Our plan is to have each factor owned by one of the team members. We study the datasets and try to find the relationship between the factor and depression and anxiety.

Our initial observation of the dataset BRFSS is that, there are two sets of data. A set belongs to years 1995 to 2010 [1] and another set from 2011 to 2017 [2]. The reason for distinguishing the dataset for these range of years is that the questionnaire changed after 2011. It gathered not only from land line but also from cellphone. Questions where updated to contain more recent related topics such as internet usage. And some territories such as Preto Rico where also included in the list of states (territories) to conduct the questionnaire.

We started our research by providing the dataset to all team member and individually investigating the dataset with the topic that research member owned. The combined data is around 1GB and we have used multiple tools to observe the dataset. We used Microsoft Excel, Numpy [4], and opendatanetowrk.com [3] to scan and skim the data.

We noticed that BRFSS are questions about the state itself than individuals. Which means that, if someone answered that they are smoker, we cannot connect this person to another question such as having depression and, consequently, draw conclusion on that particular person. Instead, the BRFSS is a questionnaire for the state and shows how many people of the sampled population from a particular state are smoker or are depressed. Hence, our conclusion would have similar nature too.

BRFSS has two sets of questions. The core question and optional one. The core questions are asked by all state participating in the BRFSS program but the optional one is not mandatory on states. Hence, some of the questions that were very interesting and important to our research, were not used in all states. We think that some topics may end up more generalized or tweaked in order to get some answers from the dataset. For example, social media is an important factor but this is not part of BRFSS questionnaire. We did find that the dataset after 2011 does have data on internet usage which we thought we can bring internet usage as a factor that encompasses social media. We have these factors tweaked or generalized:

Eating disorder to eating vegetable and fruits, social media usage to internet usage, and technology and education to education grade level.

To make this research more challenging, the depression factor is not called out in the dataset from prior 2011. Depression and anxiety are optional questions which we may combine “overall health status” as a combined factor with depression to understand how the six factors affects human health. And to make the matter worse, anxiety was not mentioned in both datasets. It is also an optional topic for BRFSS.

As for physical activity, we found data on both datasets that relates to physical activity. Our initial analysis showed that low physical activity led to low level of overall health and increase of depression. We found similar initial results with respect to alcohol and smoking consumption and overall health status.

With respect to social media factor, as per the data, we will be finding the relationship between internet vs depression and anxiety. Using the dataset, we can analyze the categories of internet use based off of several factors such as gender, race, income, age and several more. We can see whether there is a direct correlation between internet use and depression and anxiety by comparing the depression rate of those factors. If they are very similar, we can see that internet use is a direct cause of depression and if they are not, there may be other factors that were not accounted for.

Regarding the correlation between education level and the chance of having depression, we picked some state that use the depression we examined information such as whether an individual has been informed to have depressive disorder, including depression, major depression, dysthymia, or minor depression, and the level of education one attained. Based on initial analysis, it seems the higher the level of education one completed, the higher the chance of having mental issues. We are going to put extra attention on this factor to understand the nature of this initial results.

*Education/technology (education grade level) … we are going to do.*

*Eating disorder (eating vegetable and fruit) … we are going to do.*