Abstract: **(April 19th)**
Introduction: **(March 29th)**

- Background
    - Literature review
- Exploratory Data Analysis
    - Include overall data missing values and distribution, check variable types

Specific Aim 1: **(April 12th)**
Do participants with different sleep conditions have different activity levels/scores

- Exploratory analysis and visualization
    - Missing values and correlation
    - Box plots of different sleep condition groups to generate initial prediction of final results
    - Visualization of accelerometry data in order to transform numerous accelerometry data into 3 activity values
- Assumptions and Test
    - Multiple t-test or anova tests will be conducted to compare differences in physical activity performance between different sleep condition groups.
    - For each test, our hypothesis assumes that there is no significant difference in the average physical activity value between the two groups being compared.
- Outputs and results
    - Using the t-statistics and p-value to draw inferences about the specific range of activity intensity that may impact particular sleep activities.
- Potential problems and Alternative Approches
    - Wilcoxon rank-sum test
    - Welch's test

Specific Aim 2: **(April 12th)**
Identification of statistical significant variables based on the full model.

- Design and Experience
    - Rationale
    - Experimental approach
- Exploratory Analysis and visualization
    - Missing data check and imputation
    - Correlation Heatmap
        - check the correlation between all included outcome variables
    - Distribution plots
        - Check distributions of all outcome variables, check if the data have enough observations, data should be structured
- Analysis
    - Random Forest analysis
- Outputs and results
    - Visualization of decision tree
    - Visualization of feature importance

- Testing
    - Validation of the model using test data
- Potential problems and alternative approaches
    - Add more variables

Conclusion: **(April 19th)**