




PSY 710
DR. FENGQING ZHANG
DATA ANALYSIS II
BY TINASHE M. TAPERA


"UNDERSTANDING MY CRAZY SLEEPING HABITS WITH SLEEP CYCLE"

“Waking up easy is all about timing. Sleep Cycle alarm clock tracks your sleep patterns and wakes you up during light sleep. Waking up during light sleep feels like waking up naturally rested without an alarm clock.”


–[SLEEPCYCLE.COM](https://www.sleepcycle.com)

 **Sleep Cycle**


[Start](#) [How it works](#) [News & Press](#) [Support](#)



Place your phone on your nightstand.
(Android? Place in bed)



While you sleep, Sleep Cycle analyzes your sleep.



When it's time, it wakes you up in your lightest sleep phase.

[Download for iPhone](#) [Download for Android](#)

FRAMING THE PROJECT

- I've developed very unusual sleeping habits
- A lot of people tell me they're unhealthy. *But are they really that bad?*
- I purchased SleepCycle in 2011 and have been collecting data ever since
- Over arching questions to ask:
 - How has my sleep changed over time?
 - Has my sleep quality improved? Decreased?
 - If my sleep is bad, why? What can be done to improve it?
 - What can I infer from this data?
- Approach:
 - Export a .csv of all of my sleep data
 - Analyse using R and apply the statistical tools we've learnt to find something interesting in the data
 - Report and visualise important findings

EXPLORATORY FINDINGS (PROVIDED BY SLEEP CYCLE)

- Data ranges from late 2011 to present (May 2017)
- Each observation is one night, with variables measuring sleep quality, wake up mood, heart rate and activity for the previous day, and sleep notes that the user specifies.

AVERAGE SLEEP QUALITY	58%
AVERAGE BED TIME	01:35AM
AVERAGE LENGTH OF SLEEP	5H 45M
BEST DAY FOR SLEEP	FRIDAY
TOTAL TIME IN BED	8.9 MONTHS

EXPLORATORY FINDINGS (NOT PROVIDED BY SLEEP CYCLE)

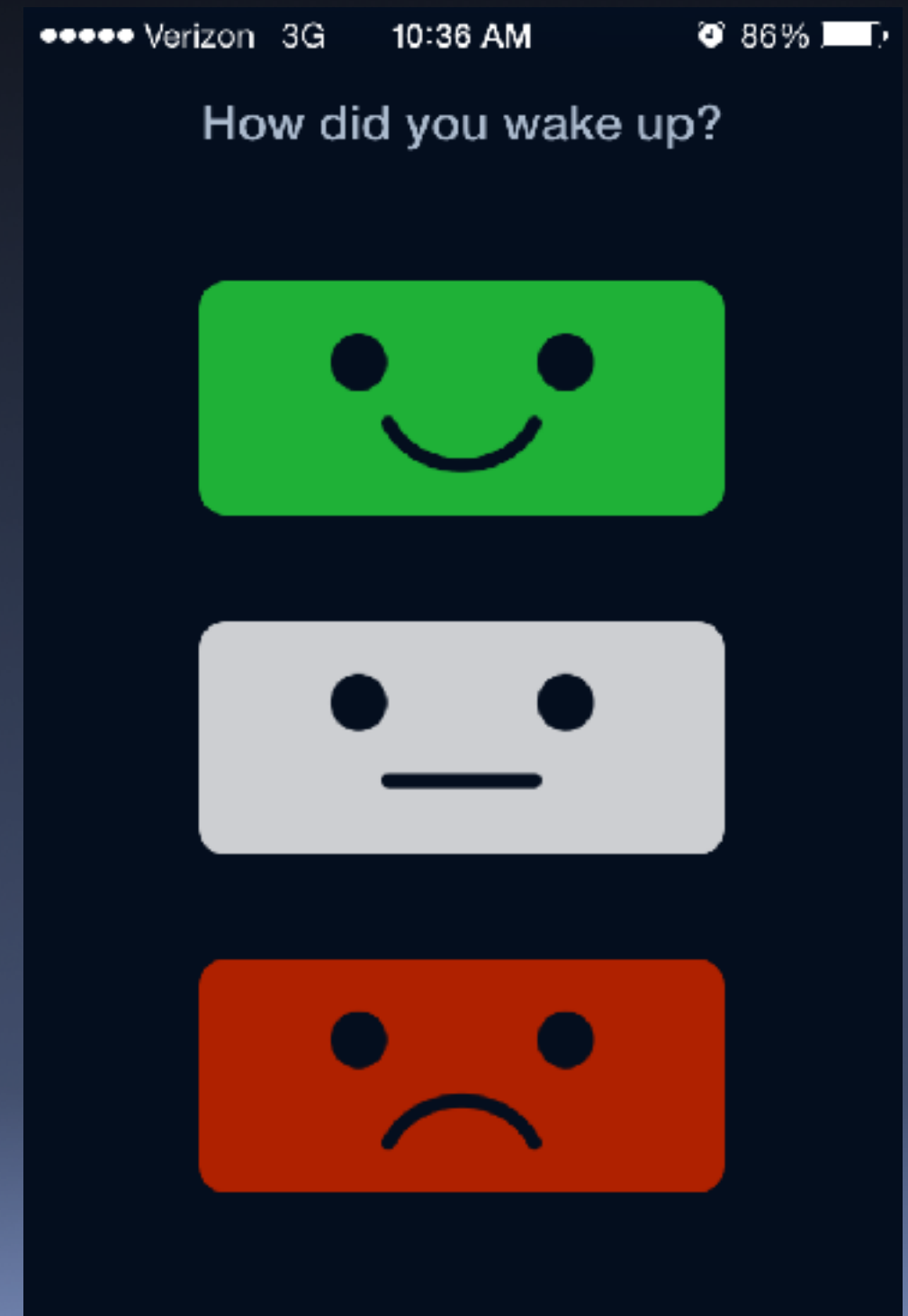
- Wake-up Moods

- I've recorded wake up moods for the past ~4 years. Ratio of moods:

:)	:	:(
1	8	1

- Sleep Notes

- User-defined variables indicating the conditions you went to sleep under
- Recorded before the night begins
- Therefore, mood may be causally inferred from sleep notes





CAN I PREDICT MY WAKE UP MOOD?

WHAT VARIABLES ARE IMPORTANT FOR PREDICTING WAKE UP MOOD?

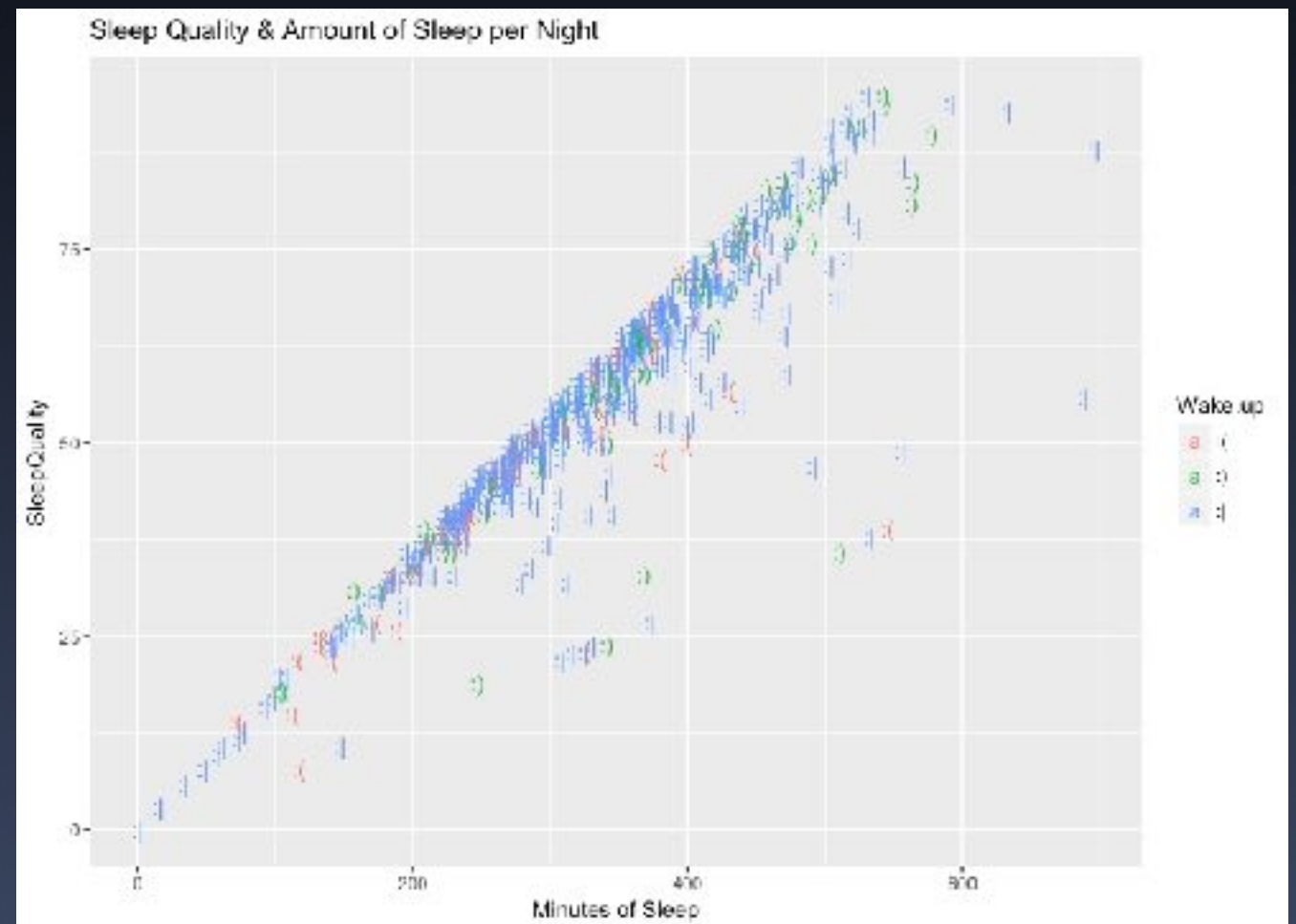
CAN I INFER THE BEST WAY TO IMPROVE MY MOOD?

HOW ACCURATE CAN THIS PREDICTION/INFERENCE BE?

WOULD I BE ABLE TO IMPROVE MY HEALTH?

DATA PREP & CLEANING

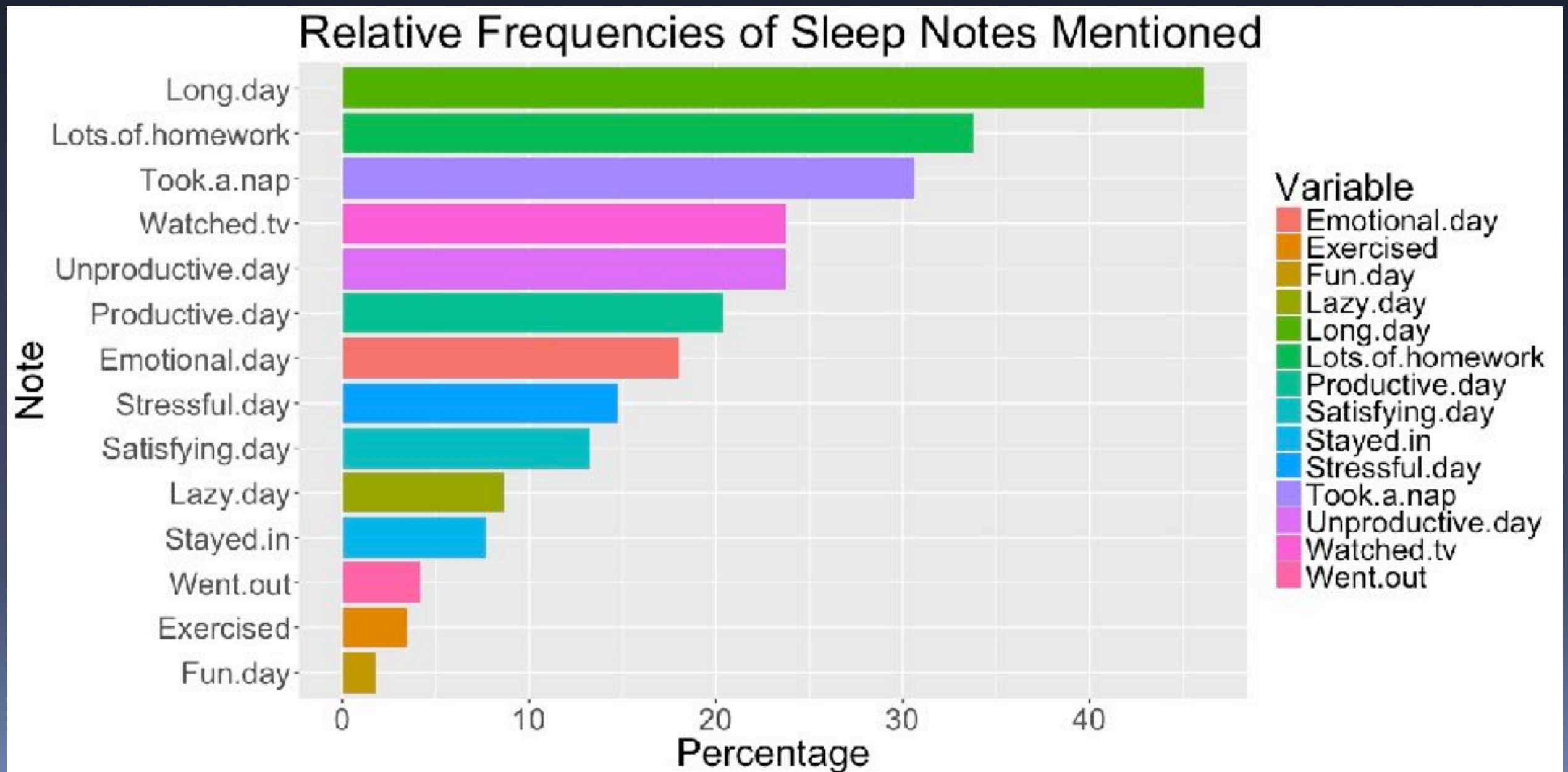
- Remove the variables for heart rate and motion tracking
 - Limited
 - Inaccurate
- Remove outliers (recordings of < 2 hours of sleep)
 - Technically a nap, not enough time for a full sleep cycle
 - Keep outliers with large values as I do have days when I sleep for 12+ hours
- Convert sleep notes from strings to a set of binary dummy variables



EMOTIONAL DAY	LONG DAY	LOTS OF HOMEWORK	SATISFYING DAY	STRESSFUL DAY	LAZY DAY	UNPRODUCTIVE DAY
WENT OUT	PRODUCTIVE DAY	STAYED IN	WATCHED TV	FUN DAY	TOOK A NAP	EXERCISED

ANALYSIS OF SLEEP NOTES

WHAT IS THE LIKELIHOOD OF A SLEEP NOTE ON AN AVERAGE NIGHT?



FITTING A MULTINOMIAL MODEL

IN THEORY...

- To predict wake-up mood, fit a logistic classifier with three outcome values
- Use sleep notes as binary conditions
- Also include sleep quality & duration (although they are highly collinear)
- Can predict the likelihood of increasing wake-up mood by understanding the probabilities of change given by the coefficients in the model.
- Evaluation:
 - AIC
 - Overall prediction accuracy on unseen data

$$\begin{aligned}\Pr(Y_i = 1) &= \frac{e^{\beta_1 \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot \mathbf{X}_i}} \\ \Pr(Y_i = 2) &= \frac{e^{\beta_2 \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot \mathbf{X}_i}} \\ &\dots\dots\dots \\ \Pr(Y_i = K - 1) &= \frac{e^{\beta_{K-1} \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot \mathbf{X}_i}}\end{aligned}$$

FITTING A MULTINOMIAL MODEL

THE MODEL IN R

- `multinom(formula = Wake.up ~ ., data = train)`
- Residual Deviance = 414.7438
 - "How well the response is predicted by the model when the predictors are included."
- AIC = 482.7438
 - "A measure of the relative quality of statistical models for a given set of data"

FITTING A MULTINOMIAL MODEL

MODEL COEFFICIENTS

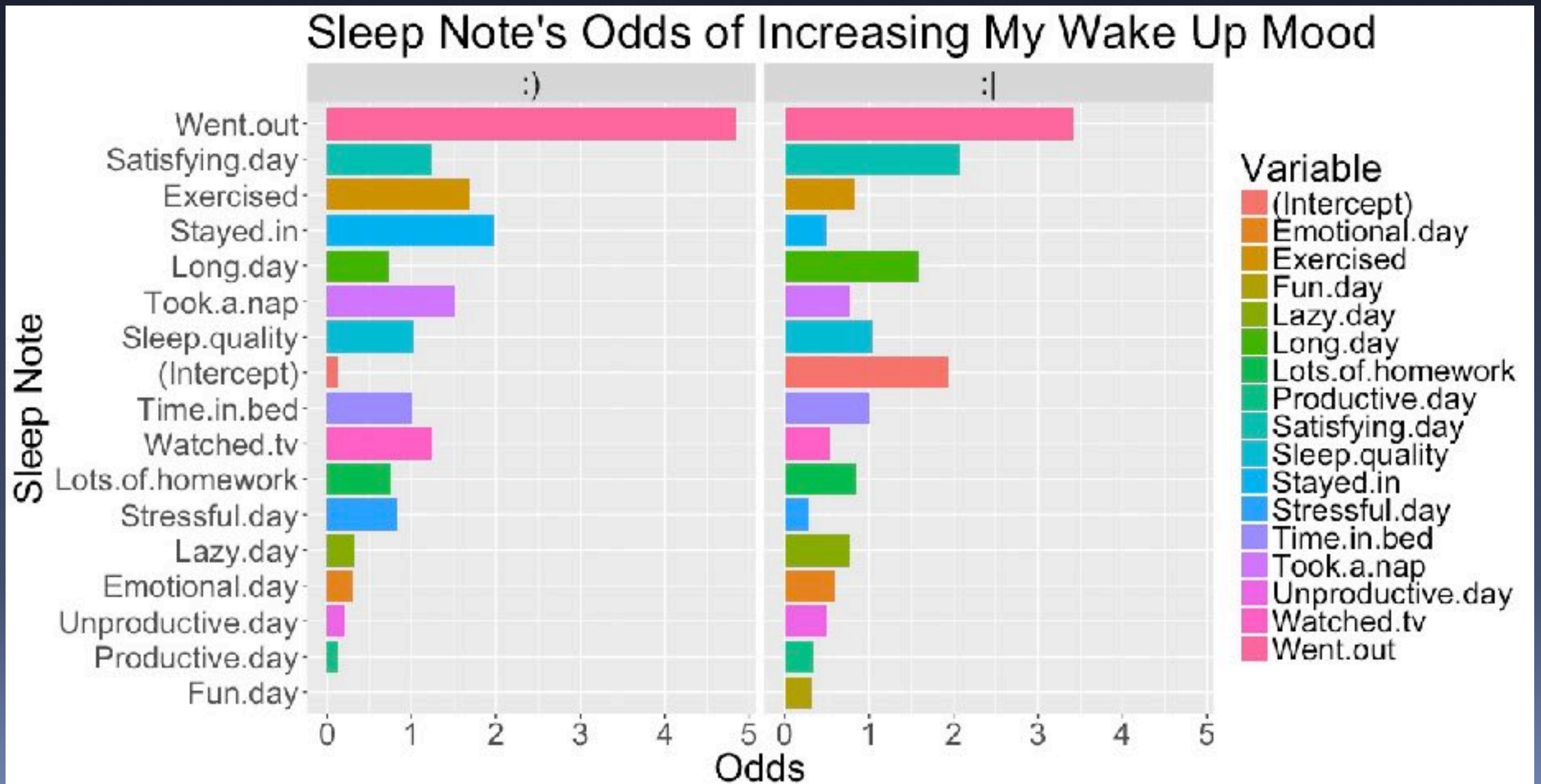
Coefficients:

	(Intercept)	Sleep.quality	Time.in.bed	Emotional.day	Long.day	
:)	-2.0351907	0.02934627	0.004723354	-1.2019508	-0.3094883	
:	0.6567463	0.03308257	0.002052723	-0.5274769	0.4602935	
	Lots.of.homework	Satisfying.day	Stressful.day	Lazy.day	Unproductive.day	
:)	-0.2819415	0.1999768	-0.1829924	-1.1491519	-1.5497615	
:	-0.1691941	0.7273376	-1.2294273	-0.2593855	-0.7029491	
	Went.out	Productive.day	Stayed.in	Watched.tv	Fun.day	Took.a.nap
:)	1.575614	-2.128149	0.6763165	0.2155899	-15.884250	0.4188251
:	1.230660	-1.107775	-0.7265632	-0.6414547	-1.134042	-0.2485300
	Exercised					
:)	0.5206860					
:	-0.1862393					

Interpretation: given that the expected wake-up mood was :(that night, a one unit change in sleep quality gives a 0.033 average increase in log-odds of having a :| mood, and a 0.029 average increase in log-odds of having a :) mood.

FITTING A MULTINOMIAL MODEL

WHAT ARE THE ODDS OF A SLEEP NOTE INCREASING MY WAKE UP MOOD?



FITTING A MULTINOMIAL MODEL

PREDICTOR SIGNIFICANCE & MODEL ACCURACY

- Using Wald's Z-test to find the p -values:
 - (Intercept) is significant for :) mood ($p = 0.0231$, $b = -2.04$)
 - Stressful.day is significant for :| mood ($p = 0.0139$, $b = -1.23$)
 - Productive.day is significant for :) mood ($p = 0.0209$, $b = -2.13$)
 - Fun.day is significant for :) mood ($p = 0$, $b = -15.88$)
- Using this model, predict the wake-up mood for 40% of the unseen data:
 - Overall Accuracy: 75%

ATTEMPT TO IMPROVE THE MODEL

FIT A WEIGHTED CLASSIFIER

- For those minority groups, over sample observations and assign a weight of 1
- For majority group, under sample observations and assign a weight in proportion to under sampling ratio
- Using a weight of $n = 10$:
 - Residual Deviance = 385.0152
 - AIC = 425.0152
 - Prediction accuracy = 71%

CONCLUSIONS

SUMMARY OF RESULTS

- SleepCycle produces a handy dataset for analysis
- Sleep notes have varying frequencies of being reported:
 - Greatest frequency: "Long day"
 - Lowest frequency: "Fun day"
 - Unsurprising frequency: "Lots of homework" [2nd]
 - Surprising frequency: "Exercised" [2nd from last]
- Multinomial logistic regression can be used to predict wake-up mood using sleep notes, sleep duration and quality, with 75% accuracy
 - Stressful, productive, & fun days have significant coefficients for the model
 - Significant coefficients are counterintuitive and do not align with frequencies
- Weighted regression does not improve accuracy
 - However, model fit by AIC is higher (weighted = 425; unweighted = 483)

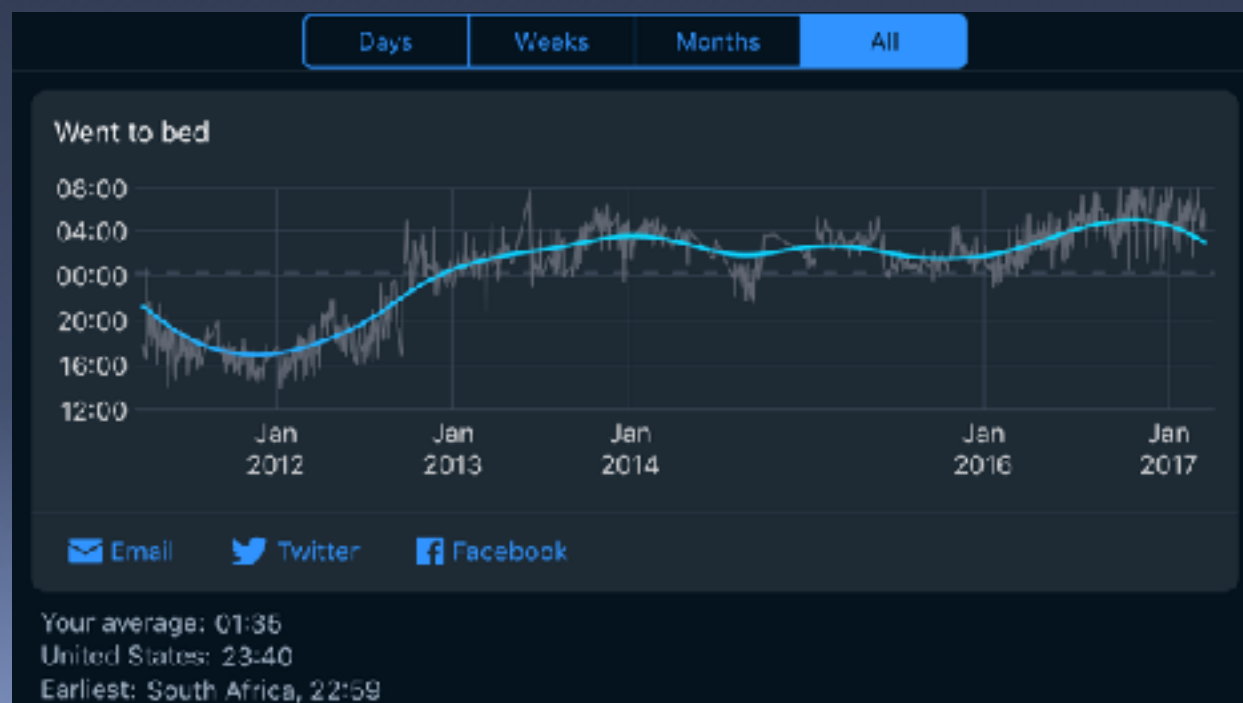
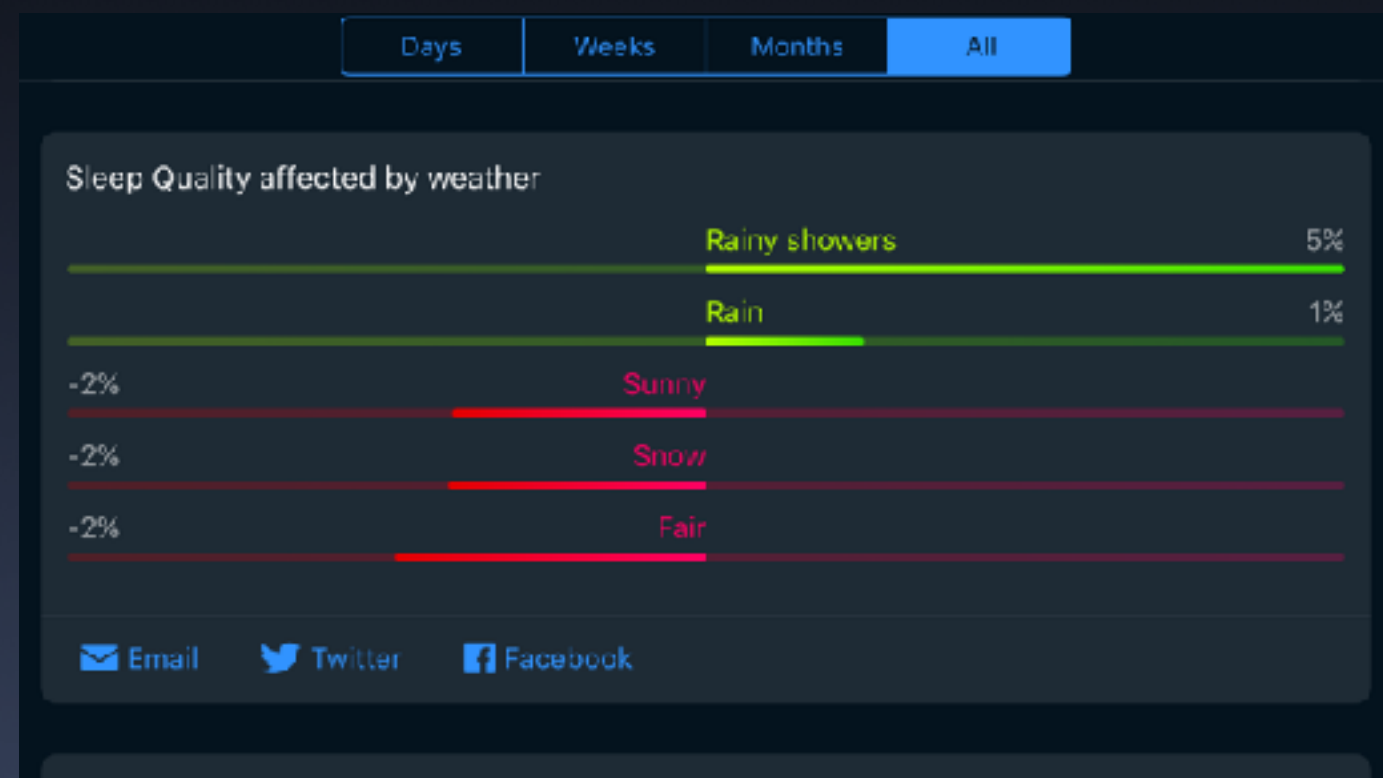
CONCLUSIONS

LIMITATIONS AND IMPROVEMENTS

- **THE DATA PROBABLY VIOLATES THE INDEPENDENT OBSERVATIONS ASSUMPTION**
 - Probably more appropriate for Growth Curve Modeling
- Sleep Quality & Sleep Duration were highly collinear
- Class imbalance problems are a huge challenge to classification
 - Weighted regression is helpful but somewhat complicated; I may have done it wrong
 - Change point detection or outlier detection may be a more appropriate approach
- On most mornings my mood is within a very wide range that I consider :|, and in order for me to put in a :) or :(, something pretty significant must have happened to my mood

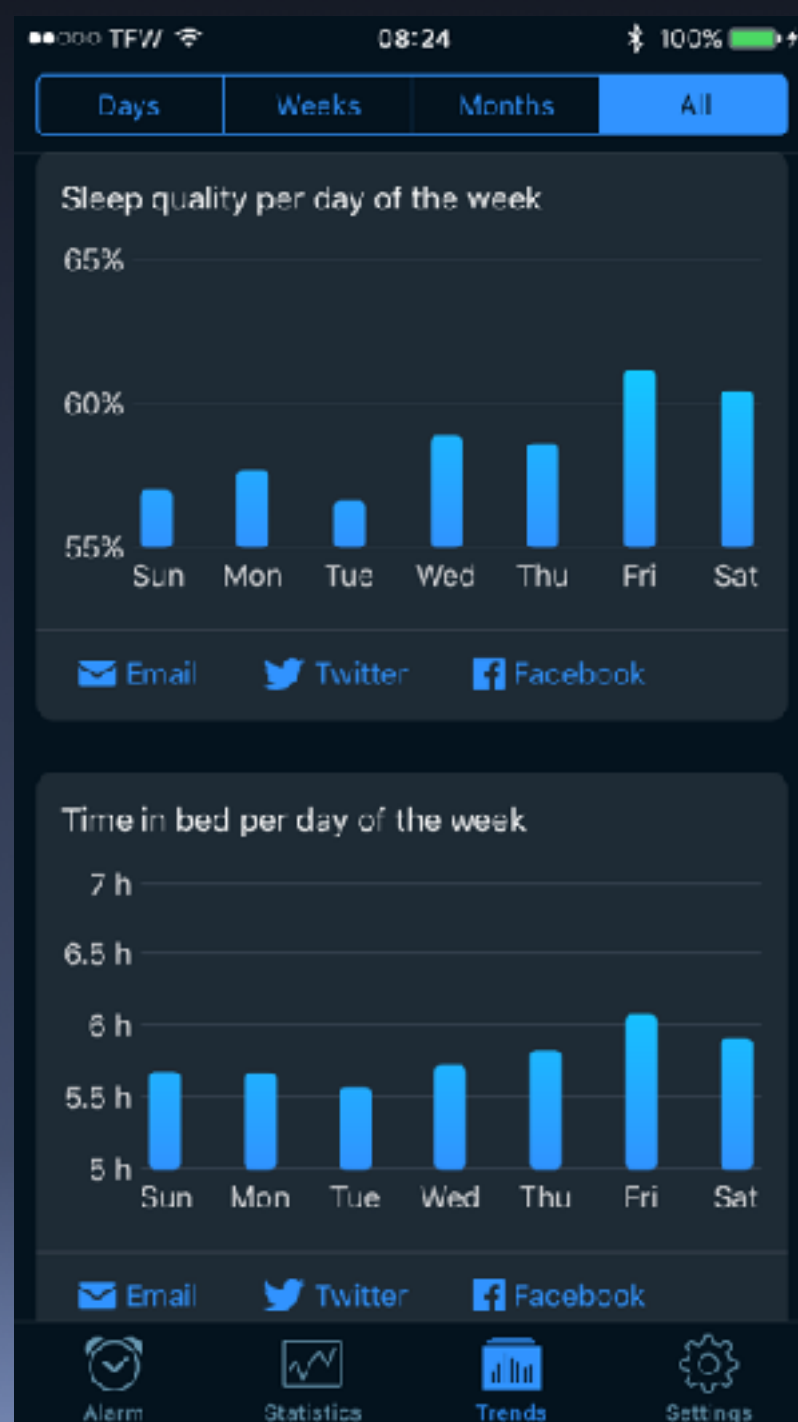
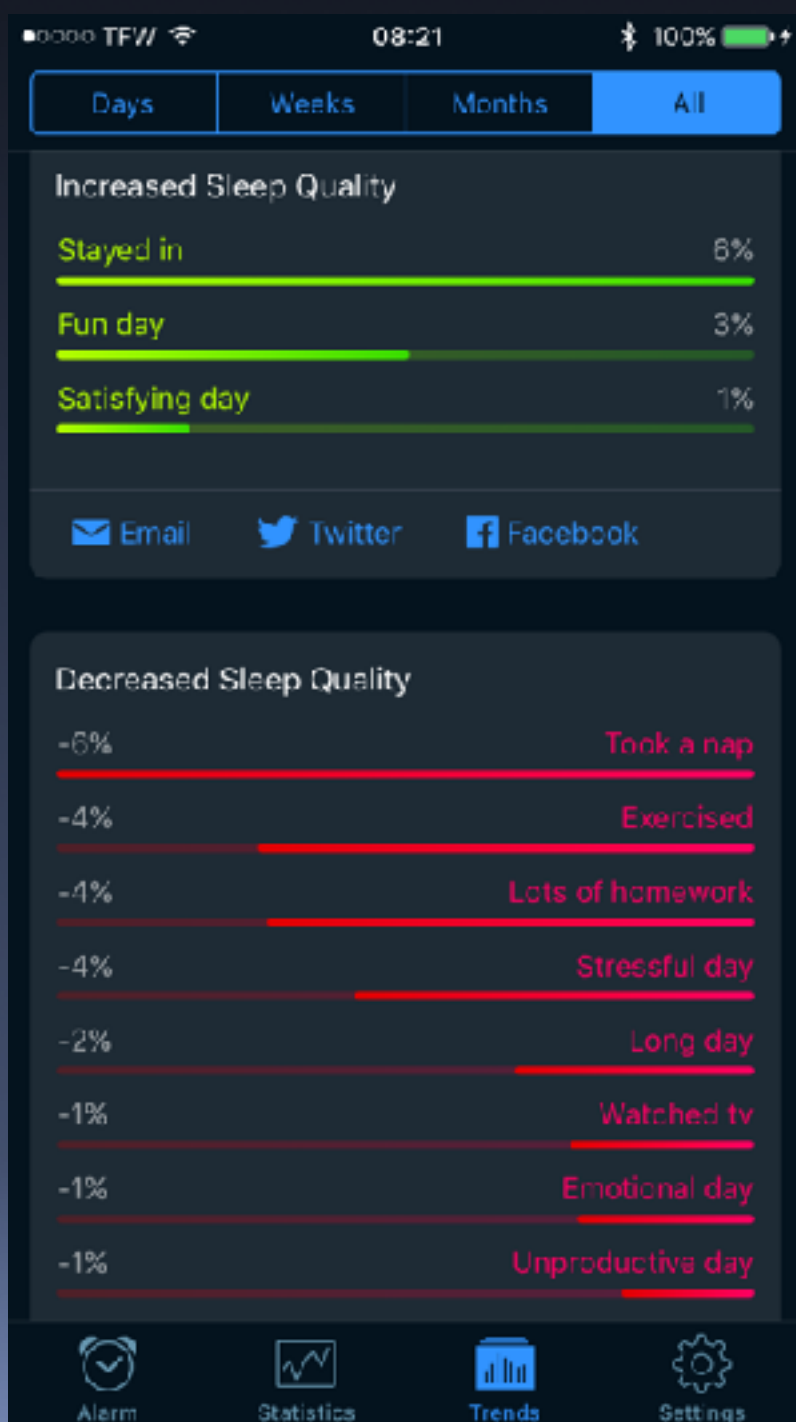
CONCLUSIONS

SLEEP CYCLE'S ACTUAL STATISTICS



CONCLUSIONS

SLEEP CYCLE'S ACTUAL STATISTICS



WORKS CITED

- Brownlee, J. (2017). 8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset - Machine Learning Mastery. Machine Learning Mastery. Retrieved 16 March 2017, from <http://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>
- Hair, J. (2014). Multivariate data analysis (1st ed.). Edinburg: Pearson.
- King, G. & Zeng, L. (2003). Logistic Regression in Rare Events Data. Journal Of Statistical Software, 8(2). <http://dx.doi.org/10.18637/jss.v008.i02>
- Sleep Cycle alarm clock. (2017). Sleep Cycle alarm clock. Retrieved 16 March 2017, from <https://www.sleepcycle.com/>
- Wenseleers, T. (2012). Getting p-values for "multinom" in R (nnet package). Stats.stackexchange.com. Retrieved 16 March 2017, from <http://stats.stackexchange.com/questions/63222/getting-p-values-for-multinom-in-r-nnet-package>