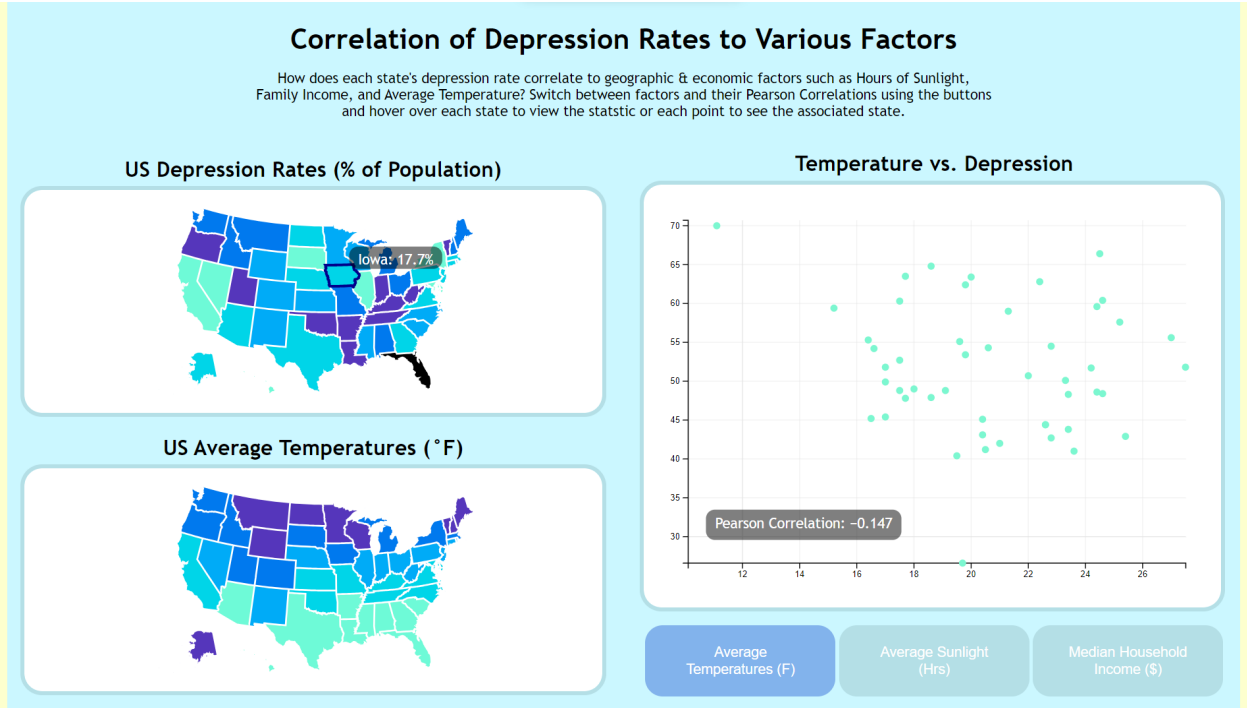
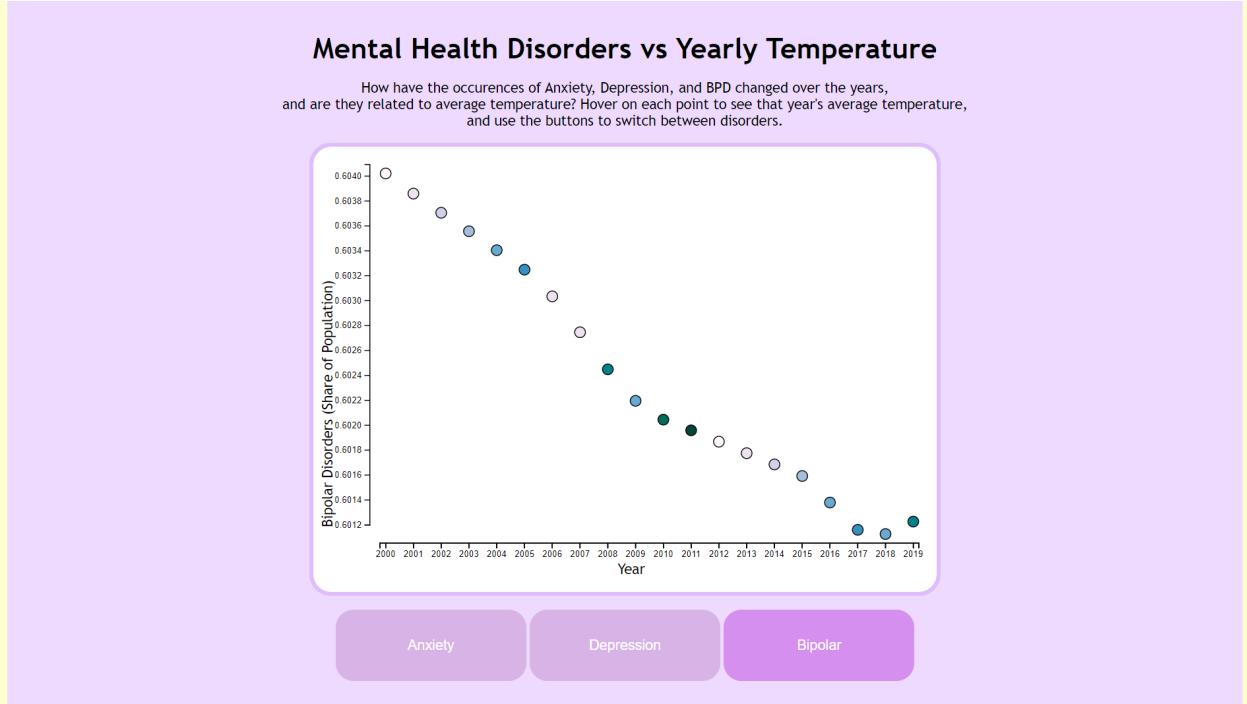
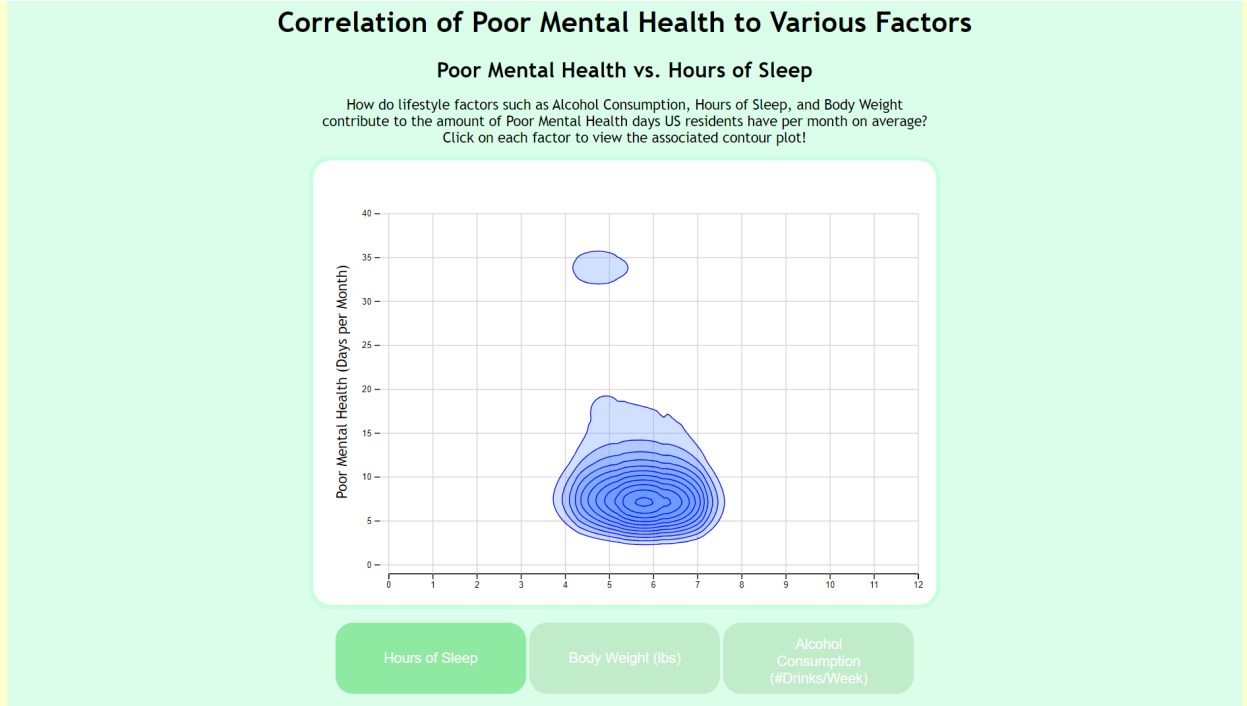
Aditya Kakade (apk67), Bhadra Bejoy (bb584), Gabby Fite (gff29), Richie Sun (rs929)

INFO 3300 Project 2: Mental Health in the United States

Visualization:







Description of Data:

Depression Rates vs Various Geographic + Economic Factors Choropleth:

The choropleth graphs and Pearson correlation scatterplot pull from four datasets: a dataset with each state's depression rate, one with each state's average hours of sunlight each year, one with each state's average temperature, and one with each state's average family income. While there was not much data cleaning, we did have to add columns to each dataset that included the state's standard ID, to ensure our choropleth map properly linked each state with the correct data value. All data was obtained from The World Population Review's database.

Mental Health vs Temperature Graph:

The data set for this graph was created by combining a dataset from a global survey on mental health, specifically, Anxiety, Depression, Bipolar, Schizophrenia, and Eating Disorders, and the yearly temperature for the year that it was recorded. We reduced the results of the survey to only include respondents from America. The temperature data was curated from the National Center for Environmental Information's yearly Climate Reports for the United States.

Mental Health vs Other Health Aspects Graph:

The dataset for this graph was obtained from the US Centers for Disease Control and Prevention (CDCP), which conducts an annual survey (BRFSS) that collects state data about

U.S. residents regarding health-related risk behaviors, wellness habits, and chronic health conditions. According to the CDCP, the data is obtained annually from state health departments that utilize in-house interviewers, contracts with telephone call centers, and universities to administer BRFSS surveys throughout the year, which consist of a standardized questionnaire. This analysis specifically utilized the BRFSS from 2020, and to process the data, we selected the columns of interest which included: hours of sleep, poor mental health days per month, average alcoholic drinks per day, and average alcoholic drinks per day. To further process the data, we removed all missing rows and data points for each of the columns and converted all string inputs into numbers.

Visual Design Rationale:

Depression Rates vs Various Geographic + Economic Factors Choropleth:

We decided to use a choropleth visualization since we wanted users to be able to compare depression rates with the other factors (average temperature, average hours of sunlight, and average family income) on a state-by-state basis. Since darker colors correspond with higher depression rates, and lower temperature, sunlight, and income to ensure consistency with most users' preconceived notions about correlations, one can visually see whether darker states on the depression rates choropleth correspond with darker states on the second, customizable choropleth. The tooltips also allow users to gain specific insight into each statistic on an individual, state-by-state basis.

We had not included the Pearson correlation scatterplot at first – however, our rationale for adding this was for transparency about our correlations. While using the choropleth, users might get an overall notion of correlation from only comparing data for individual states, which does not generalize to the entire US. The Pearson correlation scatterplot allows us to see that some factors might be more correlated with depression rates than others and that overall, no geographic or economic factor has a very strong correlation with depression rates.

We decided not to add legends or axis titles as the purpose of our visualization was to simply display a comparison between the various factors and the actual numbers were not the main focus. Displaying these would have made the visualization look too crowded and distract from the main purpose. That is why we decided to add the mouseover tooltips to display the data if the viewer wished to see it.

Mental Health vs Temperature Graph:

For our visualization, we decided to make a scatter plot since we wanted to demonstrate the shifts in data across two numerical variables, the mental health statistic and the year. Moreover, we decided to create 3 separate plots for the mental health disorders that are the most prominent in the US, Anxiety, Depression, and Bipolar Disorder. It was important to give each of these statistics their space since they had vastly different values. For example, in 2000, about 6.7 share of the population was indicated to suffer from anxiety while about 0.604 share of the

population was denoted as having Bipolar disorder. Thus, displaying these two statistics on the same graph with the same y-axis would have made the visual difficult to decipher.

Although we were interested in how the mental health statistics were impacted by temperature, directly comparing the two would not be an accurate analysis since there are many other factors in the years spanning the dataset that would lead to changes in mental health. Hence we decided to represent temperature in the color scale. We made an ordinal scale with sequential colors to demonstrate the highs and lows of the temperature.

Mental Health vs Other Health Aspects Graph

For our visualization of this dataset, we decided to utilize a contour plot, since we wanted to see if there were any trends between mental health and any of the other health variables like hours of sleep, weight, and alcohol consumption. We initially wanted to utilize a scatter plot in the visualization, but after observing the number of data points as well as the spread of the graph, we decided that it was a better idea to utilize a contour plot since many of the points were concentrated around each other and others were spread out and did not form any trends in the scatter plot itself. Where the individual points in the scatter plot did not provide much information about the relationship between the variables, the contour plot allows us to see density trends as well as provides us with a 2D readable representation of data distribution.

To create the visualization, I utilized a linear scale for both the x and y variables and utilized a d3.contourDensity to obtain the contour density data by mapping each of the individual x and y data points on the scales created, and utilized de.geoPath() to create contour lines based on the density data calculated. From our visualization alone, it was not sufficient to determine any distinct correlations between any of the variables, but rather, the contour plot of the datasets mainly seemed to reflect overall trends in the data. For example, in terms of the sleep and mental health graph, there did not seem to be any indication that sleep contributes to mental health positively or negatively, but rather, the comparison of the two datasets seemed to mainly reflect the overall average for each dataset independently, since the point of highest density seemed to fall around the given average amount of sleep of U.S residents.

General

We made sure to keep similar styling between all three visualizations. We used pastel colors of different hues. The borders of the SVG elements are slightly darker than the background, as are the buttons. The button corresponding to the visualization being displayed is darker to signal which option is picked. The button color becomes darker and the cursor becomes a pointer on button hover to signal that it is an interactive element and can be clicked.

Interactive Elements and Their Design Rationale:

Depression Rates vs Various Geographic + Economic Factors Choropleth:

This selection has three interactive elements. The first one is the ability to hover over individual states on each choropleth and view the state's value for the statistic that is currently

being viewed. We added these tooltips because while the color scale on the choropleth was informative, it only provided information on each state's standing relative to the other states' colors. The tooltips remedy this by exposing the state's actual data.

We also added mouseover functionality to the Pearson Correlation Scatterplot, allowing users to view what state each point corresponds to. This again serves the purpose of exposing the specific data behind each state. Users can match points on the Pearson Correlation Scatterplot to states on the choropleth, allowing for more insight than an overall correlation for the entire country.

Finally, we have three buttons that allow the user to switch the changeable choropleth graph and the Pearson Correlation Scatterplot between visualizing Average Temperature, Average Hours of Sunlight, and Average Family Income vs. Depression Rates for each state. These buttons were useful to ensure that our visualization was not too overwhelming – visualizing one factor at a time rather than all of them.

Mental Health vs Temperature Graph:

For this graph, we had two interactive elements. The first interactive elements are the buttons that denote which mental health statistics the user is viewing. These buttons were useful in separating the data and allowing the user/viewer to seamlessly alternate between the different statistics. These buttons also served as an indicator for the graph to change its data, specifically the y-axis.

Additionally, we utilized a mouseover function for all of the individual data points to depict the average temperature of that year. We chose to make this a mouseover function instead of a permanent label since all of the temperatures displayed together made the graph look cluttered. However, this meant that we lost some of the clarity in what we were trying to examine. Nevertheless, we gain some insight through the color scale which is always visible.

Mental Health vs Other Health Aspects Graph

In terms of interactivity, we decided to include buttons that allow the user to switch between datasets to view how other health aspects and habits may influence mental health. Since the overall purpose of the visualization was to compare and contrast the influence of different health aspects on a person's mental health, the interactivity allows the viewer to switch between datasets to facilitate comparison and analysis.

With a click of each button, the dataset representing the x-axis and x-grid lines would change, but keep the y-axis the same (since we are comparing all aspects to mental health), and a different density plot is generated based on the dataset selected.

The Story:

General:

As a whole, the correlations between the variables we were examining were not as strong as we were expecting. This conveys how mental health is a complex issue with several factors coming into play. Our group chose to examine multiple variables across multiple graphs because we wanted to see the extent to which any one of the variables impacted mental health statistics. Moreover, this allowed us to examine new datasets, so we were not limited to the potential bias of a single dataset.

Furthermore, since mental health in general is not an easy factor to measure, for each of our analyses we utilized a different method of measuring mental health/ depression to better encompass the complexity of the variable being considered. Additionally, the term "mental health" relates to a wide range of different conditions, hence with our second visualization, we investigated how some of these factors impacted different mental illnesses.

Depression Rates vs Various Geographic + Economic Factors Choropleth:

This visualization considered the impact of various geographic and economic factors on depression rates, on a state-by-state basis. As students, we often have conversations about the link between how cold it is, how sunny it is, and how our mental health is. As a result, we wanted to take a look at this factor, in addition to an economic factor like Average Family Income, and see if there was any correlation across the country as a whole. Are fewer people depressed in more sunny states, or are there other factors we can't see?

Through our Pearson correlation scatterplot for each independent variable we considered in relation to depression rates, we ended up finding that the correlations were much weaker than expected. On the map though, we were able to see that there were multiple cases where our initial hypotheses held – where states with very high temperatures, incomes, or sunlight exposure also had lower depression rates compared to others. Because of this, this specific visualization can reveal that while there are some states where the weather and socioeconomic status influence depression rates, there is most likely another, underlying group of factors that also contributes.

Mental Health vs Temperature Graph:

This visualization demonstrated how different mental health disorders had different trends throughout a twenty-year period. Surprisingly, all of the mental health statistics decreased significantly throughout the study. However, the paths for which each statistic decreased varied. Moreover, when examining the temperature values there is no correlation between the average temperature of the year and the value of any of the mental health statistics. Thus, we cannot say that there is any relation between national temperatures and the share of individuals with mental health disorders.

We want viewers to observe the different trends displayed in our graph between the different mental health statistics. Moreover, our interactive component of the temperatures encourages viewers to select the years they are curious about and see the average temperature for that year. Thus, through our visualization, we want to convey that not all mental health disorders follow the same trends and thus will not be affected by the same factors.

Mental Health vs Other Health Aspects Graph:

In this visualization, we considered the possible influence of other health aspects like hours of sleep, body weight, and alcohol consumption on mental health. In this analysis, mental health was measured in the # of days of poor mental health experienced in a month, and a contour plot was utilized to determine the spread of the data in terms of both variables. When considering the results of the visualization, there does not exist a strong correlation between poor mental health and any of the other aspects measured, but rather, each of the visualizations better reflect the overall trends that exist independent of each other. For example, in terms of hours of sleep and mental health, there did not seem to be any positive or negative correlation, but rather, the highest density of the contour plots seemed to hover around the national averages for each of the datasets rather than any dependent correlation between them.

Team Contributions

Aditya	Worked with Bhadra to create visualization allowing the comparison of Average Temperature, Sunlight, and Family Income to Depression Rates on two choropleth maps, as well as button functionality. Participated in group planning meetings and dataset collecting. Added styling, and worked on the final report.
Bhadra	Worked with Aditya to create visualization allowing the comparison of Average Temperature, Sunlight, and Family Income to Depression Rates on two choropleth maps, as well as button functionality. Took part in meetings concerning project planning and data collection. Made Pearson correlation scatterplot and state mouseover tooltips that interacted with buttons. Added styling, and worked on the final report.
Gabby	Created visualization regarding yearly temperatures and different mental health statistics. Added interactive elements to the graph, including buttons that change the datasets and axis of the graph, as well as mouse-over functions to display temperatures. Filtered dataset for this graph to only include US statistics. Worked on a report and planned Zoom meetings for collaborative sessions.

Richie	Created contour plot visualization for mental health and other health aspects. Implemented a contour density plot for mental health vs each of the other health aspects including hours of sleep, body weight, and alcohol consumption. Implemented interactivity to allow users to swap between graphs which will change the dataset being presented as well as the x-axis scale and gridlines. Worked on the final report and participated in group collaboration sessions.