CBSC 185 Final Project: How Does Physical Activity Impact Mood and Sleep Patterns?

Libby Ford

4/18/2024

# Project Introduction

For this project, I am examining the influence of vigorous physical activity on an individual’s mood and sleep patterns, as well as how these variables shift throughout the semester for the whole sample. This data is taken from CBSC 185: Trends Over Time, specifically the demographics and daily survey data self-reported from students. For this study I focused on the positive and negative mood indicators, time and satisfaction with physical activity, and hours and quality of sleep. My predictors in this study are physical activity and satisfaction with physical activity, and my outcomes are mood scores and quality and amount of sleep.

Using the variables discussed above, I performed various tests and visualizations to explore my two research questions: How does physical activity impact mood and sleep? And, does this impact change throughout the semester?

I made two predictions based on these research questions. First, that students participating in regular physical activity will have higher positive mood scores and more sleep on average. Then, students who are satisfied with their physical activity levels will have better quality sleep and higher mood scores. In my analysis, I also factored in student athlete status, as student athletes more regularly participate in physical activity and should experience better and more sleep, as well as higher moods on average.

# Part One: Data Wrangling

To begin this project, I first need to load in the raw data from our daily class survey and begin to wrangle this data to make it more manageable and in the format I need it in to easily manipulate it into visualizations and tests. Part one of this project details my thought and code process through the wrangling and combining of my key variables and the relevant identifiers.

## Loading Libraries

First, I must load the necessary libraries to complete my desired tasks.

## Importing Raw Data

Then, I need to load in the raw data set to begin wrangling and manipulating this data.

dsraw <- read.csv("~/fordcbsc185/final\_project/Daily Survey Raw Data.csv")

## Create a New Dataframe from Raw Data

Next, I need to create a new dataframe selecting only the variables I need for this project.

surveyraw <- select(dsraw, c("RecordedDate", "ID", "Q33\_1", "I.PANAS.C.P\_1", "I.PANAS.C.P\_2", "I.PANAS.C.P\_3", "I.PANAS.C.P\_4", "I.PANAS.C.P\_5", "I.PANAS.C.P\_6", "I.PANAS.C.P\_7", "I.PANAS.C.P\_8", "I.PANAS.C.P\_9", "I.PANAS.C.P\_10", "Q7\_1", "Q7\_2", "Q7\_3", "Q8", "Q9\_1","Q17\_1", "Q17\_2", "Q29\_1\_1", "Q29\_1\_2", "Q29\_2\_1", "Q29\_2\_2"))

## Write a new csv using the new dataframe and read that into this file

Now, I will create a new csv file of the dataframe I created above so that this is saved within my R project instead of just in my environment. Then, I will read that new csv file in as a dataframe to begin working with these selected variables.

write.csv(surveyraw,"~/fordcbsc185/final\_project/Newsurvey.csv")  
survey <- read.csv("~/fordcbsc185/final\_project/Newsurvey.csv")

## Rename the raw variables to more intuitive names

Now I am renaming all of my variables from the numbers and codes they were labelled to names that are more intuitive for the questions these variables are related to on my daily survey.

### Removing the rows that are tests from the beginning of the survey (filled with observations that say “delete me) and replacing mis-entered values

Here I am first removing the rows that were empty from students not wanting to fill out the survey that day as well as duplicates from students accidentally filling it out twice, and a sunday variable when someone filled this survey out on a non class day. I am also removing the test entries from this dataset and clarifying mis-entered data that was entered in the wrong format so that it does not skew my data.

## Mutate Variables

Here I am removing the time stamp at the end of the date variable to make it easier to understand and turn into a character.

### Mutate Variables to Integers

Now I am mutating all of the variables part from date and ID to integers, as they are numerical measurements of time spent doing activity or sleeping, as well as attitude scales.

### Read in Student-Athlete and In-Season Data

Here I am reading in the CSV that contains information on whether or not participating students are student athletes.

athlete1 <- read.csv("~/fordcbsc185/final\_project/W2024Ind Diff\_athlete.csv")

Now I need to remove the first row in this dataframe to get rid of what the different codes mean:

After looking at this dataframe, it appears we are missing the entry from ID 10 to identify if they are a student athlete or not. After considering this matter and looking at ID 10’s patterns of my key variables throughout the term, I decided they must not be a student athlete as they did not feel the need to add this demographic to their survey data, so I am going to create a new row in this dataframe for ID 10 and give them the “not student athlete” identifier. I then needed to put this created ID 10 in the correct place in my dataframe so that it could become the key variable to join to my survey data.

##### Investigating Missing Data

Here I will look at the missing data for the whole dataset and determine if rows need to be removed (and then remove them), if variables are all missing the same days, and if the data is missing completely at random to see how I should deal with this data.

miss\_var\_summary(survey)

## # A tibble: 25 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 restless\_scale 37 14.2   
## 2 time\_modact\_today 16 6.15   
## 3 hours\_sleep 15 5.77   
## 4 time\_vigact\_today 14 5.38   
## 5 rested\_scale 9 3.46   
## 6 time\_modact\_yesterday 9 3.46   
## 7 tired\_scale 5 1.92   
## 8 phys.act\_satisfaction\_today 4 1.54   
## 9 tvigact\_y 3 1.15   
## 10 positive\_scale 1 0.385  
## # ℹ 15 more rows

Here is where I am running the missing completely at random test.

survey %>%  
 select(-ID, -date)%>%  
 mcar\_test()

## # A tibble: 1 × 4  
## statistic df p.value missing.patterns  
## <dbl> <dbl> <dbl> <int>  
## 1 551. 458 0.00181 23

The p-value is less than 0.005 in this case, so the data is not missing completely at random. I believe this is because people are self-reporting this data and choose not to fill out certain questions on certain days for numerous reasons, not at random. I will separate my variables of interest and deal with the NAs in these categories separately based on the data and how I feel is best to deal with it.

## Subset mood variables and pivot to wide format

The next thing I need to do is select only the identifier and positive and negative mood variables and create a new dataframe for each. Then, I will pivot these mood dataframes into wide format to make it easier to create two new variables for average positive mood score and average negative mood score.

survey\_pmood <- select(survey,"date", "ID", "joyful", "cheerful", "happy", "lively", "proud")  
survey\_pmood\_wide <- pivot\_wider(survey\_pmood, names\_from = date, values\_from = c(joyful, cheerful, happy, lively, proud))  
  
survey\_nmood <- select(survey,"date", "ID", "miserable", "mad", "afraid", "scared", "sad")  
survey\_nmood\_wide <- pivot\_wider(survey\_nmood, names\_from = date, values\_from = c(miserable, mad, afraid, scared, sad))

##### Investigating Missing Data

Here I will look at the missing data in this subset to see how I should deal with these observations. This missing variable summary shows me which variables have missing data and what percentage of that data is missing.

miss\_var\_summary(survey\_nmood\_wide)

## # A tibble: 81 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "miserable\_3/19/2024 " 8 40  
## 2 "mad\_3/19/2024 " 8 40  
## 3 "afraid\_3/19/2024 " 8 40  
## 4 "scared\_3/19/2024 " 8 40  
## 5 "sad\_3/19/2024 " 8 40  
## 6 "miserable\_2/8/2024 " 6 30  
## 7 "miserable\_2/20/2024 " 6 30  
## 8 "miserable\_3/7/2024 " 6 30  
## 9 "mad\_2/8/2024 " 6 30  
## 10 "mad\_2/20/2024 " 6 30  
## # ℹ 71 more rows

miss\_var\_summary(survey\_pmood\_wide)

## # A tibble: 81 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "joyful\_3/19/2024 " 8 40  
## 2 "cheerful\_3/19/2024 " 8 40  
## 3 "happy\_3/19/2024 " 8 40  
## 4 "lively\_3/19/2024 " 8 40  
## 5 "proud\_3/19/2024 " 8 40  
## 6 "joyful\_2/8/2024 " 6 30  
## 7 "joyful\_2/20/2024 " 6 30  
## 8 "joyful\_3/21/2024 " 6 30  
## 9 "joyful\_3/7/2024 " 6 30  
## 10 "cheerful\_2/8/2024 " 6 30  
## # ℹ 71 more rows

Now I will visualize this missing data to have a better understanding of how to manage this data. Negative mood missing variable visualizations:

Positive mood missing variables: looking at the same problems here.

After investigating the NAs for both of these dataframes, I believe I need to replace the NAs from these two new dataframes so that I can more easily create average positive and negative mood variables to use for my manipulations and visualizations later in my project. I plan to replace these missing variables with ones because one is the lowest number on the mood scale. I am doing this because students are choosing not to enter their mood variables on these days, indicating that they are not feeling strong feelings of these moods because they do not feel the need to fill out these questions. I also believe that this will interfere the least with the overall data, as changing them to other numbers may skew the data in a way I do not want for my analysis.

survey\_nmood\_wide<- survey\_nmood\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), 1)))  
  
survey\_pmood\_wide<- survey\_pmood\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), 1)))

## Create Average Mood Variables

The new variables I am going to create are the average positive emotion score and the average negative emotion scores using the IPANAS mood data. I will use these in my visualizations as well as correlation and LMER tests and models. I will average the five different positive and negative mood variables for each individual timepoint to make a positive and negative mood score for each day. Then I will only need one variable to demonstrate positive and negative moods for each participant on each day.

Now I will do this same process for the positive mood variables:

Now I am going to recreate the pmood and nmood wide format dataframes with only these averages so that my data is simplified and I do not have to worry about these individual scores anymore.

survey\_nmood\_wide <- select(survey\_nmood\_wide, "ID", "3/19/2024\_Nmean":"1/23/2024\_Nmean")  
survey\_pmood\_wide <- select(survey\_pmood\_wide, "ID", "3/19/2024\_Pmean":"1/23/2024\_Pmean")

## Create new Dataframes for Physical Activity and Sleep

### Physical Activity Dataframes

#### Time Spent Doing Physical Activity

Here I am going to make new dataframes for my other key variables, starting with amount of physical activity so I can evaluate the missing data and arrangement of this variable.

survey\_time\_activity <- select(survey, "ID", "date", "tvigact\_y", "time\_vigact\_today", "time\_modact\_yesterday", "time\_modact\_today")  
time\_activity\_wide <- pivot\_wider(survey\_time\_activity, names\_from = date, values\_from = c(tvigact\_y, time\_vigact\_today, time\_modact\_yesterday, time\_modact\_today))

##### Investigating Missing Data

Here I will look at the missing data in this subset to see how I should deal with these observations. This missing variable summary shows me which variables have missing data and what percentage of that data is missing.

miss\_var\_summary(time\_activity\_wide)

## # A tibble: 65 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "tvigact\_y\_3/19/2024 " 8 40  
## 2 "time\_vigact\_today\_3/19/2024 " 8 40  
## 3 "time\_modact\_yesterday\_3/19/2024 " 8 40  
## 4 "time\_modact\_today\_3/19/2024 " 8 40  
## 5 "tvigact\_y\_3/7/2024 " 7 35  
## 6 "time\_vigact\_today\_3/7/2024 " 7 35  
## 7 "time\_modact\_yesterday\_2/8/2024 " 7 35  
## 8 "time\_modact\_yesterday\_3/7/2024 " 7 35  
## 9 "time\_modact\_today\_2/8/2024 " 7 35  
## 10 "time\_modact\_today\_3/7/2024 " 7 35  
## # ℹ 55 more rows

Now I will visualize this missing data to have a better understanding of how to manage this data.

Next, I am going to label the NAs in this dataset as zero because I am assuming if participants did not fill out these time boxes they did not complete any sort of moderate or vigorous physical activities, as a lot of my peers skipped these questions when they did not have anything to input in the boxes. I also believe that this will interfere the least with the overall data, as changing them to other numbers may skew the data in a way I do not want for my analysis.

time\_activity\_wide<- time\_activity\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), 0)))

#### Satisfaction for Physical Activity

Now I am going to create a dataframe for satisfaction with physical activity so I can evaluate the missing data and arrangement of this variable.

survey\_satis\_activity <- select(survey, "ID", "date", "phys\_sat\_y", "phys.act\_satisfaction\_today")  
satis\_activity\_wide <- pivot\_wider(survey\_satis\_activity, names\_from = date, values\_from = c(phys\_sat\_y, phys.act\_satisfaction\_today))

##### Investigating Missing Data

Here I will look at the missing data in this subset to see how I should deal with these observations. This missing variable summary shows me which variables have missing data and what percentage of that data is missing.

miss\_var\_summary(satis\_activity\_wide)

## # A tibble: 33 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "phys\_sat\_y\_3/19/2024 " 8 40  
## 2 "phys.act\_satisfaction\_today\_3/19/2024 " 8 40  
## 3 "phys\_sat\_y\_2/8/2024 " 6 30  
## 4 "phys\_sat\_y\_2/20/2024 " 6 30  
## 5 "phys\_sat\_y\_3/7/2024 " 6 30  
## 6 "phys.act\_satisfaction\_today\_2/8/2024 " 6 30  
## 7 "phys.act\_satisfaction\_today\_2/20/2024 " 6 30  
## 8 "phys.act\_satisfaction\_today\_3/7/2024 " 6 30  
## 9 "phys\_sat\_y\_2/13/2024 " 5 25  
## 10 "phys\_sat\_y\_3/12/2024 " 5 25  
## # ℹ 23 more rows

Now I will visualize this missing data to have a better understanding of how to manage this data.

Next, I am going to label the NAs in this dataset as one (the lowest number on this scale) because I am assuming if participants did not fill out this piece they are not satisfied with their physical activity or did not perform any sort of activity the day or the day before they filled out this survey. I also believe that this will interfere the least with the overall data, as changing them to other numbers may skew the data in a way I do not want for my analysis.

satis\_activity\_wide <- satis\_activity\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), 1)))

### Sleep Dataframe

#### Sleep Satisfaction Dataframe

Here I will create the sleep satisfaction dataframe so I can evaluate the missing data and arrangement of this variable.

survey\_satis\_sleep <- select(survey, "ID", "date", "Squality", "restless\_scale", "tired\_scale", "rested\_scale")  
satis\_sleep\_wide <- pivot\_wider(survey\_satis\_sleep, names\_from = date, values\_from = c(Squality, restless\_scale, tired\_scale, rested\_scale))

##### Investigating Missing Data

Here I will look at the missing data in this subset to see how I should deal with these observations. This missing variable summary shows me which variables have missing data and what percentage of that data is missing.

miss\_var\_summary(satis\_sleep\_wide)

## # A tibble: 65 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "restless\_scale\_2/8/2024 " 9 45  
## 2 "restless\_scale\_3/19/2024 " 9 45  
## 3 "Squality\_3/19/2024 " 8 40  
## 4 "restless\_scale\_2/22/2024 " 8 40  
## 5 "tired\_scale\_3/19/2024 " 8 40  
## 6 "rested\_scale\_3/19/2024 " 8 40  
## 7 "restless\_scale\_2/20/2024 " 7 35  
## 8 "restless\_scale\_3/21/2024 " 7 35  
## 9 "restless\_scale\_3/7/2024 " 7 35  
## 10 "rested\_scale\_2/8/2024 " 7 35  
## # ℹ 55 more rows

Now I will visualize this missing data to have a better understanding of how to manage this data.

Next, I am going to label the NAs in this dataset as one (the lowest number on this scale) because I am assuming if participants did not fill out this piece they did not have high quality sleep and were not willing to share this information on their survey. I also believe that this will interfere the least with the overall data, as changing them to other numbers may skew the data in a way I do not want for my analysis.

satis\_sleep\_wide <- satis\_sleep\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), 1)))

#### Sleep Amount Dataframe

Here I will create the sleep satisfaction dataframe so I can evaluate the missing data and arrangement of this variable.

survey\_time\_sleep <- select(survey, "ID", "date", "hours\_sleep")  
survey\_time\_sleep <- mutate(survey\_time\_sleep, hours\_sleep = as.integer(hours\_sleep))  
time\_sleep\_wide <- pivot\_wider(survey\_time\_sleep, names\_from = date, values\_from = c(hours\_sleep))

##### Investigating Missing Data

Here I will look at the missing data in this subset to see how I should deal with these observations. This missing variable summary shows me which variables have missing data and what percentage of that data is missing.

miss\_var\_summary(time\_sleep\_wide)

## # A tibble: 17 × 3  
## variable n\_miss pct\_miss  
## <chr> <int> <dbl>  
## 1 "3/19/2024 " 10 50  
## 2 "2/22/2024 " 7 35  
## 3 "3/7/2024 " 7 35  
## 4 "2/8/2024 " 6 30  
## 5 "2/13/2024 " 6 30  
## 6 "2/20/2024 " 6 30  
## 7 "3/21/2024 " 6 30  
## 8 "2/15/2024 " 5 25  
## 9 "3/12/2024 " 5 25  
## 10 "1/30/2024 " 4 20  
## 11 "3/5/2024 " 4 20  
## 12 "1/25/2024 " 2 10  
## 13 "2/1/2024 " 2 10  
## 14 "2/6/2024 " 2 10  
## 15 "3/14/2024 " 2 10  
## 16 "1/23/2024 " 1 5  
## 17 "ID" 0 0

Now I will visualize this missing data to have a better understanding of how to manage this data.

Next, I am going to label the NAs in this dataset as 8, the most common entry of this column or mode of this column, because i do not want to label NAs as 1 or zero because that would skew my data, so I am labeling them as the mean so that they remain in place but they do not skew the data. I also believe that this will interfere the least with the overall data, as changing them to other numbers may skew the data in a way I do not want for my analysis.

time\_sleep\_wide<- time\_sleep\_wide %>%  
 mutate(across(everything(), ~ replace(.x, is.na(.x), "8")))

## Remake long-formatted dfs from these mutated and manipulated wideformat dataframes

### Select Variables of Interest

Before pivoting all of these into long formatted dataframes, I need to select only the variables I want pivotted. I eliminated some variables in this study based on time as well as a further evaluation of what was necessary and possible for this study.

time\_activity\_wide <- select(time\_activity\_wide, "ID","tvigact\_y\_1/23/2024 ":"tvigact\_y\_3/19/2024 " )  
satis\_activity\_wide <- select(satis\_activity\_wide, "ID", "phys\_sat\_y\_1/23/2024 ":"phys\_sat\_y\_3/19/2024 ")  
satis\_sleep\_wide <- select(satis\_sleep\_wide, "ID", "Squality\_1/23/2024 ":"Squality\_3/19/2024 " )

### Pivot Dataframes

Now I will pivot these dataframes back to long format so that the changes I made to help wrangle and manage the data will also help when dealing with the data in long format instead of just using the old, unchanged long formatted dataframe.

survey\_nmood\_long <- pivot\_longer(survey\_nmood\_wide, cols = `3/19/2024\_Nmean`:`1/23/2024\_Nmean`,names\_to = "date", values\_to = "avg\_Nmood")  
survey\_pmood\_long <- pivot\_longer(survey\_pmood\_wide, cols = `3/19/2024\_Pmean`:`1/23/2024\_Pmean`,names\_to = "date", values\_to = "avg\_Pmood")  
time\_activity\_long <- pivot\_longer(time\_activity\_wide, cols = `tvigact\_y\_1/23/2024 `:`tvigact\_y\_3/19/2024 `, names\_to = "date", values\_to = "timevigact\_y")  
satis\_activity\_long <- pivot\_longer(satis\_activity\_wide, cols = `phys\_sat\_y\_1/23/2024 `:`phys\_sat\_y\_3/19/2024 `, names\_to = "date", values\_to = "phys.sat\_y")  
time\_sleep\_long <- pivot\_longer(time\_sleep\_wide, cols = `1/23/2024 `:`3/19/2024 `, names\_to = "date", values\_to = "hours\_sleep")  
satis\_sleep\_long <- pivot\_longer(satis\_sleep\_wide, cols = `Squality\_1/23/2024 `:`Squality\_3/19/2024 `, names\_to = "date", values\_to = "quality\_sleep")

### Prepare dataframes for joins

#### Mood preparation for Join

Now I need to change the date variables so that the unwanted characters are not connected to column names.

Now I need to arrange these dataframes by date so they can be combined with the other long format dataframes and share rows with the observations across different dates.

#### Physical Activity Preparation for Join

Now I will do the same for the time and satisfaction with vigorous physical activity long formatted variables.

Here I am arranging the rows in this dataframe by date to make for an easier join and a more organized dataset.

#### Sleep Preparation for Join

Now I will do the same with the quality and hours of sleep variables.

Here I am arranging the rows in this dataframe by date to make for an easier join and a more organized dataset.

### Join All of these into one larger long-formatted dataframe

Here I will use the key variables of ID and date that are the same across all of my separate datasets to rejoin all of these variables together in a long-formatted, larger dataframe so I can work with multiple of these variables at once.

long\_mood <- left\_join(survey\_nmood\_long, survey\_pmood\_long, by = c("ID" = "ID", "date" = "date"))  
  
long\_activity <- left\_join(satis\_activity\_long, time\_activity\_long, by = c("ID" = "ID", "date" = "date"))  
  
long\_sleep <- left\_join(satis\_sleep\_long, time\_sleep\_long, by = c("ID" = "ID", "date" = "date"))  
  
long\_survey <- left\_join(long\_activity, long\_sleep, by = c("ID" = "ID", "date" = "date"))

Join for my average mood variables did not work with left\_join so after different troubleshooting techniques I just pasted in my average mood variables into the new long survey dataset with the permission of Professor Shablack.

long\_survey$avg\_Pmood <- long\_mood$avg\_Pmood  
long\_survey$avg\_Nmood <- long\_mood$avg\_Nmood

Now I will join the athlete dataframe to this dataframe so that I can utilize this demographic variable in my analysis

long\_survey <- left\_join(long\_survey, athlete, by = c("ID" = "ID"))

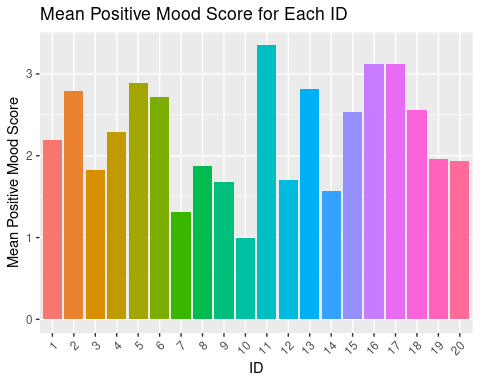
# Part Two: Visualizations and Models

Here I will make both exploratory, initial visualizations as well as visualizations assessing the relationships between my key variables. The initial visualizations will highlight the scope and variability of my key variables as well as if there were any issues with my wrangling and joining. The visualizations assessing the relationships between my key variables will demonstrate if there are possible correlations or predictive relationships between these variables and if these align with my initial hypotheses and research questions. I used bar graphs to demonstrate the variability of the data when looking at individual level ID averages, as looking at this data in this format is easier to understand than if it were in a dot or line graph. I used line graphs to demonstrate trends over time because they better visualize the shifts in overall averages for the entire sample, and I returned to bar graphs to demonstrate the relationships between my variables because the bars better demonstrate the relationships between two continuous or scaled variables.

## Mood Initial Visualizations

Here is my initial visualization for the average positive mood score for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_pmood = mean(avg\_Pmood),  
 n = n(),  
 sdBase = sd(avg\_Pmood))%>%  
ggplot(aes(x= ID, y= mean\_pmood, fill= factor(ID))) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Mean Positive Mood Score for Each ID") + xlab("ID") + ylab("Mean Positive Mood Score")

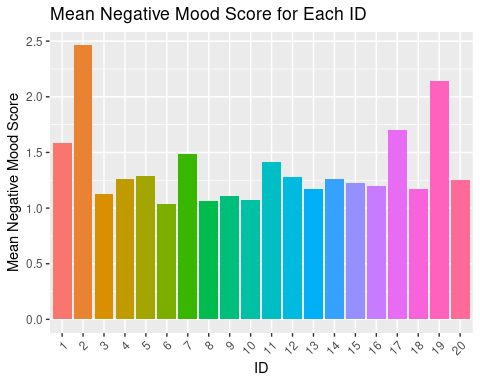


#### Analysis

This graph demonstrates the variability of the positive mood scores between each individual ID when averaging all of their survey responses across the semester. The different average positive mood scores demonstrate that individuals responded differently than their peers on a given day about their positive mood feelings. This indicates that different factors contributed to these positive mood reports everyday for each individual student ID, one of which I hypothesize being physical activity.

Here is my initial visualization for the average negative mood score for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_nmood = mean(avg\_Nmood),  
 n = n(),  
 sdBase = sd(avg\_Nmood))%>%  
ggplot(aes(x= ID, y= mean\_nmood, fill= factor(ID))) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Mean Negative Mood Score for Each ID") + xlab("ID") + ylab("Mean Negative Mood Score")



#### Analysis

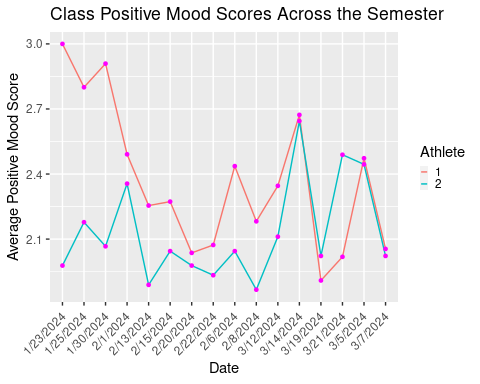
This graph demonstrates the variability of the negative mood scores between each individual ID when averaging all of their survey responses across the semester. The different average negative mood scores demonstrate that individuals responded differently than their peers on a given day about their negative mood feelings. However, there is less variability in this visualization than in the positive mood score graph, indicating that students reported more similar negative mood responses on a given day than positive. Overall, this graph shows that different factors contributed to these negative mood reports everyday for each individual student ID, one of which I hypothesize being physical activity.

## Mood Trend Over Time including Student Athlete Status

To make this trend for the entire class, I have to find the mean positive mood across all of the IDs for each date the survey was taken. To do this, I will create a new dataframe with just the positive mood means and the dates.

Now I will graph these averages over time.

ggplot(Pmood\_means, aes(x = date, y = mean\_Pmood, group = Athlete, color = Athlete)) + geom\_line() + geom\_point(color = "magenta", size = 1) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Class Positive Mood Scores Across the Semester") + xlab("Date") + ylab("Average Positive Mood Score") + theme(legend.key.size = unit(.2, 'cm'))



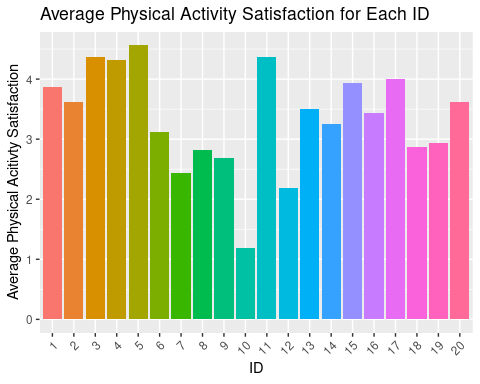
#### Analysis

This graph shows the trend of positive mood scores over time across the semester for student athletes (1) and non-student athletes (2). The trend line displays two findings that aid in proving my hypothesis. First, student-athletes had higher positive mood scores for the majority of the semester, accept for 3/19 and 3/21. This demonstrates the positive correlation between physical activity and average positive mood score. At the beginning of the semester, the average positive mood score difference was significantly higher than the middle and end of the semester, where these averages almost evened out. The reasoning behind this large disparity in January would be an interesting subject to pursue in a future study about the schedules and moods of student athletes, focusing on things like amount of practice, success of seasons, and overall team morale.

## Activity Initial Visualizations

Here is my initial visualization for the average physical activity satisfaction for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_physsat = mean(phys.sat\_y),  
 n = n(),  
 sdBase = sd(phys.sat\_y))%>%  
ggplot(aes(x= ID, y= mean\_physsat, fill= ID)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Average Physical Activity Satisfaction for Each ID") + xlab("ID") + ylab("Average Physical Acitivty Satisfaction")

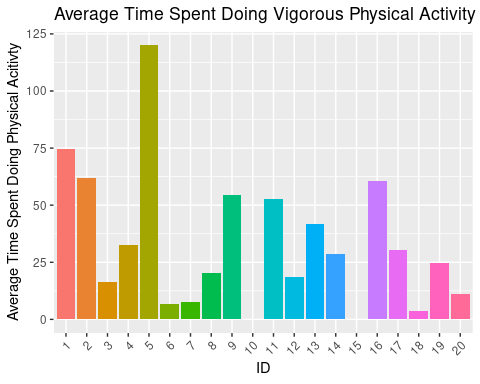


#### Analysis

This graph demonstrates the variability of the physical activity satisfaction between each individual ID when averaging all of their survey responses across the semester. The different average physical activity satisfaction levels demonstrate that individuals responded differently than their peers on a given day about their physical activity satisfaction from the previous day. This indicates that different factors contributed to these activity satisfaction levels reports everyday for each individual student ID, and that this variable could have varying effects on sleep and mood for each ID.

Here is my initial visualization for the average time spent doing vigorous physical activity for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_timeact = mean(timevigact\_y),  
 n = n(),  
 sdBase = sd(timevigact\_y))%>%  
ggplot(aes(x= ID, y= mean\_timeact, fill= ID)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Average Time Spent Doing Vigorous Physical Activity") + xlab("ID") + ylab("Average Time Spent Doing Physical Acitivty")



#### Analysis

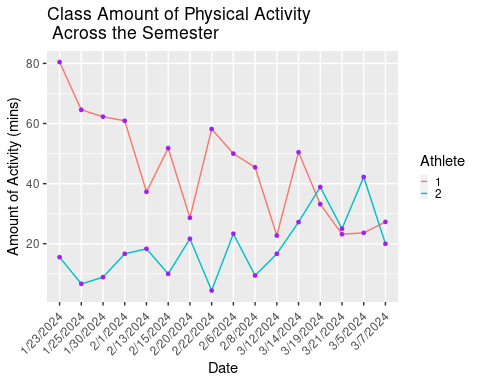
This graph demonstrates the variability of the time spent participating in vigorous physical activity for each individual ID when averaging all of their survey responses across the semester. The different average physical activity amounts demonstrate that individuals responded differently than their peers on a given day about their physical activity from the previous day. This indicates that different factors contributed to these activity level reports everyday for each individual student ID, and that this variable could have varying effects on sleep and mood for each ID based on how active these students are. It appears that #10 and #15 do not have computed activity means in this graph, indicating that they have too many inputs of zero from not answering this survey or they do not participate in vigorous physical activity in general during this winter semester.

## Activity Trends Over time including Student Athlete Status

To make this trend for the entire class, I have to find the mean activity across all of the IDs for each date the survey was taken. To do this, I will create a new dataframe with just the activity means and the dates.

Now I will graph these averages over time.

ggplot(activity\_means, aes(x = date, y = mean\_activity, group = Athlete, color = Athlete)) + geom\_line() + geom\_point(color = "purple", size = 1) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Class Amount of Physical Activity \n Across the Semester") + xlab("Date") + ylab("Amount of Activity (mins)") + theme(legend.key.size = unit(.2, 'cm'))



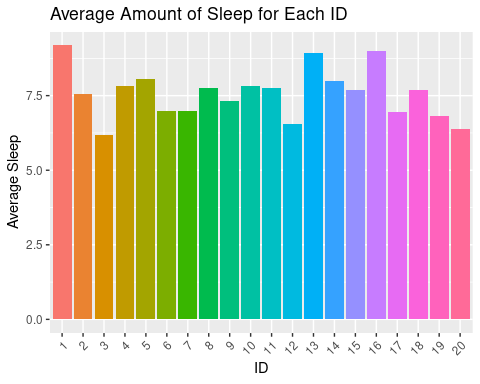
#### Analysis

This graph shows the trend of the average amount of physical activity over time across the semester for student athletes (1) and non-student athletes (2). The trend line displays two findings that aid in proving my hypothesis. First, this graph shows that the average physical activity for student athletes is higher than that of non-student athletes throughout the semester, with the exception of 3/5, 3/19, and 3/21. These days could have been off days for student athletes or days with limited or lighter practices. This demonstrates that student-athlete status is another indicator of physical activity level on average for IDs in this study, as they tend to participate in more physical activity than non athletes for the majority of days this survey was recorded. Similar to the mood score trend graph, there is a larger disparity between athletes and non athletes at the beginning of the semester than in the middle and end. This could be due to preseason athlete training, more practices at the beginning of the semester, or fewer non athletes finding time to exersize at the beginning of the semester with a new class schedule.

## Sleep Initial Visualizations

Here is my initial visualization for the average amount of sleep for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_sleep = mean(hours\_sleep),  
 n = n(),  
 sdBase = sd(hours\_sleep))%>%  
ggplot(aes(x= ID, y= mean\_sleep, fill= ID)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Average Amount of Sleep for Each ID") + xlab("ID") + ylab("Average Sleep")

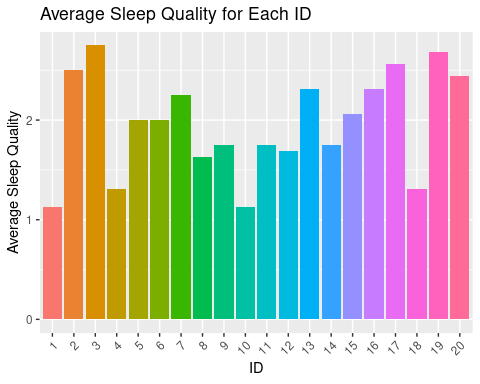


#### Analysis

This graph demonstrates the variability of hours of sleep between each individual ID when averaging all of their survey responses across the semester. The different average hours of sleep demonstrate that individuals responded differently than their peers on a given day about their hours of sleep from the previous night. This indicates that different factors contributed to the amount of sleep each individual student ID reports everyday, and I hypothesize that physical activity satisfaction and amount heavily impact this variable.

Here is my initial visualization for the average sleep quality for Each ID averaged from the whole semester.

long\_survey %>%   
 mutate(ID = fct\_inorder(ID)) %>%  
 group\_by(ID) %>%   
 summarise(mean\_satissleep = mean(quality\_sleep),  
 n = n(),  
 sdBase = sd(quality\_sleep))%>%  
ggplot(aes(x= ID, y= mean\_satissleep, fill= ID)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Average Sleep Quality for Each ID") + xlab("ID") + ylab("Average Sleep Quality")



#### Analysis

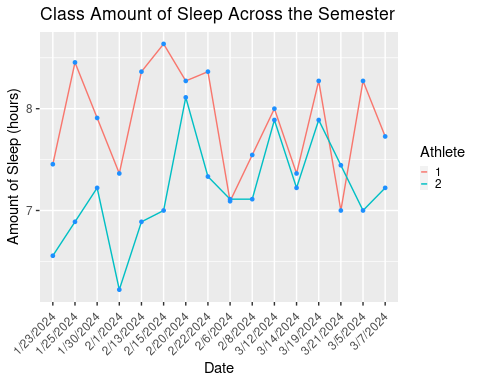
This graph demonstrates the variability of the sleep quality between each individual ID when averaging all of their survey responses across the semester. The different average sleep quality levels demonstrate that individuals responded differently than their peers on a given day about their sleep quality from the previous night. This indicates that different factors contributed to these sleep quality levels reports everyday for each individual student ID, and I hypothesize that physical activity level and satisfaction are some of the factors impacting this sleep quality.

## Sleep Trends Over Time including Student Athlete Status

To make this trend for the entire class, I have to find the mean sleep across all of the IDs for each date the survey was taken. To do this, I will create a new dataframe with just the sleep means and the dates.

Now I will graph these averages over time.

ggplot(sleep\_means, aes(x = date, y = mean\_sleep, group = Athlete, color = Athlete)) + geom\_line() + geom\_point(color = "dodgerblue", size = 1) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Class Amount of Sleep Across the Semester") + xlab("Date") + ylab("Amount of Sleep (hours)") + theme(legend.key.size = unit(.2, 'cm'))



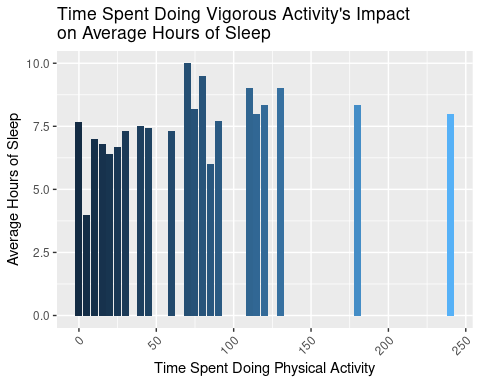
### Analysis

This graph shows the trend of amount of sleep (hours) over time across the semester for student athletes (1) and non-student athletes (2). The trend line shows that student-athletes had more average sleep for the majority of the semester, accept for 2/6 and 3/21. These lower sleep amounts could be due to longer practice or gamedays, a more difficult balance between athletics and a heavier school work load, or a variety of other factors. Overall, this trend demonstrates the positive correlation between physical activity, as we saw above that student athlete status seemingly predicts activity levels, and average sleep. Similar to average positive mood score and physical activity amount, the difference between athletes and non athletes was more drastic at the beginning and then these two trends became more similar as the semester progressed. This could be due to numerous factors, including more taxing preseason days, making athletes more tired, adjusting to being in or out of season with a new class schedule, or demands of practice in general that make student athletes sleep for longer during this period. The trend seen throughout this project of this large disparity in January would be interesting to investigate further with more data and survey questions.

## Activity’s Impact on Sleep

Here I will visually demonstrate the impact of vigorous physical activity on the time students spend asleep using the data from the amount of time student’s spent doing physical activity from the day before as well as their sleep from last night.

long\_survey %>%   
 group\_by(timevigact\_y) %>%   
 summarise(mean\_sleep = mean(hours\_sleep),  
 n = n(),  
 sdBase = sd(hours\_sleep))%>%  
ggplot(aes(x= timevigact\_y, y= mean\_sleep, fill= timevigact\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Time Spent Doing Vigorous Activity's Impact \non Average Hours of Sleep") + xlab("Time Spent Doing Physical Activity") + ylab("Average Hours of Sleep")

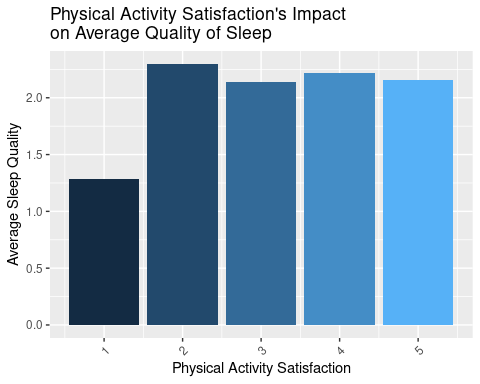


### Analysis

The above graph shows the relationship between time spent doing vigorous physical activity in minutes and average amount of sleep in hours. According to this graph, the majority of students who spend less than 50 minutes doing vigorous physical activity had fewer hours of sleep on average. Students participating in 1-2 hours of vigorous physical activity had more hours of sleep on average, but this increase in average sleep decreases when students participate in activity for more than 2 hours. This suggests that 1-2 hours of physical activity promotes more sleep, and this correlation between these variables will be tested using a correlation test in the next section of this project. The effect too mych physical activity has on hours of sleep would be interesting to explore further, as it appears to almost have the inverse effect has amount of physical activity exceeds this 1-2 hour mark.

Next, I will demonstrate satisfaction with physical activity and quality of sleep in a visualization using the variables from sleep the night before and physical activity satisfaction yesterday.

long\_survey %>%   
 group\_by(phys.sat\_y) %>%   
 summarise(mean\_qsleep = mean(quality\_sleep),  
 n = n(),  
 sdBase = sd(quality\_sleep))%>%  
ggplot(aes(x= phys.sat\_y, y= mean\_qsleep, fill= phys.sat\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Physical Activity Satisfaction's Impact \non Average Quality of Sleep") + xlab("Physical Activity Satisfaction") + ylab("Average Sleep Quality")



### Analysis

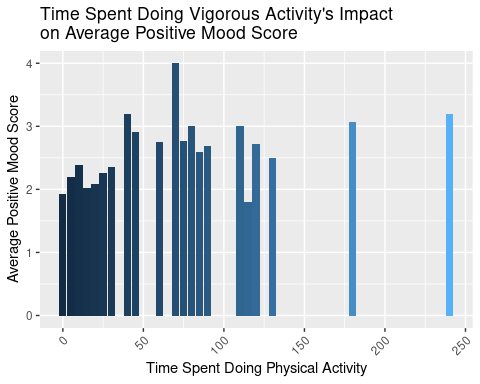
This graph indicates physical activity satisfaction’s influence on average sleep quality across the data set. The x-axis indicates students’ physical activity satisfaction from 1 (extremely unsatisfied) to 5 (extremely satisfied). Based on this graph, it appears that students who responded “extremely unsatisfied” tend to have a lower average sleep quality, whereas those with more satisfaction with their physical activity tend to have higher average sleep quality, suggesting that satisfaction with physical activity leads to higher sleep quality. It is interesting that only the “extremely unsatisfied” column carries a large difference in sleep quality, as the rest of these reports are very similar in their average sleep qualities, with “moderately unsatisfied” with physical activity actually predicting the highest amount of sleep quality.

## Activity’s Impact on Mood

### Positive Mood Score

Here I will visually demonstrate the impact of vigorous physical activity on student’s overall mood using the data from the amount of time student’s spent doing physical activity from the day before as well as their mood from today.

long\_survey %>%   
 group\_by(timevigact\_y) %>%   
 summarise(mean\_pmood = mean(avg\_Pmood),  
 n = n(),  
 sdBase = sd(avg\_Pmood))%>%  
ggplot(aes(x= timevigact\_y, y= mean\_pmood, fill= timevigact\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Time Spent Doing Vigorous Activity's Impact \non Average Positive Mood Score") + xlab("Time Spent Doing Physical Activity") + ylab("Average Positive Mood Score")

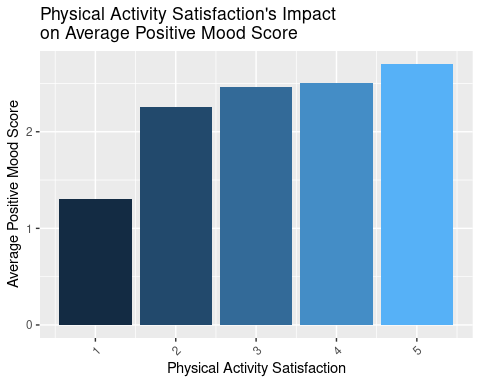


#### Analysis

This graph demonstrates the relationship between time spent doing vigorous physical activity in minutes and average positive mood score. According to this graph, as students increase their physical activity to around 60/70 mins, their mood increases. After physical activity exceeds this timepoint, average positive mood scores fluctuate and no longer increase, suggesting that around an hour of vigorous physical activity promotes higher positive mood scores, but more activity than this does not. This positive correlation between physical activity amount and positive mood score will be investigated later in a correlation test.

Here I will visually demonstrate the impact of physical activity satisfaction on student’s overall mood using the data from student’s physical activity satisfaction level the day before as well as their mood from today.

long\_survey %>%   
 group\_by(phys.sat\_y) %>%   
 summarise(mean\_pmood = mean(avg\_Pmood),  
 n = n(),  
 sdBase = sd(avg\_Pmood))%>%  
ggplot(aes(x= phys.sat\_y, y= mean\_pmood, fill= phys.sat\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Physical Activity Satisfaction's Impact \non Average Positive Mood Score") + xlab("Physical Activity Satisfaction") + ylab("Average Positive Mood Score")



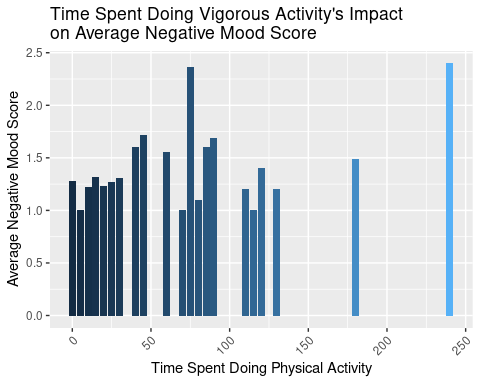
#### Analysis

The above graph demonstrates physical activity satisfaction’s influence on average positive mood score across the data set. The x-axis indicates students’ physical activity satisfaction from 1 (extremely unsatisfied) to 5 (extremely satisfied). Based on this graph, it appears that there is a clear, positive correlation between average positive mood score and physical activity satisfaction, as average positive mood score increases as students become more satisfied with their physical activity. This trend is the most clear correlation I have seen thus far in my project, and I will further explore the relationship between these variables in a correlation test below.

### Negative Mood Score

Here I will visually demonstrate the impact of vigorous physical activity on student’s overall mood using the data from the amount of time student’s spent doing physical activity from the day before as well as their mood from today.

long\_survey %>%   
 group\_by(timevigact\_y) %>%   
 summarise(mean\_nmood = mean(avg\_Nmood),  
 n = n(),  
 sdBase = sd(avg\_Nmood))%>%  
ggplot(aes(x= timevigact\_y, y= mean\_nmood, fill= timevigact\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Time Spent Doing Vigorous Activity's Impact \non Average Negative Mood Score") + xlab("Time Spent Doing Physical Activity") + ylab("Average Negative Mood Score")

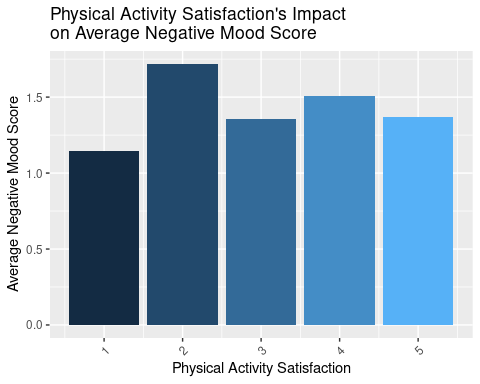


#### Analysis

This graph demonstrates the relationship between time spent doing vigorous physical activity in minutes and average negative mood score. According to this graph, as students increase their physical activity to around 60/70 mins, their mood decreases. After physical activity exceeds this timepoint, average negative mood scores fluctuate and no longer increase, but rather decrease until the last time point of around 250 minutes, where this variable spikes dramatically. This positive correlation between physical activity amount and negative mood score indicates that those participating in physical activity tend to have strong feelings in general than those who are not, as they experience more positive and negative moods and have higher mood scores on average than those with lower activity amounts. This is an interesting trend to investigate, as this graph seemingly suggests that more active people report stronger mood feelings in general.

Here I will visually demonstrate the impact of physical activity satisfaction on student’s overall mood using the data from student’s physical activity satisfaction level the day before as well as their mood from today.

long\_survey %>%   
 group\_by(phys.sat\_y) %>%   
 summarise(mean\_nmood = mean(avg\_Nmood),  
 n = n(),  
 sdBase = sd(avg\_Nmood))%>%  
ggplot(aes(x= phys.sat\_y, y= mean\_nmood, fill= phys.sat\_y)) +   
 geom\_bar(stat = "summary", fun="mean", show.legend = F) + theme(axis.text.x = element\_text(angle=45, vjust=1, hjust=1)) + ggtitle("Physical Activity Satisfaction's Impact \non Average Negative Mood Score") + xlab("Physical Activity Satisfaction") + ylab("Average Negative Mood Score")



#### Analysis

This graph demonstrates the relationship between physical activity satisfaction and average negative mood score. According to this graph, students with “moderately unsatisfied” physical activity satisfaction levels have the highest negative mood scores, followed by those with moderately satisfied, neutral, and extremely satisfied, with extremely unsatisfied physical satisfaction level coming in last for average negative mood score. This trend is interesting, as it again correlates to the trend of positive mood scores when compared to physical activity satisfaction. This indicates that students in this survey who participate in more physical activity and have more than “extremely unsatisfied” physical activity satisfaction report stronger mood feelings, both positive and negative, on average across the semester.

## Correlation Tests

Here I am going to test the correlation of my predictors and my outcomes to see if there is a predictive relationship between these variables. I will then factor in the progress over time in Linear Mixed Effects Models.

### Correlation test for time spent doing vigorous physical activity and hours of sleep:

cor.test(long\_survey$timevigact\_y, long\_survey$hours\_sleep)

##   
## Pearson's product-moment correlation  
##   
## data: long\_survey$timevigact\_y and long\_survey$hours\_sleep  
## t = 2.8722, df = 318, p-value = 0.00435  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.05025195 0.26405226  
## sample estimates:  
## cor   
## 0.1590159

#### Analysis

A correlation test between these variables demonstrated that there is a correlation coefficient of 0.159 with a p-value > 0.005 when comparing these two variables. This signifies that there is a positive correlation between these variables, as seen in my visualizations and as I predicted in my introduction. This discovery shows the positive influence time spent doing physical activity has on the amount of sleep an individual gets.

### Correlation test for physical activity satisfaction and sleep quality:

cor.test(long\_survey$phys.sat\_y, long\_survey$quality\_sleep)

##   
## Pearson's product-moment correlation  
##   
## data: long\_survey$phys.sat\_y and long\_survey$quality\_sleep  
## t = 6.9971, df = 318, p-value = 1.553e-11  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2662882 0.4566176  
## sample estimates:  
## cor   
## 0.365264

#### Analysis

A correlation test between these variables demonstrated that there is a correlation coefficient of 0.365 with a p-value > 0.001 when comparing these two variables. This signifies that there is a strong positive correlation between these variables, as seen in my visualizations and as I predicted in my introduction. This dicovery demonstrates the positive influence physical activity satisfaction has on sleep quality, as in this dataset, higher physical activity satisfaction has a positive influence on sleep quality.

### Correlation test for physical activity satisfaction and average positive mood score:

cor.test(long\_survey$phys.sat\_y, long\_survey$avg\_Pmood)

##   
## Pearson's product-moment correlation  
##   
## data: long\_survey$phys.sat\_y and long\_survey$avg\_Pmood  
## t = 11.216, df = 318, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4489613 0.6066250  
## sample estimates:  
## cor   
## 0.5323947

#### Analysis

A correlation test between these variables demonstrated that there is a correlation coefficient of 0.532 with a p-value > 0.001 when comparing these two variables. This signifies that there is a strong positive correlation between these variables, the strongest within this project, as seen in visualizations and predicted in my introduction. This discovery indicates that physical activity satisfaction has a strong positive effect on positive mood feelings within individual students within this study.

### Correlation test for time spent doing vigorous physical activity and average positive mood score:

cor.test(long\_survey$timevigact\_y, long\_survey$avg\_Pmood)

##   
## Pearson's product-moment correlation  
##   
## data: long\_survey$timevigact\_y and long\_survey$avg\_Pmood  
## t = 6.6535, df = 318, p-value = 1.252e-10  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2494928 0.4422601  
## sample estimates:  
## cor   
## 0.3495706

#### Analysis

A correlation test between these variables demonstrated that there is a correlation coefficient of 0.349 with a p-value > 0.001 when comparing these two variables. This correlation coefficient signifies that there is a positive correlation between these two variables, as seen in my visualizations and predicted in my introduction. This discovery indicates that time spent doing vigorous physical activity has a positive effect on average positive mood score and overall mood feelings within this study. We know from previous visualizations that physical activity satisfaction and amount likely positively correlates to both positive and negative mood feelings within this dataset, indicating that activity overall incites stronger emotional reactions and reports within students participating in this study.

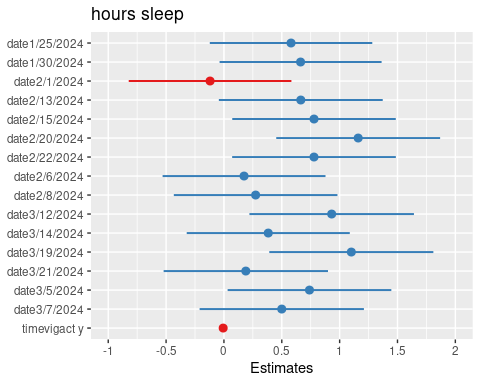
## Linear Mixed Effects Model

### Sleep and Physical Activity Regression Models

Now I will run a series of linear mixed effects regression models to demonstrate the correlations between sleep, mood, and physical activity for this dataset.

#### Regression of Time Spent Doing Physical Activity and Hours of Sleep:

plot\_model(m1)

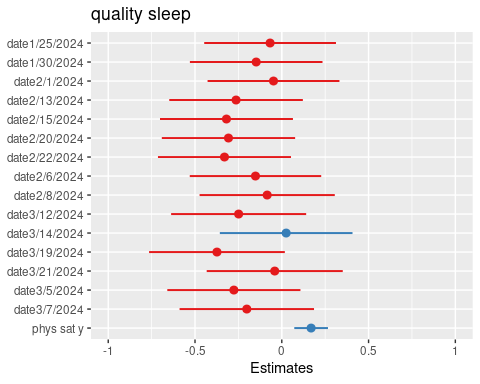


##### Analysis

This model and coinciding plot demonstrate the results of an LMER ran for the relationship between hours of sleep, and date, and it also factored in amount of physical activity. This LMER visualization shows estimates for how date predicts amount of sleep students get on average each night. The red values within the plot indicate negative correlations between date or vigorous activity and hours of sleep. The coefficient between activity amount and hours of sleep was -0.01 with a p-value of 0.15 for this regression, showing that when date is factored in for the average amount of sleep, the effect of vigorous activity seems to decrease. 2/1 was the only negative predictor for hours of sleep, meaning more individuals had less sleep on this date, which could be due to a variety of different factors like a school-wide event the night before, midterm studying, popular sports game streaming, and many others.

#### Regression of Physical Activity Satisfaction and Sleep Quality:

plot\_model(m2)



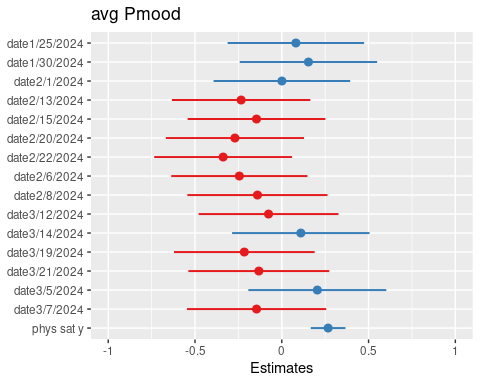
##### Analysis

This model and coinciding plot demonstrate the results of an LMER ran for the relationship between sleep quality, and date, and it also factored in physical activity satisfaction. This LMER visualization shows estimates for how date predicts sleep quality students get on average each night. The red values within the plot indicate negative correlations between date and sleep quality. 3/14 is the only date here with a positive estimate, meaning this date is the only one that predicts higher sleep quality and has a positive impact on this variable. I am not sure why this date had a positive impact on sleep quality, but it appears that the majority of students in this study had high sleep quality on this night, which could be due to various factors concerning the weather, activities at school, or overall workloads. Physical activity satisfaction also had a positive estimate, with a coefficient of 0.17 and a p-value of 0.

### Mood and Activity Models

#### Regression of Physical Activity Satisfaction and Average Positive Mood Score:

plot\_model(m3)

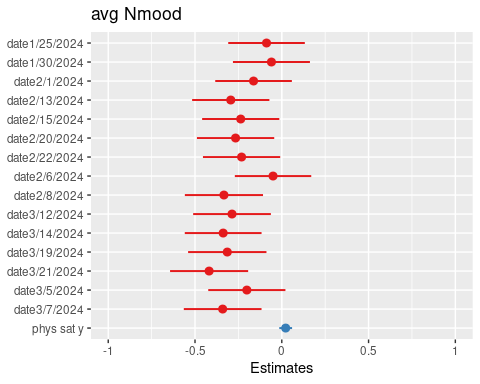


##### Analysis

This model and coinciding plot demonstrate the results of an LMER ran for the relationship between average positive mood and date, and it also factored in physical activity satisfaction. This LMER visualization shows estimates for how date predicts average positive mood, and red values indicate negative correlation between date and average positive mood score. 1/25, 1/30, 2/1, 3/14, and 3/5 all had positive estimates, meaning these dates were positive predictors for positive mood scores. The positive impact of these dates on average positive mood score could be due to the weather of these days, events on campus, stress levels, and many other factors. The correlation between activity satisfaction and positive mood was 0.27 with a p-value of zero for this regression, showing that this variable is still a positive predictor when date is factored in to the relationship.

#### Regression of Physical Activity Satisfaction and Average Negative Mood Score:

plot\_model(m4)



###### Analysis

This model and coinciding plot demonstrate the results of an LMER ran for the relationship between average negative mood and date, and it also factored in physical activity satisfaction. This LMER visualization shows estimates for how date predicts average negative mood, and red values indicate negative correlation between date and average negative mood score. All of the dates in this study have negative estimates for negative mood, meaning they negatively predict negative moods. Physical activity satisfaction still has a positive estimate in this model, however, with a coefficient of 0.02 and a p-value of 0.25, meaning it is not statistically significant.

# Conclusions

This case study provided insights into the ways that physical activity influences sleep and overall mood. These insights are useful to inform future studies expanding on the correlations and predictive relationships explored here.

## Activity’s Impact on Sleep Conclusions

The graphs and correlations conducted in this focus indicate that my initial hypotheses were correct. Students participating in more physical activity on average seem to get more hours of sleep on average. However, there appears to be a mixed effect after physical activity surpasses 1-2 hours. From this we can hypothesize that too much physical activity could have the reverse effect on hours of sleep. This would be interesting to explore in a future study. The correlation coefficient of 0.365 is statistically significant, meaning the positive correlation found here is not due to random chance or skewed data. This correlation itself proves my hypothesis. Similarly, students who report better than extreme dissatisfaction with physical activity have better average sleep quality, suggesting that satisfaction according to individuals and their definition of “good” exercise provides higher sleep quality, again according to individual level assessments. The correlation test between these variables had a coefficient of 0.159, meaning there is less of a correlation here than between the previous two variables, but it is still backed by statistical significance and proves my hypothesis.

## Activity’s Impact on Mood Conclusions

The graphs and correlations here again prove my hypotheses to be correct. There is a clear increase in average positive mood score when physical activity satisfaction increases, meaning the more satisfied individuals are with their exercise amount, the better their mood will be. This trend is backed by a statistically significant correlation coefficient of 0.532, demonstrating there is a strong positive correlation between these variables and proving that physical activity has an impact on mood. Amount of physical activity has similar results, with positive mood increasing as physical activity increases up to around 60/ 70 mins. This trend again suggests that too much physical activity could have a negative impact on positive mood, just as it did for sleep, and these findings would be interesting to explore in future studies.

## Trends Over Time Conclusions

The trends in this portion of my study indicate the benefits of being a student athlete at Washington and Lee. Athletes appear to have had higher mood scores, hours of sleep, and amount of activity across the semester. These trends are slightly surprising, as I was unsure how student athlete status would effect these variables, especially sleep. The days in which non athletes have higher averages would be interesting to examine in future studies; looking at days off, shorter practices, or off season activities. The days that predicted lower averages for the whole study would also be interesting to examine, including looking at temperature, campus activities, and world news that could contribute to low mood, sleep, or activity on specific days.

# Citations

Class Daily Survey, CBSC 185 Winter 2024, Professor Shablack Class lecture, CBSC 185 Winter 2024, Professor Shablack