Eurographics Conference on Visualization (EuroVis) 2020
M. Gleicher, T. Landesberger von Antburg, and I. Viola
(Guest Editors)

Volume 39 (2020), Number 3

ARGUS: An Interactive Visual Analytics Framework For the Discovery of Disruptions in Bio-Behavioral Rhythms

Abstract

Human Bio-Behavioral Rhythms (HBRs) such as sleep-wake cycles (Circadian Rhythms), and the degree of regularity of sleep and physical activity have important health ramifications. Ubiquitous devices such as smartphones can sense HBRs by continuously analyzing data gathered by built-in sensors, providing important clues about the degree of regularity and disruptions in behavioral patterns. Human behavior is complex and the volume and multi-channel nature of smartphone data makes it challenging to make meaningful observations, detect unhealthy HBR deviations and most importantly pin-point the causes of disruptions. Prior work has largely utilized computational methods such as machine and deep learning approaches, which while accurate, are often not explainable and present few actionable insights on HBR patterns or causes. To assist analysts in the discovery and understanding of HBR patterns, disruptions and causes, we propose ARGUS, an interactive visual analytics framework. As a foundation of ARGUS, we design an intuitive Rhythm Deviation score that analyzes users' smartphone sensor data, extracts underlying 24 hour rhythms and quantifies their degree of irregularity. This score is then visualized using a glyph that makes it easy to recognize deviations and disruptions in the regularity of HBRs. ARGUS also facilitates deeper HBR insights and understanding of causes by linking multiple visualization panes that are overlaid with objective sensor information such as geo-locations and phone state (screen locked, charging), and user-provided or smartphone-inferred ground truth information. This comprehensive array of visualization overlays in ARGUS enables analysts to gain a more complete picture of HBRs, behavioral patterns and deviations from regularity. The design of ARGUS was guided by a goal and task analysis study involving experts versed in HBR and smartphone sensing. To demonstrate its generalizability, two different datasets were explored using ARGUS in conjunction with expert feedback and input.

CCS Concepts
• ***Visualization*** → *Visualization systems and tools;* • ***Visualization application domains*** → *Visual analytics;*

1. Introduction

Humans are creatures of habit, and disruptions in those habits can have significant medical ramifications such as mental health, obesity, and heart disease [Vet18]. Humans Bio-Behavioral Rhythms (HBRs) such as *Circadian Rhythms* [Vet18] (sleep-wake cycles), and the regularity of sleep and physical activity are particularly important measures of health. Continuous monitoring of HBRs to detect deviations from normalcy can inform timely interventions and improve overall health. However, capturing and mining human behavior data to discover important patterns is challenging, especially in the real world. Smartphones are uniquely suited for capturing such data because they are ubiquitous and are equipped with multiple sensors such as accelerometers, gyroscopes, ambient light and GPS that can capture data to provide important clues about a person's behavioral rhythms. For instance, it is a common pattern for people to not interact with their phone overnight while they sleep. This lack of interaction can be captured by the smartphone and provide a useful measure of how much or little sleep people get [AMM*14]. The ability to detect, monitor, and mine such patterns can make the smartphone a *proxy* for human behaviors. Smartphone sensing of human behavior has been used to reliably detect important behavioral changes [WHW*18] and monitor subjects' health and wellness, including mental health [RACB11] and academic performance and stress situations [WCC*14].

While smartphones can gather very rich human behavior data quickly, the difficulty of discovering meaningful patterns and making sense of such data becomes more difficult as the number of participants, their duration of participation and the number of sensor streams analyzed increase. Prior work has largely utilized computational methods such as machine and deep learning approaches [VELW18, VWL18], which often are not explainable and present few actionable insights on HBR patterns or causes. Interactive Visual Analysis (IVA) is a powerful approach for analysts to make sense of large multivariate data. In this work, we focus on leveraging IVA to enable human behavior analysts to discover and monitor disruptions in HBRs using autonomously-collected smartphone behavioral data, with the aim to generate actionable insights. We are interested in not only providing a way to identify HBR disruptions but equally importantly to *explain* those disruptions. We introduce ARGUS, a visual analytics framework to represent multi-stream, heterogeneous data using intuitive visual metaphors that enable analysts to make sense of the data with ease.

Existing visual analytics frameworks for circadian rhythm are effective at detecting deviations in rhythm, but are not interactive and lack the ability to generate explainable insights [GRBG19,FVR16]. Existing interactive visual analytics for human behavior disruptions generally focus on specific features of human behavior [PXQ*11, SMB*17], such as mobility, but are not able to integrate the vast, multi-channel nature of smartphone-collected human behavior. In contrast, our approach leverages *interactive analysis* to provide linked visualization panes that leverage multiple channels to enable behavioral experts to discover deviations in HBRs as inferred from smartphone sensor data *and* explain them. We develop an intuitive Rhythm Deviation score, which then we visualize using a custom glyph which is based on an established visual metaphor called the z-glyph [CLGD18]. Smartphone gathered data is also visualized in several linked panes that highlight how often sensor streams and certain extracted features *co-occur*. For instance, a locked smartphone screen combined with sensed low light and quiet conditions could indicate nights during which the user slept well.

Overall, our contributions are:

1. A novel intuitive HBR "*Rhythmicity*" score computed from autonomously gathered smartphone data using the Lomb-Scargle periodogram. It captures HBR disruptions that can be visualized for effective, fast, and reliable analysis.
2. The *ARGUS* IVA platform that visualizes this rhythmicity score using a glyph metaphor, while also linking other behavioral panes that contextualize the rhythmicity score. Together, ARGUS captures a complete, explainable picture of users' HBRs. ARGUS aims to not only visualize deviations in behavioral rhythms but also provide opportunity for analysts to uncover easily potential *explanations* of those deviations.
3. A comprehensive evaluation of *ARGUS* using both task and goal analysis expert feedback is conducted, which also determines feasible use cases for HBR disruption analysis.
4. An insightful walk-through along well-designed use cases by specialists in behavior studies involving two real world datasets illustrates our approach and demonstrates its effectiveness for discovering, visualizing and explaining HBR disruptions.

2. Related Works

2.1. Human Behavior Patterns

Circadian Rhythm refers to how cyclical/regular a person's sleep-wake cycle and routine is, which has important health ramifications [Vet18]. As such, monitoring these rhythms and intervening, if necessary, is crucial. However, gathering and analyzing human behavior data continuously for long periods is challenging *in-the-wild*. Smartphone sensing, wherein data from the smartphone's sensors are continuously gathered and analyzed to infer its owners' behaviors, has recently emerged as an inexpensive and scalable method of human behavioral analysis. Smartphone sensing has been effectively used in inferring depression levels [SLS*16], students' Grade Point Averages [WCC*14] and the smartphone owner's current context/situation [VELW18]. Abdullah *et al* [AMM*14] even successfully used time periods during which a person's smartphone screen was locked as a reliable proxy for determining when they were asleep.

Visualizations of human behavioral data is useful for discovering, contextualizing, and understanding disruptions in circadian rhythms. Fischer *et al* [FVR16] attempted to visualize "circadian misalignment" in shift workers using data from wrist-worn actimeter sensors combined with their sleep logs. They devised an intuitive method called "Composite Phase Deviation" which enabled them to generate and visualize density plots where the area and shape connote the extent of misalignment and variability in user behavior. They introduced the concept of "islands" and "pancakes" to refer to certain areas of the density plots to find variability in sleep data.

Geissmann *et al* [GRBG19] created "Rethomics", a framework in the R language to analyze circadian rhythms. Their framework also implemented intuitive data visualizations to present circadian information about animals.

However, unlike our work, these frameworks are not able to leverage multiple channel-linking to produce interactive analysis that facilitates *explainability*.

2.2. Interactive Visual Analysis (IVA) to Detect Anomalous Human Behavior

Deviation from normal human behavior may be indicative of health problems [Vet18]. Data visualizations can be an effective method of identifying anomalous human behaviors and making deviations from normalcy clearer, which in turn enables timely interventions. Examples include the detection of mal-intents such as spreading unverified rumours [RCP*14, SCFM16, SCV*18], committing financial fraud [vdEHBvW13] or scamming people [KFS*18]. Cao *et al* [CSL*15] created TargetVue, an intuitive tool to detect anomalous behaviors on social media and detect bots. They introduced the concept of a *z-glyph*, which is an intuitive visual metaphor, expanded further in a subsequent work by Cao *et al* [CLGD18] to highlight deviations from normal behaviors. These prior IVA works are particularly relevant for detecting changes in human behavior and behavioral rhythms (gathered through smartphones) as they show how novel interactive visual solutions enable analysts to quickly generate insights and find useful patterns in digital human data. Our work adds to this field by leveraging IVA for *in-the-wild* human behavior disruption that has direct implications on health. Specifically, we use IVA to find disruptions in natural 24-hour behavioral cycles.

2.3. Interactive Visual Analysis of Mobile Gathered Data

The vast proliferation of mobile phones has created many opportunities to gather rich datasets about human behaviors such as their mobility patterns and social interactions [CFB15]. IVAs can be useful for mining such data, contextualizing and explaining human behaviors. Pu *et al* [PXQ*11] leveraged IVA that combined established visualization techniques such as parallel coordinates plots with intuitive, novel techniques such as a "Voronoi-diagram-based" data visualizations to analyze the mobility patterns of three users. Senaratne *et al* [SMB*17] use an IVA approach to analyze spatial and temporal similarities in human movements using a passively gathered mobile phone dataset. They employed matrix visualizations of the user movements. These prior works illustrate the usefulness of IVA to explore and understand human movement (a very important facet of human life), its variations, patterns and disruptions.

IVA techniques can be further augmented with novel glyphs and visual metaphors that are useful for representing complex mobile phone data. Shen *et al* [SM08] created MobiVis, an IVA tool that implemented the "Behavior Ring", a radial metaphor to represent individual and group behaviors compactly. Their tool enabled intuitive visual data mining by semantic filtering to facilitate effective analysis of "social-spacial-temporal" data that phones gather. This approach illustrates the utility of compact visual metaphors and IVA to understand complex phone data.

Unlike our method, these methods do not incorporate the concept of *cycles* and *rhythms* and the disruptions thereof. Specifically, we focus on intra-personal behavioral comparisons while these works typically focus on large, inter-personal and inter-group analysis.

3. Goal and Task Analysis: Interactive Visual Analytics To Monitor Bio-Behavioral Rhythms

Given that this domain contains such diverse and heterogeneous datasets from which many features are typically extracted, Interactive Visual Analytics (IVA) can assist in making sense of this data and identifying and monitoring people with deviations in their behavioral patterns.

We conducted goal and task analysis sessions with an expert in bio-behavioral rhythms who was also experienced in analyzing human behavior gathered using ubiquitous sensing devices. The expert was particularly interested in rhythms related to sleep behavior and how certain events may disrupt sleep patterns. For instance, the buildup of stress due to uncontrollable, external factors may cause lost sleep, which may have health ramifications.

We discussed how a smartphone may collect data that can indicate disruptions and breakages in patterns. Some of her suggestions included conceptualizing smartphone data as *channels* of information, which may provide important clues about a person's *contextual* information. Examples of channels are the state of the smartphone including screen locked, battery charging, apps being used and its GPS location. In order to derive how rhythmic a person is in their daily routines, it may be useful to find rhythms, disruptions and breaks in these channels.

Given that human behavior is complex and the smartphone channels used to make behavioral inferences can become overwhelming, IVA can be useful for making sense of, and correlating these information channels. Given that a correlation in these channels may be meaningful, (the expert suggested that correlating darkness with a lack of sound and screen being locked might be a useful method of detecting sleep). Such meaningful linkages and correlations across different channels would be difficult to show using non-visual statistics. IVA may powerfully augment an analyst's ability to make sense of people rhythms by intuitively overlaying and correlating complex channel data to increase its interpretability. The expert suggested two broad goals that she would have as someone analyzing smartphone collected human behavior data:

- **G1: Discover overall levels of behavioral rhythms:** and times during which breaks occur
  - Synthesizing an overall numerical measure or *score* to capture and quantify a person's bio-behavioral rhythm, which can then be represented visually to reveal deviations from normalcy.
- **G2: Explain and contextualize causative factors:** that led to deviations from normal rhythms across multiple channels:
  - The multiple channels of smartphone-gathered data can provide a multi-faceted view into a person's behaviors. The expert wanted several linked views of different channels such as a person's geo-location or the intensity of their smartphone interactions, which may explain the reasons for disruptions and breakage in their patterns. Such multi-view context may also enable the disambiguation of harmful disruptions in behavior patterns patterns such as staying up all night (a sign of depression) versus benign pattern disruptions such as a person who is travelling for vacation.

We also discussed the specific tasks that the analyst would like to be able to perform to achieve the goals described above. Given our collective knowledge of the data in this domain, we devised the following tasks to achieve the goals described.

- **T1, Population-level meta-analysis**: Get a quick overview of the level of rhythmicity of the behavior patterns of all participants in a study to quickly find the ones with most and least rhythms
- **T2, Anomalous HBR identification**: Identify participants with significant disruptions in their behavioral rhythms quickly
- **T3, Cross-channel HBR exploration**: Examine and contextualize individual participants' rhythmicity levels across multiple "channels". For instance, visualize physical activity vs geolocation disruptions.
- **T4, HBR contextualization**: Highlight contextual factors that might have bearing on rhythm such as day of the week, or weekday vs weekend.
- **T5, Raw sensor value drill-down**: Visualizing the values of sensor readings, which may hold important clues about what a person did around the time of disruptions in their HBR. For instance, the amount of time they spent interacting with their smartphone's screen.

Extant work has shown that smartphone channels such as phone state measurements are indicative of bio-behavioral trends that have significant implications for a person's health. Abdullah *et al* [AMM*14] were able to predict a research participants' sleeping habits using data about whether their smartphone's screen was locked. Other works have used ambient light sensor data from smartphones to detect the occurrence of sleep [MDW*14, CLC*13]. These channels' ability to detect very important bio-behaviors (sleep for instance), make their co-occurrence, lack of co-occurrence and disruptions in those channels very interesting.

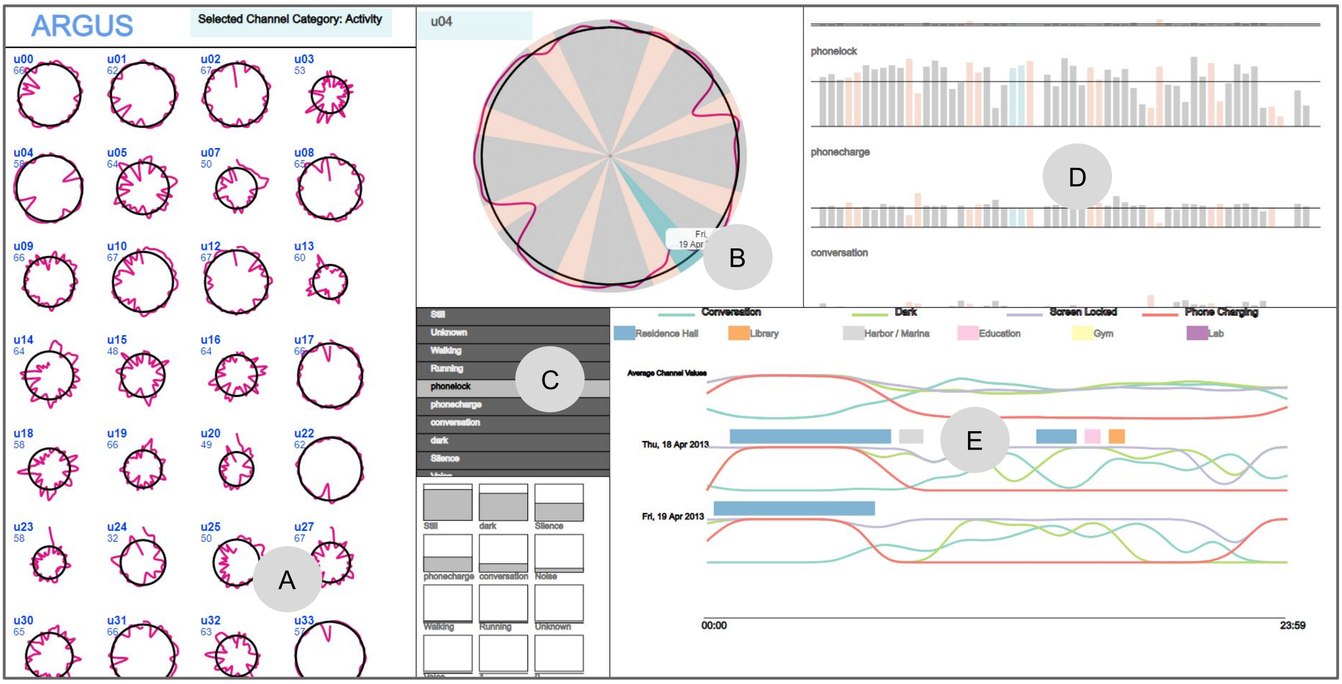

Figure 1: *ARGUS: A multi-pane visualization framework to find and explain deviations in bio-behavioral rhythms.* ***A: Eyes of ARGUS****. This view provides a quick overview of the levels and breaks in bio-behavioral rhythms across all participants.* ***B: Magnified Eye of Argus****. A user can select an eye to magnify and gain a clearer look into the rhythm changes. The lighter color represents weekends and the rest are weekdays.* ***C: Co-occurrence View****. Provide an overview of how consistently channel values co-occur to enable analyst to explain behavior in terms of a lack of co-occurrence of frequently co-occurring channels.* ***D: Duration View****. A summary for how much a channel had positive values throughout the day.* ***E: Explainability View****. Visualizing and linking multiple channels to gain a greater understanding about the causes of breaks in rhythms.*

4. ARGUS

To achieve the goals and tasks identified above, we researched and developed *ARGUS*, an interactive multi-pane visualization tool. Since the data that *ARGUS* would be dealing with is highly multi-variate and heterogeneous, we adopted the “Visual Analytics Mantra" proposed by Keim *et al* [KMS*08], which states: “Analyse First - Show the Important - Zoom, Filter and Analyze Further - Details on Demand”. Here we describe our Rhythm Deviation Score for quantifying the level of regularity of HBRs, and visualizations we created and the rationale behind our choices.

4.1. Rhythmicity Deviation Score

We now expound on our novel Rhythmicity Deviation Score, a single number we synthesized to quantify the degree of regularity of a person’s circadian rhythm based on data gathered from their smartphone sensors. The visualization panes of *ARGUS* then visualize this rhythmicity deviation score along with useful contextual information. Our Rhythmicity Deviation Score is based on the Lomb-Scargle periodogram [Lom76, Sca82], a classic method for finding periodicity in irregularly-sampled data.

Definition: Lomb-Scargle Periodogram

Let $X = \Big((y_0,t_0),(y_1,t_1),\ldots,(y_k,t_k)\Big)$ be a time series, such that $\sum_{i=0}^{k} y_i = 0$. The Lomb-Scargle periodogram, P_{LS}, of X for frequency ω is given by:

$$\mathrm{P}_{LS}(\omega) = \frac{1}{\sum_{i=0}^{k} y_i^2}\left\{\frac{\left[\sum_i y_i\, cos\,(\omega(t_i-\hat{\tau}))\right]^2}{\sum_i cos\,(\omega(t_i-\hat{\tau}))} + \frac{\left[\sum_i y_i\, sin\,(\omega(t_i-\hat{\tau}))\right]^2}{\sum_i sin\,(\omega(t_i-\hat{\tau}))}\right\},$$

where $\hat{\tau}$ is a time delay parameter such that the sinusoids are mutually orthogonal at each sample time t_i. $\hat{\tau}$ is given by:

$$tan\,(2\omega\hat{\tau}) = \frac{\sum_i sin\,(2\omega t_i)}{\sum_i cos\,(2\omega t_i)}$$

An example of the Lomb-Scargle periodogram for 1 channel (Sleeping) is shown in Figure 2. The peak of the periodogram oc-

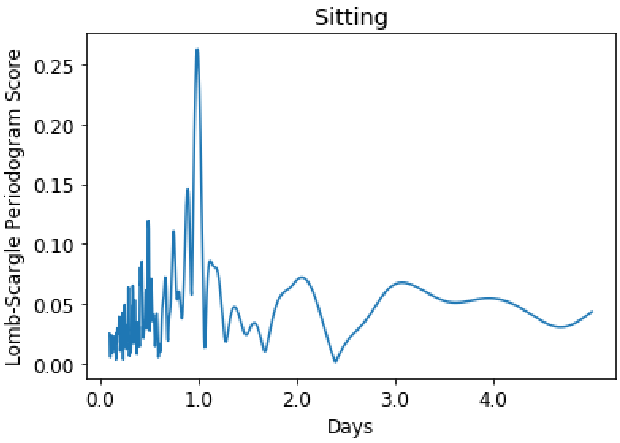

Figure 2: *The graph of the Lomb-Scargle periodogram of a particular user's sleep data. The x-axis represents the time scale (in days), and the y-axis represents the value of the Lomb-Scargle periodogram evaluated at frequencies corresponding to the time scales. We see a peak in the periodogram near 1 day, indicating that this user is relatively cyclic in their sleep habits on a 24-hour bases.*

curs at 1 day, indicating that this user is relatively cyclic in their sleep habits on a 24-hour cycle.

In order to use the Lomb-Scargle periodogram to identify disruptions in user behavior, we apply the periodogram on each individual *channel* of user behavior.

Definition: Channel

A *channel* C_i is given by a time series: $\big((c_{i,0},t_{i,0}),(c_{i,1},t_{i,1}),\ldots,(c_{i,k},t_{i,k})\big)$, such that each $c_{i,j} \in \{0,1\}$.

Channels are sequences of binary variables that can represent inferred or self-reported behavioral indicators such as physical activity (i.e. instances of time where the user was walking, sitting, etc.) or objective sensor readings such as whether the smartphone was locked, connected to a wireless network, or charging. In order to quantify changes in channel behaviors we first must define the *occurrence ration*, the length of time for which the channel was a positive instance (i.e., there *were* walking) over a certain time scale.

Definition: Occurrence Ratio

The *occurrence ratio* O_r of channel C_i for day D_n is the ratio of positive instances of C_i that occurred during day D_n. Let $C_{i,D_n} = \{(c_{i,j},t_{i,j} | t_{i,j} \in D_n)\}$. Then, the occurrence ratio is given by:

$$O_r(C_i,D_n) = \frac{\sum_{c_{i,j}\in C_{i,D_n}} c_{i,j}}{\|C_{i,D_n}\|}$$

As we are interested in investigating the users circadian rhythm, which is a 24-hour cycle, we choose 1 day to be the time scale over which this value is calculated. We can also define the *average occurrence ratio*, in order to typify the user's usual behavior.

Definition: Average Occurrence Ratio

The *average occurrence ratio* $\overline{O}_r$ of channel C_i is the average ratio of positive instances of channel C_i over all days in the set of days $\mathcal{D}$:

$$\overline{O}_r(C_i) = \frac{\sum_j O_r(C_i,D_j)}{\|\mathcal{D}\|}$$

We can now define the *circadian rhythm* of a given channel:

Definition: Circadian Rhythm

The *circadian rhythm* of channel C_i measures how cyclic C_i is on a frequency of 1 day. The circadian rhythm is defined as:

$$R(C_i) = \frac{\int_{\frac{1}{D}-\Delta t_1}^{\frac{1}{D}+\Delta t_1} P_{LS}^{C_i}(\omega)\,d\omega}{\int_{\frac{1}{D}-\Delta t_2}^{\frac{1}{D}+\Delta t_2} P_{LS}^{C_i}(\omega)\,d\omega},$$

such that $\Delta t_1 < \Delta t_2$ and $P_{LS}^{C_i}$ is the Lomb-Scargle periodogram of of C_i.

This definition is based off of Wang *et. al.'s* definition of circadian rhythm [WHW*18], though ours differs in that we use the Lomb-Scarcle periodogram rather than the *power spectral density* as we deal with unevenly sampled data. While this definition may seem obtuse, it has an intuitive basis. The circadian rhythm is simply the integral of the periodogram for a small region around the the frequency associated with 24 hours divided by a similar integral taken over a larger range of values. Intuitively, if nearly all the power of the periodogram is concentrated at the 24 hour mark (that is, their rhythm is nearly perfectly described by a 24-hour cycle), then this value should be close to 1. Otherwise, this ratio goes to 0 (indicating that the user is *not* rhythmic on a 24 hour cycle for the given channel).

Having defined the circadian rhythm, we now define disruptions in the circadian rhythm.

Definition: Channel Rhythm Disruption

The *channel rhythm disruption CD* of C_i for day D_j measures the difference in occurrence ratio of C_i for D_j and the average occurrence ratio of C_i, weighted by the circadian rhythm of C_i. CD is given by:

$$CD(C_i,D_j) = R(C_i)\cdot\|O_r(C_i,D_j) - \overline{O}_r(C_i)\|$$

The *channel rhythm disruption* is simply the change in behavior of a particular channel (that is, changes in duration of positive instances of this channel) weighted by the circadian rhythm of this channel. The reason why we weight the change in behavior by the circadian rhythm is that we only want to identify *meaningful* disruptions in behavior. For instance, assume that the behavior of the channel C_i differs from the average for a given day. However, the average behavior is meaningless if the user is not rhythmic for that channel. By weighing the change in behavior by the circadian rhythm of that channel, then deviations in non-rhythmic channels will result in only a small channel disruption score.

Additionally, we wish to quantify disruptions in user behavior on a scale larger than single channels. For this reason, we define *channel categories*.

Definition: Channel Category

A *channel category* G_k is a set of channels. Any given pair of channel categories have no elements in common. Let $\mathcal{G}$ be the set of all channel categories and let $\mathcal{C}$ be the set of all channels. Then, for a pair of channel categories G_k and G_j,

$$G_k \cap G_j = \varnothing \; \forall \; G_k \neq G_j \in \mathcal{G}$$

Additionally, we require that all channels belong to a category. Thus:

$$\bigcup_{G_k \in \mathcal{G}} \bigcup_{C_j \in G_k} C_j = \mathcal{C}.$$

For our purpose, we group our channels into 3 categories: objective sensors, geo-location, and activities. More details about the channels in each category is given in Figure 3.

Finally, we can now define the *rhythm deviation*, which is the metric we use to identify individuals who have experienced significant changes in behavior.

Definition: Rhythm Deviation

The *rhythm deviation RD* of category G_i for day D_j measures the average *channel rhythm disruption* for each channel in G_i on day D_j.

$$RD(G_i, D_j) = \frac{\sum_{C_k \in G_i} CD(C_k, D_j)}{\|G_i\|}$$

Next, we will discuss our IVA framework for investigating rhythm deviation and channel rhythm deviations.

4.2. Eyes of ARGUS

The pane shown in Fig 1 A, contains the Eyes of ARGUS (EA). Every glyph plots a person's average Rhythm Deviation (RD) score as a black circle (a circle with a larger circumference represents more overall rhythm) against the daily RD scores ordered in a clockwise direction. The strength of deviations is encoded by inwards distance from the black circle i.e. the more the purple line is towards the center, the higher the level of deviation (T2). Towards the top left of every EA is the user id and the number of days of participation. EA shows the deviations from rhythm, one circle per day, for all the days of participation. This glyph is based off the Z-glyph developed by Cao et.al. [CLGD18] to effectively visualize *deviations* from the norm. Their analysis showed significant improvements in discerning deviations when using the z-glyph over traditional line glyphs. This is particularly applicable for this problem as deviations from normal rhythmicity or punctures in rhythms are exactly what ARGUS tries to highlight. Figure 1A is gives the analyst a quick overview of the level of bio-behavioral rhythmicity of users in general (G1,T1) and any interesting users that they might want to explore further (T2).

	Activities	Sensors	Geolocation
StudentLife	App inferred activity - 4 activities	Phone locked, phone charging, conversation and dark	Geo-location gathered every 10 mins (if available)
Study1b	Participant provided labels - 18 labels	Phone locked, phone charging and wifi	Geo-location gathered after every minute(if available)

Figure 3: *Description of data channels across the 2 datasets that we use. Among the most significant differences between datasets is that activities are inferred in StudentLife while they are user-provided in Study1b. Additionally, Study1b contains a richer set of activities. StudentLife also posses information on the amount of time that each user was having a conversation with another individual, which is something Study1b lacks.*

Clicking on any EA will magnify it in the Magnified Eye of ARGUS (MEA) pane (Figure 1B). The underlying circle has the maximum possible circumference (i.e. if a person is perfectly rhythmic). The circle is divided into slices with each slice representing one day of participation. The slices with the lighter color are weekends (G2,T4). The user is guided by the RD score and can select multiple days by clicking on the slices to view them in more detail in Channel Duration view (Figure 1D) and the Explainability View (Figure 1E). Multiple day views are useful as events from a preceding day may have had an affect on the current day (G2).

The user can also select the channel category to visualize in the EA and MEA by clicking on the "Selected Channel Category" bar at the top of Figure 1A (G2,T3). The mean rhythm score across the three channels is available for visualization. The specific channels for the 2 datasets are described in Figure 3 and will be discussed in more detail in the use cases.

4.3. Duration View

The Duration View (DV) (Figure 1D) shows the overall "duration" of occurrences of a channel for every day of the participant's data. By duration, we mean the amount of time during each day that the channel was "on". For instance, the duration of time a phone was plugged in or in a particular location cluster. The duration across days is shown separately for every channel present, as bar plots. Every vertical bar represents a single day of participation and the length of the bars represents the overall duration per day for the channel. The horizontal line for every channel is a visual indicator for the mean duration. It follows the same coloring scheme as the pie layout in the MEA. In the example, in Figure 1D, the user's phone is locked throughout the day generally more than their phone is charging. This is meant to aid the analyst in explaining rhythm disruptions by providing overall duration of various levels of channel disruption (G2, T2, T5).

4.4. Co-occurrence View

The co-occurrence prevalence between certain channels is also interesting for researchers as a break in co-occurrence may indicate or explain a break in behavioral rhythm. Clicking on a users EA shows the Co-occurrence View (Figure 1C). This view has a list of all channels available for the selected user. Clicking on a channel bar shows in an ordered format (left to right, top to bottom) the *most commonly co-occurring channels* bars, that is channels that were "on" at the same time as the selected channel. The gray fill in the bars is proportional to the frequency of co-occurrence. For instance, for this person, phone-lock was mostly coincided with "still", "dark" and "silence". This is meant to let the analyst understand further the causes of break in rhythms as the absence of channel co-occurrence among mostly co-occurring channels in interesting (G2,T5). To link the occurrence of co-occurring channels, hovering over a commonly co-occurring bar highlights the durations of the selected channel and the hovered over channel in the Explainability view. An example of how this occurs is illustrated in Figure 4.

4.5. Explainability View

The Explainability View (EV) (Figure 1E) aims to provide a finer day-level view of the collected data to assist analysts in figuring out potential causes of a change in a subject's rhythm (G2,T5). The lines show the variation in the channel data available for the phone data. The height of the lines represents if the channels average channel values per hour. The first plot shows the average channel values across the entire period of collection. Every plot after that represents a specific day chosen in the MEA. For instance on 14th April (Figure 1E), the phone was charging through the night into the morning. As this is laid out on a horizontal timeline against a common scale, changes over hours are easily noticeable. The colored bars over the lines represent the durations of time for which they were in the same geographic cluster. Both the datasets we utilize in our case studies contained geo-locations of the participants throughout the day (whenever available). We ran a clustering algorithm called DBSCAN [EKS*96] to find geo-clusters for the participants. The clusters are encoded with colored bars

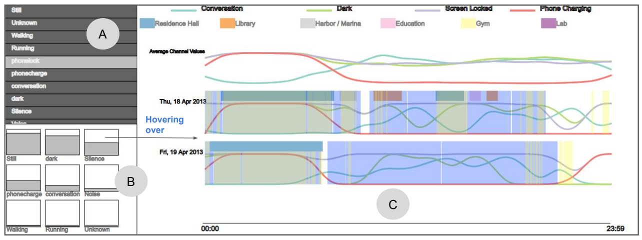

Figure 4: *In the Co-occurrence View, the user can select a channel in A to see its most commonly co-occurring channel bars (shown in B), ordered from left to right, top to bottom in terms of frequency of co-occurrence. The vertical grey fill in bars in B is proportional to frequency of co-occurrence with A. Hovering over any of the bars in B highlights the duration of the channel selected in A across the Explainability View in light blue while the hovered over bar's channel gets highlighted in light yellow. The durations in the Explainability View with a light green overlay are when the 2 channels co-occur.*

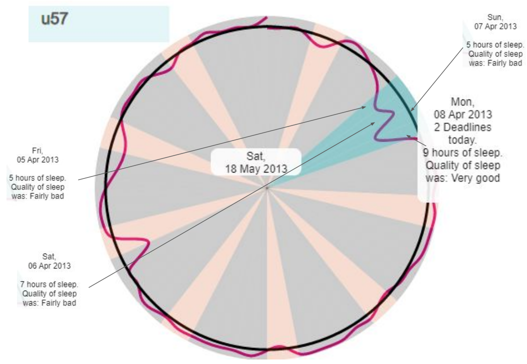

Figure 5: *Hovering over a day slice brings up a tool-tip that shows the human readable date and any other day related information (if any provided). The analyst can clearly see a disruption in rhythm and the reduction of sleep duration and quality leading up to a day with two deadlines.*

(legend shown at the top of the EV. Given limited visual real estate we only show the top 6 clusters in which the participant was present for most of the time. The legend (Figure 1E) shows the colors for the lines and the clusters. The colors were selected using a 10-class qualitative palette from ColorBrewer [BH09] to ensure that they were discernible. The human understandable categories for the clusters were gathered by running the cluster coordinates against the Foursquare API [Fou].

5. Illustrative Use Cases

To illustrate the usability of ARGUS, we introduce Emma, a graduate student in psychology who specializes in human behavioral rhythms and their effects on human health, especially for college

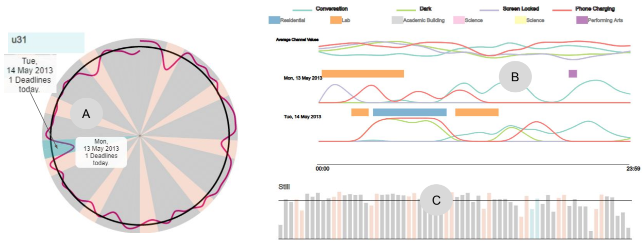

Figure 6: *May 13th and May 14th have disruptions in rhythm and both days have deadlines. Exploring these days further in the Explainability View reveals that the participant's dark, conversation, screen locked and phone charging channels are all off along with the fact that they are in a lab from around the beginning of day on May 13th.*

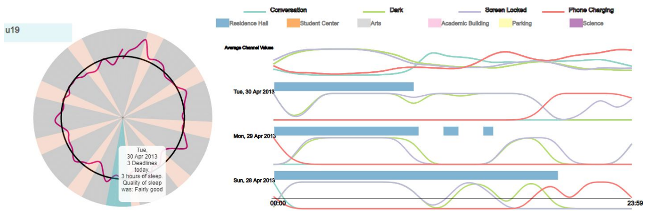

Figure 7: *There are 3 Deadlines on 30th April. In the preceding days, we is much less conversation than average. Such breaks in social activities are highly interesting for Emma.*

students. Emma has access to two different real world datasets that she is able to visualize using *ARGUS*.

5.1. Dataset 1: StudentLife

The first dataset we used for evaluating *ARGUS* an open source dataset gathered from a smartphone sensing project called StudentLife [WCC*14]. Smartphone sensor data was collected and analyzed to infer various participant behaviours including their GPA and physical state (e.g. still vs walking). The audio of scenes the user visited were also analyzed to make inferences such as whether the person was in a silent environment versus a noisy environment, or conversing. The dataset contains information for 49 students. This dataset records only significant durations (>= 1 hour) when the phone was in a dark environment, charging and the screen being locked. The data contains geo-location (whenever available) of the device throughout the day.

The StudentLife project also collected subject responses to mental health questionnaires with wellness information such as their sleep duration, sleep quality, and stress levels. Data was gathered from the beginning of a 10-week academic term and the entire collection time period was a little over two months. Analysis of this dataset provides a clearer understanding of how student behavior changes over the course of an academic term.

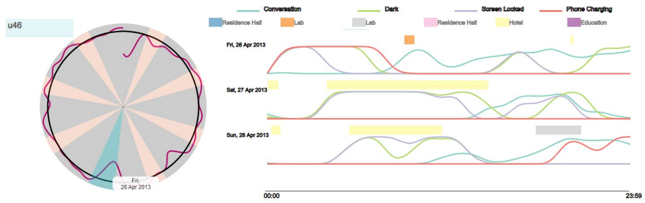

Figure 8: *There is a major disruption in bio-rhythm starting from Apr 26th, which is a Friday. Towards the end of this day, the available geo-location indicates that this person was at a hotel. The subsequent two days which are Saturday and Sunday, there are longer time periods of geo-location indicating that they are at a hotel.*

Emma's visualization of the bio-rhythms of StudentLife students using *ARGUS*:

- *Identifying and contextualizing bio-rhythm disruptions caused by deadlines:* Emma is particularly interested in exploring changes in student behavior around more stressful times such as project due dates and deadlines (G1). The students were asked to provide the academic deadlines that occurred during their the days they participated in the StudentLife study. Emma takes a look at the eyes to see if there is any participant that sticks out (T1). She notices that participant u57 has a large inward spike (T2), indicating a large deviation of their bio-rhythm from normalcy. She clicks on it to magnify it in the MEA (Figure 5). When she hovers over the day slices for which the RD score was high, she noticed that the participant had low sleep duration and quality for three straight days (April 5*th*, 6*th* and 7*th*) leading up to April 8*th*, which had two deadlines (G2, T4). After these two deadlines passed on the 8*th*, their sleep duration and quality, as well as their RD score all improved. This leads Emma to believe that the stress caused by imminent deadlines disrupted the participant's bio-rhythm (G2). Similarly, as these deadlines passed, participants bio-rhythms returned to normal. Visual overlay of multiple panes human understandable data along with objective rhythmicity scores and objective calculations in *ARGUS*, made this insight easy for Emma.
- *Explanations of bio-rhythm shifts using the Explainability View (EV):* After observing the effect that deadlines had on students' sleep patterns, Emma wants to see if she can observe a reduction in the quantity and quality of their sleep during days when participants did not respond to sleep questions. While exploring the data for u31, she notices a disruption in bio-rhythm towards the end the term (T2). Hovering over the 2 days during which the bio-rhythm was disrupted revealed that there was a deadline on both days (Figure 6A). Emma clicked on the 2 day slices to view them in the Explainability View (EV) and the duration average view. In the duration view, she notices that the participant had lower levels of the "Still" state on May 13th (Figure 6C). The participant had much lower levels of screen locked and being in a dark environment in EV (Figure 6B). She also noticed that the participant was in a "Lab" in the very early hours of the morning. The participant was again recorded as being in the lab

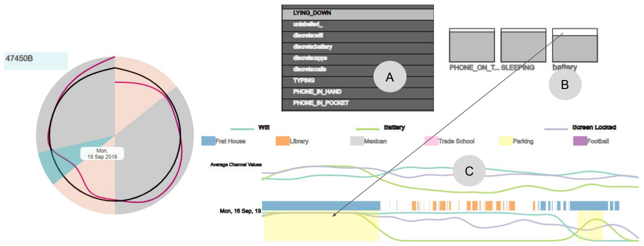

Figure 9: *Clicking the "Lying down" bar shows the most frequently co-occurring channels which are "Phone on table", "Sleeping" and "battery". Hovering over the "battery" bar shows the occurrence of that channel but also see no coincidence with "Lying down". However, the screen is locked throughout the highlighted period indicating a high probability of "Sleeping". Showing co-occurrences like these highlights the issues with such datasets and enables analysts make smarter decisions.*

for early hours of the morning and then being in a "Residential" location where they plugged their phone and were in a dark environment for a significant amount of time. The detailed and comprehensive system of overlaying various channel data in *ARGUS* enabled Emma to pinpoint, contextualize and understand a potentially concerning disruption in the students' bio-rhythm.

- *Detecting changes in other sensor channels:* Emma is also interested in the social behaviors of students around stressful times. She notices u19 who does not have a high overall rhythm in their data (T2). As she explores their data, she notices a day with three deadlines. She explores the day and the days leading up to it in the EV and notices a steep drop in the amount of conversation this student was having (Figure 7) (T4, T5). Emma is interested in discovering changes in social behavior rhythms caused by academic stress (G2).
- *Exploring the relationship between geo-location and bio-rhythms:* Small disruptions in students' bio-rhythm during a term may not be a major cause for concern [Vet18]. Emma wants to investigate how students' geo-locations affect the rhythmicity of their bio-scores. She selects "Geo-location" (T3)in the Rhythm Selector (in the top pane in Figure 1A) and notices u46 had a large deviation in their geo-location rhythm. Based on the color of the underlying slides, Emma notices that this large deviation occurred on the weekend (T4). She clicks on the non-rhythmic days (Friday, Saturday and Sunday) and notices in the EV that the participant had a small geo-location recording showing that they were at a hotel and more geo-location readings for the 2 weekend days (Figure 8). This leads Emma to believe that this was not a cause for major concern as it was an isolated incident related to travelling. The visual overlay of these various channels allowed Emma to disambiguate a potentially concerning rhythm disruption as merely being an innocuous one.

5.2. Smartphone sensor data gathered locally from our campus - Study1b

The second dataset we utilized to evaluate *ARGUS* was gathered around our own campus and will be referred to as Study1b. It contains smartphone sensor data for 103 people. Our approach was different because we did not constrain ourselves to a term like StudentLife and gathered data over a number of time different periods. Our participant population was also more diverse demographically and included teenaged undergraduates to middle-aged campus office workers. We also had a shorter average period of participation (two weeks). We used a modified version of the ExtraSensory Android application, developed by Vaizman *et al* [VELW18]. The application gathered sensor data for 20 seconds of every minute. Unlike StudentLife, this app did not suggest any inferred activity and just lets the user provide activity labels for themselves in the wild as they lived their lives. Users could provide 18 different labels for activities such as "Walking" and "Sitting", as well as phone location such as "Phone in Pocket" or "Phone in Hand". The participants varied in terms of the number of labels they provided. The application collects several similar channels such as screen locked, battery charging and geo-location.

Emma's visualization of the bio-rhythms of study1b participants using *ARGUS*:

Emma visualizes this dataset in ARGUS. As this study was conducted in the wild, subjects had to continuously label their smartphone-sensed data with ground-truth labels of their actual activity to facilitate supervised machine learning later. Emma believes that students' bio-rhythms affected the quality/accuracy of labels they provided and used *ARGUS* to explore this hypothesis.

- *Investigating the effectiveness of analyzing sensor rhythm values in the absence of human provided ground truth labels:* She analyzes participant 47450B. She notices quickly in the Duration View (DV) that the participant has done an inconsistent job of providing self-reported labels for their data. This means that she will have to rely on objective sensor values to make sense of this data. She clicks on the channel rhythm selector (top pane of Figure 1A) to select "Sensors" as the channel category (T3) as these channels do not require human labelling. She notices a day where the rhythm is off (Figure 9). She visualizes it in the EV and notices that the user was in the "Frat House" cluster (a type of on campus residence common in universities in USA). She notices in the DV that the participant provided no labels for "Lying down" or "Sleeping". She clicks on the "Lying down" bar in the co-occurrence view and notices that the top co-occurring positive values for "Lying down" are "Sleeping", "Phone on table" and "battery" (Figure 9B). She hovers over the co-occurring bars for "Sleeping" and "Phone on table" and notices these labels were also not provided. She hovers over battery and notices the light yellow overlay (Figure 9C) for "battery" but no light blue overlay for "Lying down". She notices that the screen was also locked and that the rhythm disruption for objective sensor values in this day was caused by other deviations later on in the day, which may not be that interesting (T4,G2). Overlaying and linking this multi-faceted data allowed her to dismiss this day from her concern, which may not have been possible using traditional statistical analyses.
- *Detecting erroneous labels based on unlikely co-occurrence using ARGUS*: Emma is also aware that some people may have carelessly provided labels which would make it difficult for her to accurately determine rhythms and deviations therein. She

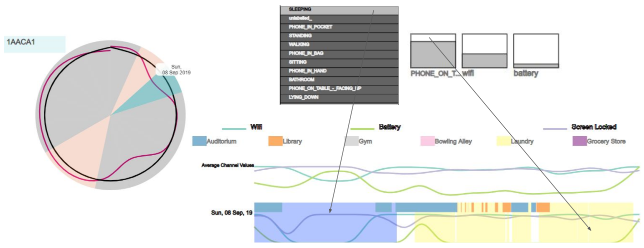

Figure 10: *Study participants may also do a poor job of self-labelling smartphone gathered data. This person usually co-labels "Phone on table" and "Sleeping" but not for this day. In addition, there is time period after the start of midnight for which the screen was unlocked that is unlikely to co-occur with "Sleeping". IVA makes these important aspects of the data human understandable.*

chooses the sensors category (T3) for the rhythms and notices a user 1AACA1 who has some bio-rhythm disruptions. She clicks on the slice for September 8*th* and she notices in the EV that the user had their screen unlocked for some time after midnight. She clicks the "Sleeping" bar in the co-occurrence view and can see that the most commonly co-occurring channels are "Phone on table", "WiFi" and "battery" (Figure 10). Hovering over "'Phone on table" show that the two channels do not co-occur on this day. Further, sleeping and phone usage while the screen was unlocked is unlikely to have occurred. This view leads Emma to believe that this may have been an instance of mislabelling and there was in fact a disruption in this person's bio-rhythm (T4, G2). Calculating the rhythm score separately for various objective sensor values enabled this as humans are error prone.

6. Evaluation

We invited 5 volunteers to interact with ARGUS and provide feedback. The volunteers were all graduate students. They were given a short tutorial about the project, datasets and ARGUS. They were then asked interact with ARGUS and go through the same use cases that Emma did. At the end of the evaluation session, they were asked to fill out a questionnaire about the ease of use of ARGUS' visual metaphors. The questionnaire and results presented in figures 11 and 12 demonstrate the efficacy of *ARGUS*.

7. Limitations and Future Work

While promising, the work we have presented has several limitations including:

- *Limits of person-level monitoring:* As the number of participants increases in size, person level monitoring of behavioral rhythms may become unfeasible. One approach of grouping people together for cohort analysis might be to sort people into *chronotypes* which is basically a person's sleep-wake cycle and the habit of going to bed and waking up at particular times each day.
- *Missing human-provided ground truth activity labels"* (as in the case of our Study1b dataset) may also become an issue as larger deployments of studies may not be able to depend upon participants to provide accurate labels. In future, we will investigate using existing machine learning-based human behavior models on accelerometer and gyroscope data collected to detect and visualize users' physical activities.

On a scale of 1 - 7 (1 being not at all and 7 being very easily), please answer these questions:

Q1: Were you able to easily find people with high and low rhythms using the Eyes of ARGUS and the Magnified Eyes of ARGUS view?
Q2: Were deviations from rhythm easy to identify both in terms of overall quantity and magnitude?
Q3: Were you able to compare the channel duration per day against the average channel duration?
Q4: Were you able to gain more insight and a more complete picture by interacting with the hover over feature on the cooccurrence view and the duration summary view?
Q5: Were you able to see changes in daily channel values in the line displays in the Explainability View?

Figure 11: *The questionnaire to evaluate ARGUS. We use a standard 7 point Likert scale [Lik32] for the possible responses. Each question was designed to assess the effectiveness and ease of use of each of our visual metaphors.*

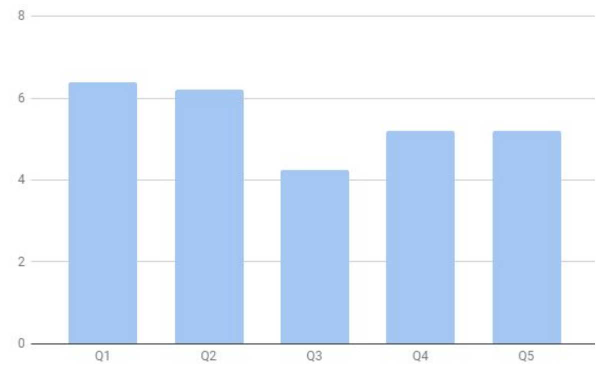

Figure 12: *Questionnaire results.*

8. Conclusion

In this paper, we presented our work on *ARGUS*, a visual analytics framework that allows analysts to not only identify disruptions in smartphone gathered bio-behavioral rhythms but also to contextualize and explain them. To guide our designs, we conducted a detailed goal and task analysis with an expert. We devised an intuitive Rhythm Deviation score that quantified the degree of rhythmicity in participants bio-rhythm, which was then visualized using a glyph visual metaphor that enabled easy identification of disruptions in bio-rhythms. *ARGUS* provided additional additional overlays including multi-sensor channel and geo-location overlays and multiple linked visualization panes, which facilitated contextualization and reasoning about participants' bio-rhythm scores. We provided an extensive walk through of illustrative use cases to show how multiple linked panes provided a clearer look into the occurrences and causes of disruptions in bio-behavioral rhythms.

References

[AMM*14] Abdullah S., Matthews M., Murnane E. L., Gay G., Choudhury T.: Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (2014), ACM, pp. 673–684. 1, 2, 3

[BH09] Brewer C. A., Harrower M.: Colorbrewer 2.0: color advice for cartography. *The Pennsylvania State University. http://colorbrewer2.org/. Accessed 6*, 02 (2009), 2010. 7

[CFB15] Calabrese F., Ferrari L., Blondel V. D.: Urban sensing using mobile phone network data: a survey of research. *Acm computing surveys (csur) 47*, 2 (2015), 25. 2

[CLC*13] Chen Z., Lin M., Chen F., Lane N. D., Cardone G., Wang R., Li T., Chen Y., Choudhury T., Campbell A. T.: Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare* (2013), ICST (Institute for Computer Sciences, Social-Informatics and ..., pp. 145–152. 3

[CLGD18] Cao N., Lin Y.-R., Gotz D., Du F.: Z-glyph: Visualizing outliers in multivariate data. *Information Visualization 17*, 1 (2018), 22–40. 2, 6

[CSL*15] Cao N., Shi C., Lin S., Lu J., Lin Y.-R., Lin C.-Y.: Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 280–289. 2

[EKS*96] Ester M., Kriegel H.-P., Sander J., Xu X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (1996), vol. 96, pp. 226–231. 7

[Fou] Foursqaure:. URL: `https://developer.foursquare.com/`. 7

[FVR16] Fischer D., Vetter C., Roenneberg T.: A novel method to visualise and quantify circadian misalignment. *Scientific reports 6* (2016), 38601. 2

[GRBG19] Geissmann Q., Rodriguez L. G., Beckwith E. J., Gilestro G. F.: Rethomics: An r framework to analyse high-throughput behavioural data. *PloS one 14*, 1 (2019), e0209331. 2

[KFS*18] Koven J., Felix C., Siadati H., Jakobsson M., Bertini E.: Lessons learned developing a visual analytics solution for investigative analysis of scamming activities. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 225–234. 2

[KMS*08] Keim D. A., Mansmann F., Schneidewind J., Thomas J., Ziegler H.: Visual analytics: Scope and challenges. In *Visual data mining*. Springer, 2008, pp. 76–90. 4

[Lik32] Likert R.: A technique for the measurement of attitudes. *Archives of psychology* (1932). 10

[Lom76] Lomb N. R.: Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science 39*, 2 (1976), 447–462. 4

[MDW*14] Min J.-K., Doryab A., Wiese J., Amini S., Zimmerman J., Hong J. I.: Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2014), ACM, pp. 477–486. 3

[PXQ*11] Pu J., Xu P., Qu H., Cui W., Liu S., Ni L.: Visual analysis of people's mobility pattern from mobile phone data. In *Proceedings of the 2011 Visual Information Communication-International Symposium* (2011), ACM, p. 13. 2

[RACB11] Rabbi M., Ali S., Choudhury T., Berke E.: Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011), ACM, pp. 385–394. 1

[RCP*14] Resnick P., Carton S., Park S., Shen Y., Zeffer N.: Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In *Proc. Computational Journalism Conference* (2014), vol. 5. 2

[Sca82] Scargle J. D.: Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal 263* (1982), 835–853. 4

[SCFM16] Shao C., Ciampaglia G. L., Flammini A., Menczer F.: Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web* (2016), International World Wide Web Conferences Steering Committee, pp. 745–750. 2

[SCV*18] Shao C., Ciampaglia G. L., Varol O., Yang K.-C., Flammini A., Menczer F.: The spread of low-credibility content by social bots. *Nature communications 9*, 1 (2018), 4787. 2

[SLS*16] Saeb S., Lattie E. G., Schueller S. M., Kording K. P., Mohr D. C.: The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ 4* (2016), e2537. 2

[SM08] Shen Z., Ma K.-L.: Mobivis: A visualization system for exploring mobile data. In *2008 IEEE Pacific Visualization Symposium* (2008), IEEE, pp. 175–182. 3

[SMB*17] Senaratne H., Mueller M., Behrisch M., Lalanne F., Bustos-Jiménez J., Schneidewind J., Keim D., Schreck T.: Urban mobility analysis with mobile network data: a visual analytics approach. *IEEE Transactions on Intelligent Transportation Systems 19*, 5 (2017), 1537–1546. 2, 3

[vdEHBvW13] van den Elzen S., Holten D., Blaas J., van Wijk J. J.: Reordering massive sequence views: Enabling temporal and structural analysis of dynamic networks. In *2013 IEEE Pacific Visualization Symposium (PacificVis)* (2013), IEEE, pp. 33–40. 2

[VELW18] Vaizman Y., Ellis K., Lanckriet G., Weibel N.: Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 554. 1, 2, 9

[Vet18] Vetter C.: Circadian disruption: What do we actually mean? *European Journal of Neuroscience* (2018). 1, 2, 9

[VWL18] Vaizman Y., Weibel N., Lanckriet G.: Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1*, 4 (2018), 168. 1

[WCC*14] Wang R., Chen F., Chen Z., Li T., Harari G., Tignor S., Zhou X., Ben-Zeev D., Campbell A. T.: Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (2014), ACM, pp. 3–14. 1, 2, 8

[WHW*18] Wang W., Harari G. M., Wang R., Müller S. R., Mirjafari S., Masaba K., Campbell A. T.: Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2*, 3 (2018), 141. 1, 5