THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

COLLEGE OF INFORMATION SCIENCE AND TECHNOLOGY

A LARGE-SCALE ANALYSIS OF PHONE USAGE PATTERNS TO UNDERSTAND RHYTHMS OF HUMAN ACTIVITIES

RYAN JAEGER
SPRING 2019

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Mathematics
with honors in Information Science and Technology

Reviewed and approved* by the following:

Saeed Abdullah
Assistant Professor of Information Science and Technology
Thesis Supervisor

David Fusco
Assistant Teaching Professor of Information Science and Technology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Smartphones are becoming increasingly prevalent around the world, and time spent on these devices is rising each year. For that reason, smartphones offer particularly useful insights into the timing of human activity, and have become an alternative approach to tracking sleep behaviors in individuals. Social jet lag - the difference in sleep timing between work days and free days - is a prevalent sleep disruption and has been associated with a number of health risks. Previous studies of social jet lag been small in scale or based on self-reported sleep information, which can be unreliable. Here we show how smartphone usage patterns can be harnessed to quantify social jet lag in a global population. To do so, we used a dataset on mobile device usage events called Device Analyzer collected by researcher at the University of Cambridge. The dataset has over 27,000 devices from countries around the world, and therefore this study is the first smartphone-based study of social jet lag with global implications. We found that smartphone usage patterns can be useful sources of information on human activity as they reflect daily routines. Additionally, we found that the median social jet lag across a global population of Android users was approximately 60 minutes, and over 50% of individuals experience social jet lags greater than 1 hour, indicating social jet lag to be a widespread disruption. Our quantification of social jet lag aligns with previous studies, and suggests that the phone use patterns could be a potentially rich source of information for future circadian computing research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The completion of this thesis marks my greatest academic achievement at Penn State. I am extremely proud of work I have done, and extremely grateful to the people who helped make it possible. In particular, I'd like to thank the following people:

> My thesis supervisor, Dr. Saeed Abdullah, for his guidance and support throughout the entirety of the project. From introducing me to the field of circadian computing to suggesting final edits to my thesis, you have been an amazing collaborator and I am grateful for your dedication. You taught me how to ask meaningful and interesting research questions, and reminded me that it's okay to take a break sometimes. Thank you for being an incredible teacher, mentor, and friend.

> My honors advisor for this thesis, Dr. David Fusco, for agreeing to oversee this project. As an honors student outside the College of IST, I was wary of pursuing this research, so thank you for taking the additional time and energy to help make this project a reality. I'll remember our genuine encouragement and excitement long after graduation.

> My advisor within the Math Department, Dr. Nate Brown, for encouraging my interests, even those outside of math itself. Thank you for caring about my growth as a person, beyond my growth as a mathematician. Your ability to help me realize what really matters in life was always a breath of fresh air.

> The members of the WHI Lab in the College of IST for their support and feedback for the past two years. I'll always be grateful for your encouragement, friendship, and love for food.

> My family and friends who listened to my stress-fueled rants about this project and who provided distractions when I needed it. I couldn't have kept my cool without you.

> The researchers at the University of Cambridge who collected the Device Analyzer dataset. Without the data, this project would literally not be possible. Furthermore, I would like to thank Daniel Hintze and his team for their session extraction parser.

From start to finish, I learned so much about the research process, collaborating with others, and my own work ethic. Writing this thesis has been such a meaningful learning experience, and I will carry these lessons with me as I leave Penn State and enter the rest of my life.

Chapter 1

Introduction

Study Overview

Mobile devices have become increasingly entrenched in the daily lives of people around the world in recent years. Many people use their smartphone from the moment they wake up in the morning to the moment they go to bed. Because of the ubiquity of phones in our lives, it raises the question: can we utilize phone usage data as a proxy for human activity? In 2018, the average US adult spent over 3.5 hours on their mobile device each day, according to a report from eMarketer [1]. As more and more time is spent on mobile devices each year, these devices become increasingly useful in reflecting actual human activity. Thus, smartphones may offer a particularly valuable insight into sleep patterns in humans.

In this study, we sought to understand how mobile phone usage patterns can be harnessed to quantify sleep disruptions in a large global population. To do so, we studied data from the Device Analyzer dataset, one of the largest available datasets on mobile phone usage statistics. Literature has shown that indirect methods of measuring human sleep habits through mobile devices have strong validity. Our study makes novel contributions in terms of scale—the Device Analyzer dataset has data from many countries and time zones, making this the first mobile phone based sleep study with the ability to speak to global behaviors. For the scope of an undergraduate thesis, this study is limited to one specific sleep disruption: social jet lag. Social jet lag refers to the difference in timing between sleep on free days and sleep on work days, and has been linked to a number of health concerns.

Through this study, the following research questions guided our work: How can data on mobile phone usage help us understand human behavior in a large, diverse population? How do daily patterns and routines manifest in longitudinal smartphone usage? Specifically, how do phone usage patterns reflect the difference in sleep timing between work days and free days, known as “social jet lag”? The contributions of this thesis are a computational framework for determining human activity patterns based on smartphone usage, as well as a large-scale analysis of social jet lag across a heterogeneous population of smartphone users.

Defining Social Jet Lag

While humans suffer from a variety of circadian disruptions, this thesis will investigate one such disruption known as “*social jet lag*.” The term social jet lag was first coined by Wittmann et al. [2]; in a 2006 publication in *Chronobiology International*. They defined social jet lag “the discrepancy between social time and biological time.” Having a chronic difference between one’s preferred sleep and one’s actual sleep is akin to the jet lag experienced when traveling across time zones. As such, this discrepancy is known as *social* jet lag since social schedules, primarily timing of work and school, are the primary culprit for this difference.

Though sleep is a universal human need, the preferred timing of sleep varies from person to person. An individual’s preference for sleep timing is based on both daily circadian activity and environmental signals, or “zeitgebers” [3]. By identifying changes in sleep time across east/west longitudes across Germany, a study by Roenneberg et al. [3] showed that environmental light is one of the strongest factors in entrainment, the process by which the human body syncs to a 24 hour cycle. When studying sleep and sleep timing, we consider *chronotype*, which refers to “the individual differences in when people sleep in reference to local time” [4] Chronotypes are typically described as either early or late, and the people with such chronotypes are referred to as “larks” or “owls” respectively [5]. We often quantify chronotypes using the *mid-sleep* point (i.e., onset + duration/2).

Social timing, such as that of work or of school, can interfere with an individual’s preferred sleep time. Individuals who have late chronotypes often suffer the worst from this imbalance of timing, since they naturally go to sleep late and are forced to wake up early. When compared to their early chronotype counterparts, late chronotype individuals tend to report higher rates of low sleep quality and daytime tiredness [6]. For that reason, they accumulate sleep debt throughout the week and try to compensate for the debt on the weekend [6]. This creates a noticeable difference in the timing between an individual’s work day sleep and their free day sleep. It is this difference in timing that we describe as social jet lag. Whereas jet lag caused by travel is transient and short-lived, social jet lag is a chronic issue that affects nearly 70% of the population [7] and has major health implications for individuals.

Wittman et al. [2] in the seminal paper where they coined “social jet lag,” studied the association of chronotype with sleep quality and psychological well-being. The researchers combined information from surveys about quality of sleep and mental well-being with the Munich ChronoType Questionnaire, a survey developed in [5] that asks individuals about demographic information, sleep timing, and self-reported chronotype. With this information at their disposal, they analyzed the correlations between chronotype and sleep quality, psychological well-being, and stimulant consumption to understand how these vary by sleep timing. Among their findings are how late chronotypes experience more severe evening tiredness and report facing more psychological issues than their early counterparts. One of the most significant correlations was between chronotype and stimulant consumption, where late chronotypes are significantly more likely to smoke, consume caffeinated beverages, and consume alcohol. Beyond their analysis of chronotype, they also look at the relationship between social jet lag and smoking behavior. Here, they calculate social jet lag by looking at the difference in mid-sleep timing between work days and free days. When they examine social jet lag versus smoking behavior, there is a strong linear relationship between the hours of social jet lag experienced by an individual and the percentage smokers. In fact, the correlations between social jet lag and smoking are even stronger than correlations between chronotype and smoking. This study is meaningful, in its establishment of the working definition of social jet lag, the difference between social timing and biological timing. It also provides a first method of calculating social jet lag for an individual, the difference between mid-sleep point on work days and free days. Finally, it also develops a clear link between this particular circadian disruption and health implications, which is in this case smoking behavior.

While social jet lag was first tied to an increased risk of smoking behavior, more recent studies have also established significant links between this phenomenon and other health and well-being risks [7]. Specifically, social jet lag has been strongly associated with the increased risk of obesity [7]. Using a large epidemiological database, Roenneberg et al. [7] looked at the association between sleep timing and social jet lag with the body mass index of an individual. They found that social jet lag was effective in predicting whether or not a given individual was of normal weight or overweight. Furthermore, social jet lag is positively associated with weight increase. Overall, the large-scale study found that living “against the clock” was an important factor in the obesity epidemic. Beyond smoking and obesity, social jet lag has been tied to cardiometabolic risk [8], risk of developing depression in rural populations [9], and decreased academic performance in undergraduates [10].

When considering the findings of these studies, it is evident that discrepancies in sleep timing are associated with health issues. For that reason, gaining new insight into social jet lag globally will be a valuable undertaking and one with far-reaching health and wellness implications.

Smartphone-Based Sleep Sensing

Whereas the previously mentioned studies of sleep behavior have focused on self-reported survey information, tracking one's sleep can be obtrusive and challenging to maintain. The potential issues of self-reporting have inspired other researchers to develop alternative ways of sensing sleep in humans. One of the most promising of these avenues is the use of smartphones for sleep sensing. Since smartphones are becoming more and more ubiquitous in the daily lives of individuals, they may provide a previously untapped insight into human behaviors.

Published in 2014 in *Ubicomp*, Abdullah et al. [11] demonstrates the use of smartphone data to detect circadian misalignment in individuals. For a sample of nine undergraduate students, this study used an Android app to collect data on calls, text messages, browser search, application use, and screen use for these individuals' smartphones. From here, they generated a rule-based sleep algorithm that predicted sleep sessions based on extracting the longest non-usage session beginning between the hours of 10PM and 7AM local time. The low-cost algorithm was successful in predicting sleep within 45 minutes of ground truth for each individual, when compared to sleep journals kept by each subject as ground truth on their sleep behavior. Using this algorithm, Abdullah et al. conducted a series of analyses on sleep debt, sleep inertia, and social jet lag for individuals of various chronotypes. Specifically, they detected evidence of social jet lag in all individuals and found late chronotypes to be the worst sufferers of social jet lag. While the sample size of 9 for this study is small, it developed an inexpensive, reliable, and unobtrusive way of inferring circadian patterns in individuals via their smartphones. Our study can be seen as a natural extension of this publication, as we are looking to examine social jet lag in a large population through a similar data source.

Chen et al. [12] offers a methodology for inferring sleep sessions using data native to the smartphone. In this publication in *Pervasive Health* in 2013, Chen et al. present their “Best Effort Sleep (BES)”, a sensor-based computational model for daily automated sleep duration monitoring. In short, the app combines phone usages metrics, such as the time and length of usage and of recharge, with environmental factors, such as periods of prolonged silence, darkness, and stillness, to infer an individual’s sleep. The pilot study of BES showed promising results, with the ability to predict sleep within 42 minutes, while still remaining unobtrusive to the user. Chen et al.’s incorporation of sensor data from a phone’s camera, microphone and accelerometer, differentiate this sleep algorithm from others based solely on usage information, similar to Abdullah et al. While our study does not incorporate these environmental factors, this paper offers another example of using phone session as a means of extracting key information about sleep behavior, a crucial step in our methodology.

Wahl and Amft [13], in a paper published in *ACM IMWUT* in 2018, expand upon data driven models for predicting sleep behavior by incorporating an “expert model” containing knowledge regarding human circadian and homeostatic processes. When exploring the efficacy of sleep estimation over 98,000 simulated sleep days, they found that a model incorporating an expert model improved estimation by 47% for sleep onset and 58% for wake up. One of the most significant features of this model is the incorporation of a user feedback to help personalize the sleep estimation for each individual. The use of personalization helped to improve estimation by 59% for sleep onset and 57% for wake up. The ability to include knowledge about human biology and personalize predictions to each user offers promise in future development of sleep sensing systems.

Another recent paper [14] investigated the ability to use smartphones to track sleep, but through touchscreen interactions rather than application usage or social media. By tracking the timing of touchscreen taps for Android users, Borger et al. developed a basic method for identifying sleep by looking at the longest period of no activity during a 24-hour period. They compared this value to measures of physical activity, measured by accelerometers worn on the wrists of participants. Based on these methods, they found that the sleep onset and wake-up time estimates from the touchscreen interactions matched very closely ($R^2 = 0.9$) to the estimates from actigraphy, a significantly more established method. These findings suggest that touchscreen-based estimation works as well as actigraphy and therefore should be explored as a new approach to measuring sleep. While this study only focused on 84 participants and focusing on sleep in general rather than social jet lag, Borger et al.'s study is extremely relevant to our research as it further validates that smartphones offer an accurate way of sensing sleep in users.

Social media can serve an alternative source of measuring an individual's interactions with their mobile device, and other studies, such as Murnane et al. [15], have harnessed social media to study sleep. They measured social jet lag in 9 undergraduate students based on phone usage, activity on social media apps, and posts on Facebook. Together, they synthesized all this information of an individual’s phone use and predict their sleep onset, duration, and wake time. For 7 out of the 9 individuals studied, Murnane et al. was able to closely estimate the actual sleep duration, which was recorded in a sleep journal. When analyzing social jet lag, Murnane et al. found that late chronotypes were the greatest sufferers. They continued to research behavior regarding attention, cognitive performance, and mood following adequate or inadequate sleep, and discuss design criteria for technologies meant to track our innate biological patterns. The incorporation of social media information to support phone usage data in predicting sleep is a valuable addition and has since been expounded upon by other researchers.

A recent study by Till Roenneberg [16] looks at how an active social media account can act as a proxy for human activity timing. Specifically, Roenneberg shows how an active Twitter account, in this case @realdonaldtrump, can be used to characterize an individual’s chronotype. Using tweets from December 2014 to March 2017, Roenneberg parses the device used to make each tweet to isolate which tweets are coming from Donald Trump directly and which ones are being sent from his staff. By plotting a time series of the average activity of the tweets, they look for the troughs in activity and use that as a measurement of sleep in the individual. By doing so, Roenneberg characterized Donald Trump as being of an extreme early chronotype, with having an average sleep midpoint of 1:30AM, as compared to 3:30AM average. While this study is limited to a single individual and his active Twitter account, it is evidence that sleep-wake behavior can be characterized by one’s online activity. Furthermore, this study was designed to understand an individual’s sleep behavior without having the ground truth available. Intuiting one’s sleep behavior based on the time of minimal activity is a central idea to our study.

Large-Scale Studies of Human Activity

One of the challenges of investigating sleep behavior in humans is scale. It is difficult to create formal sleep studies for a large population, and self-reported data can be unreliable. Some studies that have used mobile phone data to understand human activity on a population scale rather than on the individual scale, as explored in the previous section. The scientific value of understanding sleep behavior on a population scale is immense, so this avenue of research is especially ambitious.

Differing from other studies by focusing on call records, the paper by Monsivas et al. [17] published in *PLOS Computational Biology* in 2017 used mobile phone calling patterns to understand human activity in urban environments. The dataset consisted of anonymized call detail records from within the same time zone in Southern Europe. By examining at the timing of calls at different times of the year, Monsivas et al. determined that the onset and end of users' calling activity is related to environmental light. Similar to [3], these researchers found evidence of an east-west shift in call timing, indicating that the sun time has an impact on human activity. Furthermore, Monsivas et al. showed that call activity follows seasonal patterns and varies based on time of year. By looking at the periods of low calling activity, they estimated the midsleep time of different age groups, hence giving a chronotype characterization for these groups. Our study employs a similar methodology of interpreting a period of low activity as a proxy for sleep of an individual, however on a global rather than national scale.

In a 2016 study, Walch et al. [18] sought to create a "global quantification of normal sleep schedules" from smartphone data. The researchers developed a mobile app titled, *ENTRAIN*, where users can submit data on normal sleep times, time zones, and typical lighting for the ostensible purpose of overcoming jet lag from travel. Using this submitted data from over

5000 global app users, Walch et al. provided a characterization of sleep behavior by demographic. They found that users who experience more outdoor light than indoor light tend to go to sleep earlier and sleep more. The team investigated societal pressures on sleep, finding that, on a country level, bedtime is significantly more predictive of sleep duration is. This conclusion makes intuitive sense, as social constructs, like timing of work and school, are more rigid and often define when an individual must wake, so changing the bed time would do more to increase overall sleep. Like other studies mentioned previously, this study validates the use of mobile technology to collect massive datasets regarding human activity at little cost. However, unlike the studies in [11] and [15], the scale of the data collection is much greater, so the scope of the conclusions is also grander.

Another recent and relevant study explores how Twitter behavior on the county level can be used to characterize the social pressures that exist on daily human activity [19]. Specifically, this paper provided an in-depth analysis of social jet lag in the United States at the county level. Leypunskiy et al. [19] aggregates tweets for over 1800 counties in the US, and examined average timing of activity. Using the troughs in activity as a means of identifying approximate sleep, they looked for shifts in trough position between work days and free days to indicate social jet lag. After quantifying "Twitter social jet lag" in the US as a whole, the researchers explored how the time of year affects social jet lag and found that it is most severe in February and least severe in June and July. They also show that the social jet lag in a county can be tied to the K-12 school calendar, with higher rates of social jet lag happening on breaks and holidays. While the conclusions about sleep behavior are limited to the county level rather than the individual level, the paper provided a valuable insight into how social media data can be harnessed to understand human activity at scale. Furthermore, the methodology and analyses conducted in this paper act as direct inspiration for how we characterize social jet lag from an individual's phone data for our study.

With this background, we have established a definition of social jet lag, confirmed the validity of using smartphones to track sleep, and contextualized other large-scale studies of sleep. Our study combines these three threads to provide a large-scale smartphone-centric quantification of social jet lag. The value of our study will be a newfound insight into social jet lag in the general population, providing the groundwork for future interventions to mitigate this specific circadian disruption.

Chapter 2

Data Description

Smartphones are used widely and extensively in modern life, so these devices offer insights into human activity that have been previously untapped. One of the largest dataset of smartphone activity — the Device Analyzer dataset — was collected by researchers at the University of Cambridge [20]. In our study, we employ data from Device Analyzer to understand timing of human activity by identifying peaks and troughs in users' smartphone usage.

The Device Analyzer dataset is the world's largest dataset on smartphone usage in terms of duration of data collection, diversity of devices included, and the granularity of detail collected [20]. To collect this data, the researchers at Cambridge developed an Android application that allows users to track their phone activity and understand their usage patterns. The application collects data continuously in the background after initial set up, and creates a minimal burden on user experience. This application captures a comprehensive time-series log of hundreds of different events. Over 100,000 data points on average are collected each day for a given device, including data on network connectivity, application usage, screen activity, call records, sensor readings, and battery levels among many others. See Table 1 below for a full breakdown of category and number of events collected. This data is stripped of personal identifiable information, salt hashed to preserve privacy, and periodically submitted to a central "Device Analyzer" repository.

Category	Event types collected
Device settings	33
Installed Applications	17
System Properties	29
Bluetooth devices	21
WiFi networks	11
Disk storage	6
Energy & Charging	5
Telephony	20
Data usage	38
CPU & memory	11
Alarms	10
Media & Contacts	8
Sensors	15

Table 1: Device Analyzer Events

Today, the Device Analyzer dataset has over 100 billion data points and over 1,900 years of collected time. For the copy of the Device Analyzer dataset that we use for this study, there are 3.3 terabytes of phone usage data for over 27,000 users from hundreds of countries around the world.

Using the Device Analyzer dataset to understand human activity has a number of benefits. First, the sheer size of the dataset allows for broader conclusions to be drawn about smartphone activity and hence broader conclusions about human sleep and wake behavior. With over 27,000 users to analyze, our findings will be highly generalizable. While other studies ([19], [3], [17]) have looked at widespread sleep behavior on a national level, having global data allows us to understand sleep without any national or cultural influences. Phone behavior can change day-to-day and be subject to outside influences. Since many of the users in the Device Analyzer dataset have collected data for extended lengths of time, we can focus on broader trends in behavior and avoid short-term novelty effects [20]. Finally, since the data collection method is so unobtrusive to the user, this dataset closely reflects actual usage behavior.

While other papers have used the Device Analyzer dataset to understand mobile device usage characteristics [21], app usage prediction [22], and Android advertising libraries [23], our study is the first to harness this rich dataset to understand human sleep activity. Because of the size, scale, and granularity of the Device Analyzer dataset, our study has the potential to be the most extensive sleep sensing study to date.

Chapter 3

Methodology

A key challenge of smartphone-based sleep studies is designing a robust methodology to transform smartphone activity to timing of individual behavior. In this chapter, we describe our methodology in depth. We first explain the process of extracting sessions from the Device Analyzer dataset. Then we describe generating plots of mobile activity over time we coin as "mobograms", as a substitute for traditional "actograms" meant to display activity in formal laboratory settings. Finally, we explain calculating social jet lag from mobograms for weekday and weekend activity.

Sessions

In our analysis, we use phone usage sessions as our means of understanding human activity. Hintze et al. [21] defines mobile phone usage sessions as "consecutive periods of time during which a user interacts directly with their mobile device." They differentiated between unlocked sessions, typical user interactions such as application use and text messaging, and locked sessions, which may include checking the time, battery or notifications, all of which do not require the device to be unlocked. For the purposes of this study, we studied only unlocked sessions, since, on average, the majority of time spent on devices is with unlocked sessions for the Device Analyzer dataset [21] . Usage sessions serve as the link between phone logs and human activity.

As described in Chapter 2, the Device Analyzer dataset stores key-value pairs and associated timestamps for hundreds of different events for a phone, ranging from application use to sensor readings. However, for the purpose of detecting usage sessions, we only need to check the time stamps of a certain subset of keys. For the sake of computational ease, we started by selecting only these keys (alarm, hf, phone, screen, shutdown, system). The keys of interest included information about alarms, lock state, telephone usage, screen power and brightness, shutdowns, and system preferences. For a complete description of keys used, see [21].

With the relevant keys, we employ a parser developed by Hintze et al. [21] to extract usage session from Device Analyzer data. Based on the keys provided, the method identifies sessions of usage by looking at when the phone exits a state of being locked with the display off and when it returns to that state. In short, if a user unlocks their phone, the time from unlock to relock would be an "unlocked usage session". A complete state machine for this method created by Hintze et al. [21] can be found in Figure 1 below that models how the parser incorporates lock status and call information into its calculation of locked and unlocked usage sessions. Again, we only are concerned with unlocked sessions for the purpose of this study.

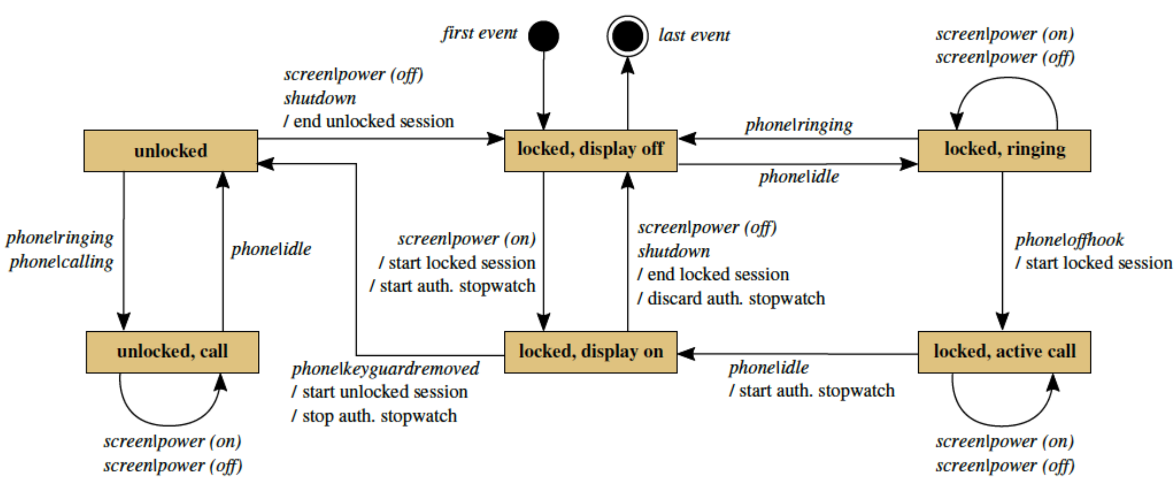

Figure 1: State Machine for Session Detection [20]

We deployed the parser to calculate sessions for each day of data for each device in our study. For a given device, we had the timing of usage sessions for each day, modeled below in Figure 2, where the horizontal axis is hour of the day, and the segments shaded in red are periods of active usage for each date shown on the left.

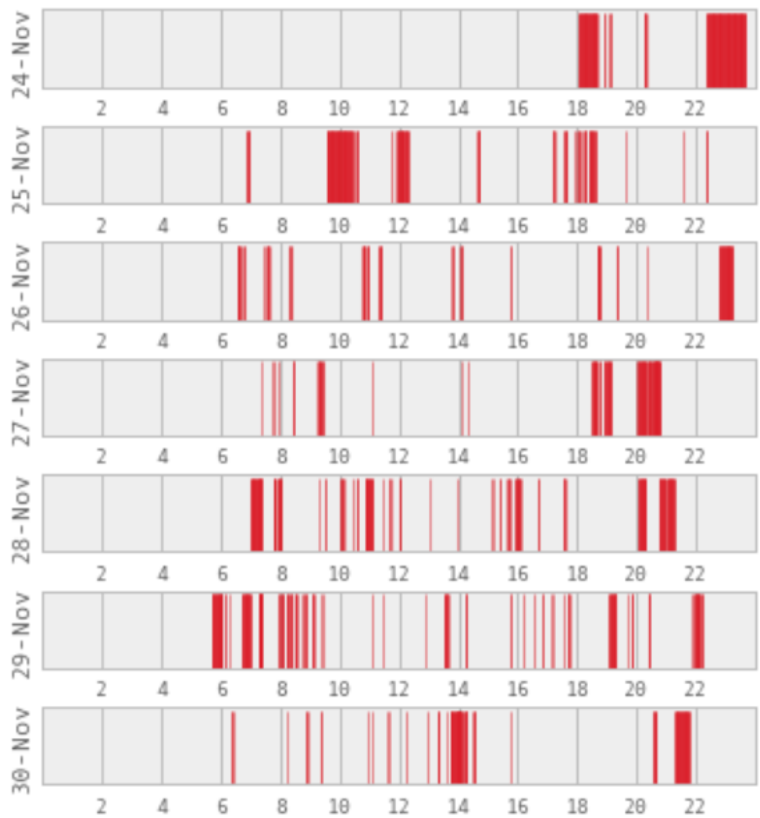

Figure 2: Sample Device Usage by Day

Mobograms

From the set of usage sessions for each day for each device, there was a need to understand general trends in sleep habits. To this end, we generated a plot of average activity over time for each individual. We refer to these as *mobograms*, which represent the average activity of a device over a 24-hour day in local time.

For each minute of the day, we aggregate the number of sessions across all days that occur during that minute and divide by the total days in the dataset for that given device, and we refer to this value as "activity." That is,

$$activity = \frac{\#of\ days\ with\ session\ during\ this\ minute}{total\ \#\ of\ days\ for\ device}$$

The result is a mobogram that shows time of the day on the horizontal axis, and the activity on the vertical axis. This mobogram indicates the activity for a device over the course of an average day.

Since social jet lag refers to a difference in sleep timing, we wanted to study the intersection of one day's activity with the next. For that reason, we found it helpful to look at 48 hours of activity rather than a single day. Using double-plotted mobograms in this way is common in circadian research. To create this, we appended the mobogram to the end of itself to get a 2-day average activity for an individual in local time. Since the mobogram represents the average day of activity, appending the mobogram to itself will represent the average 2 days of activity. See Figure 3 for a sample double-plotted mobogram.

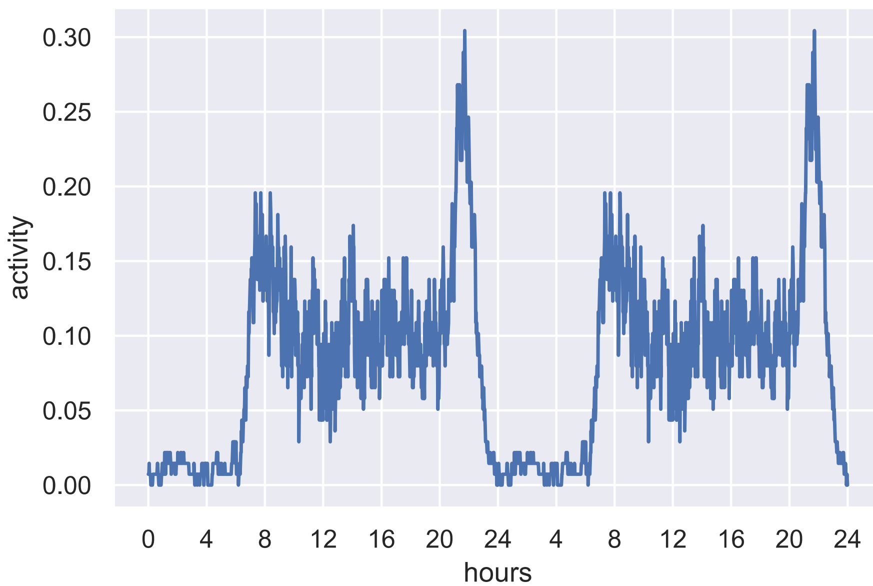

Figure 3: Sample Double-Plotted Mobogram

The above mobogram shows the average activity for a device for an average day of use. The horizontal axis shows the time in hours, and since we append the mobogram to itself to understand the interaction between days, we label the corresponding hours accordingly. The vertical axis shows the proportion of days in total that have a usage session over a particular minute, which we label as "activity".

However, since social jet lag refers to the difference in sleep timing between work days and free days, we must examine mobograms for weekdays and weekends separately as shown in Figure 4.

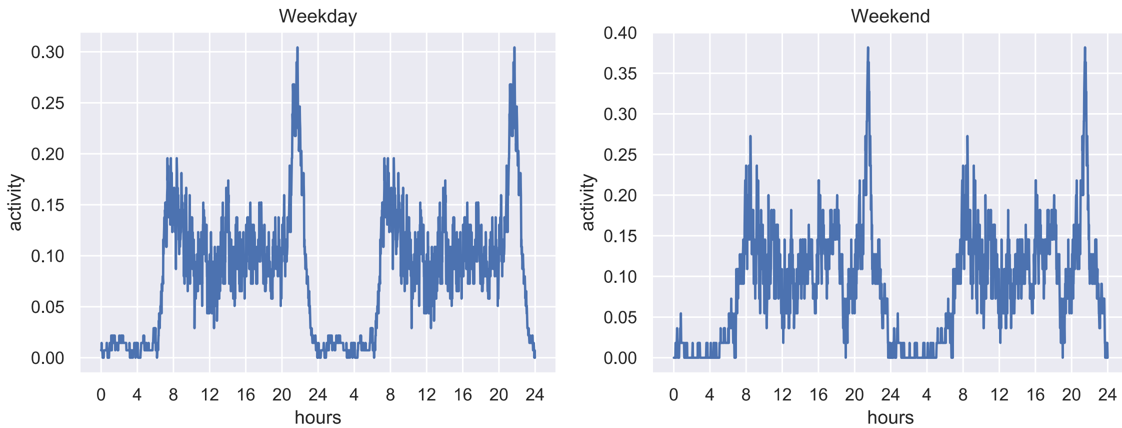

Figure 4: Sample Double-Plotted Mobograms for Weekday and Weekend

Social Jet Lag

To quantify social jet lag, we want to determine the difference of timing of activity on weekdays and weekends. Since individuals do not interact with their devices as they sleep, we expect to find a trough in activity overnight corresponding to sleep. Our methodology for calculating social jet lag is to fit a parabola to the nighttime trough in activity for both weekdays and weekends, and calculating the shift in parabola vertex position.

We estimate the coefficients a, b, and c via a least square fit to the data such that the equation for the parabola is given by:

$$y = ax^2 + bx + c$$

Then, the coordinates for the vertex of the parabola are given by:

$$\left(\frac{-b}{2a}, \frac{4ac - b^2}{4a}\right)$$

Since we expect a trough in activity during the night time window, parabolas were our choice for identifying a local minimum. The vertex of a parabola here serves as our means of measuring minimum activity in each window. Other studies [19] have used a similar methodology for fitting parabolas to identify the midpoint in activity troughs. The shift in the vertex position will correspond to the shift in sleep timing, and thus acts as our metric for social jet lag.

The first challenge in a curve-fitting approach to calculating social jet lag is to define a window for what we consider “nighttime.” Other literature [19] has used 2AM to 10AM as a window for calculating social jet lag. However, upon preliminary analysis, we found that the 8 hour time window was not sufficient to fully capture changes in nighttime activity, so instead we expanded to two windows: 12AM to 10AM (10 hours), and 10PM to 10AM (12 hours).

We examined the mobogram segment in each window and fit a parabola to the segment using functions from the Python package SciPy. Figure 5 below shows the parabola fit to a weekday mobogram and a weekend mobogram for a given device.

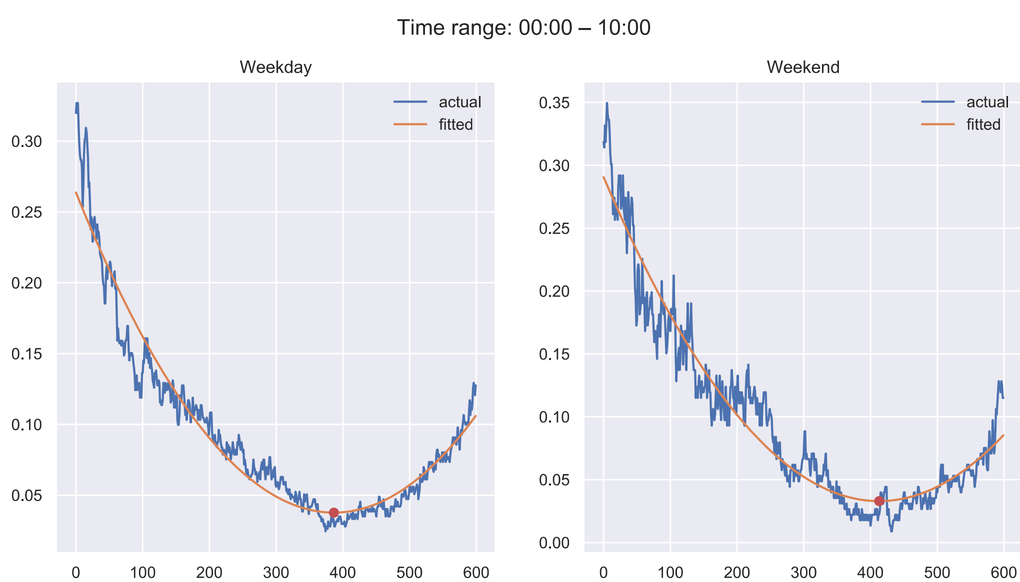

Figure 5: Fitting Parabolas to Weekday and Weekend Mobograms

From the fitted parabola, we calculated the position of the vertex, measured in minutes since the beginning of the time window. We interpret the shift in vertex position between the weekend mobogram and the weekday mobogram to be social jet lag. Specifically, we define:

$$social\ jet\ lag = |v_{weekend} - v_{weekday}|$$

where $v_{weekend}$ and $v_{weekday}$ are the vertex positions of the fitted parabolas for weekends and weekdays respectively measured in minutes.

Together, these steps serve as our computational framework for measuring social jet lag from phone usage data from Device Analyzer. The ability to convert logs of phone events to measurements about social jet lag is one of the key contributions of this study.

Chapter 4

Results

Filtering

With a methodology established for converting phone usage patterns into information about human activity and sleep, we then had to filter out devices that would be unsuitable for the analysis. To begin, the first filter was to ensure that there was both weekday and weekend days for a given device, which lowered the number of usable devices from 27,058 to 16,760. Next, our second filter was to ensure that there was enough activity in a given day. If there is very limited phone usage within a given day, it is not beneficial to understanding human activity over that day. Therefore, we selected 90 minutes to be the minimum threshold for activity use within a given day for that day to be included in all our analyses. Since the average American adult spends over 3 hours on their phone daily, 90 minutes is a reasonable and modest threshold. Requiring that all days considered have at 90 minutes of activity dropped our number of usable devices lower from 16,760 to 8,218. While this number is significantly lower than the initial 27,058 we had at our disposal, the sample size is still large enough to draw meaningful conclusions from, and the data we selected has higher quality, which removes noise from our analysis.

Average Mobogram

Before analyzing the results of social jet lag in a global population, we first sought to understand what the average daily activity looks like for a device user in our dataset. To this end, we aggregated all devices in our filtered dataset and calculated a mobogram for the average daily activity. The resulting mobogram can be seen below in Figure 6.

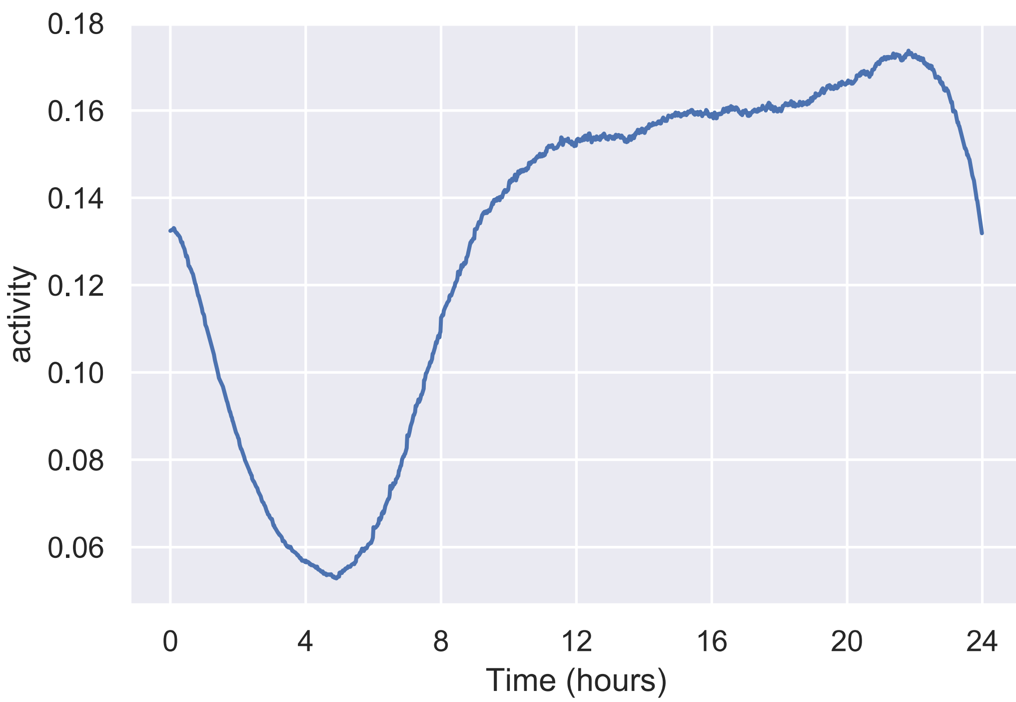

Figure 6: Average Mobogram Across All Devices

The above curve represents the trends in timing of device activity across all devices in our dataset. There is a trough in activity in the early hours of the day, approximately 1AM-8AM, and a peak in activity in the later hours, approximately 3PM – 10PM. The sharp increase in activity following the trough may be attributed to wake time. As individuals wake for the day, their device activity may increase quickly and then increase more slowly as the day progresses.

One of the central research questions of this study was: How do daily patterns and routines manifest in longitudinal smartphone usage? The above average mobogram for all devices serves to answer that question. The presence of a clear peak and clear trough indicate that there are common routines among device users and these routines are reflected in their smartphone usage. If there were no daily routines manifesting in phone usage data like this, we could expect a roughly uniform distribution. Because of the distinct shape of the average mobogram, we can confirmation that daily activity patterns of individuals are indeed reflected in their smartphone usage.

Distribution of Social Jet Lag

In order to quantify social jet lag in the global population, we instituted a set of additional filters to further ensure the quality of our data. Beyond the existing filters , we also wanted to verify that the number of days was sufficient to draw reasonable insights about differences between weekdays and weekends.

Our choice was to assert a 7-day minimum of phone usage records to be included in our analysis of social jet lag. In order to rightfully understand shifts in timing throughout a week, we require at least one week of data to be a usable device for analysis. Furthermore, preliminary analysis showed similarities in the effect of requiring 7, 30, or 90 days as the minimum threshold, so we elected to select 7 as our cut-off to include the most devices. This filter limited the number of devices for social jet lag analysis to 6,141.

Additionally, for each time window, we excluded device with irregular shaped mobograms which cause irregularities in the social jet lag calculation, and outliers in social jet lag, as defined by a standard IQR outlier detection method. For the 12AM-10AM time window, the final number of usable devices was 4,902, and for the 10PM-10AM time window, the final number of usable devices was 5,000.

With all these filters in place and outliers removed, the distribution of social jet lag, as calculated with a 12AM-10AM time window, is shown in Figure 7 below.

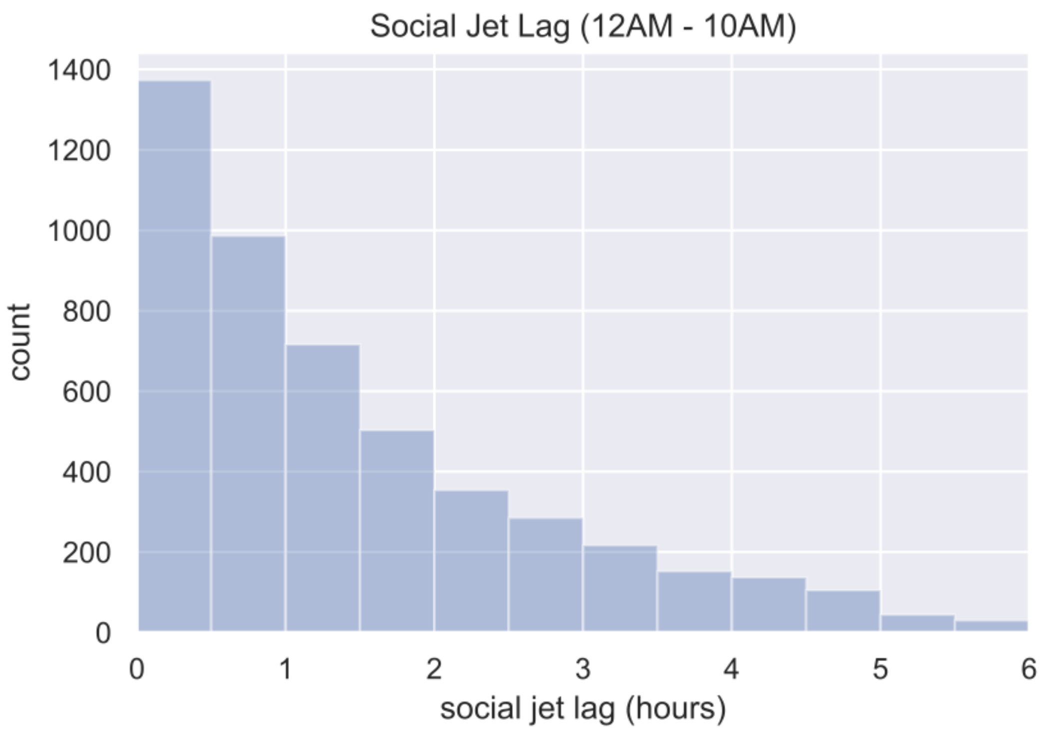

Figure 7: Social Jet Lag Distribution (12AM-10AM)

This histogram shows the distribution of social jet lag using bins of 30 minutes. The median social jet lag here is 62.3 minutes, with a standard deviation of 78.0 minutes. The distribution is right skewed, indicating that the majority of values are centered around lower social jet lag values, with fewer large values. Also, 51.8% of devices indicate a social jet lag greater than 1 hour, and 27% indicate a social jet lag greater than 2 hours.

The distribution for the social jet lag calculated with a 10PM – 10AM time window is shown in the histogram of Figure 8 below.

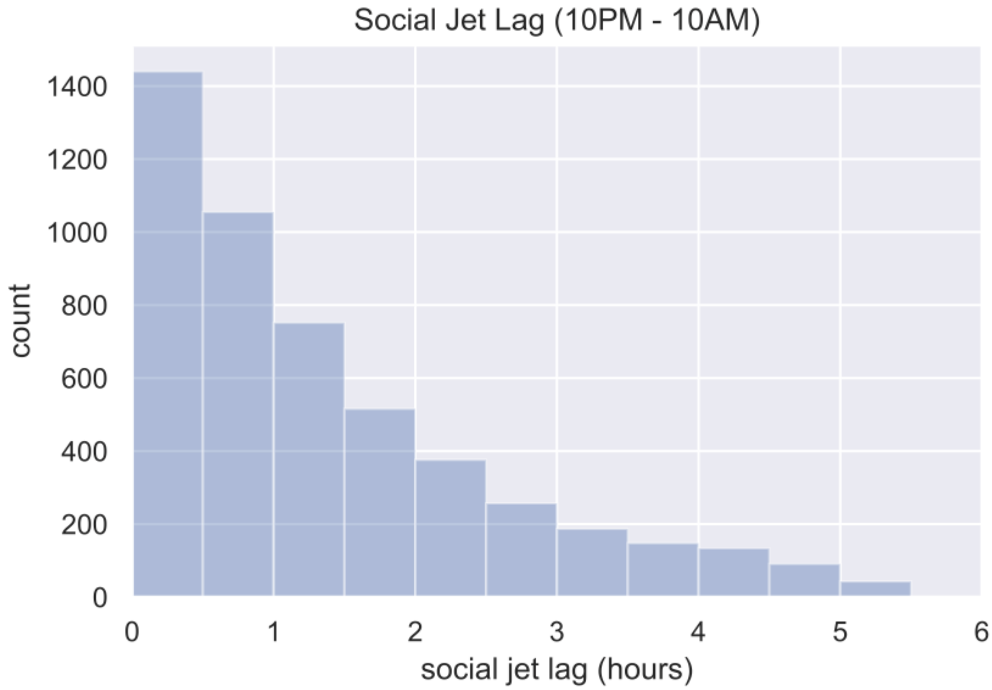

Figure 8: Social Jet Lag Distribution (10PM-10AM)

The results using the two different time windows are very similar. The social jet lag for 10PM – 10AM is similarly right skewed, with the majority of devices falling within the first two hours. The median social jet lag is 60.3 minutes, and the standard deviation is 73.6 minutes. Using this time window, our findings suggest that 50.1% of devices have social jet lag greater than 1 hour, and 24.7% have social jet lag greater than 2 hours. Since the median social jet lag value between the two time windows differs by only 3 minutes, we consider these results to be very robust. For the sake of clarity, we will refer to our median social jet lag for our analysis as 60 minutes moving forward.

Regardless of which time window we consider, the results are consistent with previous studies on quantifying social jet lag. Through a study of county-level Twitter activity, Leypunskiy et al. [19] found an average "Twitter social jet lag" among counties to be approximately 75 minutes. Our median of 60 minutes is in good agreement with that finding. However, Leypunskiy et al's [19] distribution of Twitter social jet lag is symmetric with the center around 75 minutes, whereas our social jet lag distributions are obviously right skewed. Roenneberg et al. [7] found that 69% of the working population experienced more than 1 hour of social jet lag and 33% experienced 2 hours or more. While our corresponding proportions are slightly lower, both Roenneberg et al.'s findings and ours indicate that most people experience social jet lag on the order of hours. Overall, our findings are consistent with prior research despite coming from a very different data source.

Seasonality of Social Jet Lag

While the quantification of social jet lag is the central purpose of this study, we also sought to understand how social jet lag varies by time of year. Since the human circadian rhythm entrains to sun time [3], we expect sleep timing to differ based on the timing of daylight hours, and therefore vary by time of year. Because the Device Analyzer dataset has data from Android devices around the globe, there was a need to limit our analysis of seasonality to only devices from the United States. If we included all global devices, we would have devices from both the Northern and the Southern Hemispheres, so the seasonal variations in sun exposure would be confounded.

Since most devices do not have country listed in our copy of the dataset, when limiting to only devices from the US, we were limited to 1,265 devices. For all devices in this set, we aggregated activity by month to generate average mobograms for each month of the year for both weekdays and weekends. Using these constraints, we then calculated a social jet lag value on a monthly level. The monthly values of social jet lag for each time window are shown in the line graphs below in Figure 9.

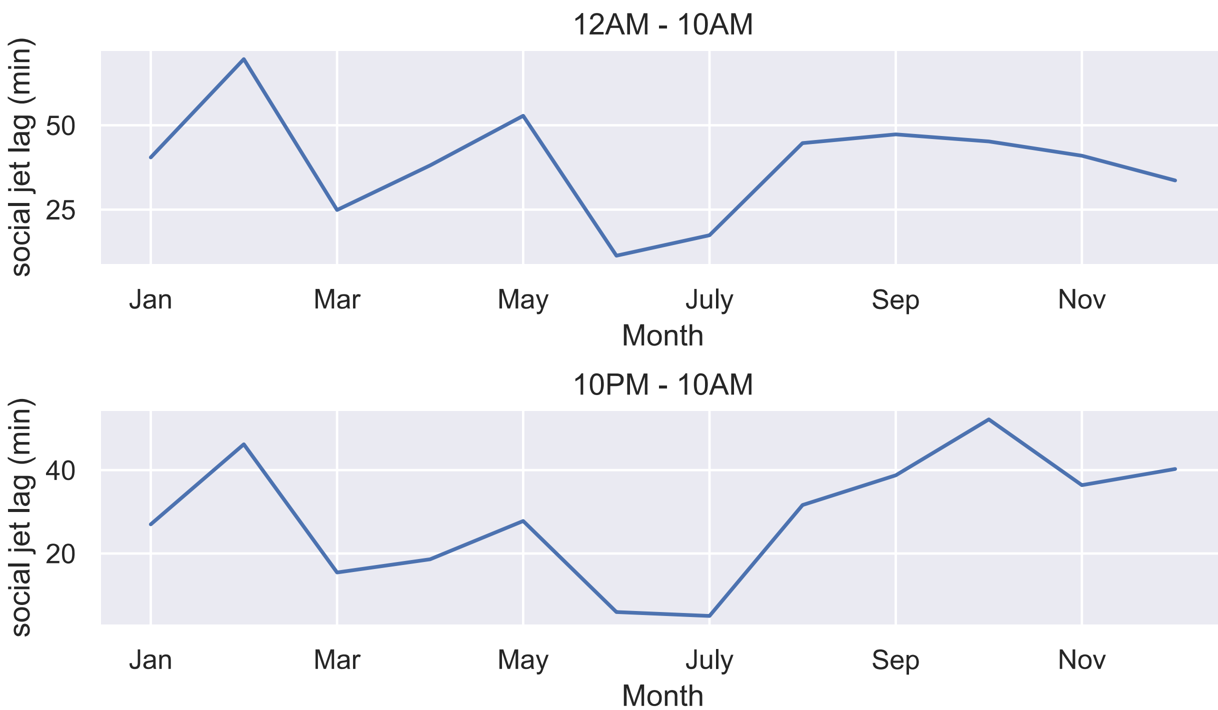

Figure 9: Seasonal Social Jet Lag in US Devices

While there are variations between the two time windows, both time windows show minimum social jet lag in June and July. Additionally, both time windows show a peak in February, with the 10PM – 10AM window also showing a peak in October. The values of social jet lag for every month are all noticeably less than the median of the above distributions. Since the monthly calculations are the aggregates of 1,739 devices, we expected less variability between the weekend and weekday behavior, and hence lower social jet lag values.

These findings are in close agreement with a similar analysis conducted by Leypunskiy et al. [19] regarding their county-level Twitter data, describing a "broad pattern of lowest social jet lag in the summer with a maximum in late winter or early autumn in general." Leypunskiy cites changes in weekday schedules, namely school holiday schedules, to explain the low social jet lag in the summer months rather than seasonal changes in daylight time. We cannot draw the same conclusions about our findings with the data at hand, but further research could be conducted to formally explore the relationship between seasonal changes in social jet lag and seasonal day length.

Chapter 5

Discussion

To gain a broader understanding of sleep disruptions in daily life, this study analyzed phone usage logs to act as an intermediary for individual behavior. More traditional studies of quantifying sleep patterns rely on formal sleep studies or self-reported data, such as the Munich ChronoType Questionnaire, for example. However, there are a number of challenges with these methods; self-report data can be unreliable, sleep studies can be intrusive, and neither of these methods are easily scalable. For that reason, obtaining a large sample size for population-wide sleep disruptions is challenging. Our review of literature established the validity of using smartphones to understand patterns of human activity. In this study, we constructed a quantification of social jet lag through usage records of smartphones from across the globe.

To do so, we employed data on phone usage from the Device Analyzer dataset; this dataset is particularly well-suited for our goals because of its granularity, its large scale, and its geographical diversity. The unique qualities allowed us to draw more conclusions about human activity in a previously unattainable way. From here, we developed a methodology for understanding device usage from the keys collected in Device Analyzer and modeling the average activity over weekdays and weekends. The shift in position between the weekday curve and the weekend curve indicates social jet lag. The complete end-to-end framework for extracting social jet lag from Device Analyzer records proves the efficacy of smartphones to inform scientific understanding of human behavior, therefore answering the first research question.

To answer the second research question, there was a need to understand how daily patterns and routines appear in this kind of data. Upon examination of the average mobogram for all devices, we gained an insight into the average activity of humans across this set of devices. This mobogram has minimum activity in the early hours of the day and peaks in the later hours. We interpret the minimum activity to correspond to average time spent sleeping, and rises throughout the day before dropping off again. The curve's definitive, non-constant shape means that phone activity is not distributed randomly and follows the kind of daily patterns we expect. The shape of the average mobogram aligns with our expectation of daily routines of individuals using these devices, indicating that our proposition of extracting data on human activity from smartphone data is valid.

The final research question focused on how phone usage patterns reflect social jet lag using this computational framework. After extensive filtering to ensure data quality in terms of sufficient activity per day, sufficient number of days, and removal of outliers, we developed histograms for social jet lag. By looking at two different time ranges as our windows for nighttime behavior, we found very similar distributions of social jet lag. Across both time windows, we found the median social jet lag experienced to be 60 minutes with a standard deviation of approximately 75 minutes. The distribution of social jet lag was right skewed, meaning most devices had associated values close to 0, with fewer devices having large values of social jet lag. Based on our findings, 50% of the population experiences a social jet lag of 1 hour or more, and 25% experience more than 2 hours. When we examined the seasonal variations in social jet lag, we found it to be lowest in the summer months, namely June and July, and highest in February and October. While we expected social jet lag to be tied to seasonal variations in sun exposure, our findings only partially support this. The lowest social jet lag occurs in the summer months when the days are longest, however, the highest social jet lag does not happen in December when days are shortest. Overall, these results indicate that social jet lag is a widespread issue facing a global population. Because of the negative health effects associated with higher levels of social jet lag, our findings suggest the need for further research and design surrounding how to minimize this sleep disruption.

Beyond confirming that social jet lag is prevalent across individuals around the globe, our findings also align with previous work on the subject; Roenneberg et al. [7] quantified social jet lag using the Munich ChronoType Questionnaire in 2012 and similarly found it to be a pervasive issue. In their study, Roenneberg et al. found that 69% of individual experienced at least an hour of social jet lag, and 33% experienced more than 2 hours [7]. Our proportions were slightly lower, but suggest the same shape and the same conclusion that social jet lag is a common sleep disruption.

When compared with other methods that use smartphone sensing to understand social jet leg, our results align well. Abdullah et al. [11] identified sleep sessions by longest period of minimal activity and found social jet lag values between 20 and 100 minutes for 9 participants. Employing a similar methodology but incorporating social media data for 9 participants, Murnane et al. [15] has social jet lag values between 40 to 80 minutes. Both of these studies have ranges that align well with our findings, with our center of 60 minutes. The most comparable study to ours in terms of method and analysis is from Leypunskiy et al. [19] which focuses on social jet lag at the county level based on troughs in aggregated Twitter activity. Their "Twitter social jet lag" has a symmetric, bell-shaped distribution with a center of 75 minutes. Our distribution has a notably different shape, but the center values between the two studies only differ by 15 minutes. Leypunskiy et al. also examined seasonal variations in social jet lag and came to very similar conclusions, finding the lowest values in June and July and highest in both February and October [19]. While we expected our seasonal variations in social jet lag to be explained by changes in daylight time, Leypunskiy et al. found the variations were closer tied to changes in sleep timing on work days as the K-12 school calendar changes [19]. Further research could be done to verify if the same relationship to K-12 school timing exists in our social jet lag findings.

While our results are in general agreement with previous studies, it is important to consider the unique qualities of our dataset that explain the subtle differences. First, the Device Analyzer dataset is well-suited for our purpose of looking at sleep disruptions on a large scale because of the many devices and the variety of geographic locations. There are over 27,000 devices in the current version of the Device Analyzer dataset and the size of the dataset will only grow in coming years. Within the dataset, there are devices from dozens of countries and nearly every time zone. Other large-scale studies focus on changes across a country [17] or may have to rely on self-reported data to understand global sleep behavior [18]. Ours is the first study of its type to have the sample size and diversity of users to be able to draw meaningful conclusions about global sleep behavior.

Few other studies of social jet lag have employed a dataset with such specific information about phone usage. While Abdullah et al. [11] and Murnane et al. [15] employed usage session detection methods based around phone activity like we do, the session extraction method in our study is more specific and robust. When compared with social-media-based method, our method encompasses more of all smartphone activity. Since smartphone users today have such intimate relationships with their devices, employing device usage to understand individual behavior is particularly useful. While individuals may not be reliable at tracking their own sleep habits, smartphone data collection can be very reliable, and extracted information about activity can be even more accurate. Perhaps one of the greatest strengths of this dataset is its granularity, and therefore its ability to speak to individual behaviors. Other studies like Leypunskiy et al.'s [19] can only speak to trends in social jet lag at the county level. While this study has extensive analysis of social jet lag in terms of distribution in seasonality, based on their dataset, they are not able to speak to individual behavior as our study does. Our study is uniquely situated to speak to social jet lag experienced by individuals in a global population, which can be of great value to chronobiologists.

While the dataset and methodology we use are well-suited to understand sleep disruptions globally, we must also recognize the limitations of this particular study. To begin, we lack any ground-truth for true timing of human activity. Since the users are anonymized, we have no way of understanding their actual daily behaviors and routines. Second, our analysis is based around the assumption that a device is used by a single individual. Social jet lag is centered around an individual's sleep timing, therefore it does not make sense for multiple users. While we presume this is the case for most devices, we have no information about user identities, so we have to work under this assumption. Also, we only look at unlocked sessions of device usage since they constitute the majority of interactions with a device. However, incorporating locked sessions may change our understanding of device usage behaviors. For example, using our session parser, an active call that began in a locked state is considered a locked session, and therefore not accounted for in our analysis. Studying both unlocked and locked usage sessions may be an area of exploration for future studies. Finally, the Device Analyzer dataset only includes Android devices. While we predict smartphone usage to be similar across all platforms, we only have

Android devices at our disposal, which may limit our about ability to generalize to all kinds of mobile devices.

Chapter 6

Conclusion

In this study, we sought to explore how smartphone usage patterns can be harnessed to reflect human sleep patterns. Specifically, we asked how longitudinal smartphone usage can serve as a proxy for human activity and how daily patterns and routines are reflected by that usage. Furthermore, we looked to quantify social jet lag, the difference in sleep timing between work days and free days in a global population based on this type of smartphone data. Whereas other studies have focused on social media usage or smartphone usage in a small number of participants, our study is the first to employ smartphone usage for large-scale analysis of sleep disruptions. Using the Device Analyzer dataset of Android smartphone usage, we deployed a parser to extract sessions of phone usage. Then we examined the average phone usage for weekdays and weekends and computed the difference in timing between the two; this difference is indicative of social jet lag.

Based on our method, we have clear evidence that smartphone usage is indeed indicative of daily patterns and routines, meaning that extracting timing of smartphone usage can act as a reasonable metric of timing of an individual's activity. We also found that social jet lag is a prevalent problem in the population, with a median social jet lag of 60 minutes. From this, we found that 25% of the population experience more than 2 hours of social jet lag. Upon a seasonal analysis of social jet lag, we found social jet lag highest in late winter and early fall, and lowest in the summer months. While these findings are not explained by seasonal variations in daylight time, they are consistent with previous work.

The contributions of this study are: the development of a computational framework for extracting sleep disruptions from smartphone usage patterns; and an analysis of social jet lag across the global population of Android devices. The findings of this study confirm that social jet lag is a prevalent sleep disruption that significantly impacts a large proportion of the population. Because social jet lag is associated with multiple health issues, understanding this sleep disruption further has important implications. With an increasingly clear understanding of social jet lag in the population, it creates future design opportunities to minimize it and develop interventions to maintain consistent sleep schedules. Finally, our work contributes to the growing body of research focusing on how smartphones can be harnessed to aid large-scale understanding of human health.

Future Work

By the nature of an undergraduate thesis, this study is limited in its scope and the extent of its analysis. However, we foresee a number of valuable avenues for future work of circadian research using the Device Analyzer dataset.

Our analysis of social jet lag in this thesis is valuable, though not exhaustive. We see a number of opportunities to extend and improve upon our work on social jet lag. First, for the purpose of ensuring the integrity of the data, we created a number of filters to require a minimum number of days and minimum activity per day. For that reason, we removed many devices and days of usage from our analysis. Incorporating this data may yield a clearer picture of social jet lag. Other papers such as [15] and [11] have studied chronotype in relation to social jet lag, and found that people with later chronotypes suffer more from social jet lag than their early chronotype counterparts. Another opportunity to extend this study would be to quantify chronotype from this data and analyzing its effect on social jet lag. Since the social jet lag was not found to be tied to sun time, further research could look for explanatory factors for the seasonal variations in social jet lag. Finally, the Device Analyzer dataset has extensive information about time zones, so studying the effect of time zone on social jet lag, or differences within a given time zone would be valuable analyses.

Beyond exclusively studying social jet lag, the ability to measure human activity based on smartphone usage opens opportunities to explore many topics in circadian research. Other studies have shown that human sleep behavior entrains to sun time [3]. Wake times and sleep times could be inferred from this data, then referenced against seasonal variations in sun time, and potentially confirm this finding on a global scale. Another avenue of research could explore the impact of daylight savings time on sleep. The shift in timing associated with daylight savings time can impact the alignment of social time and biological time within individuals. Examining the impact of daylight savings time on the timing of smartphone activity could offer a window into how sleep is disrupted in the spring and the autumn each year. Overall, the results of our study indicate that the longitudinal smartphone usage data from the Device Analyzer dataset offer valuable insight into human activity at a population level, and can be an integral data source for future research in circadian computing.

BIBLIOGRAPHY

[1] "Mobile Time Spent 2018 - eMarketer Trends, Forecasts & Statistics." [Online]. Available: https://www.emarketer.com/content/mobile-time-spent-2018. [Accessed: 10-Mar-2019].

[2] M. Wittmann, J. Dinich, M. Merrow, and T. Roenneberg, "Social Jetlag: Misalignment of Biological and Social Time," *Chronobiology International*, vol. 23, no. 1–2, pp. 497–509, Jan. 2006.

[3] T. Roenneberg, C. J. Kumar, and M. Merrow, "The human circadian clock entrains to sun time," *Current Biology*, vol. 17, no. 2, pp. R44–R45, Jan. 2007.

[4] T. Roenneberg, "What is chronotype?: Preface," *Sleep and Biological Rhythms*, vol. 10, no. 2, pp. 75–76, Apr. 2012.

[5] T. Roenneberg, A. Wirz-Justice, and M. Merrow, "Life between Clocks: Daily Temporal Patterns of Human Chronotypes," *Journal of Biological Rhythms*, vol. 18, no. 1, pp. 80–90, Feb. 2003.

[6] F. Giannotti, F. Cortesi, T. Sebastiani, and S. Ottaviano, "Circadian preference, sleep and daytime behaviour in adolescence," *Journal of Sleep Research*, vol. 11, no. 3, pp. 191–199, Sep. 2002.

[7] T. Roenneberg, K. V. Allebrandt, M. Merrow, and C. Vetter, "Social Jetlag and Obesity," *Current Biology*, vol. 22, no. 10, pp. 939–943, May 2012.

[8] P. M. Wong, B. P. Hasler, T. W. Kamarck, M. F. Muldoon, and S. B. Manuck, "Social Jetlag, Chronotype, and Cardiometabolic Risk," *The Journal of Clinical Endocrinology & Metabolism*, vol. 100, no. 12, pp. 4612–4620, Dec. 2015.

[9] R. Levandovski *et al.*, "Depression Scores Associate With Chronotype and Social Jetlag in a Rural Population," *Chronobiology International*, vol. 28, no. 9, pp. 771–778, Nov. 2011.

[10] R. Á. Haraszti, K. Ella, N. Gyöngyösi, T. Roenneberg, and K. Káldi, "Social jetlag negatively correlates with academic performance in undergraduates," *Chronobiology International*, vol. 31, no. 5, pp. 603–612, Jun. 2014.

[11] S. Abdullah, M. Matthews, E. L. Murnane, G. Gay, and T. Choudhury, "Towards circadian computing: 'early to bed and early to rise' makes some of us unhealthy and sleep deprived," 2014, pp. 673–684.

[12] Z. Chen *et al.*, "Unobtrusive Sleep Monitoring using Smartphones," in *Proceedings of the ICTs for improving Patients Rehabilitation Research Techniques*, Venice, Italy, 2013.

[13] F. Wahl and O. Amft, "Data and Expert Models for Sleep Timing and Chronotype Estimation from Smartphone Context Data and Simulations," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–28, Sep. 2018.

[14] J. N. Borger, R. Huber, and A. Ghosh, "Capturing sleep-wake cycles by using day-to-day smartphone touchscreen interactions," *bioRxiv*, Nov. 2018.

[15] E. L. Murnane, S. Abdullah, M. Matthews, T. Choudhury, and G. Gay, "Social (media) jet lag: how usage of social technology can modulate and reflect circadian rhythms," in

Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15, Osaka, Japan, 2015, pp. 843–854.

[16]T. Roenneberg, "Twitter as a means to study temporal behaviour," *Current Biology*, vol. 27, no. 17, pp. R830–R832, Sep. 2017.

[17]D. Monsivais, A. Ghosh, K. Bhattacharya, R. I. M. Dunbar, and K. Kaski, "Tracking urban human activity from mobile phone calling patterns," *PLOS Computational Biology*, vol. 13, no. 11, p. e1005824, Nov. 2017.

[18]O. J. Walch, A. Cochran, and D. B. Forger, "A global quantification of 'normal' sleep schedules using smartphone data," *Science Advances*, vol. 2, no. 5, pp. e1501705–e1501705, May 2016.

[19]E. Leypunskiy *et al.*, "Geographically Resolved Rhythms in Twitter Use Reveal Social Pressures on Daily Activity Patterns," *Current Biology*, vol. 28, no. 23, pp. 3763-3775.e5, Dec. 2018.

[20]D. T. Wagner, A. Rice, and A. R. Beresford, "Device Analyzer: Understanding Smartphone Usage," in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, vol. 131, I. Stojmenovic, Z. Cheng, and S. Guo, Eds. Cham: Springer International Publishing, 2014, pp. 195–208.

[21]D. Hintze, P. Hintze, R. D. Findling, and R. Mayrhofer, "A Large-Scale, Long-Term Analysis of Mobile Device Usage Characteristics," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–21, Jun. 2017.

[22]H. Cao and M. Lin, "Mining smartphone data for app usage prediction and recommendations: A survey," *Pervasive and Mobile Computing*, vol. 37, pp. 1–22, Jun. 2017.

[23]S. Demetriou, W. Merrill, W. Yang, A. Zhang, and C. A. Gunter, "Free for All! Assessing User Data Exposure to Advertising Libraries on Android," in *Proceedings 2016 Network and Distributed System Security Symposium*, San Diego, CA, 2016.

ACADEMIC VITA

RYAN JAEGER

EDUCATION

THE PENNSYLVANIA STATE UNIVERSITY, SCHREYER HONORS COLLEGE University Park, PA

BACHELOR OF SCIENCE IN MATHEMATICS, MAY 2019

- **Honors in Information Science and Technology, Minor in Statistics**
- **Thesis:** A Large Scale Analysis of Phone Usage Patterns to Understand Rhythms of Human Activities
- **Coursework:** Linear Algebra, Stochastic Modeling, Time Series Analysis, IT Project Management, Applied Data Sciences, Advanced Data Management, Critical Thinking for Leadership

EXPERIENCE

GRENZEBACH GLIER & ASSOCIATES Chicago, IL

ANALYTICS/CONSULTING INTERN *June 2018 - Present*

- Enhanced wealth screening software by analyzing over 2 million records to validate donor rating system
- Automated and streamlined reporting process for donor insight surveys, reducing cost for clients
- Presented recommendations on fundraising objectives for past client to senior consultants as case study

WELLBEING & HEALTH INNOVATIONS LAB University Park, PA

RESEARCH ASSISTANT *January 2018 - Present*

- Quantified population-wide sleep habits through analysis of longitudinal phone use data
- Created computational framework to import, clean, process and analyze 18+ terabytes of phone activity data
- Contextualized, synthesized, and communicated findings in lab meetings towards completion of Honors Thesis

MERCK & CO., INC. West Point, PA

MANUFACTURING ANALYTICS INTERN *May 2017 - July 2017*

- Restored 500+ employee laptops as part of global recovery effort following massive cyber attack
- Developed series of business intelligence dashboards to promote efficiency on vaccine production line
- Coordinated and facilitated training sessions on utilizing data visualization software for manufacturing

NITTANY DATA LABS University Park, PA

PROJECT MANAGER *February 2016 - May 2017*

- Led team of 4 analysts to conduct exploratory data analysis for local EdTech startup as pro-bono project
- Completed 10-week training course in principles of machine learning and data visualization
- Collaborated and negotiated with Penn State alumni and corporate partners to acquire speakers for seminars

LEADERSHIP

PRESIDENTIAL LEADERSHIP ACADEMY University Park, PA

MEMBER, CLASS OF 2019 *April 2016 - Present*

- Coauthored 58-page policy proposal outlining technology-based solutions to college mental health crisis
- Selected as one of 30 members from highly competitive pool of 200+ applicants from Penn State Class of 2019
- Fostered critical thinking through three-year leadership experience of seminar classes, field trips, & speakers

PENN STATE MARCHING BLUE BAND University Park, PA

CLARINET GUIDE/HEAD LIBRARIAN *August 2015 - Present*

- Led 26-member clarinet section to achieve musical and marching excellence in all performances
- Directed team of 3 librarians to coordinate the organization and distribution of music for 280+ instrumentalists
- Rehearsed 20+ hours per week to learn challenging new marching routines for each home football game

TECHNICAL SKILLS

- Python, SQL, Tableau, R, NoSQL, MATLAB, C++, SAS, TIBCO Spotfire

AWARDS AND HONORS

- Phi Beta Kappa Honor Society, *Lambda of Pennsylvania Chapter* *May 2018*
- Evan Pugh Senior Award, *The Pennsylvania State University* *March 2018*
- Dean's List, *Eberly College of Science* *December 2015 - May 2018*
- Eagle Scout Award, *Boy Scouts of America* *June 2015*