Investigating Association between Physical Activity and Sleep Quality with Accelerometry Data

Team 4: Chengcheng Feng, Weiyi Li, Jingman Li, Jiyoon Lee, Xinyue Qiu
BIS 687 Final Project Report
May 3, 2023

Introduction

The ability to predict human sleeping conditions is of significant importance. Firstly, sleep is a critical component of overall health and well-being, and disruptions in sleep patterns have been linked to a range of negative health outcomes, including cardiovascular disease, obesity, and mental health issues. Accurately predicting sleep patterns could therefore help identify individuals who are at increased risk of these health problems and allow for early intervention and prevention strategies. Secondly, by predicting sleep conditions, healthcare providers can more accurately diagnose and treat sleeping disorders, improving patients' quality of life and reducing the risk of negative health outcomes.

With the emerging needs, we have identified that prior research relied heavily on self-reported measures of activity intensity and lacked the precision needed to accurately establish the association between physical activity and sleep quality in clinical trials. The difficulty in quantifying total physical activity across different levels of intensity further complicates the issue. To overcome these limitations, objective measures of physical activity have been developed and complement self-reported assessments. One such method is the use of accelerometry with wearable devices, which has gained widespread acceptance in recent years.

Accelerometers are sensors that detect the acceleration of objects in motion along reference axes, and they provide a reliable and objective means of measuring physical activity. Accelerometry can be employed with specific signal processing techniques to extract physical activity information, which can then be utilized to characterize activity levels (Willetts et al., 2018; Ravi et al., 2005; Yang & Hsu, 2010). Further, several studies have confirmed the associations between accelerometry-derived measures of physical activity or sleep and certain diseases such as obesity, cardiovascular disease, cardio-metabolic disease, breast cancer, psychiatric disorder, and Parkinson's disease (Guo et al., 2019; Guo et al., 2020; Ramakrishnan et al., 2021; Barker et al., 2019; Cassidy et al., 2018; Dennison et al., 2021; Nikbakhtian et al., 2021).

Provided with previous research, we have decided to utilize data from UK Biobank. UK Biobank data contains self-reported biomedical variables with summary acceleration data, as well as a variety of self-reported sleep evaluations. This will allow us to explore the relationship between activity intensity and sleep quality with the aid of accelerometers. In this study, we will employ a hybrid approach that combines statistical analysis and machine learning techniques to the UK Biobank dataset.

Our hypothesis is that there is a substantial variation in the frequency and intensity of physical activity among individuals with varying sleep conditions. A further hypothesis is that using accelerometry data can be a supplement to predict human sleep conditions in conjunction with other self-reported biomedical variables. Thus we will test these hypotheses in the following Specific Aims:

Specific Aim 1. To determine if participants with different sleep conditions have different activity levels/scores.

Specific Aim 2. To evaluate whether the inclusion of accelerometry data can enhance the precision of predicting sleeping conditions.

Specific Aim 1

Hypothesis and Rationale

In aim 1, we are trying to figure out whether different sleep qualities will have significantly different performance in each intensity of accelerometry-derived physical activity.

As described in the previous section, numerous studies have explored the link between physical activity and sleep, and most conclude that certain types of physical activity improve sleep quality and duration. However, different exercise intensities may have opposite effects on sleep quality. For instance, some studies have found that moderate exercise training over the course of several weeks can improve sleep quality and duration for adolescents, whereas vigorous exercise during the same timespan has been shown to decrease sleep duration for some teens.

Experimental Approach

In order to quantify the intensity of physical activity, only patients with accelerometers will be considered in our experiments. Accelerometers will record the values for each patient during activities, which can be used to classify activity intensities. Multiple two-sample t tests will be conducted to compare differences of physical activity performance between different sleep condition groups. In each test, a categorical variable which measures sleep quality will be used as the outcome variable, and the independent variable will be the sum of values at each intensity for each patient.

Due to the positive correlation between activity intensity and acceleration, acceleration can be utilized to classify activities into low, moderate, and high intensity levels. The summation of all fractions within each intensity level reflects the proportion of acceleration generated by different types of activities. For instance, the total fraction of acceleration within the range of 1 to 19 milli-gravities corresponds to the amount of low-intensity activity.

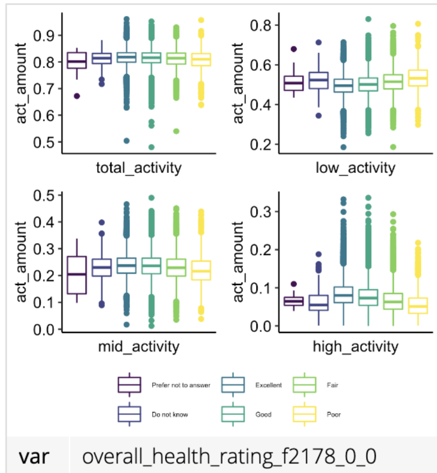
Acceleration	Activity
1 - 19 milli-gravities	Low-intensity Activity
20 - 95 milli-gravities	Mid-intensity Activity
100 - 1900 milli-gravities	High-intensity Activity
1 - 1900 milli-gravities	Total Activity

Table 1: Acceleration and Activity Intensity

Interpretation of Results

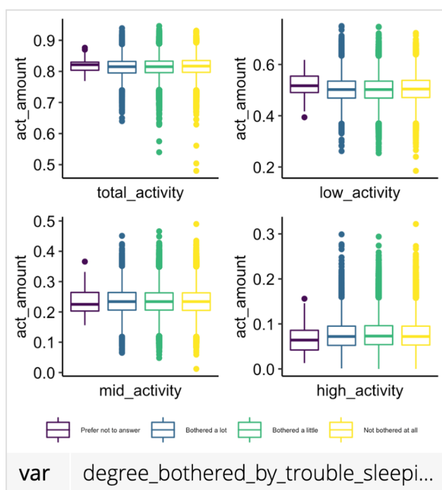
We conducted data-preprocessing and exploratory data analysis for the full dataset. We checked for missing values across different levels and stages and visualized the results by variable fields, which consisted of 470 distinct values. For instance, we used the alcohol drinker status variable to illustrate the visualization of missingness. The histogram displayed the study stage on the x-axis and the percentage of missing values on the y-axis for each stage. We observed that the percentage of missing values at baseline was comparatively negligible when compared to the follow-up period, a trend observed for most of the 470 values. Consequently, we used baseline variables for subsequent analysis. Moreover, we summarized the counts and percentages of missing values by category for the baseline data.

Subsequently, we generated box plots to display the distributions of activity scores among different sleep conditions for each categorical variable, with each subplot representing an intensity level. A total of 15 variables were included, comprising overall health rating, neuroticism score, and 13 sleep-related variables. The boxplots indicated that there was a significant difference in physical activity scores across different sleep condition groups in most cases. For instance, regarding the overall health rating, the means and distributions of activity amount differed across the six levels of health rating for all low, mid, and high activity levels. Notably, at the high activity level, the distribution of people with excellent overall health rating was significantly distinct from those with poor health rating.



Graph 1: Box plot of activity level for overall health rating

However, there are exceptions to this trend. Upon examining the mid activity subplot for the degree of bother caused by trouble sleeping in the last 3 months, we can observe that there is minimal variation across the four levels.

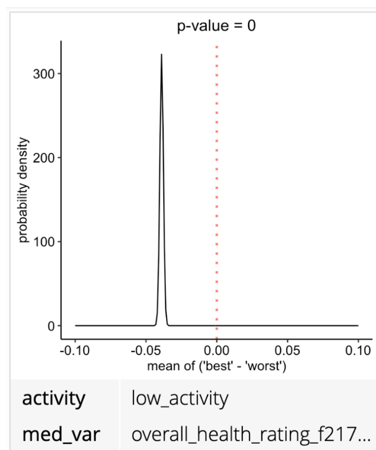


Graph 2: Box plot of activity level for degree bothered by trouble sleeping in the last 3 months

To verify our observations from the boxplots, we selected categorical variables pertaining to sleep conditions and employed several tests to evaluate our hypothesis. Considering our goals and purposes, we chose the t-test to compare the differences in physical activity performance among various sleep condition groups. Our null hypothesis was that there is no difference in the means of sleep quality among the three activity groups, while the alternative hypothesis was that there is at least one difference in means of sleep quality score among the three activity levels. We used the p-value to derive the final outcome.

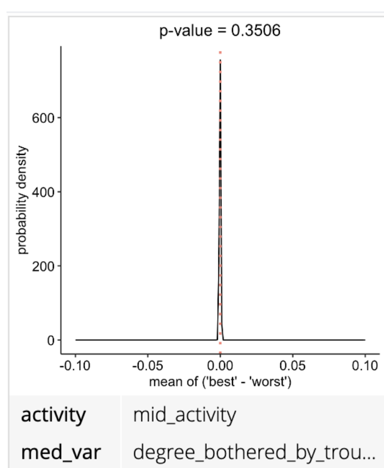
According to the Central Limit Theorem, the mean of users' activity amount with a medical condition ('best' or 'worst') approximates a normal distribution. Thus, the difference in mean activity amount between 'best' and 'worst' should also have a normal distribution. The p-value here was the minimum value between the p-values of testing if the mean difference was larger than 0 or smaller than 0. If the p-value was smaller than 0.05, then the activity amount difference between 'best' and 'worst' cases in sleep conditions was significant.

For instance, considering the overall health rating variable, we observed that the p-value was smaller than 0.05 for all three activity levels, indicating a significant difference in means and distributions of the three activity groups between individuals with excellent health ratings and those with poor health ratings. This finding was consistent with the boxplot results.



Graph 3: p-value and distribution of overall health rating at low activity level

On the other hand, for mid-activity values of "degree bothered by trouble sleeping in the last 3 months", the p-value is larger than 0.05. We do not have enough power to reject the null hypothesis and there is no significant difference between the means and distributions of mid activity across different cases in "degree bothered by trouble sleeping".



Graph 4: p-value and distribution of degree bothered by trouble sleeping in the last 3 months at mid activity level

Conclusion & Limitations

By analyzing all of the results of the t-tests, we can conclude that there is a large difference in the means and distributions of different activity groups across different sleep condition groups for most of the categorical variables listed, but not for sleep duration, neuroticism score, and degree bothered by trouble sleeping in the last 3 months.

While the dissimilarities among activity groups have been acknowledged in relation to the majority of sleep conditions, causality cannot be definitively attributed to a single factor. Further research is needed to establish a more comprehensive understanding of the relationship between activity levels and sleep conditions.

Specific Aim 2:
Hypothesis and Rationale

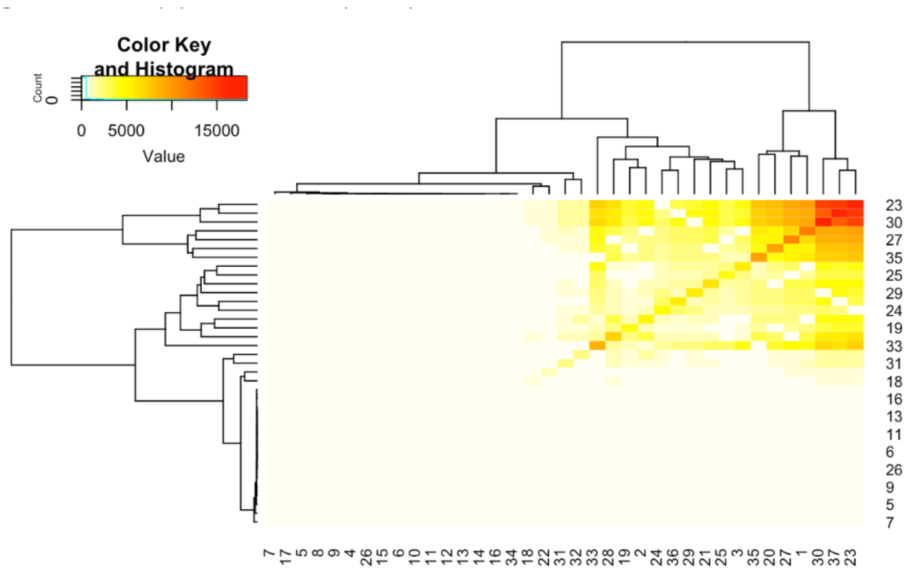
In aim 2, we are trying to answer the research question: Does including the accelerometry data enhance the accuracy of predicting sleeping conditions? We hypothesize that the activity level scores measured by accelerometers can significantly predict sleep quality. This idea is based on the assumption that there is a relationship between physical activity and sleep quality, proved in aim 1. To investigate this relationship, the proposed experimental approach involves two steps: Firstly, a dependent variable called "sleep profile" will be constructed using a Louvain Clustering model, with input variables related to sleep profiles. This is because our dataset contains a variety of sleep quality related variables, which does not suit our requirement of predicting sleep quality through the random forest model. Thus we are creating a new variable named 'sleep profile' through Louvain clustering. Secondly, a Random Forest model will be used to predict the sleep profile based on accelerometer data. The independent variables in this step will include the activity scores and demographic variables, while the dependent variables will be the clustering-based sleep profile variables. Through this approach, the study aims to uncover the extent to which physical activity levels can be predictive of sleep quality.

Dependent Variable Construction:
Experimental Approach

In step 1, we are using louvain clustering to construct our dependent variable. The Louvain method is a useful unsupervised algorithm used to detect communities within large networks. The mechanism used to execute the Louvain method involves a hierarchical clustering algorithm that recursively merges communities into a single node. The algorithm then executes the modularity clustering on the condensed graphs. Modularity quantifies the quality of an assignment of nodes to communities, which is calculated based on the random forest algorithm. This algorithm maximizes the modularity score for each community. One of the best features of the Louvain method is that there is no need for community numbers input, making it highly efficient for clustering large networks. In our analysis, the clusters identified from louvain clustering will be used as factors of the 'sleep profile' variable, which will serve as our new dependent variable in step 2.

Interpretation of Results

In the Louvain clustering model, we used 13 sleep quality related variables in the UK Biobank dataset, that includes sleep duration, getting up in the morning, nap, snoring, etc. These variables are all self-evaluated. Before fitting the model, we checked the correlation map of selected variables to visualize their hierarchical structure and identify the fact that our variables are organized by its local structure(Chen et al, 2015). Given Louvain clustering is based on correlation, plotting the correlation matrix will allow us to see the correlation structure among these variables.



Graph 5: Correlation Matrix of Factorized Sleep Quality Related Variables

The analysis results confirmed the presence of a hierarchical structure in our dataset, as evidenced by several pairs of highly correlated variables (Graph 6 in the Appendix). For more detailed information on the variables tested and their corresponding numbers, please refer to Table 2 in the Appendix. Following this, we proceeded with clustering analysis using the full set of variables.

	Cluster Number	Frequency
1	9	3257
2	2	2578
3	7	2388
4	3	2108
5	10	1930
6	1	1817
7	6	1399
8	14	1096
9	8	1019
10	18	973

Showing 1 to 10 of 548 entries

	Cluster Number	Frequency
11	13	964
12	5	883
13	16	583
14	19	540
15	12	435
16	4	422
17	130	3
18	260	3
19	20	2
20	23	2

Showing 11 to 20 of 548 entries

Table 2: Cluster number and Cluster size

After clustering, we obtained a total of 548 clusters. Given the limited number of participants in our dataset, we chose to select the top 11 clusters with the highest number of participants for constructing the new dependent variable. The threshold was determined based on the number of participants in each cluster and the number of participants remaining after removing null values. Subsequently, we constructed a new dataset that only included participants assigned to these selected clusters.

Random Forest Model:
Experimental Approach

In step 2, we used random forest models to predict sleep quality. The Random Forest is an ensemble learning method that works by building multiple decision trees at training time. For classification, the output is the class selected by the most number of trees (Wikipedia contributors, 2023). The mechanism of how random forests are constructed is that each individual tree in the random forest outputs a class prediction, and the class with the most votes becomes the model prediction. There are two prerequisites for random forests to perform well. The first is there needs to be some actual signal in our features so that models built using those features do better than random guessing. The second is the predictions (and errors) made by the individual trees need to have low correlations with each other (Yiu, 2021).

Interpretation of Results

We built two random forest models to predict sleep quality, one model with demographic variables and the second model with demographic variables and accelerometer variables.

The first random forest model was built using clustering based sleep profile variables from step 1 as the independent variables and demographic variables as the dependent variables. The number of trees used to build the model was 500 and the number of variables tried at each split was 2. The out-of-bag (OOB) estimate of error rate was 83.46% and the accuracy was 16%.

The second random forest model was built using clustering based sleep profile variables from step 1 as the independent variables and demographic variables along with accelerometer variables as the dependent variables. The number of trees used to build the model was 500 and the number of variables tried at each split was 2. The out-of-bag (OOB) estimate of error rate was 79.74% and the accuracy was 19%. Adding accelerometer variables increased accuracy rate by 3%.

The variable importances of the dependent variables were calculated for the second prediction model. The most important variable was weight (2118.4141), followed by low_activity (2058.26818), mild_activity (2051.46715), high_activity (1946.05916), standing height (1691.86763), year of birth (1498.19034), and sex (92.05787). All activity score variables were ranked significant predictors of sleep quality.

Conclusion & Limitations

Random Forest prediction models that predict sleep quality increases in accuracy rate when the independent variables include demographic and accelerometer variables. Based on variable significance from random forest, activity score variables are more significant predictors of sleep quality than demographic variables

There were two main limitations for aim 2. The first one was that the sleep profile variables constructed were not as accurate. We are limited to 11 levels because of restrained sample size in each level, because NAs were removed and the dataset was split into test and train data. A potential improvement for this issue could be to include more samples for a broader coverage of sleep profiles.

Our second limitation was that there was loss of significant information during the aggregation process. Independent variables (especially the activity scores) represent summary data, which could have resulted in the loss of significant information during the aggregation process. This could be improved by taking full advantage of the raw accelerometer data.

Conclusion

In summary, this study has demonstrated the usefulness of accelerometer data in predicting sleep conditions, and the potential of using machine learning techniques to predict sleep quality based on UK Biobank dataset. Our results from the analysis have suggested that activity scores are significantly correlated with the sleep qualities among participants, and the incorporation of accelerometry data improved the model's performance. This indicates that accelerometer data could be a valuable supplement to self-reported biomedical information for predicting sleep conditions. We could also learn that when predicting sleep quality, activity scores are statistically significant variables along demographic variables.

However, we also learned that our analysis is limited from various aspects. Our random forest model's predicting accuracy remained relatively low, even after incorporating activity scores as its independent variables. Several explanations underlie: summarization of accelerometry data using activity levels could lead to loss of information, and the self-reported sleep quality related variables are also subject to severe imprecision. This issue could be addressed by using raw accelerometry data, more precise and objective data of sleep quality, and employing more deep learning models for prediction.

In the long term, we hope this report serves as an enlightening article to researchers analyzing sleep quality and activity levels, as well as the potential benefits of using accelerometry data in predicting sleep conditions. Further studies could explore the use of more advanced analytical methods and larger datasets to improve the accuracy of sleep condition prediction. Ultimately, we hope this research can serve as a base to developing more effective interventions to promote healthy sleep habits and improve overall health outcomes.

References

Willetts, M., Hollowell, S., Aslett, L., Holmes, C. & Doherty, A. (2018), ‘Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 uk biobank participants’, *Scientific reports* **8**(1), 1–10.

Ravi, N., Dandekar, N., Mysore, P. & Littman, M. L. (2005), Activity recognition from accelerometer data, *in* ‘Aaai’, Vol. 5, Pittsburgh, PA, pp. 1541–1546.

Yang, C.-C. & Hsu, Y.-L. (2010), ‘A review of accelerometry-based wearable motion detectors for physical activity monitoring’, *Sensors* **10**(8), 7772–7788.

Guo, W., Key, T. J. & Reeves, G. K. (2019), ‘Accelerometer compared with questionnaire measures of physical activity in relation to body size and composition: a large cross-sectional analysis of uk biobank’, *BMJ open* **9**(1), e024206.

Guo, W., Fensom, G. K., Reeves, G. K. & Key, T. J. (2020), ‘Physical activity and breast cancer risk: results from the uk biobank prospective cohort’, *British journal of cancer* **122**(5), 726–732.

Ramakrishnan, R., Doherty, A., Smith-Byrne, K., Rahimi, K., Bennett, D., Woodward, M., Walmsley, R. & Dwyer, T. (2021), ‘Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the uk biobank cohort study’, *PLoS medicine* **18**(1), e1003487.

Barker, J., Smith Byrne, K., Doherty, A., Foster, C., Rahimi, K., Ramakrishnan, R., Woodward, M. & Dwyer, T. (2019), ‘Physical activity of uk adults with chronic disease: cross-sectional analysis of accelerometer-measured physical activity in 96 706 uk biobank participants’, *International Journal of Epidemiology* **48**(4), 1167–1174.

Cassidy, S., Fuller, H., Chau, J., Catt, M., Bauman, A. & Trenell, M. I. (2018), ‘Accelerometer-derived physical activity in those with cardio-metabolic disease compared to healthy adults: a uk biobank study of 52,556 participants’, *Acta diabetologica* **55**(9), 975–979.

Dennison, C. A., Legge, S. E., Bracher-Smith, M., Menzies, G., Escott-Price, V., Smith, D. J., Doherty, A. R., Owen, M. J., O’Donovan, M. C. & Walters, J. T. (2021), ‘Association of genetic liability for psychiatric disorders with accelerometer-assessed physical activity in the uk biobank’, PloS one 16(3), e0249189.

Nikbakhtian, S., Reed, A. B., Obika, B. D., Morelli, D., Cunningham, A. C., Aral, M. & Plans, D. (2021), ‘Accelerometer-derived sleep onset timing and cardiovascular disease incidence: a uk biobank cohort study’, European Heart Journal-Digital Health 2(4), 658–666.

Chen, Z., He, Z., Liang, H., Zhao, C., & Liu, Y. (2019). Correlation-Based Community Detection. arXiv preprint arXiv:1904.04583.

Wikipedia contributors. (2023). Random forest. *Wikipedia*
https://en.wikipedia.org/wiki/Random_forest

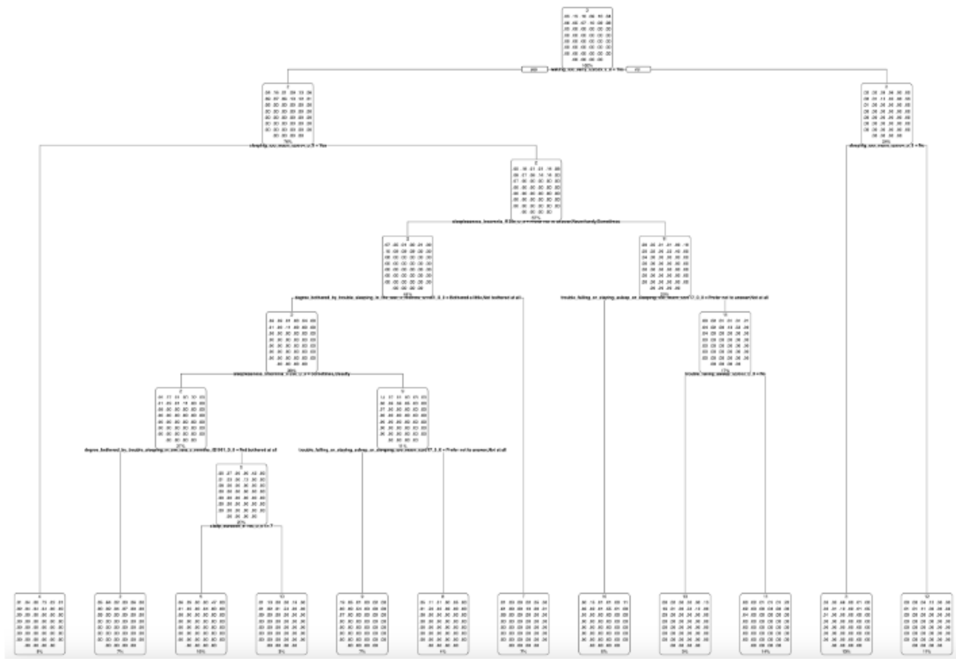
Yiu, T. (2021, December 10). Understanding Random Forest - Towards Data Science. *Medium*.
https://towardsdatascience.com/understanding-random-forest-58381e0602d2

Appendix

```
 [1] "degree_bothered_by_trouble_sleeping_in_the_last_3_months_f21061_0_0_Bothered a little"
 [2] "degree_bothered_by_trouble_sleeping_in_the_last_3_months_f21061_0_0_Bothered a lot"
 [3] "degree_bothered_by_trouble_sleeping_in_the_last_3_months_f21061_0_0_Not bothered at all"
 [4] "degree_bothered_by_trouble_sleeping_in_the_last_3_months_f21061_0_0_Prefer not to answer"
 [5] "sleep_duration_f1160_0_0_-1"
 [6] "sleep_duration_f1160_0_0_-3"
 [7] "sleep_duration_f1160_0_0_10"
 [8] "sleep_duration_f1160_0_0_11"
 [9] "sleep_duration_f1160_0_0_12"
[10] "sleep_duration_f1160_0_0_13"
[11] "sleep_duration_f1160_0_0_14"
[12] "sleep_duration_f1160_0_0_15"
[13] "sleep_duration_f1160_0_0_16"
[14] "sleep_duration_f1160_0_0_17"
[15] "sleep_duration_f1160_0_0_2"
[16] "sleep_duration_f1160_0_0_3"
[17] "sleep_duration_f1160_0_0_4"
[18] "sleep_duration_f1160_0_0_5"
[19] "sleep_duration_f1160_0_0_6"
[20] "sleep_duration_f1160_0_0_7"
[21] "sleep_duration_f1160_0_0_8"
[22] "sleep_duration_f1160_0_0_9"
[23] "sleeping_too_much_f20534_0_0_No"
[24] "sleeping_too_much_f20534_0_0_Yes"
[25] "sleeplessness_insomnia_f1200_0_0_Never/rarely"
[26] "sleeplessness_insomnia_f1200_0_0_Prefer not to answer"
[27] "sleeplessness_insomnia_f1200_0_0_Sometimes"
[28] "sleeplessness_insomnia_f1200_0_0_Usually"
[29] "trouble_falling_asleep_f20533_0_0_No"
[30] "trouble_falling_asleep_f20533_0_0_Yes"
[31] "trouble_falling_or_staying_asleep_or_sleeping_too_much_f20517_0_0_More than half the days"
[32] "trouble_falling_or_staying_asleep_or_sleeping_too_much_f20517_0_0_Nearly every day"
[33] "trouble_falling_or_staying_asleep_or_sleeping_too_much_f20517_0_0_Not at all"
[34] "trouble_falling_or_staying_asleep_or_sleeping_too_much_f20517_0_0_Prefer not to answer"
[35] "trouble_falling_or_staying_asleep_or_sleeping_too_much_f20517_0_0_Several days"
[36] "waking_too_early_f20535_0_0_No"
[37] "waking_too_early_f20535_0_0_Yes"
```

Table 2: Factorized Sleep Quality related Variables

Clustering Treemap



Graph 6. Treemap of Louvain Clustering