Depression Analysis

Introduction

Depression is a serious mental illness that manifests as a general lack of initiative and interest in activities, along with symptoms including melancholy, a sense of emptiness, anxiety, and sleep disturbances. In addition, characteristics including a sense of shame or worthlessness, low energy, difficulties concentrating, suicidality, and psychotic symptoms may be present. The quantity, intensity, and length of symptoms, as well as the effects on social and occupational function, all contribute to how severe depression is. Another severe psychiatric disease known as bipolar disorder frequently includes depression. The fundamental distinction between unipolar depression and bipolar illness is the latter's tendency to periodically experience mania, which is characterized by elevated self-esteem, impulsivity, increased activity, decreased sleep, and goal-directed behavior. Both diseases are hereditary abnormalities, which can be thought of as a genetic susceptibility to environmental disruptions that could possibly set off mood swings. In addition to being linked to lifestyles associated with diurnal rhythms that are out of sync with the natural daylight cycle, depression is associated with disrupted biological rhythms brought on by environmental disturbances like seasonal changes in daylight and alteration of social rhythms due to shift work or longitude travel. In addition to alcohol and drug misuse, other variables that may contribute to the development of depressive symptoms include physical health problems, medical side effects, life events, and social circumstances. All people may have these symptoms of depression as a result of these other reasons. Around 15% of people experience depression at some point in their lifetime, although the occurrence of episodes whose intensity does not match the criteria for a depressive diagnosis is much more common. Actigraph recordings of motor activity are regarded as an objective way to observe depression, yet psychiatric research on this subject is far from complete.

Related literature

In the paper Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing (1) researchers proposed a new approach to predicting depression using passive sensing data from students' smartphones and wearables. They proposed a set of symptom features that proxy the DSM-5 defined depression symptoms specifically designed for college students. They identified a number of important new associations between symptom features and student self reported PHQ-8 and PHQ-4 depression scores. The study captured depression dynamics of the students at the beginning and end of term using a pre-post PHQ-8 and week by week changes using a weekly administered PHQ-4. They showed that symptom features derived from phone and wearable sensors can predict whether or not a student is depressed on a week by week basis with 81.5% recall and 69.1% precision.

In the paper Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors (2) the study evaluates the feasibility and performance of assessing depressive symptom severity by using behavioral and physiological features obtained from wristband and smartphone sensors. They used three machine-learning models estimating depressive symptom severity, one combining features from smartphone and wearable sensors, one including only features from the smartphones, and one including features from wrist sensors and evaluated in two different scenarios. They concluded that monitoring of MDD patients through smartphones and wrist sensors following a clinician-rated HDRS assessment is feasible and may provide an estimate of changes in depressive symptom severity.

In the paper Current Advances in Wearable Devices and Their Sensors in Patients With Depression (3) a literature survey was conducted of research into the development and use of wearable devices and sensors in patients with depression. They collected 18 studies that had investigated wearable devices for assessment, monitoring, or prediction of depression. In the report they examined the sensors of the various types of wearable devices (e.g., actigraphy units, wristbands, fitness trackers, and smartwatches) and parameters measured through sensors in people with depression. In addition, they discussed future trends, referring to research in other areas employing wearable devices, and suggested the challenges of using wearable devices in the field of depression. Real-time objective monitoring of symptoms and novel approaches for diagnosis and treatment using wearable devices will lead to changes in management of patients with depression. During the process, it is necessary to overcome several issues, including limited types of collected data, reliability, user adherence, and privacy concerns.

Problem Formulation

In this project the goal is to find out how depression can be analyzed using actigraph data. We try to understand whether the activity level of the subject contributes to the depression of the subject. The data used in this project contains condition and control group and the activity measurements for each individual subject. Using this data it is assessed whether there is significant difference between the activity level of the two groups and also between the individuals.

Dataset Description

The dataset for this project is obtained from kaggle. The dataset contains two folders, whereas one contains the data for the controls and one for the condition group. For each patient a csv file has been provided containing the actigraph data collected over time. The columns are: timestamp (one minute intervals), date (date of measurement), activity (activity measurement from the actigraph watch). In addition, the MADRS scores are provided in the file "scores.csv". It contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

Methods

The problem is assessed by using exploratory data analysis techniques. The methods contain data cleaning and different types of data visualization and data statistics to support the observations. These methods are described below and the plots are provided in the results section. The plots are also available in the jupyter notebook which also contains different summary statistics related to the observations.

Data cleaning

Many of the subject records contain a long period of several days where no or almost no activity is recorded. This may be because the sensor was offline or it was not working properly. Those data points are removed.

Data analysis

The data analysis is divided into three parts. First the subject level difference is explored and the goal is to try to understand whether there is difference between individuals in the control and condition group.

Secondly the two groups are explored and assessed whether there is difference between the groups and their activity levels. This is done by two different methods. First the means of both groups is compared and investigated whether there is difference. Secondly the zero activity counts of both groups are explored and also examined if there is difference between the two groups.

Thirdly, the effect of different categorical features on the madrs scores is examined. This is not related to the activity data but still provides relevant information related to the problem in hand.

Results

Subject level

In figure 1 below is plotted the activity of condition subject number 1. The activity contains only the non-negative values and it is transformed into the log scale.

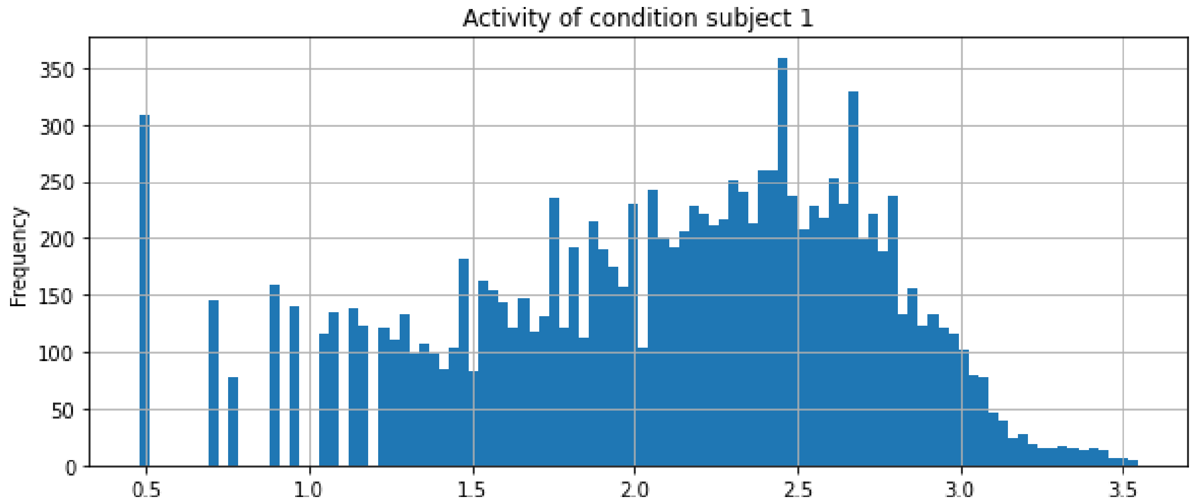

Figure 1. Activity of condition subject 1.

In figure 2 below is plotted the activity of control subject number 3. The activity contains only the non-negative values and it is transformed into the log scale.

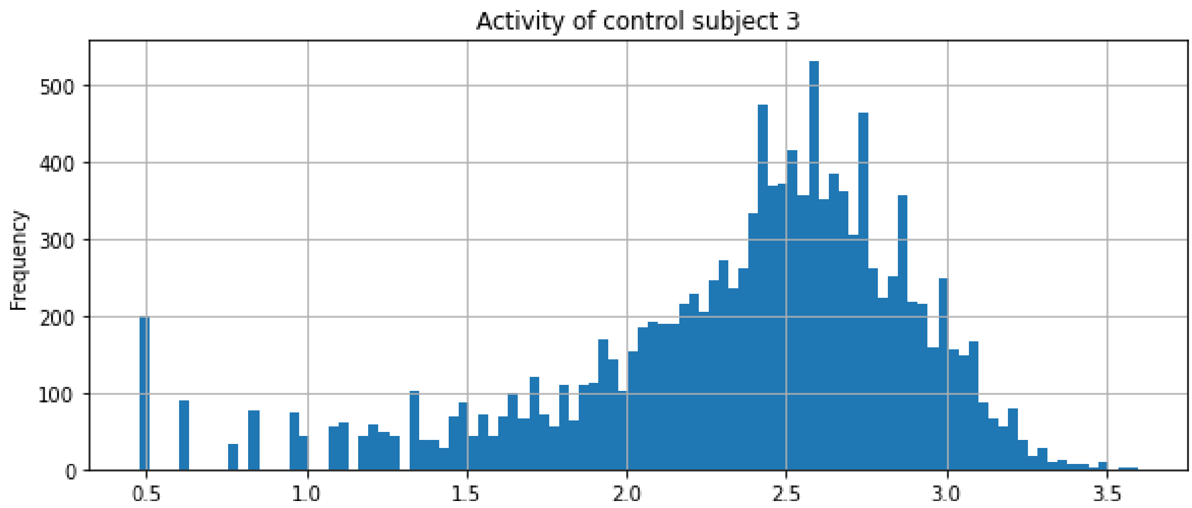

Figure 2. Activity of condition subject 3.

Already from these two plots it can be seen that the control subject has more mass in the histogram in the high values than the condition subject, meaning that the control subject has more measurements of higher activity.

Group level

The biggest observation and probably the most predicted one is that the control group has higher daily average activity than the condition group. This can be seen from figure 3 below.

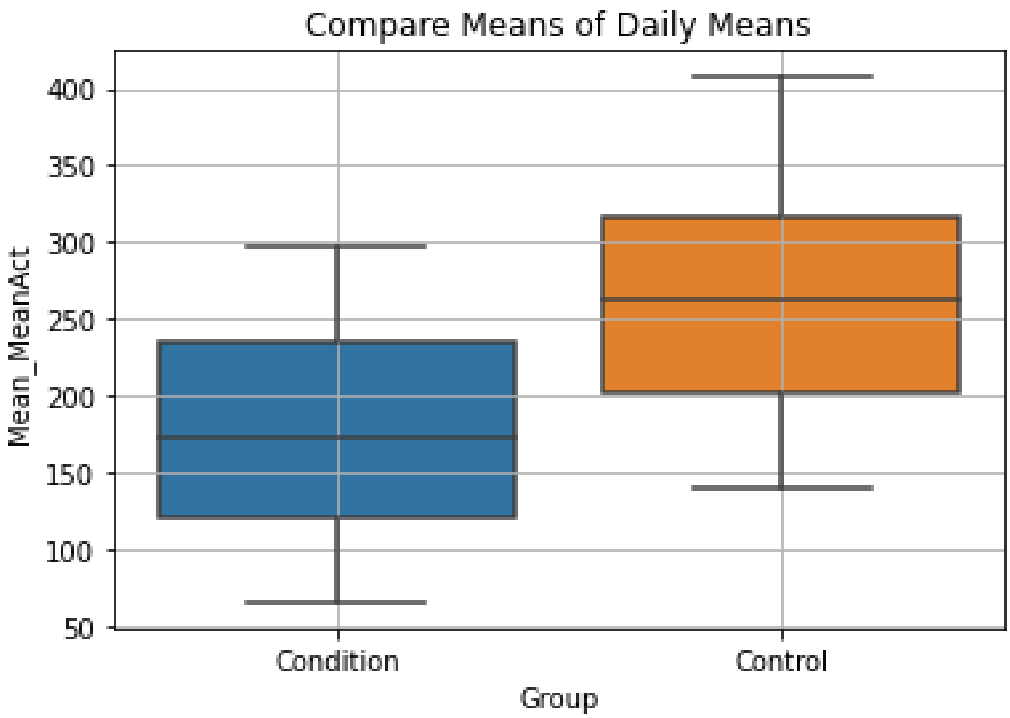

Figure 3. average daily activity means.

In figure 4 below the zero activity count of both groups is plotted. In the x axis are the different dates in the measurement period and in the y axis is the zero activity proportion. From this figure it can be seen that the condition group has a higher proportion of zero activity than the control group. This is inline with the previous observation from figure 3 which illustrates the difference between the group means.

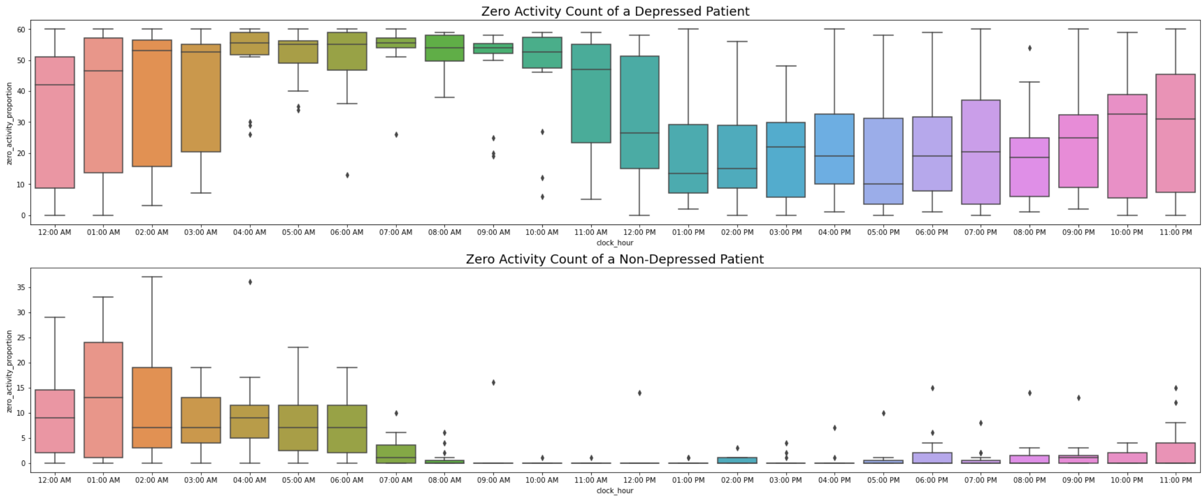

Figure 4. Zero activity count of both groups.

Effect of categorical features in the MADRS scores

The effect of the categorical features on the MADRS scores were also examined. This is done because we want to also understand which features of subjects contribute to their depression the most. This does not directly relate to the activity of an individual but is anyway good insight.

From the figure 5 below one can see that the madrs score is different between different age groups. For example the median of the madrs score is highest in the age group of 30-34.

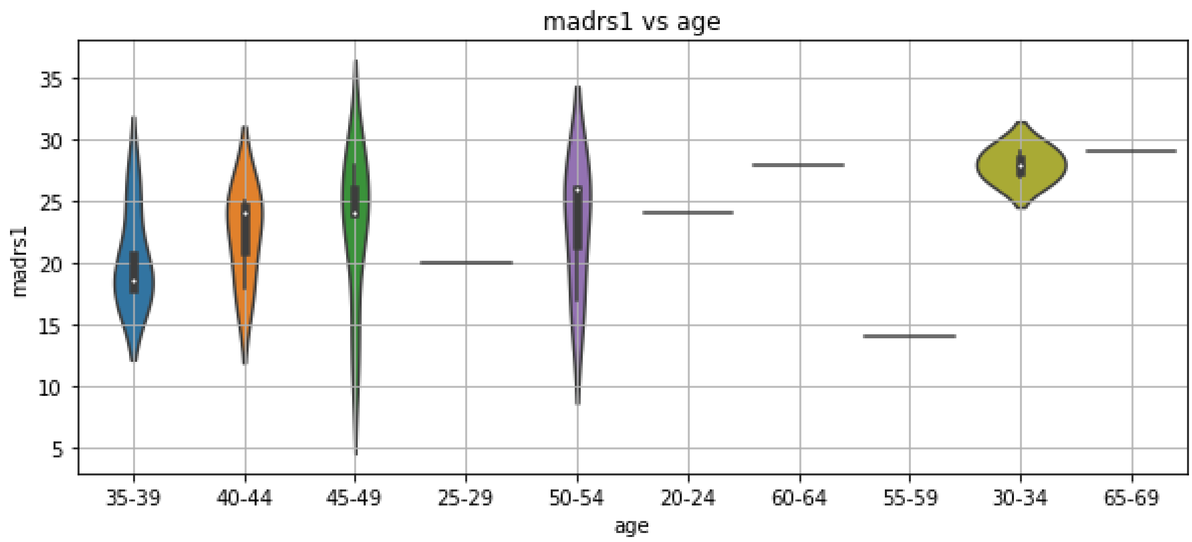

Figure 5. Madrs score between different age groups

From the figure 6 below one can see that the person working or studying hase lower median madrs score than those who are unemployed, on sick lean or in pension.

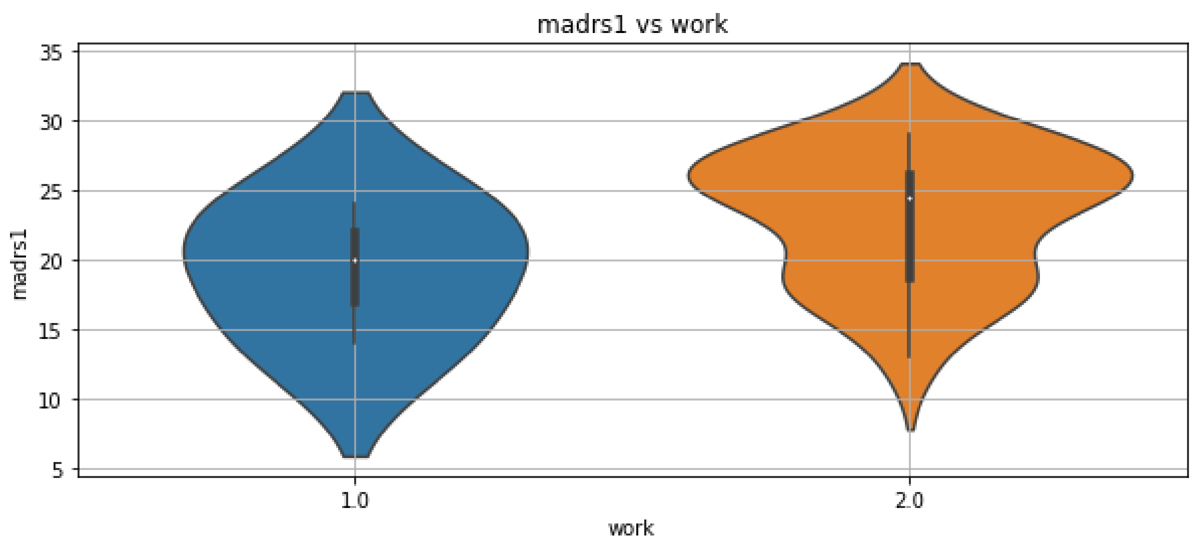

Figure 6. Madrs score between working/studying (1) and unemployed/sick leave/pension(2)

From the figure 7 below one can see that the person married or cohabiting has lower median madrs score than those who are single.

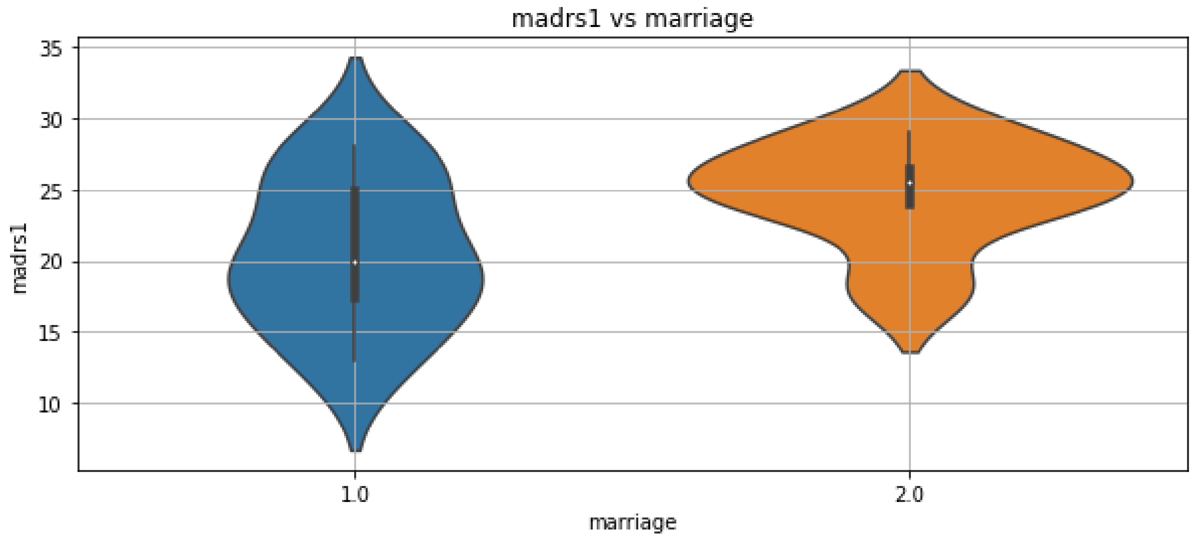

Figure 7. Madrs score between marriage/cohabiting(1) and single(2)

In the figure 8 below there seem to be no big differences between genders.

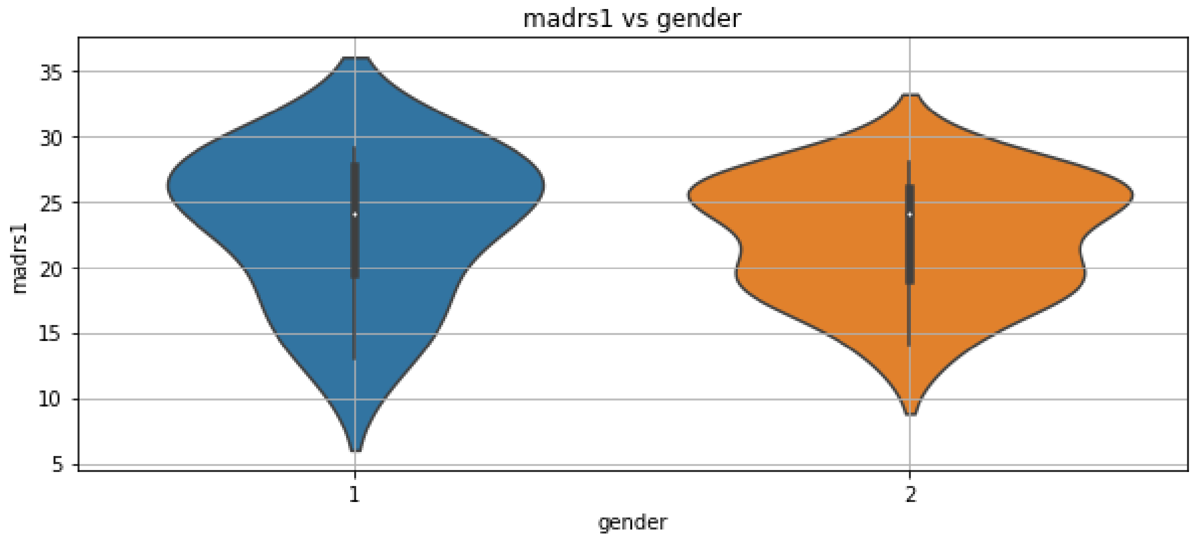

Figure 8. Madrs score between genders

Surprisingly people with higher education (16-20 years) have the highest madrs score.

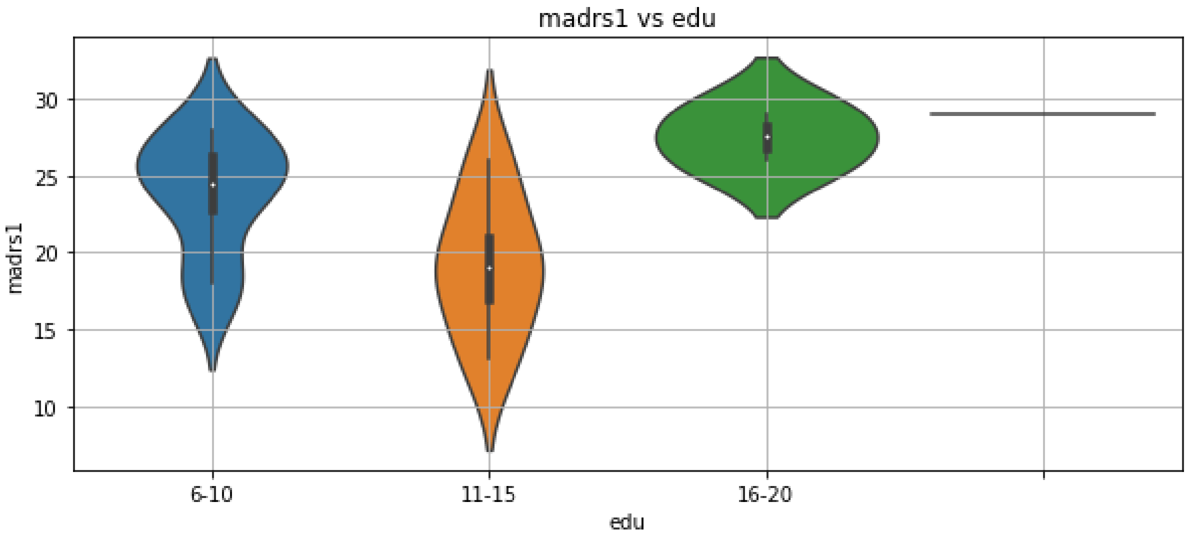

Figure 4. Madrs score between education years

Conclusion & Discussion

The most important observation from this project is that daily activity significantly contributes to depression scores. This result can most clearly be seen from figure 2. where the means of both control and condition groups are plotted. This is not perhaps a surprise since activity is always encouraged for better physical and mental health. The result can also be seen from figure 3. where the zero activity of both control and condition group is plotted. The condition group has clearly much more zero activity data than the control group. This project is mainly based on data visualization techniques and therefore limited to what can be interpreted from those figures

In the future work the use of machine learning algorithms could be investigated to predict the depression of an individual based on their activity data. Also by analyzing individuals activity data online in realtime one could suggest daily activity levels to prevent depression.

References

1. Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1, Article 43 (March 2018), 26 pages. https://doi.org/10.1145/3191775

2. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, Fisher LB, Cusin C, Nyer M, Yeung A, Sangermano L, Mischoulon D, Alpert JE and Picard RW (2020) Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors. Front. Psychiatry 11:584711. doi: 10.3389/fpsyt.2020.584711

3. Lee S, Kim H, Park MJ and Jeon HJ (2021) Current Advances in Wearable Devices and Their Sensors in Patients With Depression. Front. Psychiatry 12:672347. doi: 10.3389/fpsyt.2021.672347