

## **CS 3300 Project 1 Description - aj274, aw592, ldp54**

**Dataset** To determine the density of health support in each state, we used the May 2014 State Occupational Employment and Wage Estimate dataset from the Bureau of Labor Statistics. For each state, we collected total employment for the following jobs with job codes: Mental Health Counselors (21-1014), Child Family and School Social Workers, Healthcare Social Workers, Mental Health and Substance Abuse Social Workers, Social Workers All Other (21-1021 through 21-1029), Psychiatrists (29-1066), Clinical Counseling and School Psychologists, and Psychologists (All Other) (19-3031 through 19-3039) and then divided the total by population of the state which we found from the US Census Bureau. We collected the Employment per 1000 jobs datapoint for each job as well. We also looked the percentage of veterans in each state, which was found from the US Department of Veteran Affairs using the VetPop2014 Document. Income statistics (normalized for consumer price index and regional property price) were obtained from The Bureau of Economic Analysis. Mental health data was found through the CDC and was accrued through the Behavioral Risk Factor Surveillance System.

### ***Mapping of Datasets***

**Map** The map of the US was fairly straightforward - on a linear scale, three colors were chosen as stop points on our gradient to represent the severity of the mental health problems in each state. Each percent value was passed through the linear scaling function, and the output was set as the fill color for each state path. The top 5 and bottom 5 states in terms of percent were highlighted with a drop shadow to emphasize our focus in the radar plots.

**Scatters** The two scatter plots used linear axes to plot the income and percentage affected (scatter 1) as well as mental health worker density and percentage affected (scatter 2). The radius of circles in both scatter plots were scaled linearly from the percentage of veterans per

state, and the color of the circles were also scaled linearly using the same scale function in the map. A line of best fit was added to visualize any correlation in data.

**Radar Plot** In our radar plot, each “arm” represents a certain statistic, which is then plotted with a linear scale towards the direction of the arm. The further out the shape of the plot extends towards one arm, the higher the value that state has for the statistic. This results in a 2-dimensional geometric representation of a 4-statistic dataset. By using two colors to distinguish between the US average and the state we are plotting, we are able to visually compare each state to the average and see where they differ.

**Story** We wanted to take a look at different statistics relative to each state, and see if there was any relation between those numbers and the frequency of mental health issues. We started with a broad view of the United States based on mental distress alone, and we noticed that the top 5 and bottom 5 states in terms of mental health issues exhibited some sort of pattern - the highest rates of mental health issues seemed to belong to lower income Southern states, while the lowest rates of mental health issues seemed to be from lower population states in the Midwest (except Hawaii).

We then went to the statistical analysis side of things - we researched statistics that were related to mental health issues, and we concluded that income was a strongly correlated factor when discussing these problems. Since veterans with mental issues are also a popular point of discussion, we decided to take a look at that as well. As expected, there seemed to be a strong negative correlation between mental health issues and normalized income - our line of best fit suggests that trend as income decreases, mental health issues increase. When looking at the percentage of veterans in each state, we noticed that there was a slight grouping in our representation space - similarly sized circles seemed to be somewhat in the same area.

With this information, we decided to look at another statistic compared to mental health. It seemed logical to us that the availability of mental health workers would be heavily correlated to mental health issues, so we plotted the data in another scatter plot, where the y-axis was changed from income to mental health worker density. Surprisingly enough, there

was a very weak, if not absent correlation between the two statistics. This could have been explained by a multitude of reasons, one compelling argument of which is just sampling error for less populous states and outside, “invisible” factors.

Since mental health worker density did not seem to be correlated from our plots, we decided to take a more holistic approach. We thus focused on the top 5 best and worst states in terms of mental health issues. What made them different from the rest of the states? What does a “good” state look like, as opposed to a “bad” state, and what does the average even look like? We drew inspiration for a radar representation from Thomas Preusse’s Radar Chart Visualization and decided to take that route. By geometrically plotting a multitude of statistics which are hypothesized to be related to mental health, we can reduce the number of dimensions needed to represent our data from  $n$ -statistics to just 2 dimensions. More importantly, we can tell which statistics differ and, more importantly, how they differ, at a glance. From our radar charts, we saw that all of the most distressed states had different levels of mental health workers per thousand, suggesting that there isn’t much correlation with that statistic. However, total health workers (physical and mental) were either at or below average for the worst states. Percentage of veterans seemed to be along the average, while income was below average. Looking at the lowest distress states, we can see that most points fit almost perfectly with the national average, except for Hawaii’s lower income and North Dakota’s lower mental health workers per thousand. Minnesota, however, has greater-than-average values for every statistic. Hawaii’s status as a state with both lower income than average and yet the lowest amount of mental distress can be attributed to outside factors, including but not limited to climate and way of life.

From the data, we have learned that income is perhaps the most strongly correlated factor when talking about mental health and distress. While the availability of mental health workers might be uncorrelated in our representation, we believe that it is attributed to either outside influences or sampling error. In addition, overall health workers seem to be an important statistic - out of the 5 “worst” states, 2 had very low health workers compared to the average. Out of the 5 “best,” however, none were found to be as lacking in that department. We believe that improving mental health takes a holistic approach as implied by our data.