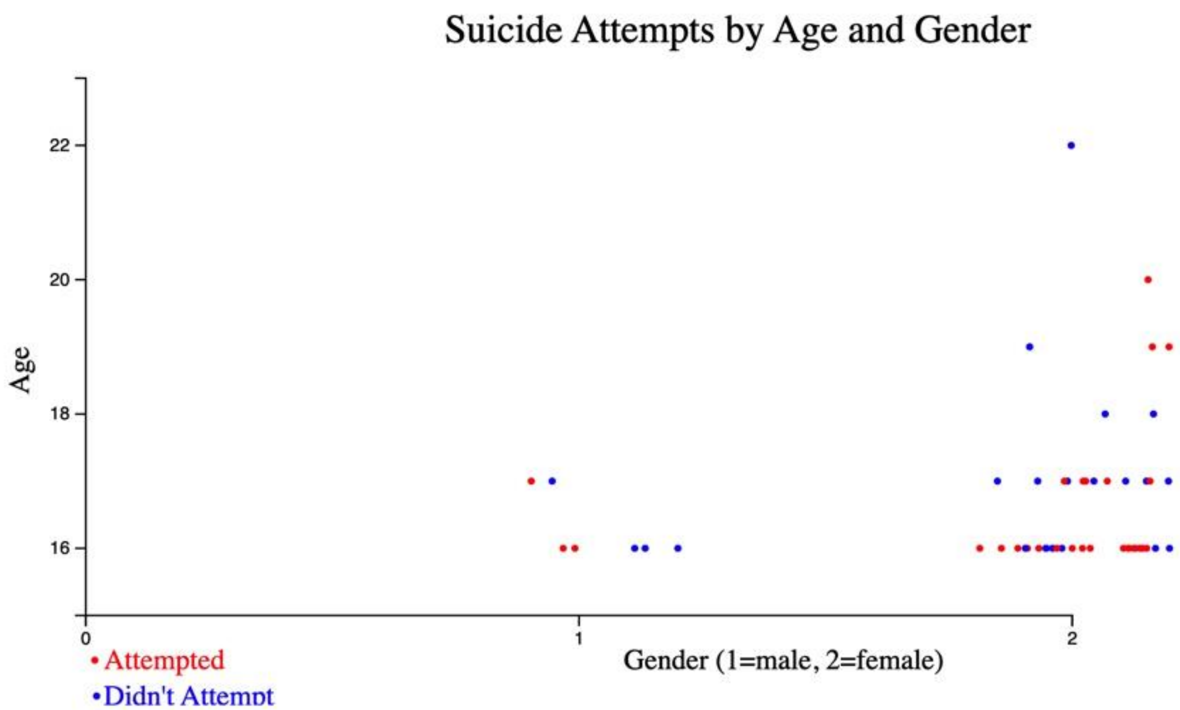
Visualizations 2 and 3 are scatterplots created using the second dataset, the Instagram Suicidality data. The first scatterplot shows a correlation between the number of Instagram followers and the number of mean likes each user gets on Instagram. These are the "Number_Followers_Instagram" and "Mean_Likes_Instagram" columns in the dataset. When I originally created this scatterplot with the normal variables, many of the points were clustered in the left corner. So, I used a log scale for the "Number of Instagram Followers" variable so that the points are more spread out, and it is easier to see a trend. The color of the dots represents whether they have or have not attempted suicide. Red corresponds to attempted suicide while black corresponds to didn't attempt suicide. A slider is included below the scatterplot for ages ranging from 10 to 30 years. Sliding this slider makes the dots bigger for users of the selected age. Clicking on a dot in the scatterplot shows the age, number of followers, and number of mean likes, in that order (some ages are not available because they were not in the dataset so they get bigger at 0).



We see from this visualization that there seems to be a slight positive correlation between the Number of Instagram followers and Number of Likes, which makes sense. It is interesting, however, that there is no clear correlation of these variables with the suicide attempts or non-attempts. It can be seen that the outliers on the right side of the plot all have attempted suicide. But the rest of the dots are a more even mix of attempts and non-attempts. When using the slider to look at ages, we notice that teens are the most likely to attempt suicide.

Visualization 3

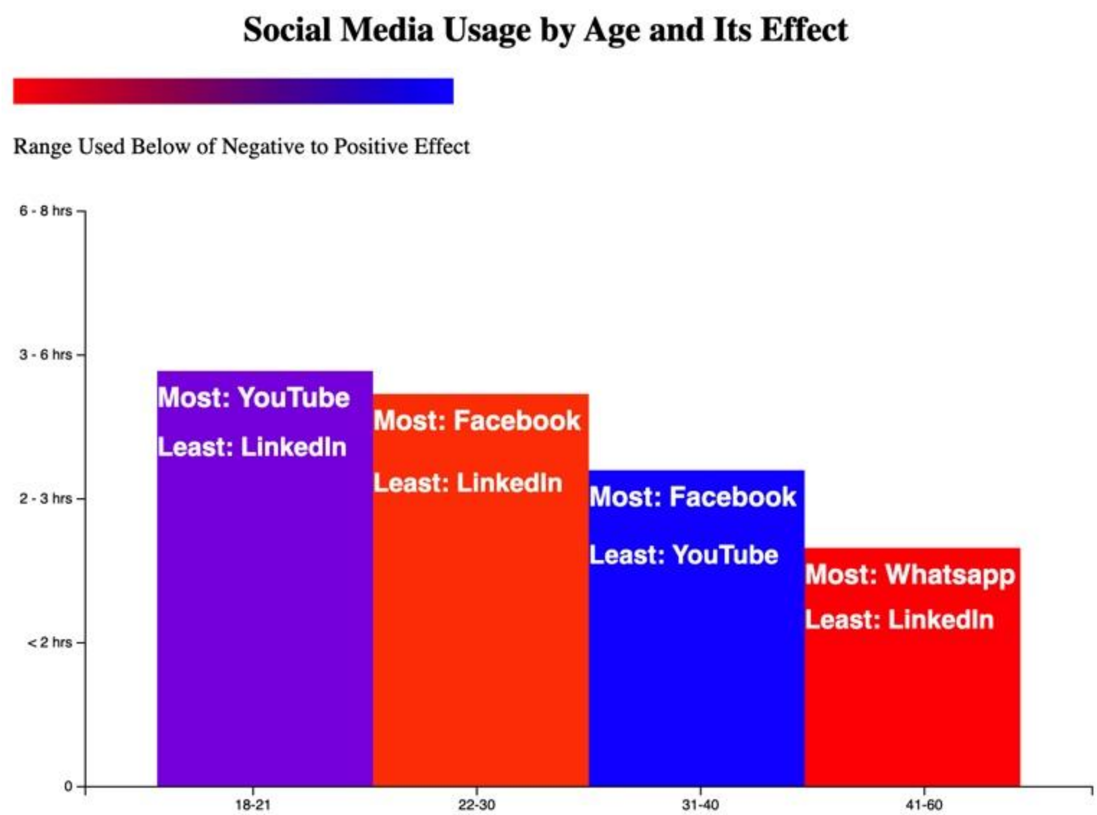
Visualization 3 shows attempts or non-attempts for suicide by age and gender. The x-axis uses the “Gender” variable from the dataset while the y-axis uses the “Age” variable. The color of the dots represents whether they have or have not attempted suicide. “Jitter” is used to spread out the points to make them easier to see.



From this visualization, it is clear that the greatest number of attempted suicides are by teens between the ages of 16-19. We also see that females are more likely to attempt suicide than males. It is important to note, however, that the number of male data points in this dataset are significantly lower than the number of female data points, so this may be skewed.

Visualization 4

The last visualization is a bar chart I made using the third dataset, “Effects of social media on mental health during Covid-19 lockdowns”. It shows the average number of hours social media is used across the different age groups. The x-axis is using the “AgeRange” column and the y-axis is using the “SocialMediaTimes” from the dataset. I used the “pandas” Python library again to find the average for each age range. The color ranges from red to blue represents a negative to positive impact of social media that was reported. The red is a negative overall effect while the blue is a positive overall effect, and no effect is halfway between the two. This uses the “OverallEffect” column from the dataset. For each age range, on the bar chart you can also see the social media platforms that people found most and least helpful while talking about mental health on average.



Looking at this visualization, we see that social media usage clearly decreases as age increases. But the reported effect on a scale of negative to positive impact does not have a clear correlation to age. The 41-60 age range reported having the most negative effect, while the 31-40 age range reported having the most positive effect. The 18-21 and 22-30 age ranges are somewhere in between with one leaning more towards positive and the other leaning more towards negative, respectively. In terms of social media platform, we see that Facebook tends to be more helpful while LinkedIn tends to be the least helpful when talking about mental health.

Conclusions

The data used in this project mostly focused on teens ages 16-21, particularly the first and second datasets. Based on this data and these visualizations, it can be seen that teens with greater social media usage tend to struggle more with mental health. Suicide attempts are the most common in this age range, although there is no clear pattern by gender. Older social media users tend to be aware of the negative effects that come with it. When looking at the data by state, there is no clear pattern in mental health by geographic region. But it is interesting to see which states have higher reported feelings of depression and suicidal thoughts. It can also be seen that mental health does not directly correspond to general health. Interacting with the visualizations on my webpage allows for further observations. As a next step, it could be interesting to look at some of the more positive impacts of social media.

Acknowledgements

I was able to complete this project because of the skills I learned from Professor Rushmeier's “Data and Information Visualization” course. I am very grateful for Professor Rushmeier's help throughout this semester as I completed this project.

Sources

1. Chaffey, Dave. “Global Social Media Statistics Research Summary 2022 [June 2022].” Smart Insights, January 30, 2023. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/#:~:text=The%20number%20of%20social%20media,growth%20of%20%2B137%20million%20users.
2. Walsh, Dylan. “Study: Social Media Use Linked to Decline in Mental Health.” MIT Sloan, September 14, 2022. https://mitsloan.mit.edu/ideas-made-to-matter/study-social-media-use-linked-to-decline-mental-health.
3. “The Effects of Social Media on Mental Health.” Lancastergeneralhealth.org. Accessed February 1, 2023. https://www.lancastergeneralhealth.org/health-hub-home/2021/september/the-effects-of-social-media-on-mental-health.
4. Braghieri, Luca, Ro’ee Levy, and Alexey Makarin. “Data and Code for: Social Media and Mental Health.” openICPSR. American Economic Association, July 18, 2022. https://www.openicpsr.org/openicpsr/project/175582/version/V1/view?path=%2Fopenicpsr%2F175582%2Ffcr%3Aversions%2FV1%2FData%2FInput%2FMentalHealthSurvey%2FMental_Health_Survey_Feb_20_22.csv&type=file.
5. Brown, Rebecca, Eileen Bendig, Tin Fischer, David Goldwich, Harald Baumeister, and Paul L. Plener. “Brown_Bendig_Instagram_Suicidality_Dataset.Xlsx.” figshare. figshare, February 24, 2019. https://figshare.com/articles/dataset/Brown_Bendig_Instagram_Suicidality_Dataset_xlsx/7763333.

6. “Effect of Social Media on Mental Health during Covid-19 Lockdowns.” Zenodo, September 16, 2021. https://zenodo.org/record/551284.