Dictionary for Scripps Data

Motion Data

1. Unix Data
    - Data type: int
    - Meaning: The float representation of date/time in the UTC time zone. It is a float that represents the number of seconds that has passed from January 1st, 1970 at UTC. It is essentially the Unix time representation of the date/time that motion is starting to be measured.
    - values/statistics:
        - Range: 1543340132-1544629892
            - 11/27/2018 @ 5:35pm (UTC) - 12/12/2018 @ 3:51pm (UTC)
        - Median: 1543985621
            - 12/05/2018 @ 4:53am (UTC)
        - IQR: (1544307360-1543662644)
            - 12/08/2018 @ 10:16pm (UTC) - 12/01/2018 @ 11:10am (UTC)
        - Number of missing values: 0
        - NOTES:
            - Measurement starts every ~30 seconds (give or take 1-2 seconds, sometimes certain cycles are missed)
2. Date
    - Data type: string
    - Meaning: Meaning: String representing time that motion is starting to be measured. Formatting is in HH:MM:SS
    - values/statistics: REFER TO UNIX DATA
3. Time
    - Data type: string
    - Meaning: String representing time that motion is starting to be measured. Formatting is in HH:MM:SS
    - values/statistics: REFER TO UNIX DATA
4. Motion Seconds
    - Data type: int
    - Meaning: MOST LIKELY the duration that the ring detects motion during the 30 second interval that it is documenting.
    - values/statistics:
        - Range: -1-30
        - Median: 2
        - IQR: 12 (12-0)
        - Number of missing values: 0
        - NOTES:
            - Skewed right, a majority of the motion is minimal
            - Motion occasionally is at negative time, which is impossible. Motion is only -1 when NTC temp is 0 AKA when the ring is off
5. NTC temp
    - Data type: float
    - Meaning:
        - Definition of NTC: ‘They are primarily used as resistive temperature sensors and current-limiting devices’
            - Read more http://www.resistorguide.com/ntc-thermistor/

        - This temperature is most likely measuring the temperature of the battery of the ring, NOT the temperature of the user. This variable is most likely important in detecting the safety/health of the ring battery/electrical components
    - values/statistics:
        - Range: 0-39.53
        - Median: 32.85
        - IQR: 8.16 (35.98-27.82)
        - Number of missing values: 0
        - NOTES:
            - Generally an unimportant value for consumers concern, might be valuable to determine whether the ring is on or off

6. Ring state
    - Data type: int
    - Meaning: MOST LIKELY values that represent ring states in a different dictionary. Interesting thing to note is that ring state goes from 1-6, but ring is never at state 2
    - values/statistics:
        - Unique Values: 1,3,4,5,6
        - Most common value: 3
            - Distribution of counts: 1:2976, 3:23071, 4:12, 5:5939, 6:10987
        - Median: 3
        - IQR: 3 (6-3)
        - Number of missing values: 0
        - NOTES:
            - From my grouping experimentation, 1 = ‘ring is off’ , 4/6 = ‘little to no motion, possibly not being worn’,
            - On average, NTC temp increases as ring state increases. Could possibly be based off of battery charge?
7. Motions low
    - Data type: int
    - Meaning: UNSURE; it’s not a measure of duration of low motion since motions low and motions high don’t add up to motions second. Most likely some kind of qualitative measurement not involving duration, could be a ranking system or something that accounts different factors of the ring (motion directions, intensity, etc.). Most likely only a measurement of the ring itself, not the motion of the user.
    - values/statistics:
        - Range: 0-35
        - Median: 1
        - IQR: 9 (9-0)
        - Number of missing values: 0
        - NOTES:
            - Motions low and motions high could possibly be distinguished by heart rate differences
            - Generally skewed right
8. Motions high
    - Data type: int
    - Meaning: UNSURE; it’s not a measure of duration of high motion since motions low and motions high don’t add up to motions second. Most likely some kind of qualitative measurement not involving duration, could be a ranking system or something that accounts different factors of the ring (motion directions, intensity, etc.). Most likely only a measurement of the ring itself, not the motion of the user.
    - values/statistics:

    - Range: 0-46
    - Median: 0
    - IQR: 0 (0-0)
    - Number of missing values: 0
    - NOTES:
        - Motions low and motions high could possibly be distinguished by heart rate differences
        - Generally skewed right

9. Regularity
    - Data type: int
    - Meaning: Boolean value. MOST LIKELY where boolean represents whether the data follows regular trend; 0 being true, 1 being false.
    - values/statistics:
        - Range: 0-1 (boolean values)
        - Median: 0
        - IQR: 0 (0-0)
        - Number of missing values: 0
        - NOTES:
            - A majority of the regularity values are 0, leading me to believe that 0 means that data point follows regular trends
10. Average Y
    - Data type: float
    - Meaning: A value representing the average Y of the ring during its 30 second measurement duration. Unsure of what the coordinate system is, the possibility that it could be average change in Y rather than the actual Y coordinate itself.
    - values/statistics:
        - Range: -1024-1016
        - Median: 0
        - IQR: 416 (192-(-224))
        - Number of missing values: 14893
        - NOTES:
            - For all data points with a missing average Y, motion seconds and motions low and motions high are all 0. If motions low or motions high has values, Motion seconds 0 still means that it has an average Y
11. Average Z
    - Data type: float
    - Meaning: A value representing the average Z of the ring during its 30 second measurement duration. Unsure of what the coordinate system is, the possibility that it could be average change in Z rather than the actual Z coordinate itself.
    - values/statistics:
        - Range: -1024-1016
        - Median: -328
        - IQR: 864(96-(-768))
        - Number of missing values: 17869
        - NOTES:
            - The same rule for average Y applies to average Z. if motion seconds and motions low and motions high are all 0, average Z is missing. However, if motion seconds is also -1, average Z but NOT average Y is missing

NOTES ON DATA:

- Data seems to be recorded every half minute (for the most part); at the very least, it's always around :30 and :00
- Data is most likely a measurement of the ring and its state at half minute intervals, most likely measures for the 30 second time interval

IBI Data (Interbeat interval)

GENERAL NOTES:

- 'IBI is a scientific term used in reference to the time interval between individual beats of the mammalian heart'
- Generally measured in units of milliseconds
- IBI values vary from beat to beat, which can be measured with heart rate variability (HRV). This information is not included in the dataset
    - Certain cardiac conditions/illnesses can cause IBI values to be nearly constant (meaning HRV is nearly zero)

SOURCE: https://en.wikipedia.org/wiki/Interbeat_interval

- As heart rate decreases, IBI increases
- HR = 60000/(mean IBI)

SOURCE:
https://support.mindwaretech.com/2017/09/all-about-hrv-part-2-interbeat-intervals-and-time-domain-stats/

1. Date
    - Data type: String
    - Meaning: String representing date that interbeat interval was measured. Formatting is in DD.MM.YYYY
    - values/statistics: REFER TO UTC TIME
2. Time
    - Data type: String
    - Meaning: String representing time that interbeat interval was measured. Formatting is in HH:MM:SS
    - values/statistics: REFER TO UTC TIME
3. Validity

Data type: Integer

Meaning: UNSURE; initially thought it was a boolean, but unique values contradict this theory. Could possibly be levels of validity or a confidence level of validity for the IBI measurement. Could also possibly be a dictionary value for validity (certain numbers could represent issues during measurement)

values/statistics:

- Unique Values: 0,1,2,3,15,-1,-2
- Most common value: 1
    - Distribution of counts: 0:1624, 1:410655, 2:57849, 3: 19573, 15: 147, -1:96, -2:139
- Median: 1
- IQR: 0 (1-1)
- Number of missing values: 0
- NOTES
    - -1, -2, don't have IBI or UTC time, only padded IBI. They could be measures of severity; -1 validity data seems to have abnormally large padded IBI, while -2 has less abnormally large padded IBI
    - 15 seems to be on the opposite spectrum compared to -1/-2. 15 has padded IBI and IBI values, but the data is **only** 0. There is nothing abnormal about time

        - There is nothing abnormal about the measurements of UTC time. 1 seems to have the least variability and the largest IBI, then 3, then 2. Both 2 and 3 have a minimum IBI/padded IBI of 0, while 1 has a minimum of 324.

4. Padded IBI
    - Data type: Integer
    - Meaning: Padded IBI is the same value as IBI as long as Validity is not a negative value (when validity is negative, IBI is NaN and Padded IBI is a large number). It's possible that padded IBI is the measurement of interbeats, and is converted into an interval if time is successfully documented. However, in general it is not different than IBI
    - values/statistics:
        - Range: 0-50074000
        - Median: 1012
        - IQR: 172 (1104-932)
        - Number of missing values: 0
        - NOTES:
            - Largely skewed right w/ quite the large variance
5. IBI
    - Data type: Float
    - Meaning: IBI is the same value as Padded IBI as long as Validity is not a negative value (when validity is negative, IBI is NaN and Padded IBI is a large number). Its possible that IBI is the definite measurement of IBI calculated by padded IBI, and is represented by NaN when padded IBI is not collected correctly or time is not collected. In general IBI is not different than padded IBI.
    - values/statistics:
        - Range: 0-2000
        - Median: 1012
        - IQR: 172 (1104-932)
        - Number of missing values: 235
        - NOTES:
            - Mean is much closer to the median (998) meaning that this has much less skew than Padded IBI
            - Graphing out the counts of the data show that it is fairly normal (especially compared to Padded IBI)
6. UTC time
    - Data type: int
    - Meaning: The float representation of date/time in the UTC time zone. It is a float that represents the number of seconds that has passed from January 1st, 1970 at UTC. It is essentially the Unix time representation of the date/time that IBI was measured.
    - values/statistics:
        - Range: 1543351057-1544628851
            - 11/27/2018 @ 8:37pm (UTC) - 12/12/2018 @ 3:34pm (UTC)
        - Median: 1543995783.5
            - (rounded up) 12/05/2018 @ 7:43am (UTC)
        - IQR: (1544321582.0-1543657935.75)
            - (rounded up) 12/09/2018 @ 2:13am (UTC) - 12/01/2018 @ 9:52am (UTC)
        - Number of missing values: 235
        - NOTES:
            - Same data points that are missing IBI are missing UTC time
            - Generally, there is little to no skewness in the data

NOTES ON DATA:

- At each time, IBI is measured 6 times. This could be 6 different sensors taking measurements.

Sleep Data

1. Date
    - Data type: String
    - Meaning: Date that sleep data was recorded. Date SHOULD be when the sleep period ended. Format is in 'DD.MM.YYYY'
    - values/statistics: VIEW BEDTIME END UNIX
2. Bedtime Start Unix
    - Data type: Float
    - Meaning: Time that sleep data began to be recorded (which is sensed by the ring). It is a float that represents the number of seconds that has passed from January 1st, 1970 at UTC. It is essentially the Unix time representation of the date/time that sleep data began to be recorded.
    - values/statistics:
        - Range: 1543350953-1544598810
            - 11/27/2018 @ 8:35pm (UTC) - 12/12/2018 @ 7:13am (UTC)
        - Median: 1543954412
            - 12/04/2018 @ 8:13pm (UTC)
        - IQR: (1544349481-1543559378)
            - 12/09/2018 @ 9:58am (UTC) - 11/30/2018 @ 6:29am (UTC)
        - Number of missing values: 0
        - NOTES:
3. Bedtime end Unix
    - Data type: Float
    - Meaning: Time that sleep data stopped being recorded (which is sensed by the ring). It is a float that represents the number of seconds that has passed from January 1st, 1970 at UTC. It is essentially the Unix time representation of the date/time that sleep data stopped being recorded.
    - values/statistics:
        - Range: 1543351793-1544628690
            - 11/27/2018 @ 8:49pm (UTC) - 12/12/2018 @ 3:31pm (UTC)
        - Median: 1543955132
            - 12/04/2018 @ 8:25pm (UTC)
        - IQR: (1544370601-1543590818)
            - 12/09/2018 @ 3:50pm (UTC) - 11/30/2018 @ 3:13pm (UTC)
        - Number of missing values: 0
        - NOTES:
4. Bedtime start
    - Data type: String
    - Meaning: Time that sleep data began to be recorded (which is sensed by the ring). Time is represented by a string that formatted in 'HH:MM:SS'
    - values/statistics: VIEW BEDTIME START UNIX
5. Bedtime end
    - Data type: String
    - Meaning: Time that sleep data stopped being recorded (which is sensed by the ring). Time is represented by a string that formatted in 'HH:MM:SS'
    - values/statistics: VIEW BEDTIME END UNIX
6. TimeZone

- Data type: Float
- Meaning: a float representing the time offset from UTC of the time zone where the sleep data was recorded. A majority of this sample data was taken in California, which is 8 hours less than the standard UTC time. This link is a helpful resource to determine timezone: https://www.timeanddate.com/time/map/
- values/statistics:
    - Range: -8-(-8)
    - Median: -8
    - IQR: 0 (-8-(-8))
    - Number of missing values: 0
    - NOTES:
        - Eventually a method may be needed to standardize everything into a regular timezone (most likely UTC to keep things consistent)

7. Debug info
    - Data type: String
    - Meaning: A string that lets the programmer know of any issues that occurred with recording sleep data. So far, the data has shown that there is either no debug info (AKA sleep detection worked correctly) or will say 'Bedtime detection failed', in which case no further data will be recorded. There may be more errors, however in our sample this was the only debug issue that occurred.
    - values/statistics:
        - Range: N/A
        - Median: N/A
        - IQR: N/A
        - Number of missing values: 35
        - NOTES:
            - May potentially require a different dictionary in the future of all potential debugs
8. Battery consumption
    - Data type: String
    - Meaning: A string that represents the battery consumption during the sleep measurement. Not all data points have a measurement, with a majority having no data on battery consumption. There is a high chance that battery consumption is not taken into consideration unless it is greater than 1%. For the data we have currently, this is relatively unimportant but might be useful to analyze if we have data points with an irregular amount of battery consumption.
    - values/statistics:
        - Range: 1-2
        - Median: 1.47
        - IQR: .5 (1.7-1.2)
        - Number of missing values: 22
        - NOTES:
            - These values are strings because of the percentage sign. In the future, it would be most useful to just have them as floats (without the percentage sign) and to just include the percentage in the column header
9. Is longest
    - Data type: Float
    - Meaning: UNSURE of what this represents. This is definitely a float that represents a boolean, however, I'm unsure of which is which. Based on certain groupby's, it seems that items with is longest valued at 1 have significantly larger times in bed, which would suggest 1 being 'True'. However, when

grouped by date, there are multiple values being 1, suggesting that it doesn't signify a longest sleep time or anything of that nature. This variable DEFINITELY needs further explanation.

- values/statistics:
    - Range: 0-1
    - Median: 1
    - IQR: 1 (1-0)
    - Number of missing values: 2
    - NOTES:
        - Is longest is included in the Oura API, but there is no explanation to what it actually is

10. Time in bed
    - Data type: Float
    - Meaning: The duration of time in bed in minutes.
    - values/statistics:
        - Range: 3-579
        - Median: 46
        - IQR: 478(498-20)
        - Number of missing values: 2
        - NOTES:
            - A majority of these points are on the lower side (data is strongly skewed right); it might be effective for analysis to split up the data into sleep data and nap data
            - It might be wise to remove points where time in bed is high but sleep minutes as 0 (therefore implying no sleep); the only reason that this data might be kept would be to analyze why no sleep occurred
11. Sleep score
    - Data type: Float
    - Meaning: 'Sleep score represents overall sleep quality during the sleep period. It is calculated as a weighted average of sleep score contributors that represent one aspect of sleep quality each. The sleep score contributor values are also available as separate parameters.' (source: website).
    - values/statistics:
        - Range: 0-94
        - Median: 26
        - IQR: 60 (83-23)
        - Number of missing values: 2
        - NOTES:
            - Score is typically 1-100, but if the score is unavailable or incalculable, score will be 0. Would be interesting to find why score would be 0
                - There is only one score that is 0; there seemed to be no issues with detecting sleep. The sleep time was 0, but there were plenty of other data points where score is 0. What could possibly make this score 0 is the fact that it was an extremely short in bed period (3 minutes) (data point 34)
            - There's a large left skewness
12. Sleep minutes
    - Data type: Float
    - Meaning: The duration of time spent sleeping in minutes. Should add up to wake minutes to total up to time in bed.
    - values/statistics:
        - Range: 0-514
        - Median: 9

        - IQR: 427 (427-0)
        - Number of missing values: 2
        - NOTES:
            - It might be useful to determine when sleep measures are determined. How does the ring understand that the user in bed, and when does it acknowledge that the user is asleep?
13. Wake minutes
    - Data type: Float
    - Meaning: The duration of time spent laying in bed not asleep in minutes. Should add up to sleep minutes to total up to time in bed.
    - values/statistics:
        - Range: 3-223
        - Median: 39
        - IQR: 41 (53-12)
        - Number of missing values: 2
        - NOTES:
            - It might be useful to determine when sleep measures are determined. Does the ring automatically stop measurements once user gets out of bed?
14. REM minutes
    - Data type: Float
    - Meaning: The duration of time spent in REM sleep cycle in minutes. REM, light, and deep sleep minutes should add up to sleep minutes
    - values/statistics:
        - Range: 0-218
        - Median: 0
        - IQR: 131 (131-0)
        - Number of missing values: 2
        - NOTES:
            - More information can be found here: https://blog.fitbit.com/sleep-stages-explained/
15. Light minutes
    - Data type: Float
    - Meaning: The duration of time spent in light sleep cycle in minutes. REM, light, and deep sleep minutes should add up to sleep minutes
    - values/statistics:
        - Range: 0-330
        - Median: 8
        - IQR: 203 (203-0)
        - Number of missing values: 2
        - NOTES:
            - More information can be found here: https://blog.fitbit.com/sleep-stages-explained/
16. Deep minutes
    - Data type: Float
    - Meaning: The duration of time spent in deep sleep cycle in minutes. REM, light, and deep sleep minutes should add up to sleep minutes
    - values/statistics:
        - Range: 0-96
        - Median: 0
        - IQR: 65.5 (65.5-0)
        - Number of missing values: 2

        - NOTES:
            - More information can be found here: https://blog.fitbit.com/sleep-stages-explained/
17. Efficiency
    - Data type: Float
    - Meaning: 'Sleep efficiency is the percentage of the sleep period spent asleep (100% * sleep.total / sleep.duration)' (source: website)
    - values/statistics:
        - Range: 0-97
        - Median: 48
        - IQR: 87 (87-0)
        - Number of missing values: 2
        - NOTES:
18. Lowest HR time minutes
    - Data type: Float
    - Meaning: UNSURE; SHOULD BE the minute marker of when the lowest heart rate time was at. However, could be the duration of time in minutes that heart rate was at its lowest during the sleep period (although unlikely, especially since the max of lowest HR time minutes is in the triple digits). If it is the alternative, it is highly likely that 'lowest HR' is simply just a minimum threshold; most of the values are much too large for it to be the number of minutes that lowest heart rate was at.
    - values/statistics:
        - Range: 0-406
        - Median: 20
        - IQR: 30 (38-8)
        - Number of missing values: 2
        - NOTES:
19. Lowest HR
    - Data type: Float
    - Meaning: The lowest heart rate achieved during the sleep period measurement. The lowest heart rate occurs during sleep and not during a wake period.
    - values/statistics:
        - Range: 0-100
        - Median: 56
        - IQR: 22 (73-51)
        - Number of missing values: 2
        - NOTES:
            - There may be some outliers, since heart rate goes down to 50. However, this may truly be the lowest heart rate measurement, since ~50 is still the 25th percentile 8
20. Wake-up count
    - Data type: Float
    - Meaning: The number of times the user woke up during the sleep period. Most likely measured by change of HR, breath, or break of sleep cycle
    - values/statistics:
        - Range: 0-5
        - Median: 0
        - IQR: 2 (2-0)
        - Number of missing values: 2
        - NOTES:
21. Sleep latency

- Data type: Float
- Meaning: ‘Detected latency from bedtime_start to the beginning of the first five minutes of persistent sleep’ (source: website). Essentially the amount of time it takes for the user to fall asleep, or the time between the start of bedtime to the beginning of sleep; MOST LIKELY measured in minutes, however could be measured in seconds or another unit of time.
- values/statistics:
    - Range: 0-25
    - Median: 3
    - IQR: 15 (15-0)
    - Number of missing values: 2
    - NOTES:
        - According to the API description on the website, latency score talks about latency being in minutes, giving further assurance that sleep latency is also measured in minutes.

22. Breath
    - Data type: Float
    - Meaning: UNSURE; could possibly be the maximum number of breaths in a minute during the sleeping duration (due to the significantly larger values than breath_v).
    - values/statistics:
        - Range: 13.75-14.375
        - Median: 14
        - IQR: 0.3125 (14.25-13.9375)
        - Number of missing values: 2
        - NOTES:
23. Breath_v
    - Data type: Float
    - Meaning: The average respiratory rate. Should be the average number of breaths per minute (a rate measurement).
    - values/statistics:
        - Range: 2.625-3.75
        - Median: 3
        - IQR: 0.3125(3.1875-2.875)
        - Number of missing values: 2
        - NOTES:
            - Unsure of the comparison between breath and breath_v, however interesting that the IQR is the same range (different by a factor of ~4.46)
24. Average HR
    - Data type: Float
    - Meaning: ‘The average heart rate registered during the sleep period.’ (source: website). Measurement is a rate and is average beats per minute during the sleep measurement duration.
    - values/statistics:
        - Range: 55.875-62.750
        - Median: 58.375
        - IQR: 3.125 (60.125-57)
        - Number of missing values: 2
        - NOTES:
25. SleepMidPoint
    - Data type: String

- Meaning: The data is a string object in the format ‘HH:MM’ or ‘H:MM’ that represents the midpoint time of sleep measurement. The time is written in military time to avoid AM/PM confusion. To compute values and statistics, the values were converted to time objects utilizing datetime.
- values/statistics:
    - Range: 02:07-23:39
    - Median: 12:41
    - IQR: (14:49-3:00)
    - Number of missing values: 2
    - NOTES:
        - On the website, the SleepMidPoint is described as ‘Detected latency from bedtime.start to the beginning of the first five minutes of persistent sleep.’ However, this is similar to the latency description, which is right above it. It’s most likely that this is an error

26. Temperature
    - Data type: Float
    - Meaning: SHOULD BE ‘Skin temperature deviation from the long-term temperature average’ (source: website). Should be in degrees celsius. However, the values are quite large, so this could potentially be the average skin temperature during the sleep measurement period.
    - values/statistics:
        - Range: 0-36.78
        - Median: 36.67
        - IQR: 0.14 (36.7-36.56)
        - Number of missing values: 2
        - NOTES:
27. restless%
    - Data type:
    - Meaning: ‘Restlessness of the sleep time, i.e. percentage of sleep time when the user was moving.’
    - values/statistics:
        - Range: 2-102
        - Median: 41.5
        - IQR: 4 (45-40)
        - Number of missing values: 15
        - NOTES:
            - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
                - Missing values could be due to no sleep, no movement, or other factors
            - There also seems to be discrepancies within the data; there’s a data point where restlessness is at more than 100%; this is either an error it means that restless % also factors in movement when not in sleep
28. Got up count
    - Data type: Float
    - Meaning: Should be the number of times that the user got up during the sleep measurement period. Unsure of how the ring measures this (most likely large change in heart rate and x/y/z coordinates, however unsure of why it continues measuring sleep duration if not asleep)
    - values/statistics:
        - Range: 0-1
        - Median: 0

    - IQR: 0 (0-0)
    - Number of missing values: 3
    - NOTES:
29. Score total
    - Data type: Float
    - Meaning: ‘Represents total sleep time's (see sleep.total) contribution for sleep quality. The value depends on age of the user - the younger, the more sleep is needed for good score. The weight of sleep.score_total in sleep score calculation is 0.35.’ (source: website) TL;DR: The part of the score purely dependent on amount of sleep. This is the largest aspect of the score.
    - values/statistics:
        - Range: 1-97
        - Median: 1
        - IQR: 76.75 (77.75-1)
        - Number of missing values: 3
        - NOTES:
            - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
            - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don’t have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
            - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?
30. Score deep
    - Data type: Float
    - Meaning: ‘Represents deep (N3) sleep time's (see sleep.deep) contribution for sleep quality. The value depends on age of the user - the younger, the more sleep is needed for good score. The weight of sleep.score_deep in sleep score calculation is 0.10.’ (source: website) TL;DR: The part of the score purely dependent on the amount of deep sleep.
    - values/statistics:
        - Range: 1-96
        - Median: 1
        - IQR: 73.25 (74.25-1)
        - Number of missing values: 3
        - NOTES:
            - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
            - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don’t have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
            - NOTE: there is no score for light sleep
            - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?
31. Score REM
    - Data type: Float
    - Meaning: ‘Represents REM sleep time (see sleep.rem) contribution for sleep quality. The value depends on age of the user - the younger, the more sleep REM is needed for good score. The weight

of sleep.score_rem in sleep score calculation is 0.10.' TL;DR: The part of the score purely dependent on the amount of REM sleep.
- values/statistics:
    - Range:
    - Median: 1
    - IQR: 97.25 (98.25-1)
    - Number of missing values: 3
    - NOTES:
        - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
        - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don't have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
        - NOTE: There is no score for light sleep
        - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?

32. Score efficiency
- Data type: Float
- Meaning: 'Represents sleep efficiency's (see sleep.efficiency) contribution for sleep quality. The higher efficiency, the higher score. The weight of sleep.score_efficiency in sleep score calculation is 0.10.' TL;DR: The part of the score purely dependent on the efficiency of sleep.
- values/statistics:
    - Range: 1-100
    - Median: 20
    - IQR: 87 (88-1)
    - Number of missing values: 3
    - NOTES:
        - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
        - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don't have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
        - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?
        - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?

33. Score latancy (should be latency?)
- Data type: Float
- Meaning: 'Represents sleep onset latency's (see sleep.onset_latency) contribution for sleep quality. A latency of about 15 minutes gives best score. Latency longer than that many indicate problems falling asleep, whereas a very short latency may be a sign of sleep debt. The weight of sleep.score_latency in sleep score calculation is 0.10.' (source: website) TL;DR: The part of the score purely dependent on how long it takes for the user to fall asleep AKA sleep latency.
- values/statistics:
    - Range: 76-98
    - Median: 81.5

    - IQR: 14 (90-76)
    - Number of missing values: 3
    - NOTES:
        - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
        - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don't have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
        - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?

34. Score Disturbances
    - Data type: Float
    - Meaning: 'Represents sleep disturbances contribution for sleep quality. Three separate measurements are used to calculate this contributor value:
        - Wake-up count (sleep.wake_up_count) - the more wake-ups, the lower score.
        - Got-up count (sleep.got_up_count) - the more got-ups, the lower score.
        - Restless sleep (sleep.restless) - the more motion detected during sleep, the lower score.

      Each of these three values has weight 0.05 in sleep score calculation, giving sleep.score_disturbances total weight of 0.15.' (source: website) TL;DR: Score that factors wake-up count, got-up count, and restless sleep. This is the second largest weighted item.
    - values/statistics:
        - Range: 66-99
        - Median: 99
        - IQR: 21.75 (99-77.25)
        - Number of missing values: 3
        - NOTES:
            - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
            - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don't have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
            - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?

35. Score Alignment
    - Data type: Float
    - Meaning: 'Represents circadian alignment's contribution for sleep score. Sleep midpoint time (sleep.midpoint_time) between 12PM and 3AM gives highest score. The more the midpoint time deviates from that range, the lower score. The weigh of sleep.score_alignment in sleep score calculation is 0.10.' TL;DR: This aspect of the score is dependent on the time of sleep. Because it is assumed that sleep is most effective between 12PM and 3AM, more sleep that occurs between that sleep period leads to a higher score.
    - values/statistics:
        - Range: 1-100
        - Median: 15
        - IQR: 97.75 (98.75-1)
        - Number of missing values: 3

- NOTES:
    - Is an int, however reading in file reads it in as a string. May lead to issues in the future in data analysis. ISSUE IS THAT NaN CANNOT BE ANALYZED, therefore NaN needs to be dropped.
    - In order for any score to be analyzed, you MUST drop any data points where sleep score is 0 (since these data points don't have any score). Realistically, this value should be 0 (according to the website), so this may need to be changed in the future
    - Could potentially use AI to learn what weights should be applied; could it be possible that some factors are more important than others for different users?

36. FW Version
    - Data type: Float
    - Meaning: There are no data points in this column. UNSURE of what FW stands for, however this is most likely just a description of a data and actually a measurement of data. Could it possibly stand for fall/winter version?
    - values/statistics:
        - Range: N/A
        - Median: N/A
        - IQR: N/A
        - Number of missing values: 37
        - NOTES:
37. Sleep phases
    - Data type: Float
    - Meaning: There are no data points in this column. There are a lot of columns that follow along this column, which suggests that the following columns could be related to sleep phases. However, the following columns could also be regarding items such as hypnograms or rMMSD. UNSURE
    - values/statistics:
        - Range: N/A
        - Median: N/A
        - IQR: N/A
        - Number of missing values: 37
        - NOTES:
38. 0-1739
    - MOST LIKELY HYPNOGRAMS: 'A string that contains one character for each starting five minutes of the sleep period, so that the first period starts from sleep.bedtime.start: - '1' = deep (N3) sleep - '2' = light (N1 or N2) sleep - '3' = REM sleep - '4' = awake'. In the dataset, the string is replaced with a number instead
    - Statistics would vary a little bit for this data. Rather than analyzing all the data, it might be better to analyze per data point or to organize more into how long it takes for patterns to begin occurring. It'd be a good move to decide how to analyze these later on
    - NOTES:
        - These are the only possible values: {'', '1', '100', '2', '200', '3', '300', '4', '400'}. The triple digit numbers ALWAYS occur near the end, and always occur after 2 spaces. Triple digit numbers could potentially mean a separate cycle, or could mean it's different data. Could triple digits potentially mean rMMSD?
        - There is ALWAYS double the number of single number digits than the number of minutes in bed; THESE ARE PROBABLY MEASUREMENTS OF SOMETHING EVERY HALF MINUTE
        - There is ALWAYS the same number of triple number digits as the number of minutes in bed; THESE ARE PROBABLY MEASUREMENTS OF SOMETHING EVERY MINUTE

- The triple number digits follow the same, if not very similar, pattern to the single digit numbers. I want to say that these triple digit numbers are generalizations of the more precise single digit numbers
- This data is useful in that it gives an idea of how long each cycle lasts. This could be useful to help users determine how long they should nap, especially if cycles are consistent

Spreadsheet tab	Element or value display name	Description	Data type	Character Length	Acceptable Values	Required?	Accepts null value?
S1	454db1	454db1 source Database searched for RNA sequences	varchar	255	n/a	y	n
S1	SGN	SGN source Database searched for RNA sequences	varchar	255	n/a	y	n
S1	All unigenes	Total number of unigenes from database searched	integer	255	All unigenes	y	n
S1	Matched Unigenes	Number of matched unigenes from database searched	integer	255	Matched Unigenes	y	n
S1	Average Length	Average length of RNA sequences obtained from 454 sequencing experiments databases	integer	255	0.0-9999	y	n
S1	Date	date	date	10	YYYY-MM-DD	y	n
S1	Standard Deviation	Standard Deviation of length of RNA sequences	integer	20	0.0-9999	y	n

Reports and Datasets – the Contents Report

Column	Value
Variable	Name of the variable
Count	Count of non-missing values
Filled	% of rows that are filled with data
NMiss	Number of missing values
Miss_Pct	% of rows with missing values
Unique	Number of unique data values
Unique_Pct	Unique / Filled formatted as % of Unique values filled
Unique_Pct_All	Unique / Count formatted as % of Unique values overall
Stats	'Y' if this can be a class value or a report variable; in other words, it's a discrete variable, and not a key or a continuous variable
Contents Data	SAS metadata values for data type, length, Format, InFormat, Label, and varnum

Dataset_Profiler.sas is detailed here. Dataset_Explorer.sas is detailed in my paper, Using Dictionary Tables to Explore SAS Datasets.

https://www.sleepfoundation.org/articles/good-fair-or-poor-how-well-do-you-sleep-0

- Delves into the factors that the national sleep foundation deem important

https://www.healthline.com/health/how-much-deep-sleep-do-you-need#sleep-deprivation

- More information on the sleep cycles themselves

https://www.dreams.co.uk/sleep-matters-club/4-alternative-sleeping-cycles-infographic/

- Different styles of sleep/different sleep patterns; may be useful for predictive modeling based off of sleep patterns

http://healthysleep.med.harvard.edu/healthy/science/variations/individual-variation-genetics

- Genetic variations that lead to different aspects of sleep; try to find some articles about whether different people need different amounts of sleep in the sleep cycles

Prepare:

- Questions on the unsure variables

Work with:

- IBI: how long in a day is IBI detected well, do own calculations with HRV
- Sleep: cycling of phases (when are you in deep sleep)
    - Model between amount of sleep and cycles of sleep

Read:

- Research paper and references