Momentary Stressor Logging and Reflective Visualizations: Implications for Stress Management with Wearables*

SAMEER NEUPANE†, University of Memphis, USA

MITHUN SAHA, University of Memphis, USA

NASIR ALI‡, University of Memphis, USA

TIMOTHY HNAT‡§, CuesHub, PBC, USA

SHAHIN ALAN SAMIEI, University of Memphis, USA

ANANDATIRTHA NANDUGUDI‡, University of Memphis, USA

DAVID M. ALMEIDA, The Pennsylvania State University, USA

SANTOSH KUMAR§, University of Memphis, USA

Commercial wearables from Fitbit, Garmin, and Whoop have recently introduced real-time notifications based on detecting changes in physiological responses indicating potential stress. In this paper, we investigate how these new capabilities can be leveraged to improve stress management. We developed a smartwatch app, a smartphone app, and a cloud service, and conducted a 100-day field study with 122 participants who received prompts triggered by physiological responses several times a day. They were asked whether they were stressed, and if so, to log the most likely stressor. Each week, participants received new visualizations of their data to self-reflect on patterns and trends. Participants reported better awareness of their stressors, and self-initiating fourteen kinds of behavioral changes to reduce stress in their daily lives. Repeated self-reports over 14 weeks showed reductions in both stress intensity (in 26,521 momentary ratings) and stress frequency (in 1,057 weekly surveys).

CCS Concepts: • **Human-centered computing → Field studies**; **Empirical studies in HCI**.

Additional Key Words and Phrases: Stress-tracking, Stressor-logging, Visualizations, Behavioral Changes, Stress Intervention, Emotion/Affective Computing, Wearable Sensors

---

*A version of this work appears in ACM CHI 2024.

†Corresponding Author

‡Work done while working at the University of Memphis

§Timothy Hnat and Santosh Kumar have financial interests in CuesHub, PBC, including stock ownership and management roles.
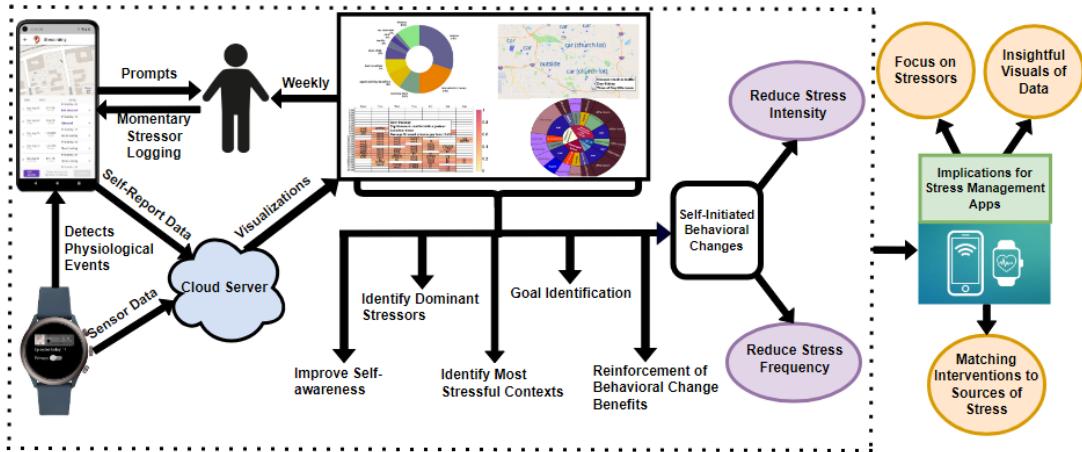

Fig. 1. Key steps in the MOODS study along with impacts on study participants and implications for stress management apps.

1 INTRODUCTION

In modern society, repeated or excessive exposure to stress is prevalent and it can have detrimental effects on physical, mental, social, and financial well-being [57]. Stress disrupts the body's homeostasis, triggering an overproduction of stress hormones such as cortisol and adrenaline [15]. This physiological imbalance elevates the risks of chronic diseases such as diabetes, hypertension, heart disease, and cancer [15, 17]. Stress also exerts a profound influence on mental well-being, elevating anxiety, depression, poor sleep, and fatigue [26]. The adverse cycle perpetuated by stress can drive individuals toward unhealthy coping mechanisms like unhealthy eating, substance use, or social isolation, corroding their overall quality of life [49]. The ramifications of stress extend to tangible consequences such as workplace incidents, absenteeism, and reduced efficiency [1]. Stress is estimated to cost over $400 billion in the United States [28] alone.

After two decades of research by the scientific community to detect stress in real-time using wearables [24, 54, 70], commercial devices such as those from Fitbit [62], Garmin [22], and Whoop [46] have recently introduced stress tracking and in-the-moment interventions prompted by the detection of physiological responses indicative of potential stress.

These new capabilities are being used to improve stress management in two ways. First, passively collected data are being used to create new visualizations for self-reflection that show stress arousal levels in different spatial and temporal contexts [40, 41, 56, 63, 71]. Qualitative evaluations have shown that these visualizations lead to new insights with a potential for self-initiated behavioral changes [41]. The therapeutic potential of such visualizations, however, is yet to be established as statistically significant changes in self-reported stress pre- and post-study were not found.

Second, the passive detection of stress has been used to deliver interventions at stressful moments [31]. Although the reduction in self-reported stress has been observed before and after some of these momentary interventions, only a negligible study-long reduction in stress has been found thus far [31, 48, 66]. We, however, note this is still an active area of exploration given that the passive detection of stress has emerged only recently and efficacious interventions can take decades to emerge, accepted, and widely used. Interestingly, participants in these studies wished the momentary interventions to match the specific source of their stress, i.e., stressor, at the current moment [31, 66].

In this paper, we investigate the *feasibility of collecting momentary stressors* and the *utility of integrating stressors in self-reflective visualizations in reducing self-reported stress* in a longitudinal study.

We address several challenges in the presented study — 1.) Although stress responses can now be detected automatically, most *stressors can't yet be identified from sensor data*. How should stressors then be collected? 2.) If self-reporting is used to collect stressors, when should participants be asked to report stressors? Participants may not have the *cognitive capacity* to engage in self-reporting at the time of stress [25] and the recall of stressors usually *fades from memory* over time [23]. 3.) How should self-reports be collected so that we can capture the *diversity* in stressors but *minimize the participant burden*? 4.) How should *self-reported stressors be integrated* into the visualizations of stress arousal data from sensors and the associated spatiotemporal contexts? 5.) How can visualizations provide *new utilities* with every week of additional data in a longitudinal study so participants see a benefit in continuing to collect data and overcome the accruing *burden, growing fatigue, and diminishing interest* over time observed in a longitudinal study?

To conduct our research, we developed a new smartwatch app, a cross-platform smartphone app, and a companion cloud software. We deployed them in a 100-day field study completed by 122 participants. During the study, called Mobile Open Observation of Daily Stressors (MOODS), participants received prompts when physiological events were detected by the smartwatch app, indicated by changes in pulse physiology measured at the wrist level. The prompts were delivered *soon after the detected event concluded* so that the memory of the stressor would still be fresh, but they would have recovered from the recent event. Participants were asked to rate their stress and log the stressor responsible (on the smartphone app) if they rated the event as stressful. We used *predictive autocomplete* to limit self-report burden. Each week, they received *new types* of visualizations that allowed them to dig deeper into their data. To assess change in self-reported stress, the intensity of stress was measured via momentary ratings, and the frequency of stress was measured via weekly surveys. Qualitative feedback was collected during the exit interview to understand the mechanisms behind changes in self-reported stress. Participant burden was assessed from app usage logs; utility was measured from app approval ratings. Importantly, no monetary incentives were provided, thereby emphasizing the intrinsic utility and burden of the app as the primary drivers of user retention.

Figure 1 provides a visual summary of this work. Our key contributions are to

(1) Show the feasibility of engaging participants to log their stressors upon the detection of a physiological event,
(2) Show that new visualizations of stress and stressors each week can keep participants engaged, and
(3) Show that momentary logging of stressors and visualizations of stress and stressors data have the potential to increase self-awareness, trigger self-initiated behavioral changes, and decrease self-reported stress.

2 RELATED WORKS

Our work investigates the feasibility of collecting self-reports of stressors triggered by sensor data, providing visualizations of stress for self-reflection, and assessing its impact on changes in self-reported stress. It builds upon and contributes to prior works on using sensor data to trigger self-report prompts, stressor logging, stress visualizations, and stress interventions. We discuss related work in each of these topics in the following.

2.1 Sensor-Triggered Prompting

Real-time detection of changes in human states such as physical activity or transitions between activities from wearable devices has been used to improve the timing of prompting for self-reports or interventions [3, 5, 20, 68].

For example, in a study to assess physiological correlates of emotions [7, 45], participants were asked in [29] to collect ECG, impedance cardiogram (ICG), and accelerometry data for 14 days. Whenever inter-beat interval (IBI) changed by

a threshold, participants were asked to rate their emotions. Our work uses a similar study design but generates prompts based on the detection of physiological responses, some of which are indicative of a potentially stressful event.

Furthermore, real-time detection of stress from physiological events [8, 13, 61] or from digital traces such as emails [31] makes it possible to deliver interventions in moments of stress [8, 13, 61]. Fitbit Sense 2 [62] recently introduced a *Body Response* feature that detects when heart rate, heart rate variability, skin temperature, and electrodermal activity indicate a potential stress event. It then prompts participants to log their mood and engage in an intervention of their choice to manage their stress in the moment. Our approach is similar but we wait until the conclusion of the physiological event when participants are likely to have recovered from the event but are likely to still remember the stressor.

2.2   Logging and Journaling of Stressful Events for Stress Management

Stressor is an event or situation, whether social or emotional (e.g., challenge, threat, opportunity) that triggers a stress response, i.e., deviation from homeostasis [15]. Successfully addressing stress usually requires activating coping mechanisms that directly address the responsible stressor [31, 66]. Therefore, to guide the development of appropriate stress interventions, many works have focused on understanding prevalent stressors and their role in PTSD [60], depression [51], anxiety disorder [11], heart diseases [53], asthma or rheumatoid arthritis [69], as well as in the general population [4, 27, 39]. These studies have shown that stressors predict daily symptoms and health outcomes.

Stressors have been collected at the end of day, via telephones [4], surveys [39], and smartphones [27]. More frequent collections (5-6 times daily) have resulted in more stressors reported, and hence are being adopted more widely [72].

It has also been shown that writing about stressful events has been shown to improve health, psychological well-being, physiological functioning, and general functioning [67]. This inspired many works on facilitating stressor journaling or even mood tracking to improve awareness. For example, the DeepMood application [64] directed participants to input their moods and activities into a smartphone app three times daily, with the goal of anticipating episodes of depression. As participants actively engaged in the mood-monitoring process through these apps, they observed an increase in emotional self-awareness. Some works have even combined logging with stress intervention. In [21], stress management techniques, gratitude journaling, a combination of both, and a control condition were investigated across four management course sections. The results indicated that students in the combined intervention and gratitude journaling groups exhibited increased levels of classroom engagement and perceived meaning.

As journaling involves a substantial participant burden, emerging technology has been increasingly leveraged to reduce participant burden [19]. In [30], data collected by a wearable was processed at the end of the day and detected stressful events were presented to the participants with an option to select from a list of 13 stressors to assist with the recall. Chatbots have been explored to engage participants in logging their moods for three weeks [37]. To reduce the burden further, a single item was developed to be delivered on smartwatches so participants can log stress ratings with a quick glance and a tap [35, 38].

Some commercial devices have also adopted these approaches. For example, Fitbit Sense 2 [62] smartwatch asks participants to rate their mood and engage in an intervention such as reflecting on their feelings and triggers upon detecting stress-related physiological events. The WHOOP Journal [46] allows participants to document their perception of stress levels, enabling them to track and analyze how they subjectively experience stress in the smartphone app.

Logging of stressors still involves two types of burden — the physical effort in entering the stressor and mental effort in recalling the stressor. We address both of these issues by prompting participants soon after the conclusion of a potentially stressful event to reduce the recall effort and by providing a predictive autocomplete to reduce the physical effort. We evaluate the former via weekly surveys and the latter via time taken to log the stressors from app usage logs.

2.3 Visualizations of Stress and Stressors

A large body of work exists on presenting visualizations of stress data collected from wearables to facilitate new insights that can improve stress management and user engagement. The majority of these works combine stress arousal [58] (sometimes expressed as body response [62], bodily reactions [56], arousal level [41]) with spatiotemporal context and self-reports [62]. These visualizations have been targeted at self-reflection by participants [40, 41, 56, 63, 71] and sometimes for experts' review [58]. In many cases, the visualizations are provided at the end of study [6, 41, 52], and in some cases, the participants had access to the visualizations throughout the data collection period [36, 40, 56, 71]. Most works targeted the visualization for private use by the participant [40, 41, 56], while some allowed visualization of data aggregated across multiple participants to be displayed in public places for reflection by the entire group or organization [71]. In [6, 36], partners were allowed to view the visualizations of each others' data together to understand the impact of each other's actions on the stress level of their partner. These works have shown the potential of stress visualizations to increase awareness, gain new insights, and improve stress management by promoting behavior change.

Prior works also introduced a wide variety of templates. They include calendar view [40], line, bar, and box plot [63], and pie chart [41]. We use the color coding of arousal levels obtained from sensors overlaid on calendars and bar charts to integrate spatiotemporal contexts with stress data from wearables, line graphs to show trends, box plots to show variability, and pie charts to show prevalence. We contribute to the body of works on stress visualization by proposing ways to integrate self-reported stressors with stress arousal and spatiotemporal contexts obtained passively. We target eight insights [14] and introduce a new protocol, i.e., new visualization types each week (for 14 weeks) that allows participants to gradually dig deeper as they collect more data each week.

2.4 Stress Interventions

As we are evaluating the impact of our visualizations on stress reduction, we review some recent works on stress intervention to report the magnitude of the study-long reduction in stress. Interventions can have a momentary reduction in stress as well as a reduction in study-long stress, with the latter indicating the potential of the intervention to have a lasting impact. A recent work [31] delivered both pre-scheduled interventions and stress-triggered interventions. They detected stress using email, calendar, camera, and heart rate. A chatbot prompted participants to practice one of three stress intervention exercises based on Cognitive Behavioral Therapy (CBT) and Dialectical Behavioral Therapy (DBT). In four weeks of using interventions, ($n = 86$) participants reported approximately 6.8% reduction in momentary stress post-intervention compared with pre-intervention (from 2.16 to 1.82 on a 5-point scale). However, no significant study-long reduction in overall stress was found.

In [66], 30 participants used a browser-based application to receive stress micro-interventions for four weeks. One experimental sub-group reported the highest reduction of 14.8% (mean reduction of 0.74 on a 5-point scale) in momentary stress between pre and post-intervention. However, no overall reduction in study-long stress was found in this study either. In a 30-day study with 47 participants [48], a calendar-mediated stressor anticipation application was deployed. It allowed participants to anticipate expected stressful events. Participants who executed a problem-focused intervention reported a 5.8% (-0.29 on a 5-point scale) reduction in stress between pre and post-event. Again, no study-long change in stress was found. A 28-day study with 65 participants [42] deployed a system that set adaptive goals in three coaching dimensions (Exercise, Meditation, and Accessibility). They reported a 22.5% study-long reduction on one stress measure but did not find any significant reductions on two other stress measures. A study [65] with 37 mental health professionals

(a) Smartwatch app home screen   (b) Smartphone app home screen   (c) Stressor annotation screen
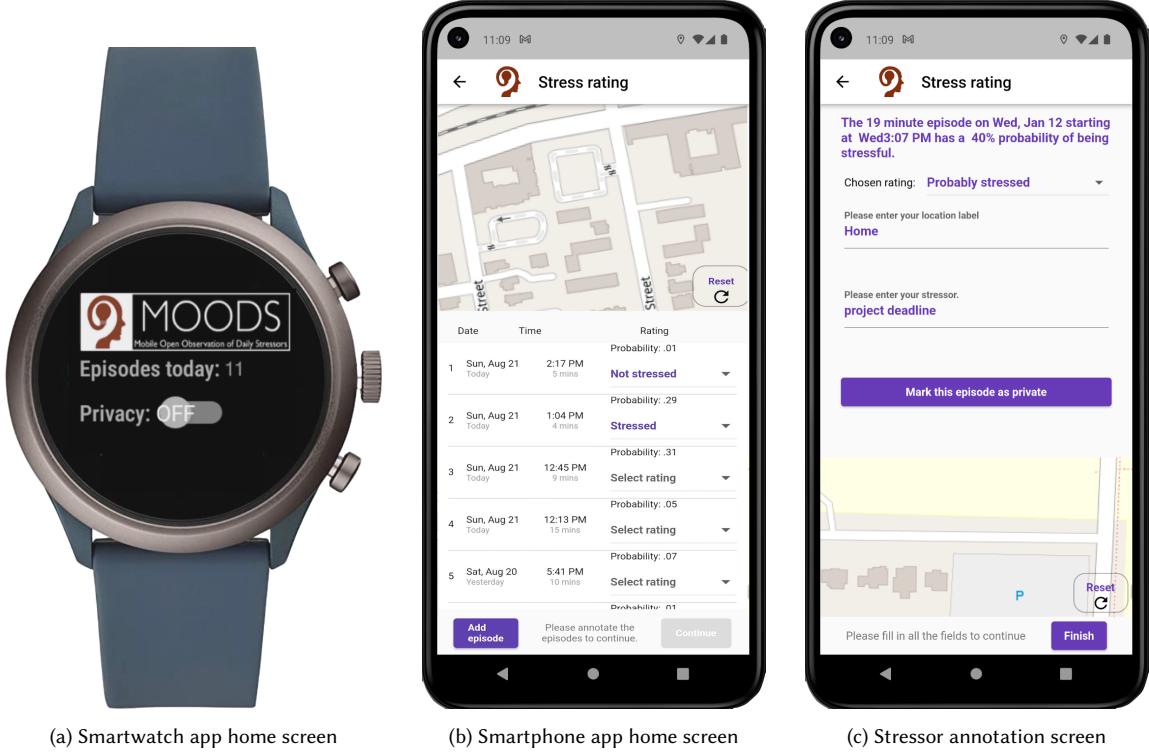
Fig. 2. MOODS app screenshots for events on smartwatch and stressor annotation, review, and edits on smartphone

who performed a 6-week meditation-based stress reduction (MBSR) program, reported a 7.6% reduction in study-long stress. In summary, study-long reduction in self-reported stress is usually low or negligible.

In contrast, our work investigates whether providing visualizations of self-reported stressors together with stress and associated spatiotemporal contexts as a personal informatics system [50] can quantitatively reduce self-reported stress and keep participants engaged in a longitudinal study.

3 METHODS

3.1 Design Rationale for the MOODS Study

Our apps (Figure 1) were designed to serve as a Stage-Based Model of Personal Informatics System [50]. Except for the *Preparation* stage when participants select what data they want to collect and using which tool (e.g., deciding to participate in our study), the MOODS Study aimed at helping participants with the remaining four stages. Many momentary stressors fade from people's memory soon after they are over unless it is a major event with a longer-lasting impact [23]. Many of these stressors may not be reported when asked later on. Our study was designed to help participants with the *Collection* stage by prompting them to easily log their stressors soon after their occurrence while limiting burden, as suggested in [48]. It helped them with the *Integration* stage by summarizing their data in visual forms. To help them with the *Reflection* stage, they were given the opportunity to review their stressor logs anytime and were presented with 16 different visualizations. These visualizations were meant to help them identify and execute
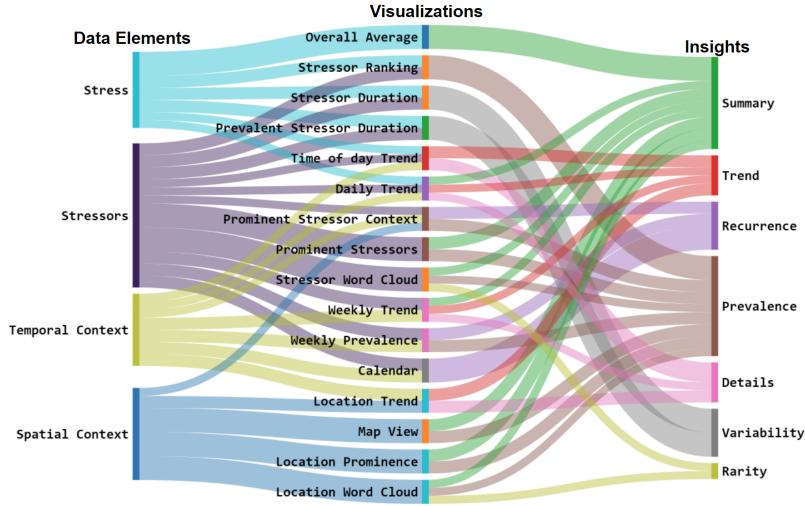
Fig. 3. Overall design of weekly visualizations, showing which visualizations (in the center column) used which data elements (left column) and aimed for what insights [14] (right column).

specific *Actions* to better manage their stress as well as to observe the impact of their self-initiated actions on their stress and stressors.

3.2 MOODS System Design

We developed an app for WearOS smartwatches, a cross-platform companion smartphone app, and cloud services.

3.2.1 Physiological Events, Stress Ratings, and Stressor Annotations. Similar to recent works that use commercial smartwatches running pre-trained proprietary models [2, 10, 18, 36], we use a pre-trained model (used in CuesHub smartwatch app [16]) to detect physiological events, indicative of physiological activation at the wrist level. Each event has a start time, end time, and duration. Similar to Garmin watches, the model we use processes wrist motion and Photoplethysmography (PPG) sensor data and produces a score between 0 and 100 that indicates the likelihood with which the current physiological event is considered to exhibit a stress response.

To limit participant burden, we selected only a subset of physiological events for prompting. We gave a higher weight to high scores as a key focus of our study was to collect momentary stressors. Specifically, all events with scores $> 95^{th}$ percentile were selected, whereas 80% of events in the $\{75^{th}, 95^{th}\}$, 10% of events in the $\{25^{th}, 75^{th}\}$, and 20% of events in the $\{0, 25^{th}\}$ percentile were randomly selected for an average of 3, 2, and 1 prompts per day, respectively. Our prompts cover all score ranges to obtain user-labeled sensor data in both stress and non-stress situations in real life.

Participants received a prompt on both their smartwatch and their smartphone. To determine which of the presented events were indeed stressful, participants were asked to choose a rating from the following Likert scale (*'Not stressed,' 'Probably not stressed,' 'Unsure,' 'Probably stressed,' 'Stressed'*). If the annotation was *'Stressed,' 'Probably stressed,'* or *'Unsure'*, the smartphone app prompted them to provide the semantic location and describe the likely stressor (see Figure 2c). Using a Likert Scale allowed us to also understand the intensity with which participants perceive different

momentary stressors. To reduce the burden, we incorporated a predictive text input module that displayed closest-matching stressors as participants typed in the input box. We prepopulated our database with 80 unique stressors (e.g., `traffic/transportation`, `work overload/demand`, `miscommunication`, `financial issues`, `interaction with boss, etc.`) from the literature [4]. This database was augmented, locally on the smartphone, with each newly-entered stressor to help participants annotate their specific stressors in a quick and consistent manner. To further assist them in recalling their stressors, participants were also shown the event location on a map along with the date, start time, duration, and the likelihood with which the app determined this event to be stressful (see Figure 2c).

Once participants completed their annotations, they returned to the stressor dashboard, which also served as the smartphone app's home screen (see Figure 2b). The dashboard displayed all previously logged events with the option to edit the stress rating, stressor description, or to mark the event as private, which excluded those events from any analysis. Finally, the dashboard allowed participants to self-report a stress event manually.

3.3 Weekly Visualizations to Increase Self-Awareness of Stress, Stressors, and their Surrounding Contexts

We designed 16 different visualizations from the stress ratings and stressor logs of participants. The goal of these visualizations was to help them self-reflect and gain 1.) better realization of stressful moments, 2.) better awareness of momentary stressors, and 3.) a better understanding of spatio-temporal contexts surrounding stressors.

3.3.1 Design Rationales for Weekly Visualizations. Figure 3 shows the overall design of our visualizations. They used the following data elements — 1.) Stress (ratings and duration), 2.) Stressors, 3.) Temporal Context (time of day, day of week), and 4.) Spatial Context (places and locations). They aimed to encode patterns over these data to facilitate the following insights — 1.) Summary, 2.) Trends, 3.) Recurrences across time and place, 4.) Prevalence, 5.) Details, 6.) Variability, 7.) Rarity, and 8.) Comparison (facilitated via interactivity in the visualizations). These insights are based on the eight categories of insights proposed for self-reflection during the *Reflection* and *Maintenance* stages of personal informatics [50] among quantified selfers [14]. We made two modifications — 'Prevalence' was used for 'Self-reflection' to discover dominant stressors and 'Recurrence' was used for 'Correlation' to discover associations of stressors with routine, place, or specific daily behavior. Figure 3 is a Sankey Diagram where *Data elements* and *Insights* are two types of source nodes and visualizations are the target nodes. Among *Data element* nodes, 'stressors' has the largest number of links (11), while 'prevalence' and 'summary' have the largest number of links among *Insight* nodes with 8 links.

Given a large number of combinations of data elements and insights (4×8) and diversity among them, we selected templates that suited each purpose. For example, for the overall summary, we used an angular gauge chart (as in fuel gauge indicators), sunburst chart to show the relationship among top stressors and their spatial and temporal context, a donut chart to show prevalence, a calendar overlay to show recurrence, line charts for trend, bubble chart to combine prevalence and trend, scatter plot for combining details with trend and prevalence, a combination of violin and box plot to show variability, and word cloud to combine prevalence with rarity. Excess abstractions in visualization sometimes hinder participants from gaining valuable insights from their data [55]. Hence, we utilized participants' own wording in most visualizations, making only minor spelling corrections. Where the input text was lengthy, we utilized abbreviations formed from the initial letters of each word to condense the information for visualization purposes.

All except three visualizations were interactive. Participants had the option to tap on any data point to unveil the complete form of abbreviations and access additional details, such as location, time of day, or day of the week associated with the stressors as shown in Figures 4b and 4c. Moreover, the colors were selected to be friendly for those with color blindness. Visualizations were generated using [34] and [32]. Finally, we ensured that the visualizations were responsive

(a) Prominent Stressors

(b) Map View

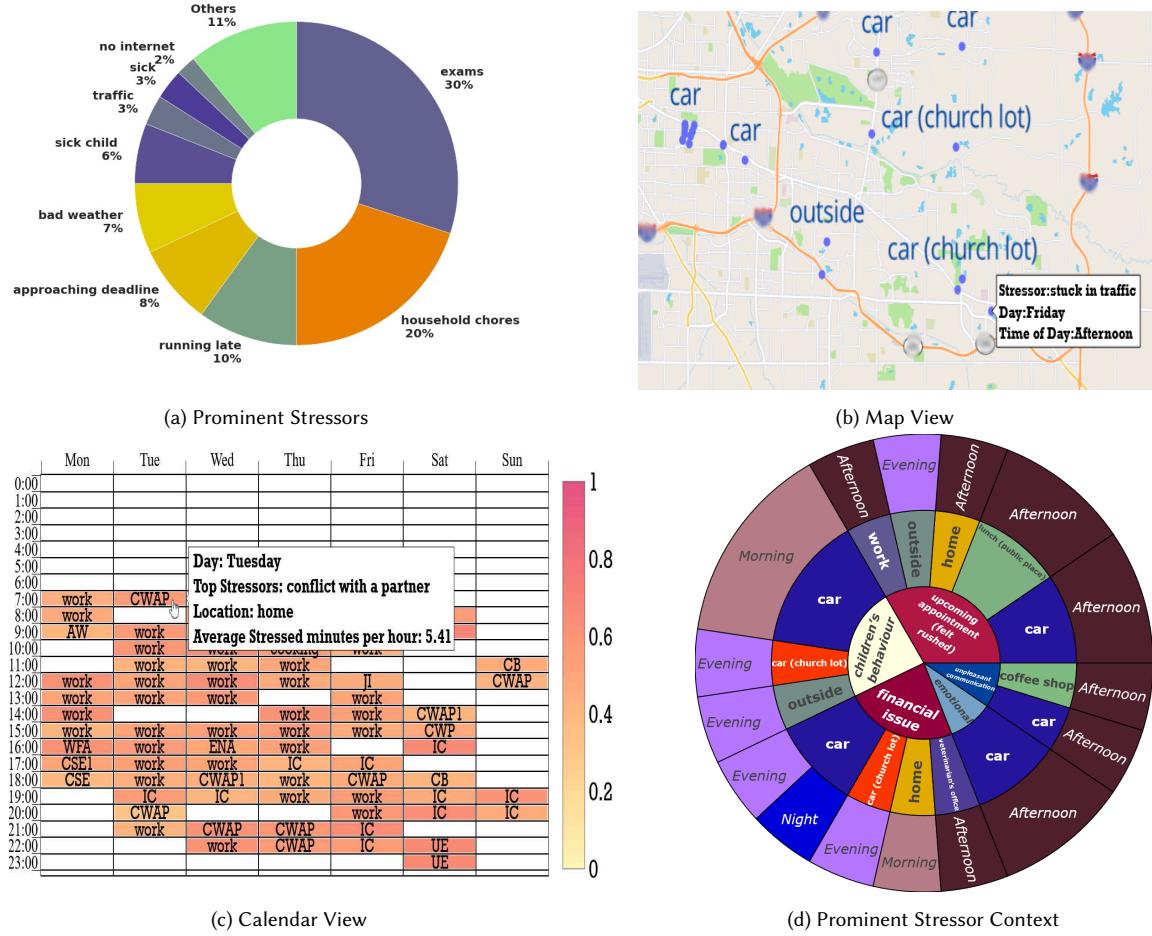(c) Calendar View

(d) Prominent Stressor Context

Fig. 4. Examples of 4 visualizations. Details on each visualization can be found in Supplementary materials.

for viewing on various screen sizes including smartphones, tablets, and desktops or laptops. Figure 4 shows four of these visualizations. More details about each visualization appear in the supplementary materials.

3.3.2 Chronological Order of Weekly Visualizations. After completing the first week of data collection, participants received two visualizations. A new visualization was added in each subsequent week. By incorporating cumulative data into each visualization, participants could see how their stress patterns evolved over time. The chronological ordering of the visualizations was largely driven by the sufficiency of data for each visualization to be meaningful. Hence, those providing an overall summary or top stressors were introduced early on, followed by those displaying prevalence, recurrence (e.g., calendar), trend, details of individual stressors (that are very sparse early on), and variability. Word clouds were saved for the end as they needed a large amount of data to display both prevalence and rarity. Figure 5a shows the order in which the 16 visualizations were introduced in the MOODS study over 14 weeks.

3.4 MOODS Study Description

3.4.1 Procedure. We obtained Institutional Review Board (IRB) approval for the MOODS study. Interested participants completed a brief screening survey via Qualtrics: Eligibility criteria included having United States residency, ability to freely and independently provide legal consent, ownership of a sufficiently modern smartphone (queried based on operating system version), willingness and ability to follow study instructions in the English language, regular and sufficient access to mobile internet service, understanding of shortened smartwatch battery life, and willingness to recharge the smartwatch throughout the day.

Participants selected for enrollment were sent an email providing personalized instructions for installing the MOODS study app on their smartphones, ensuring only selected participants could proceed with the enrollment process. Once installed, participants reviewed and completed the informed consent form via the smartphone app; participants were also provided a hyperlinked (PDF) document so that they could read the consent form on another preferred device. Participants who downloaded the MOODS study app but did not consent were afforded no further app functionality; all participants were informed they could uninstall the app with no penalty or further obligation to the study.

Consented (enrolled) participants completed a brief stress awareness assessment via the smartphone app. Participants also provided their demographic information (i.e., age, gender, occupation) and their shipping address. We then shipped them a Fossil Sport smartwatch and the instructions to configure the smartwatch with their smartphone via the MOODS study app. Every week, participants were sent weekly visualizations of their stressor patterns via their study-provided email and asked to complete brief weekly questionnaires via the smartphone app.

No ongoing or end-of-study incentive was provided, to simulate real-life usage. At the conclusion of the MOODS study, participants were allowed to retain the smartwatch. They were asked to collect data for a period of 100 days, however, they could exit the study at any time without returning the smartwatch. Even though some participants continued using the apps for up to 173 days, we limited our analysis to 14 weeks of data to maintain consistency and alignment with the planned study duration. Upon completion, participants received an email that included an invitation to either complete an end-of-study interview (virtually) or complete a questionnaire (via Qualtrics). Any participant who completed 100 days of data collection or was deemed to have stopped participating after completing at least 30 days of data collection was invited for an exit interview.

3.4.2 Exit Interview. During the exit interview, we collected participant feedback covering various aspects, including their interactions with the MOODS study apps, experiences with visualizations, and any goals or behavioral changes initiated throughout the study. Participants shared details on their preferred hand for wearing the smartwatch, devices used for reviewing weekly visualizations, potential interference of study apps with daily activities, and their overall preferences and dislikes concerning the app. Ratings were gathered regarding the utility of weekly visualizations in understanding stress patterns, visualization preferences, and the reasons behind participant likes and dislikes. Participants were also asked about any self-improvement goals established before or during the MOODS study and whether those goals were achieved. Additionally, inquiries were made into behavioral changes implemented during the MOODS study, participants' awareness of stress and stressors, their confidence in sustaining behavioral changes, and the likelihood of seeking out devices or apps for continued stress monitoring. Notably, participants were not directed to set goals or make behavioral changes pre or during the study; any such initiatives were self-initiated. The goal was to assess whether stressor logging with self-reflection aided participants in achieving their goals and managing stress through behavioral changes.

3.4.3   Participants. MOODS apps were tested by 22 pilot participants for 12 weeks to debug and improve the apps and study procedures for long-term remote operation. For the main phase, we began our recruitment by sending IRB-approved email invitations to academic groups, posting messages in professional media (e.g., LinkedIn), and posting flyers in various organizations with the intention of obtaining a diverse participant pool across occupation, geographical, and demographic attributes. We sent enrollment emails to 302 participants out of which 183 enrolled via the MOODS study app, and 143 were able to successfully configure the smartwatch with the smartphone. Out of 136 participants who began collecting data, we report data from $n = 122$ (39 men, 77 women, and 6 non-binary) who annotated at least one stressor. They come from 13 different U.S. states. Of these, 71 were professionals in occupations such as academia, information technology, administration, public safety, health, legal, sales, and construction. Fifty (50) were students pursuing higher education, and one participant did not disclose their occupation. The average age of participants was $38(SD = 13)$ years. We use '*SD*' to denote one standard deviation. The majority of participants ($n = 78$) were iPhone users. Among the 46 participants who participated in the exit interview, 7 chose Zoom, 3 opted for phone, and the remaining completed a Qualtrics survey. For the rest of the paper, we use *P#* to associate specific participants with their testimonials.
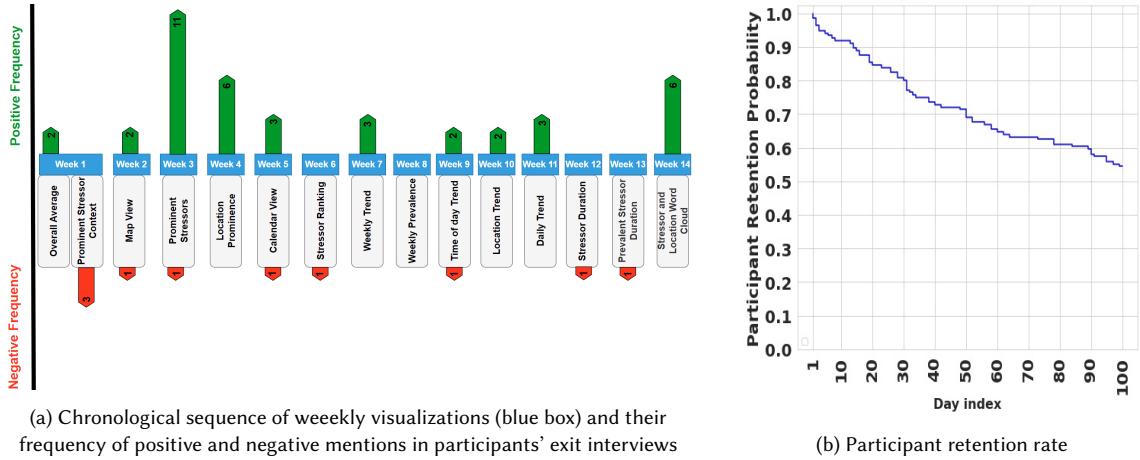
3.5   Stress Metrics and Statistical Tests

3.5.1   Stress Metrics. Participants rated their stress multiple times daily when prompted. These stress ratings were converted to a scale of 0 to 4, with 0 representing *Not Stressed* and 4 representing *Stressed.* These ratings provide us an opportunity to analyze the intensity of those prompted events, which we call *Stress Intensity*. Additionally, each Sunday morning, participants were asked to indicate the number of times they experienced stress in the past week each day, with four response options: `At most once`, `More than once but at most twice`, `More than twice but at most three times`, and `Four or more times`. These responses were mapped to numerical values ranging from 1 to 4. We use this value as a metric of *Stress Frequency*. We use weekly surveys instead of stressor logs for frequency as not all stress occurrences can be captured via prompts, e.g., when the watch is not worn, no prompts are issued.

3.5.2   Statistical Tests. To determine trends in self-reported measures of stress intensity and frequency, we used the Mann-Kendall Trend test [33], which is commonly employed for analyzing consistently increasing or decreasing trends in time series data. In the rest of the paper, we use MK test for the Mann-Kendall test, "m" to denote the coefficient of slope, and "b" for the coefficient of intercept. We used the linear mixed effects model (LMM) for regression analyses of stress intensity and frequency with fixed effects (constant intercept and slope (with respect to week) across all participants) and two random effects for each participant; a random intercept and a random slope (with respect to week). To assess the differences within or between groups of participants, first, the Shapiro-Wilk test was performed to check for the normality of the data. Since most data from this study did not fit a normal distribution, we used the non-parametric Mann-Whitney U test for independent samples and the Wilcoxon signed-rank test for dependent samples for statistical comparisons. We used Python and R to process the data and perform the statistical analyses.

4   RESULTS

We start by analyzing participant retention and data contribution to demonstrate the feasibility of sensor-prompted logging of stressors in a longitudinal study. Next, we present the like and dislike mentions for comparative evaluation of the 16 visualizations. We then statistically evaluate the effects of momentary stressor logging and visualizations on stress intensity and frequency. To understand the mechanisms behind a reduction in self-reported stress, we analyze

(a) Chronological sequence of weeekly visualizations (blue box) and their frequency of positive and negative mentions in participants' exit interviews

(b) Participant retention rate

Fig. 5. (a) Chronological sequence of weeekly visualizations (blue box) and their frequency of positive and negative mentions in participants' exit interviews and (b) Participant retention

increases in self-awareness, initiation of behavior changes, and its impact on stress. Finally, we analyze the physical and mental effort involved in logging stressors and the approval ratings of the MOODS study apps to assess their utility.

4.1 Participant Retention and Data Collection in the MOODS Study

To understand how engaging the sensor-prompted stressor logging and weekly visualizations were for the study participants, we analyze the participant retention in our study. Figure 5b shows the survival plot depicting participant retention in our study. Out of 136 participants who began collecting data, 110 remained active in the study for at least 30 days, resulting in a 30-day retention rate of 81%.

On days with at least one prompt, participants received an average of 5.2 prompts per day, for a total of 35,981 prompts. Participants responded to 3.86 (i.e., 74%) prompts per day, resulting in 26,732 stress ratings. Participants marked 211 events as private. An average of 1.62 stressors were reported per day, resulting in 11,222 stressors that consisted of 1,476 unique stressors. On average, each participant logged $83 (SD = 76.5)$ stressors and an average of $21 (SD = 17)$ unique stressors. The ten most reported stressors were *anxiety*, *approaching deadlines*, *heavy traffic*, *too much work*, *acute or chronic pain*, *work*, *unsure*, *playing video game*, *research*, and *unpleasant conversation*. A total of 1,057 weekly surveys were completed.

4.2 Comparative Evaluation of Weekly Visualizations

In their weekly survey, participants were asked what impact the most recent visualization had on them. As new visualizations were introduced each week, we could potentially use their responses for evaluating each visualization. However, the visualizations were cumulative (for the benefit of participants) in that all previously introduced visualizations gradually became more useful as they showed the cumulative data, presenting ambiguity in using this method of evaluation. We, therefore, use their exit interview responses. Participants were asked about their preferences regarding the visualizations used in the MOODS study. They were given an open-ended question to express which visualizations they liked the most and the least. We examined the frequency of mentions for each visualization and categorized them as positive or negative. Figure 5a shows the frequency of these mentions for each visualization (green and red bars

(a) Average weekly stress intensity   (b) Average weekly stress frequency
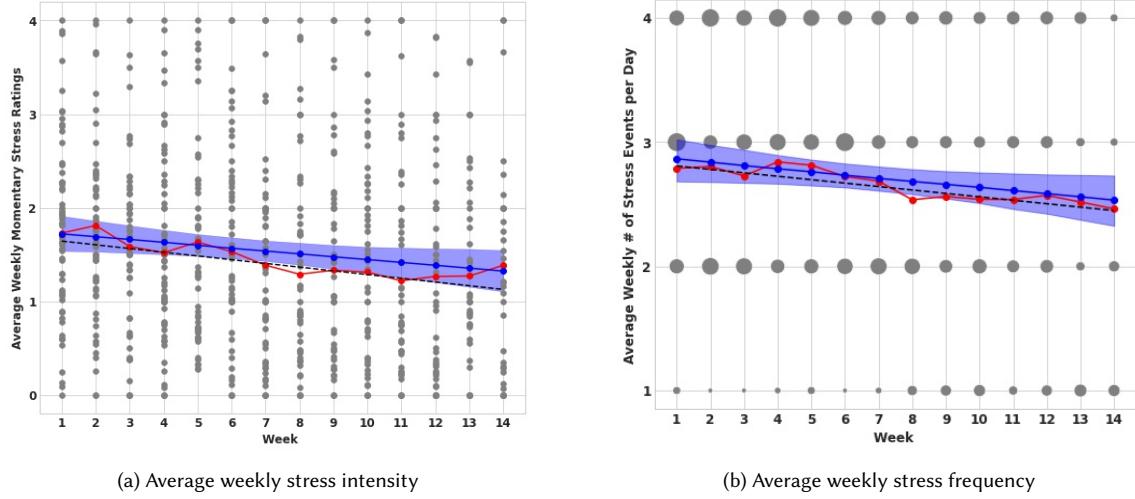
Fig. 6. Impact of self-reflection on the intensity and frequency of self-reported stress during the 14-week study. Red dots represent the population average for each week, the dashed lines represent the trend line, and the band around the blue line (population level LMM fit) represents the 5th and 95th percentiles calculated via bootstrap using resampling and replacement.

indicate positive and negative mentions respectively). Four participants mentioned that they liked all visualizations without mentioning a specific one. Hence, their responses are not included in the positive mentions of any visualization.

We now present some participant comments expressing why they like some visualizations more than others. Participants preferred visualizations that were *easier to understand*. P41 said that *"The map view (Figure 4b), prominent stressors (Figure 4a), location prominence, and calendar (Figure 4c) were visually easy to understand. The prominent stressor context (Figure 4d) was difficult to understand."* P4 mentioned that *"I liked the word cloud, as I'm a language person and it was visual and easy to understand. The prominent stressors graph was helpful."*

Another factor was related to *readability*. P26 said *"I really liked word cloud, especially bold for prevalent stressors."* *Readability* was mentioned by other participants, especially when it presented a challenge. P27 said *"Schedule grid hard was to read for calendar view."* P10 said *"I did not like the prominent stressors graph because many of the acronyms did not make sense, and I did not know what it was referring to."* Similarly, P16 said *"Sunburst chart (Prominent Stressor Context) was difficult because of small text."*

A third factor cited related to the use of *colors*. P8 said *"I noticed patterns across time of the day. Example: trouble sleeping (stressor). Chart with the different colors helped with that."*

Although there were some common themes that made some visualizations more or less liked by many participants, offering a choice of 16 visualizations that showed different aspects of their data and aimed at eight different insights helped many participants (with differing needs and preferences) find something they could benefit from. One participant (P28) even mentioned explicitly, *"It was awesome to have so many options for visualizations presented."*

4.3 Reduction in Study-long Self-reported Stress

We computed weekly averages to measure the *Stress intensity* for the past week for each participant. Grey dots in Figure 6a represent the weekly *Stress intensity* average for different participants over the 14-week study. The red dots represent the average value of *Stress intensity* across all participants for each week. The size of grey dots in Figure 6b

represents the proportion of participants who selected a specific value of *Stress frequency* in their weekly response for that week. The red dots represent the average value of *Stress frequency* across all participants for each week.

To assess whether the weekly averages of *stress intensity and frequency* (in Figures 6a and 6b) show an increasing or decreasing trend over the 14-week period, we used the MK test. We find that there is a statistically significant decline in the trajectory (depicted by dashed lines) for both *stress intensity* ($b = 1.64$ and $m = -0.039^*$, $p < 0.05$) and *stress frequency* ($b = 2.81$ and $m = -0.027^*$, $p < 0.05$), where b represents the intercept and m the slope. The estimated study-long decrease over the 14-week period in *stress intensity* and *stress frequency* are about 11% and 9.5% respectively.

We next check whether these decreases may be due to initial elevation bias and subsequent decline bias as has been observed in repeated measures [59]. Initial elevation bias refers to the tendency of participants to report higher levels or intensity of a target state in the initial survey (first few reports) than they actually experience, while decline bias refers to the tendency to report lower levels of intensity in subsequent reports (after the first few reports). In their study, the authors concluded that initial elevation bias is more pronounced than later decline bias for negative mental states in longitudinal studies of thoughts, feelings, and behaviors. To explore whether the decreasing trend in stress ratings we observed was influenced by initial elevation bias, we excluded the measurements from the entire first week for both *stress intensity and frequency*. Even after this adjustment, the MK test still revealed a significantly decreasing trend. Therefore, we can conclude that the initial elevation bias did not account for the reductions observed here.

In the MK test, we used all data from all participants. However, there is wide variability between participants in self-reported stress. To check whether the trends persist when the self-reported stress measurements from each participant are analyzed separately, we employed linear mixed-effects models (LMM). We incorporated both fixed and random effects to account for the inherent participant variability. We excluded participants who had less than 5 weeks of survey ratings and participants who provided the same responses to all of their weekly surveys, i.e., had no variability. This left us with 63 participants. For *stress intensity*, $b = 1.76(\pm0.76)$ and $m = -0.03^*(\pm0.062)$ ($p < 0.05$) for fixed effects. For *stress frequency*, $b = 2.86(\pm0.74)$ and $m = -0.022^*(\pm0.066)$ ($p < 0.05$). Although decreasing trends were statistically significant for both *stress intensity* and *stress frequency*, they were not uniform across all participants, as the random effects were statistically significant for each measure.

To see how the slope and intercept obtained in the LMM compare with that obtained via the MK test, we conducted the MK test for the same 63 participants that were analyzed in LMM. The MK test yields ($b = 1.61$ and $m = -0.041^*$) ($p < 0.05$) for *stress intensity* and ($b = 2.84$ and $m = -0.029^*$) ($p < 0.05$) for *stress frequency*, showing a similarity in both the slope and the intercept. We conclude that the participants reported a decreasing level of stress over time.

To add some meaning to the magnitude of reduction in self-reported stress, we compare the data from the first week to that from Week 14. For *stress intensity*, we see from the population-level fit (blue line) in Figure 6a that the average self-rating of momentary stress reduced from being close to *Unsure* (i.e., 1.72) to *Probably Not stressed* (i.e., 1.32).

For *stress frequency*, the population-level fit (blue line) in Figure 6b shows that the frequency of stress events shows a reduction of 0.33 (=2.86-2.53) per day, resulting in 10 fewer stress events per month.

4.4 Potential Explanations for Overall Reduction in Stress

4.4.1 Changes in Self-Awareness of Stress and Stressors During the MOODS Study. In a pre-study survey, the participants were asked: `What motivated you to join the MOODS?`, and to select one or more from five options. Figure 7a shows the distribution of their responses. Three responses received over 70% mentions. Two of these were related to improving self-awareness and the third was to help advance science.
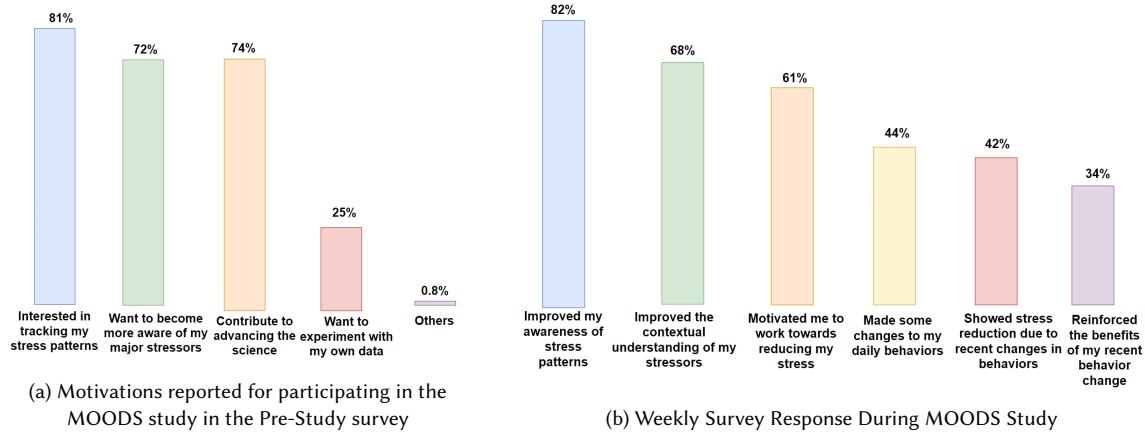
(a) Motivations reported for participating in the MOODS study in the Pre-Study survey

(b) Weekly Survey Response During MOODS Study

Fig. 7. Pre-Study and Weekly Survey Response

At the end of each week, the participants were asked: `What impact did the most recent visualizations have on you?`, and to select all that apply from the options shown on the Figure 7b, and `None`. A large majority of participants (82%) selected either the `Improved my awareness of stress patterns` or the `Improved the contextual understanding of my stressors` option. Among participants interested in tracking stress patterns (from their pre-study survey), 85% reported improved awareness, while for those aiming to understand stressors, 74% reported enhanced contextual understanding of their stressors.

This was also supported by comments from participants in exit interviews. *P*8 said *"Eye opener. I was aware before study that often stressed but now after seeing stressors and patterns, it was insightful."* *P*18 said *"I was kinda shocked to see the amount of stressors I had."* *P*4 said *"The study made me more conscious of what stresses me out and how often I feel stressed."* *P*10 mentioned *"It helped me to see how anxious I was getting over minute details."* *P*16 said *"I was more stressed in the Mornings during the week and big peak around Monday and tapered off throughout the week."* *P*41 mentioned, *"Increased mental awareness to what repetitively caused my stress."*

4.4.2 Motivation, Action, and Reinforcement towards Self-initiated Behavior Change. We first quantitatively analyzed self-report data to see how many participants reported motivation, taking action, and receiving reinforcement toward behavior change. In the weekly survey, participants reported the impact the recent visualizations had on them by selecting a response from the options shown on x-axis labels of Figure 7b, and None. Figure 7b shows the percentage of participants who reported the specific response in one or more of their weekly surveys. We observe that 61% of participants mentioned being motivated to work towards reducing their stress. Next, we observe that 44% of participants mentioned taking specific action to reduce their stress. Finally, 42% and 34% of participants respectively mentioned that the visualization helped them see the reduction in stress due to their recent behavior change and reinforced the benefit of their recent behavior change (both referred to as *Reinforcement*).

We next analyze the weekly trend in changes in the proportion of participants who reported taking an *Action* and receiving *Reinforcement* (Figure 8a). We find a significantly increasing trend (via the MK test) for *Reinforcement* ($b = 13.67$, $m = 0.99^*$, $p < 0.05$) and a non-significant, but increasing trend for *Action* ($b = 16.79$, $m = 0.63$, $p = 0.22$).
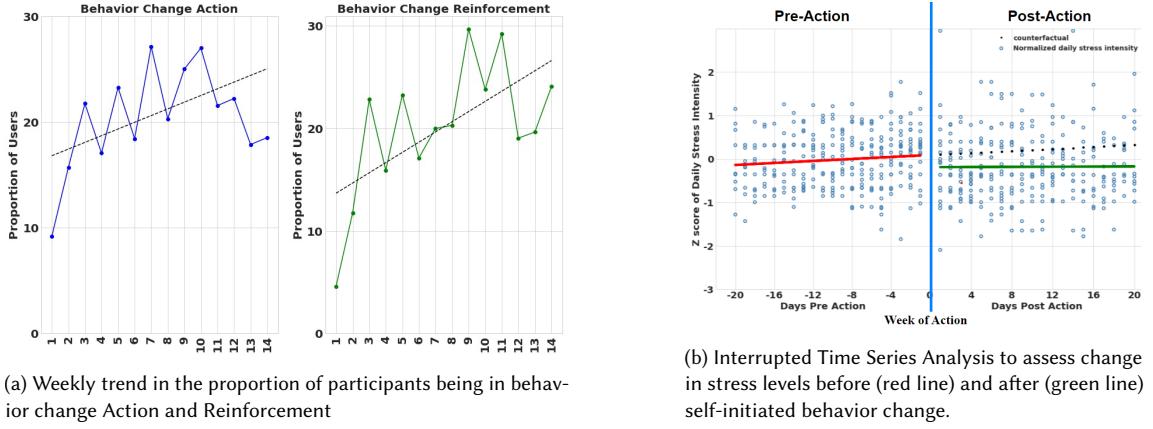
(a) Weekly trend in the proportion of participants being in behavior change Action and Reinforcement

(b) Interrupted Time Series Analysis to assess change in stress levels before (red line) and after (green line) self-initiated behavior change.

Fig. 8. Weekly Trend of Behavior Stages and Comparison of Daily *Stress Intensity* Before and After Initial Action

For participants who completed the exit survey, a significant increasing trend was found for both the *Action* ($b = 13.71$, $m = 0.64$, $p = 0.08$) and *Reinforcement* ($b = 12.79$, $m = 1.22^*$, $p < 0.05$).

We next present the four different ways that stressor logging and weekly visualizations motivated participants to move toward or reinforce behavior changes with the corresponding quotes.

1. Goal Identification: Some participants identified what to work on. *P28* said *"I found the simple visualizations that highlighted the data around the frequency of stressors, day of the week, and location to be great in identifying patterns and areas of need."* *P7* said, *"The weekly visualizations I received towards the beginning of the study were the ones I liked the most because it was very interesting to see my stress patterns on visual graphs for the first time. Toward the end, they helped me keep track of what factors were causing the most stress so I could focus on reducing them."* *P9* said, *"I started writing down things to track stressors."* *P42* highlighted, *"While I was fairly certain that I had stressors occurring in my daily life, this app helped me to acknowledge and attempt the management of those stressors. It also revealed other stress factors that I had not considered before which also informed me of how they could be contributing to a cumulative feeling of stress."*

2. Identify Dominant Stressor(s): Some participants learned about their dominant stressors. *P26* said *"I realized my biggest stressor was approaching the deadline and tried not to procrastinate as much and work on it."* *P8* said *"I noticed patterns across the time of the day. Example: trouble sleeping (stressor)."* *P4* mentioned *"Although it was hard to confront as I realized that most of my stress is caused by other people (interpersonal conflict), which is pretty annoying."* *P25* said *"Change or communicate with my advisor about being responsible for most of my stress."*

3. Identify Most Stressful Contexts: Some participants learned about their most stressful contexts. *P27* said, *"I realized I was most stressed while working and became more aware of certain situations where more stressed especially the meetings."* *P31* said *"I identified some stressors that I can work on to reduce, especially social contexts."*

4. Reinforcement of the Behavioral Change Benefits: The visualization helped some participants reinforce their behavioral changes by showing gradual stress reduction. *P7* said *"I noticed a decrease in my stress on the road and behind the wheel!"*. *P8* said, *"I noticed less stressed episodes towards the end compared to the beginning of the study."* *P10* said *"I loved the weekly trends because I could see if what I have been doing has helped in the long run."*

4.4.3 Types of Self-Initiated Behavior Changes. During the exit interview, we asked participants to describe any behavioral changes they made during the study. They reported making fourteen (14) behavioral changes that we broadly

Table 1. Fourteen (14) Self-Initiated Behavior Changes Reported by Participants in Exit Interview

Groups	Behavior Changes (14)	Supporting Quotes from Participants
Emotion Regulation	Staying calm	*P43 : "I tried to stay calm during situations or try to meditate through them."*
	Becoming less reactive by controlling their emotions	*P12 : "I am now controlling emotions in certain situations which I never thought earlier."*
	Anticipating stressful situations	*P12 : "I anticipate and think about the stress levels in advance while going through some situations."*
	Breaking away from stressful situations	*P42 : "Most of my self-improvement goals were to find moments to breakaway from stressful situations when they happen or to pre-plan physical activities that would be after a known stressful situation."*
Enhanced focus, Better planning, and Improved productivity	Slowing Down	*P8 : "I used to rush in work but now in the last month, I have started to slow down."*
	Better planning	*P26 : "I started taking more time to plan the day more so as to reduce stress."* *P35 : "I made changes to my schedules, particularly on Thursdays, to improve my daily organization and planning."*
	Goal setting	*P13 : "I set myself a goal to be more consistent with my work. I could have done better but I did improve a little."*
Better Self-care and Making Healthier Choices	Eating regulation	*P9 : "I started writing down things to track stressors, installed an app to document eating."*
	Taking more walks	*P4 : "I started working out more and taking walks to increase my step count."*
	Improving sleep	*P30 : "I tried to lower the amount of stress for myself at home. I got more sleep and made caring for myself a priority."* *P10 : "My goal was to make sure I did not stress as much with smaller details. I did my best to get adequate sleep , which lowered my stress."*
	Taking more breaks between work	*P28 and P8 : "I started taking more breaks between work through the day."*
	Deep breathing	*P28 and P13 : "I started practicing deep breathing more frequently throughout the day."*
	Meditation	*P42 : "Most of my behavioral changes centered around taking more walks or providing time to stop what I was doing and to breathe or meditate to settle out of a stressful situation."*
	Quit smoking	*P23 : "I quit smoking."*

classify into three groups — 1.) Better regulating emotions before, during, or after stressful situations, 2.) Improving productivity via enhancing focus and better planning, and 3.) Taking better self-care via developing healthier habits. These are described in Table 1, along with supporting quotes from participants.

4.4.4   Impact of Self-initiated Behavior Change on Self-reported Stress. As previously noted, approximately 44% of participants in the MOODS study reported making self-initiated changes to reduce their stress. We subsequently focused on examining whether these actions had a noticeable effect on reducing their daily stress levels. To do this, we selected those participants who initiated such actions for the first time between the fourth and eleventh week of the study and participants who had stress intensity ratings for at least one week before and one week after adopting these stress-reducing actions. This process resulted in a subset of 17 participants meeting these criteria. On average, these participants began their stress-reduction actions in the sixth week, with a standard deviation of 2 weeks.

An MK test on the weekly *Stress Intensity* data of these participants yields $b = 1.81$ and $m = -0.03$, where the intercept is slightly higher and the slope slightly lower than the population. To assess the effect of their self-initiated change on their stress ratings, we treated the first week they mentioned taking an action as the week of their intervention (action week). We organized the time series data of each of these participants so that their pre-action and post-action weeks are aligned. To account for variations in stress intensity among participants, we normalized individual stress intensity scores by converting them into z-scores for the purpose of this analysis. We calculated a z-score for each individual's overall study stress intensity. We excluded the action week's data from this analysis. Figure 8b shows the *interrupted time series* analysis, a quasi-experimental design to analyze data over an extended timeframe before and after an intervention to assess the impact of the intervention [43] in natural experiments and studies without randomization. These effects can manifest as alterations in the time series trend or shifts in the intercept following the intervention. The y-axis of Figure 8b depicts normalized daily stress intensity scores. These scores are depicted from the 20th annotation day before the action week on the left side of the blue line. Additionally, the normalized stress intensity scores from after the action week up to the 20th annotation day are shown on the right side of the blue line. Dotted

(a) Proportion of transportation-related stressors (blue) and stressors reported during commute (green) for *P*38

(b) Proportion of work-related stressors for *P*8

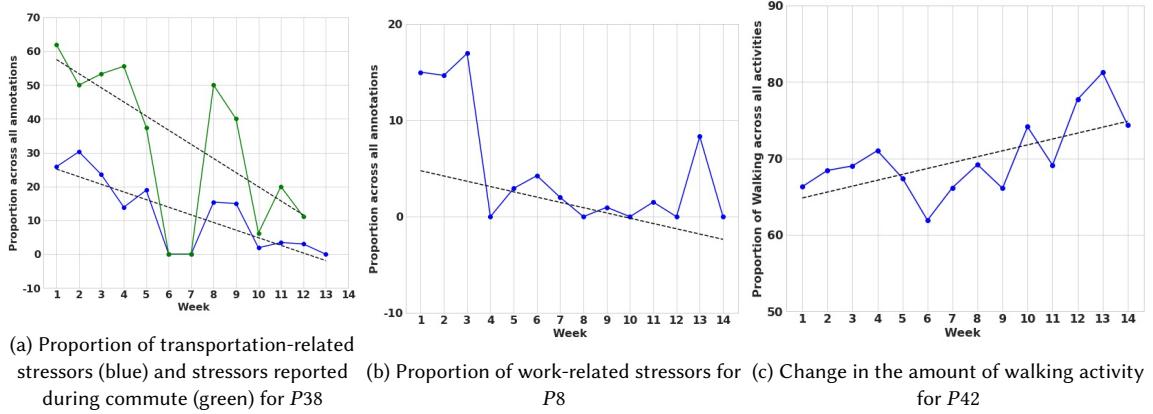(c) Change in the amount of walking activity for *P*42

Fig. 9. Data supporting evidence of specific behavioral changes reported by three participants during exit interview

lines represent counterfactual (stress levels had the self-initiated behavior change not happened.) Although the MK test showed a decreasing trend, we notice that the majority of the decrease is concentrated in the pre- to post-action time period, i.e., slope was increasing in pre-action stage and the slope is close to zero in post-action stages, with a statistically significant reduction (-0.278^*, $p < 0.05$) from pre- to post-action.

4.5 Baseline Stress Differences Between Participants with and without Self-Improvement Goals or Actions

During the exit interview, we asked participants if they had set any self-improvement goals. Out of 46 participants who completed the exit interview, 18 reported having set self-improvement goals either before or during the MOODS study. We also asked participants if they had implemented any behavioral changes over the course of the study. Out of 46 participants, 24 acknowledged making behavioral changes as a result of their study involvement. Among the cohort of participants who had self-improvement goals before the study commenced, 88.8% actively engaged in making behavioral changes throughout the study duration. In contrast, out of the participants who did not report any goals prior to the study, only 29% reported making behavioral changes during the study period.

We wanted to explore whether participants who had established goals or made behavior changes reported higher or lower stress levels at the outset compared to participants without any goals or those who did not report making any behavioral changes during the study. Since baseline stress measurements were not initially available, we estimated a baseline period for analysis. This baseline period was determined by identifying the week at which participants reported significantly different stress levels in comparison to the initial week(s). By employing the Wilcoxon signed rank test, we conducted a statistical comparison between the stress intensity levels of the first week and those of subsequent weeks. We find a significant reduction in stress intensity by the fourth week (with mean values of 1.81 in the first week and 1.57 in the fourth week, $p < 0.05$). Consequently, we identified the baseline period as spanning the first three weeks. Thus, we define baseline stress intensity for each participant as the average stress intensity in the first 3 weeks. We employed the Mann-Whitney test to comparatively assess baseline stress intensity among different participant groups based on whether they had established goals prior to the study or had pursued behavioral changes during the study.

4.5.1 Do participants who reported having specific self-improvement goals report a higher baseline stress? We examined the average baseline stress intensity among participants who had set goals before participating ($n = 18$) and those who

had not ($n = 28$). The results indicated that participants who had established goals prior to the MOODS study exhibited an average baseline stress intensity of 1.83 vs. 1.67 for those who did not set any goals. Although not statistically significant ($p = 0.32$), participants with pre-established goals tended to report a slightly higher baseline stress intensity.

4.5.2 Do participants who reported making behavioral changes during the MOODS study report a higher baseline stress? In the examination of baseline stress intensity among participants, it was found that individuals who made behavioral changes during the study ($n = 24$) exhibited an average baseline stress intensity of 1.89 vs. 1.56 for those who did not make any behavioral changes. Although not statistically significant ($p = 0.15$), participants who engaged in behavioral changes reported a slightly higher baseline stress intensity.

4.6 Illustrations of Data Supporting Behavioral Changes Reported in the Exit Survey

During the exit interview, we asked participants for more details if they mentioned having a goal or making behavior changes. We analyze how well these changes are reflected in the data of some participants for illustration purposes.

P38, whose most prevalent stressor was transportation-related, mentioned *"I had a goal of being less stressed during their commute by utilizing deep breathing and being cognizant of my listening choices on the drive."* Figure 9a shows the weekly proportion (percentage) of transportation-related stressors reported by this participant. The blue line represents the proportion across all annotations related to transportation stressors, while the green line specifically represents the proportion of annotations made during their commute. The commute data was extracted from the Google Activity API. On Weeks 6 and 7, no transportation-related stressors were reported (Figure 9a). MK test shows a significantly decreasing trend in the weekly proportion of transportation-related stressors across all annotations ($b = 27.42$, $m = -2.25^*$, $p < 0.05$) and for stressors reported during the commute ($b = 61.73$, $m = -4.47^*$, $p < 0.05$). This indicates that as the weeks progressed, this participant reported experiencing fewer transportation-related stressors.

P8, whose most prevalent stressor was work issues, stated *"I tried to work a little less and take more breaks during work, and as a result, noticed fewer episodes of stress towards the end of the study compared to the beginning."* Figure 9b shows the weekly proportion of `work issues` stressors across all annotations. The MK test shows a significantly decreasing trend ($b = 5.32$, $m = -0.54^*$, $p < 0.05$) in the frequency of `work-related` stressors over time.

Figure 9c shows the weekly walking trend of P42 who stated *"My primary behavioral changes revolved around increasing my walking activity."* As clearly depicted in the figure, there is a noticeable and significantly upward trend ($b = 64.05$, $m = 0.77^*$, $p < 0.05$) in their walking activity which was also extracted from the Google Activity API.

4.7 Burden & Utility In Using The App

4.7.1 Participant Burden. When asked in the exit interview, approximately 84% and 78% of the participants stated that the MOODS study app and smartwatch did not disrupt their daily activities respectively. We next analyze the physical and mental effort involved in using the app i.e., burden.

Aside from charging the smartwatch, participants were asked to rate their stress and annotate the stressor multiple times daily. Participants reported in [48] that typing an event repeatedly without auto-complete became burdensome. Upon their recommendation, we implemented a predictive text input module that presented potential stressors as participants typed. We used logs collected from the smartphone app to calculate the time taken to enter a stressor for every stressor entry. The predictive text input module resulted in a significant reduction in the time taken for stressor entry over time as more and more episodes were annotated ($b = 50.46$, $m = -0.58^*$, $p < 0.05$) (see Figure 11a). The gradual reduction in time commitment may have played a role in keeping the participants engaged for a long duration.
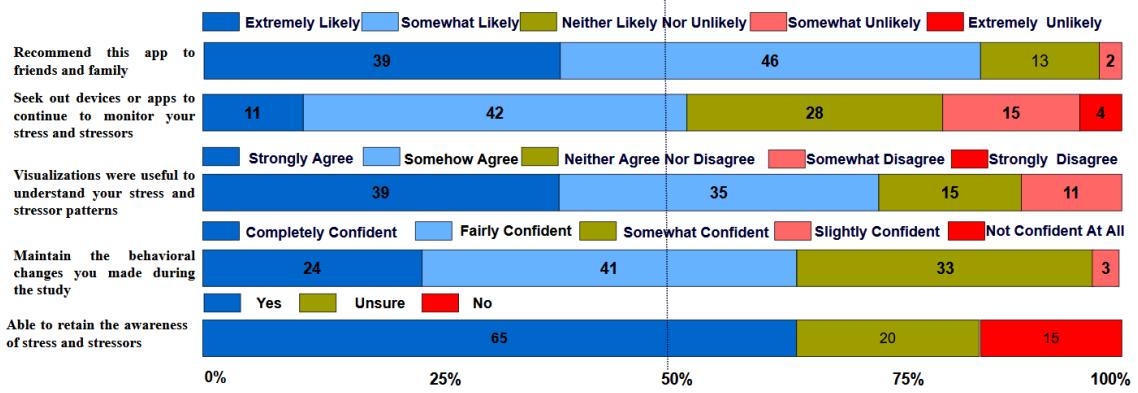
Fig. 10. Participants' Evaluation and Perception of the Stressor Logging Application

Another aspect of the burden may be the mental effort in recalling the stressors. In the weekly survey, participants were queried about the ease of recollecting their daily stressors from the past week, and their responses were quantified on a numerical scale (ranging from "Quite easy" to "Very difficult," scored from 1 to 5). As Figure 11b shows, the score remains closer to 2, indicating that recalling the stressors did not pose a burden, perhaps due to asking them right after an event occurred and showing them the time and location.

4.7.2   Participant Ratings of App Utility. We analyze the responses to five questions about the app's utility from the exit survey. Figure 10 shows the five questions and the proportion of participants selecting specific responses. The scores indicate that participants were very satisfied with the app and reported deriving tangible real-life benefits by visualizing their own data. In fact, 85% participants indicated a willingness to recommend the app to friends and family.

Comments from some participants explain their sentiments. *P38* said *"This has been an interesting experience. I encouraged others to participate as well." P28* said *"I really enjoyed participating in this study as I know it likely provides valuable information; however, it was personally rewarding as well to help me identify my stressors and areas of need. So from that personal standpoint I appreciate the opportunity to participate."*

5   DISCUSSION & LIMITATIONS

5.1   Feasibility of Logging Stressors

As stressor logging involves both physical and mental effort, engaging participants so they keep logging over a long time is a challenge. An analysis of mental health apps (with 10,000 or more installs) found that the average retention rate at the 30-day mark was 3.3%[9]. The most frequent reasons for discontinuance include too much time needed to enter data, loss of interest, hidden costs, and confusing to use [44]. Although our study did not involve any costs, it could have lost participants due to the other three factors. However, our study showed that using novel mechanisms to lower burden and improve utility resulted in a 25 times higher retention rate (see Section 4.1). We also found that 85% of participants were willing to recommend our apps (see Figure 10). We hypothesize that this high retention rate and high approval rating resulted from the following — 1) reducing the mental effort of recalling stressors by prompting them close to the time of occurrence of a potentially stressful event, but waiting until the event was over so they can recover; 2) reducing the physical effort in entering the stressor by using predictive auto-complete. As fatigue grows over time, predictive auto-complete reduces the time needed for stressor logging as most stressors reported are already

(a) Time taken to annotate a stressor
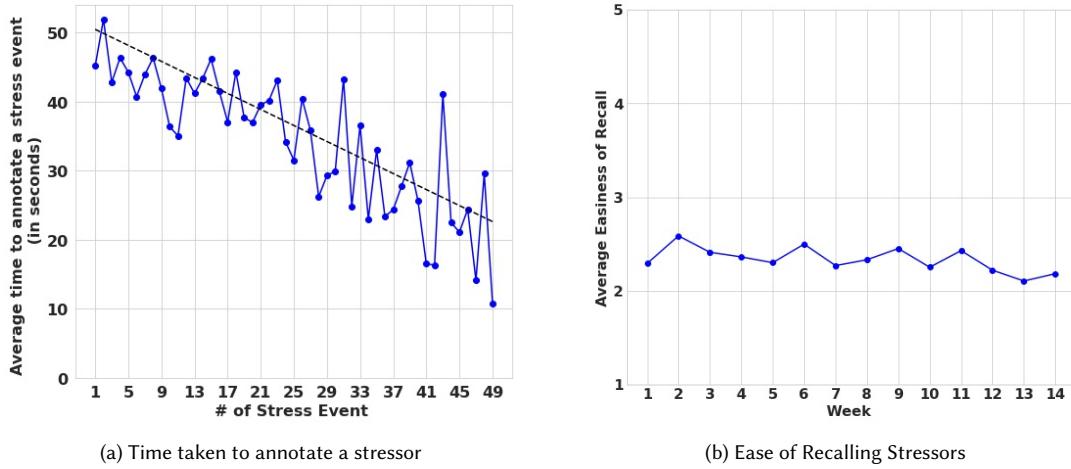
(b) Ease of Recalling Stressors

Fig. 11. Time commitment, and recall burden

in the database; 3) integrating stressors with stress arousal and spatiotemporal context in visualizations so they can identify their dominant stressors and contexts; and 4) providing new types of visualizations each week that allowed them to progressively dig deeper in their data.

5.2 Impact of Stressor Logging and Integration of Stressors in Self-reflective Visualizations

To enhance stress management, previous studies have employed various new technologies, including chatbot-triggered exercises rooted in Cognitive Behavioral Therapy (CBT) and Dialectical Behavioral Therapy (DBT) [31], a browser-based application delivering micro-stress interventions [66], a calendar-mediated stressor anticipation application [48], and a system focused on setting adaptive goals in Exercise, Meditation, and Accessibility dimensions [42]. Our study shows that sensor-prompted logging of stressors and integrating stressors with associated spatiotemporal contexts in stress visualizations for self-reflection holds a potential for a statistically significant decrease in self-reported stress that is at par or better than many stress intervention studies. We note that people report having *some* or *little control* over their daily stressors [12], implying that there is only a small subset of stressors that can be addressed via behavioral modifications. Therefore, a reduction of 10% in self-reported stress intensity and 10 fewer stressors per month reflects a meaningful change that can potentially improve long-term health outcomes [4, 67].

Exploration of the mechanisms behind the stress reduction shows that stressor logging and weekly visualizations helped participants become more aware of patterns in their stress and stressors, identify what they needed to work on (e.g., dominant stressors or most stressful spatial or temporal contexts), and evaluate the impact of self-initiated behavioral changes to discover what worked best for them. Specifically, participants reported making fourteen (14) types of behavioral changes (see Section 4.4.3). It provides support for offering a large repository of interventions [66] and provides support for anticipatory interventions [48] in addition to during and post-stress interventions, as participants reported using each type. Finally, our study shows that if stress intervention can indeed match the source of stress as participants asked for in [31], it has a chance to produce a longer-lasting reduction in stress.

5.3 Design Implications for Stress Tracking Apps

5.3.1 Focus on Stressors. One of the findings from this study is that providing an easy and quick way to log stressors right around the time a stress event may be occurring can improve the chances of engaging participants in logging their stressors. Participants mentioned the approach followed in this study to log their stressors easy to recall (see Figure 11).

5.3.2 Insightful Visual Summaries. Once participants log their stressors, helping them visualize their own data improved the awareness of stress and stressor for most participants. It led many of them to make behavioral changes, improve productivity, and adopt healthier habits. This study also showed which kinds of visualizations of stressors were found to be most and least helpful by the participants. Adopting these can improve the real-life utility of stress-tracking apps.

5.3.3 Matching Interventions to Sources of Stress. It is desirable to find stress interventions that can work for many people in a variety of situations. However, a recent study [31] found that participants wanted the intervention to relate to the specific stressor they were experiencing. Our findings support this as well. Even though our study did not recommend any intervention, engagement with stressor logging and visualization and the resulting growth in awareness led many participants to initiate several behavioral changes. When we analyze the specific behavioral changes initiated (see Section 5.2), we find substantial diversity. These show that stress interventions that include a diverse array of suggestions [66] and aim to match the intervention to the source of stress are more likely to be effective. For an intervention application to automatically match a prompted intervention to the source of stress, however, AI models are needed that can identify the stressor from sensor data. Data collected in our study that have both field-collected sensor data and stressor labels provided by participants can accelerate such research.

5.4 Limitations

This study had several limitations that can inspire future research.

5.4.1 Benefit of Using Sensor-triggered Prompts. In this study, prompts were generated across all spectrum of stress scores so sensor data could be collected to further improve the performance of AI models for detecting stress. Future studies could reduce participant burden further by only selecting higher stress arousal scores to generate prompts. To establish how useful it is to use such pre-trained AI models to generate prompts, future studies can conduct a randomized control trial (RCT), wherein one group receives prompts tailored for higher stress scores to minimize participant burden, while the other group uses an alternative mechanism to reduce participant burden, i.e., receive randomly generated prompts without having to wear any smartwatch. The objective would be to assess and compare the effectiveness of these two approaches in terms of user engagement, retention rates, and stress reduction.

5.4.2 Observational Study with No Control Arm. We observed an approximately 11% reduction in self-reported stress over the 14-week study period with momentary stressor logging and weekly visualization of self-reported stressors. This study-long reduction in self-reported stress is higher than several recent works on stress intervention. However, this should only be regarded as showing a potential for stress reduction and not definitive evidence due to several caveats. First, this was an observational study with participants' own data from the initial weeks acting as their control. Second, this being a study without any ongoing or end-of-study incentive, it may have attracted participants motivated in stress tracking. Third, there was no micro-incentive for completing self-ratings or weekly surveys. This may increase the chance that missing data may not be random. Fourth, different participants contributed data for different amounts of time potentially introducing other biases. To see if the trends persist when the analysis was limited to $4 \leq \forall k \leq 13$

weeks resulting in more participants (up to $n = 86$) available for complete case analysis, we repeated the LMM analysis and found similar trends with $-0.023 \leq m \leq -0.029$ for *stress intensity* and $-0.018 \leq m \leq -0.029$ for *stress frequency*. Finally, the magnitude of reduction in stress was estimated from a linear fit and it was not using a standard scale such as the Perceived Stress Scale (PSS) at both pre- and post-study stages. A randomized control trial (RCT) is needed to truly determine whether momentary stressor logging and weekly visualizations can indeed reduce stress and improve awareness leading to behavior change and should become an essential component in future stress interventions.

5.4.3 Disentanglement of Visualizations to Isolate their Impact. In this study, several factors could have potentially contributed to the reduction in self-reported stress intensity and frequency. They included getting prompts upon the conclusion of a potentially stressful event, logging stressors, having access to the app where they can review a timeline of their self-reported stressors, getting 16 types of visualizations that included a combination of stress, stressors, spatial context, and temporal context aimed at eight insights, the introduction of a new visualization type each week, and weekly surveys. A future study can investigate the incremental impact of one or more of these factors and identify the ones with a potential for the highest impact so they can be considered for future sensor-triggered interventions.

5.4.4 No Distinction between Productive and Unproductive Stressors. Stress is sometimes necessary to stay alert, overcome challenges, and improve work performance. Examples include challenging projects, competitions, and exams. Unproductive stressors harm well-being, causing negative impacts on health and efficiency. Examples include work overload, conflicts, and catastrophization. In this study, no distinction was made to account for this. *P*16's experience in a study exemplifies this. Engaged in a dissertation, they were annoyed by watch notifications for stress. Recognizing stress from academics, they disabled notifications, seeing stress as intrinsic to their goals. They even removed the watch to avoid interruptions. This underscores understanding of participant context for effective stress management. Future stress management apps should aim to distinguish between productive and unproductive stressors.

5.4.5 Positive vs. Negative Emotions. There might be many instances where physiological responses are triggered by excitement, eating, or substance consumption. Such instances could be interpreted as positive stress or eustress. Eustress is a positive cognitive response to a stressor, which is associated with positive feelings and a healthy physical state whereas; distress is defined as the stress associated with negative feelings and physical impairments [47]). While the identification of eustress instances could be enlightening for participants, some participants in this study raised concerns about such instances. *P*4 said *"Any time I so much as watch a TV show with drama or tension, the watch registers a suspected stress episode. Excitement, exercise, and substance use all trigger the app."* *P*41 said *"I personally was confused on how actual stress was being detected vs. just an elevated heart rate. The app would alert to stress during times of relaxation. For example, while watching TV, eating, or just general excitement about something."* Such testimonials indicate that prompting participants during eustress moments could confuse them and reduce their trust in the stress app. Developing methods to automatically differentiate eustress from distress can further improve the utility of stress apps.

5.4.6 Device Limitations. As shown in Figure 10, participants were not as enthusiastic about seeking out devices or apps. When asked about disliked aspects of the app, the most frequent complaint was the short battery life of the smartwatch as it lasted about 6 hours. Fortunately, this issue is easily addressable in commercial apps as modern smartwatches last the entire day. Some iPhone participants highlighted occasional Bluetooth connectivity issues. Intermittent problems in establishing and maintaining a stable connection between the smartwatch and iPhone frustrated these participants.

5.4.7 Stress Rating Biases. The likelihood of a presented event considered to be stressful by the smartwatch app was displayed in the smartphone app when participants were asked to annotate the event. It was added at the request of pilot participants so they could more easily remember a stressor. *P*31 said *"It influenced sometimes but many times it made me think harder to recall the stressors."* *P*35 said *"Stress likelihood didn't influence and might have been subconsciously."* Although some participants mentioned that this did not bias them (*P*26 said *"I had no bias from stress likelihood."*, *P*27 said *"It may have had a little bias but I tried to rely on my own judgment."*), this display may have created a bias in the stress ratings participants chose for the event. As our analysis to determine a trend in this data included a within-person analysis, this bias may not have had a substantial impact on the results presented here.

5.4.8 No Analysis of Clinical Significance. In our analysis, we only tested for statistical significance. Establishing any clinical significance (i.e., impact) of momentary stressor logging and visualizations will require pre and post-study assessment of clinical outcomes of interest such as addiction, depression, weight, sleep, daily symptoms, etc.

6 CONCLUSION

Real-time detection of physiological events indicative of potential stress in conveniently-worn wearables opens exciting new possibilities to improve stress management. Research presented here in a 100-day study shows the feasibility of soliciting stressors from participants upon the detection of a physiological event by a smartwatch. Our findings show that momentary stressor logging and providing simple and insightful visualizations of these data can lead to better awareness of stress and stressors. It can also lead to self-initiated behavior modifications. Future work can investigate how to integrate such capabilities in consumer wearables and conduct micro-randomized trials to determine how such capabilities may be used in conjunction with momentary interventions to help people better manage their stress.

ACKNOWLEDGEMENTS

REFERENCES

[1] 2023. https://www.stress.org/workplace-stress

[2] Fatema Akbar, Gloria Mark, Stephanie Prausnitz, E Margaret Warton, Jeffrey A East, Mark F Moeller, Mary E Reed, and Tracy A Lieu. 2021. Physician stress during electronic health record inbox work: in situ measurement with wearable sensors. *JMIR Medical Informatics* 9, 4 (2021), e24014.

[3] Ali Akbari, Jonathan Martinez, and Roozbeh Jafari. 2021. Facilitating human activity data annotation via context-aware change detection on smartwatches. *ACM Transactions on Embedded Computing Systems (TECS)* 20, 2 (2021), 1–20.

[4] David M Almeida, Elaine Wethington, and Ronald C Kessler. 2002. The daily inventory of stressful events: An interview-based approach for measuring daily stressors. *Assessment* 9, 1 (2002), 41–55.

[5] Samaneh Aminikhanghahi, Maureen Schmitter-Edgecombe, and Diane J Cook. 2019. Context-aware delivery of ecological momentary assessment. *IEEE Journal of Biomedical and Health Informatics* 24, 4 (2019), 1206–1214.

[6] Rummana Bari, Md Mahbubur Rahman, Nazir Saleheen, Megan Battles Parsons, Eugene H Buder, and Santosh Kumar. 2020. Automated detection of stressful conversations using wearable physiological and inertial sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.

[7] Lisa Feldman Barrett. 2017. *How emotions are made: The secret life of the brain.* Pan Macmillan.

[8] Samuel L Battalio, David E Conroy, Walter Dempsey, Peng Liao, Marianne Menictas, Susan Murphy, Inbal Nahum-Shani, Tianchen Qian, Santosh Kumar, and Bonnie Spring. 2021. Sense2Stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management

intervention for smoking relapse prevention. *Contemporary Clinical Trials* 109 (2021), 106534.

[9] Amit Baumel, Frederick Muench, Stav Edan, and John M Kane. 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research* 21, 9 (2019), e14567.

[10] Brandon M Booth, Hana Vrzakova, Stephen M Mattingly, Gonzalo J Martinez, Louis Faust, and Sidney K D'Mello. 2022. Toward Robust Stress Prediction in the Age of Wearables: Modeling Perceived Stress in a Longitudinal Study with Information Workers. *IEEE Transactions on Affective Computing* 13, 4 (2022), 2201–2217.

[11] Olivia Calancie, Lexi Ewing, Laura D Narducci, Salinda Horgan, and Sarosh Khalid-Khan. 2017. Exploring how social networking sites impact youth with anxiety: A qualitative study of Facebook stressors among adolescents with an anxiety disorder diagnosis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 11, 4 (2017).

[12] Eric S Cerino, Susan T Charles, Jacqueline Mogle, Jonathan Rush, Jennifer R Piazza, Laura M Klepacz, Margie E Lachman, and David M Almeida. 2023. Perceived control across the adult lifespan: Longitudinal changes in global control and daily stressor control. *Developmental psychology* (2023).

[13] Kemeng Chen, Wolfgang Fink, Janet Roveda, Richard D Lane, John Allen, and Johnny Vanuk. 2015. Wearable sensor based stress management using integrated respiratory and ECG waveforms. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–6. https://doi.org/10.1109/BSN.2015.7299369

[14] Eun Kyoung Choe, Bongshin Lee, et al. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications* 35, 4 (2015), 28–37.

[15] George P Chrousos. 2009. Stress and disorders of the stress system. *Nature reviews endocrinology* 5, 7 (2009), 374–381.

[16] CuesHub and CuesHub, PBC. 2024. *CuesHub, PBC.* https://www.cueshub.com/

[17] Shirui Dai, Yongzhen Mo, Yumin Wang, Bo Xiang, Qianjin Liao, Ming Zhou, Xiaoling Li, Yong Li, Wei Xiong, Guiyuan Li, et al. 2020. Chronic stress promotes cancer development. *Frontiers in oncology* 10 (2020), 1492.

[18] Xianghua Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. 2021. Data engagement reconsidered: A study of automatic stress tracking technology in use. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.

[19] Muna Dubad, Catherine Winsper, Caroline Meyer, Maria Livanou, and SJPM Marwaha. 2018. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. *Psychological medicine* 48, 2 (2018), 208–228.

[20] Genevieve Fridlund Dunton, Eldin Dzubur, Keito Kawabata, Brenda Yanez, Bin Bo, and Stephen Intille. 2014. Development of a smartphone application to measure physical activity using sensor-assisted self-report. *Frontiers in public health* 2 (2014), 12.

[21] Carol L Flinchbaugh, E Whitney G Moore, Young K Chang, and Douglas R May. 2012. Student well-being interventions: The effects of stress management techniques and gratitude journaling in the management education classroom. *Journal of Management Education* 36, 2 (2012), 191–219.

[22] Garmin. 2022. Stress tracking. https://www.garmin.com/en-US/garmin-technology/health-science/stress-tracking/

[23] Daniel Gilbert. 2009. *Stumbling on happiness*. Vintage Canada.

[24] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2017. Monitoring stress with a wrist device using context. *Journal of biomedical informatics* 73, 159–170.

[25] James J Gross. 2015. The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry* 26, 1 (2015), 130–137.

[26] Thomas G Guilliams and Lena Edwards. 2010. Chronic stress and the HPA axis. *The standard* 9, 2 (2010), 1–12.

[27] Andree Hartanto, Kristine YX Lee, Yi Jing Chua, Frosch YX Quek, and Nadyanna M Majeed. 2023. Smartphone use and daily cognitive failures: A critical examination using a daily diary approach with objective smartphone measures. *British Journal of Psychology* 114, 1 (2023), 70–85.

[28] Juliet Hassard, Kevin Teoh, Tom Cox, M Cosmar, R Gründler, D Flemming, B Cosemans, and K Van den Broek. 2014. Calculating the cost of work-related stress and psychosocial risks. (2014).

[29] Katie Hoemann, Zulqarnain Khan, Mallory Feldman, Catherine Nielson, Madeleine Devlin, Jennifer Dy, Lisa Barrett, Jolie Wormwood, and Karen Quigley. 2020. Context-aware experience sampling reveals the scale of variation in affective experience. (01 2020). https://doi.org/10.31234/osf.io/cvjb8

[30] Seyedmajid Hosseini, Raju Gottumukkala, Satya Katragadda, Ravi Teja Bhupatiraju, Ziad Ashkar, Christoph W Borst, and Kenneth Cochran. 2022. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. *Scientific Data* 9, 1 (2022), 1–13.

[31] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *CHI Conference on Human Factors in Computing Systems*. 1–16.

[32] J. D. Hunter. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 3 (2007), 90–95. https://doi.org/10.1109/MCSE.2007.55

[33] Md. Hussain and Ishtiak Mahmud. 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4, 39 (25 7 2019), 1556. https://doi.org/10.21105/joss.01556

[34] Plotly Technologies Inc. 2015. *Collaborative data science*. Montreal, QC. https://plot.ly

[35] Stephen Intille, Caitlin Haynes, Dharam Maniar, Aditya Ponnada, and Justin Manjourides. 2016. μEMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1124–1128.

[36] Yanqi Jiang, Xianghua Ding, Xiaojuan Ma, Zhida Sun, and Ning Gu. 2023. IntimaSea: Exploring Shared Stress Display in Close Relationships. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[37] Masamune Kawasaki, Naomi Yamashita, Yi-Chieh Lee, and Kayoko Nohara. 2020. Assessing users' mental status from their journaling behavior through chatbots. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.

[38] Zachary D King, Judith Moskowitz, Begum Egilmez, Shibo Zhang, Lida Zhang, Michael Bass, John Rogers, Roozbeh Ghaffari, Laurie Wakschlag, and Nabil Alshurafa. 2019. Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–22.

[39] Patrick Klaiber, Jin H Wen, Anita DeLongis, and Nancy L Sin. 2021. The ups and downs of daily life during COVID-19: Age differences in affect, stress, and positive events. *The Journals of Gerontology: Series B* 76, 2 (2021), e30–e37.

[40] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized stress management: enabling stress monitoring with lifelogexplorer. *KI-Künstliche Intelligenz* 29 (2015), 115–122.

[41] Rafal Kocielnik, Natalia Sidorova, Fabrizio Maria Maggi, Martin Ouwerkerk, and Joyce HDM Westerink. 2013. Smart technologies for long-term stress monitoring at work. In *proceedings of the 26th IEEE international symposium on computer-based medical systems*. IEEE, 53–58.

[42] Artie Konrad, Victoria Bellotti, Nicole Crenshaw, Simon Tucker, Les Nelson, Honglu Du, Peter Pirolli, and Steve Whittaker. 2015. Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3829–3838.

[43] Evangelos Kontopantelis, Tim Doran, David A Springate, Iain Buchan, and David Reeves. 2015. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *bmj* 350 (2015).

[44] Paul Krebs, Dustin T Duncan, et al. 2015. Health app use among US mobile phone owners: a national survey. *JMIR mHealth and uHealth* 3, 4 (2015), e4924.

[45] Peter Kuppens, Francis Tuerlinckx, James Russell, and Lisa Barrett. 2012. The Relation Between Valence and Arousal in Subjective Experience. *Psychological Bulletin* 139 (12 2012). https://doi.org/10.1037/a0030811

[46] Tatiana Kuzmowycz. 2023. Introducing stress monitor: A new way to monitor and manage stress. https://www.whoop.com/thelocker/introducing-stress-monitor-a-new-way-to-monitor-manage-stress/

[47] Richard S Lazarus. 1993. From psychological stress to the emotions: A history of changing outlooks. *Annual review of psychology* 44, 1 (1993), 1–22.

[48] Kwangyoung Lee, Hyewon Cho, Kobiljon Toshnazarov, Nematjon Narziev, So Young Rhim, Kyungsik Han, YoungTae Noh, and Hwajung Hong. 2020. Toward future-centric personal informatics: Expecting stressful events and preparing personalized interventions in stress management. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[49] Noelle R Leonard, Marya V Gwadz, Amanda Ritchie, Jessica L Linick, Charles M Cleland, Luther Elliott, and Michele Grethel. 2015. A multi-method exploratory study of stress, coping, and substance use among high school youth in private schools. *Frontiers in psychology* (2015), 1028.

[50] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.

[51] Carolyn M Mazure. 1998. Life stressors as risk factors in depression. *Clinical Psychology: Science and Practice* 5, 3 (1998), 291.

[52] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 849–858.

[53] Thomas Münzel, Johannes Herzog, Frank P Schmidt, and Mette Sørensen. 2017. Environmental stressors and cardiovascular disease: the evidence is growing. , 2297–2299 pages.

[54] Simon Ollander, Christelle Godin, Aurélie Campagne, and Sylvie Charbonnier. 2016. A comparison of wearable and stationary sensors for stress detection. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 004362–004366.

[55] Amon Rapp and Federica Cena. 2016. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies* 94 (2016), 1–17.

[56] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölinder. 2010. Mind the body! Designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM conference on designing interactive systems*. 47–56.

[57] Robert M Sapolsky. 2004. *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping*. Holt paperbacks.

[58] Moushumi Sharmin, Andrew Raij, David Epstien, Inbal Nahum-Shani, J Gayle Beck, Sudip Vhaduri, Kenzie Preston, and Santosh Kumar. 2015. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 505–516.

[59] Patrick E Shrout, Gertraud Stadler, Sean P Lane, M Joy McClure, Grace L Jackson, Frederick D Clavél, Masumi Iida, Marci EJ Gleason, Joy H Xu, and Niall Bolger. 2018. Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences* 115, 1 (2018), E15–E23.

[60] Derrick Silove, Ingrid Sinnerbrink, Annette Field, Vijaya Manicavasagar, and Zachary Steel. 1997. Anxiety, depression and PTSD in asylum-seekers: associations with pre-migration trauma and post-migration stressors. *The British Journal of Psychiatry* 170 (1997), 351.

[61] Eric N Smith, Erik Santoro, Neema Moraveji, Michael Susi, and Alia J Crum. 2020. Integrating wearables in stress management interventions: Promising evidence from a randomized trial. *International Journal of Stress Management* 27, 2 (2020), 172–182.

[62] Fitbit Staff. 2022. Learn more about Fitbit's new body response feature on sense 2. https://blog.fitbit.com/fitbit-new-stress-detection/

[63] Stefan Stepanovic, Vadym Mozgovoy, and Tobias Mettler. 2019. Designing visualizations for workplace stress management: results of a pilot study at a swiss municipality. In *International Conference on Electronic Government*. Springer, 94–104.

[64] Yoshihiko Suhara, Yinzhan Xu, and Alex'Sandy' Pentland. 2017. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In *Proceedings of the 26th International Conference on World Wide Web.* 715–724.

[65] Yang Suyi, Pamela Meredith, and Asaduzzaman Khan. 2017. Effectiveness of mindfulness intervention in reducing stress and burnout for mental health professionals in Singapore. *Explore* 13, 5 (2017), 319–326.

[66] Xin Tong, Matthew Louis Mauriello, Marco Antonio Mora-Mendoza, Nina Prabhu, Jane Paik Kim, and Pablo E Paredes Castro. 2023. Just Do Something: Comparing Self-proposed and Machine-recommended Stress Interventions among Online Workers with Home Sweet Office. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–20.

[67] Philip M Ullrich and Susan K Lutgendorf. 2002. Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine* 24, 3 (2002), 244–250.

[68] Niels van Berkel, Jorge Goncalves, Lauri Lovén, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2019. Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* 125 (2019), 118–128.

[69] Rattanjeet S Vig, Paul Forsythe, and Harissios Vliagoftis. 2006. The role of stress in asthma: insight from studies on the effect of acute and chronic stressors in models of airway inflammation. *Annals of the New York Academy of Sciences* 1088, 1 (2006), 65–77.

[70] Johnny Chun Yiu Wong, Jun Wang, Eugene Yujun Fu, Hong Va Leong, and Grace Ngai. 2019. Activity recognition and stress detection via wristband. In *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia.* 102–106.

[71] Mengru Xue, Rong-Hao Liang, Bin Yu, Mathias Funk, Jun Hu, and Loe Feijs. 2019. AffectiveWall: Designing collective stress-related physiological data visualization for reflection. *IEEE Access* 7 (2019), 131289–131303.

[72] Matthew J Zawadzki, Stacey B Scott, David M Almeida, Stephanie T Lanza, David E Conroy, Martin J Sliwinski, Jinhyuk Kim, David Marcusson-Clavertz, Robert S Stawski, Paige M Green, et al. 2019. Understanding stress reports in daily life: A coordinated analysis of factors associated with the frequency of reporting stress. *Journal of Behavioral Medicine* 42 (2019), 545–560.

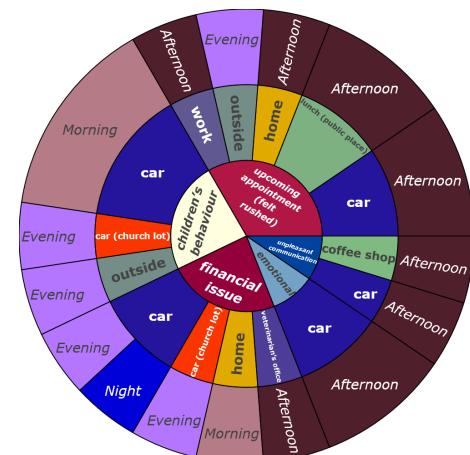# Supplementary Materials: Weekly Visualizations for Self-Reflection of Stress and Stressors

We designed 16 visualizations of stress ratings and stressor logs that participants provided to facilitate their self-reflection. Details of each visualization with an example appear below.

1 WEEK 1

1.1 Overall Average

(a) Overall and recent trends in stress and stressors

(b) Context (place and time) of five most prominent stressors

Fig. 1. Visualizations for Week 1

Participants received two sets of visualizations at the end of their first week. The first one (see Figure 1a) presented an overall summary of their data. It included

(1) Stressed Minutes per Hour (aggregated from all data collected thus far): This metric represents the average duration of stress experienced per hour (of data collected). As the amount of data collected can vary between participants and between days for the same participant, per hour instead of per day was used.

(2) Average Stressed Minutes (in the Current Week): To contrast with the overall average, this value indicates the average duration of stress experienced in the most recent week. It offers a snapshot of the ongoing stress levels during the current week.

(3) Average Number of Stressors Reported per Week (aggregated from all data collected thus far): The average number of stressors reported per week thus far indicates the typical frequency of stressors in the participant's daily life.

(4) Number of Daily Stressors (in the Current Week): It helps see how the most recent week went compared with the overall average.

Collectively, Figure 1a provides an overview of overall trends and recent changes in their stress and stressors.

1.2 Prominent Stressor Context

To help identify their most dominant stressors, the second visualization in Week 1 (see Figure 1b) showed participants the locations and times of day associated with the five most prominent stressors. By tapping on each stressor, participants could view detailed information about the specific location and time of occurrence for the selected stressor.

2 WEEK 2: MAP VIEW

Using the GPS data and participants' reported semantic locations at the time of reporting a stressor, the map view (see Figure 2a) showed semantic locations of reported stress instances on a map. Each point on the map corresponds to a specific location where stress was reported. By hovering over or tapping on each point, participants could access additional details such as the associated stressor, the day it occurred, and the specific time of day when the stressor was encountered. Furthermore, in cases where new locations were reported during the current week, participants had the option to interact with the map legend. Clicking or tapping on the legend allowed participants to access and explore the newly reported locations for the ongoing week, enhancing their understanding of stressor occurrences.
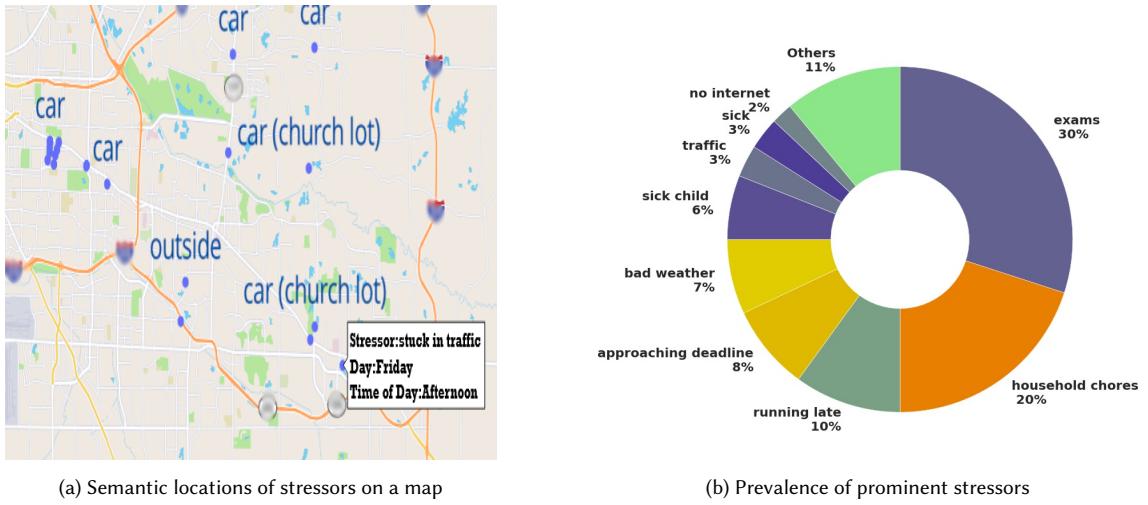

(a) Semantic locations of stressors on a map

(b) Prevalence of prominent stressors

Fig. 2. Visualizations for Weeks 2 and 3

(a) Locations most associated with stress
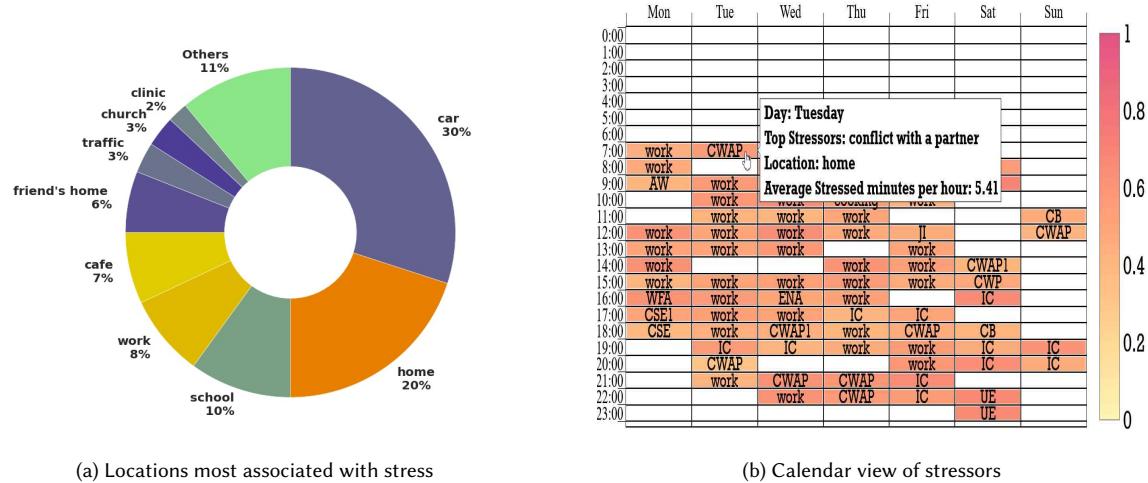
(b) Calendar view of stressors

Fig. 3. Visualizations for Weeks 4 and 5

3 WEEK 3: PREVALENCE OF PROMINENT STRESSORS

Tens of stressors logged by the participants allowed an analysis of the prevalence of prominent stressors (i.e., the amount of time attributed to each stressor). Figure 2b shows an example. Each segment of the chart represents a specific stressor, and participants had the option to interact with the visualization by hovering over or tapping on each segment. Upon interaction, the abbreviations representing the stressors expanded, revealing the corresponding percentage of stressed duration associated with that particular stressor. This interactive feature enabled the participants to gain insights into the relative impact of different stressors on their lives.

4 WEEK 4: LOCATIONS ASSOCIATED WITH STRESS (LOCATION PROMINENCE)

From the semantic location participants reported at the time of stress ratings and the duration of the event estimated by the app was used to determine the percent of total stressed time that participants were at specific locations. Figure 3a an example of this visualization. Participants could engage with the visualization by clicking on or tapping the legend, which corresponded to specific locations. This action allowed participants to view the percentage of stressed duration associated with each location. By offering this interactive functionality, the visualization aimed to enhance participants' understanding of the spatial context of their stress experiences.

5 WEEK 5: CALENDAR VIEW OF STRESSORS

A calendar view can help participants associate stressors with their calendar and make it easier to identify the stress impact of various activities. Therefore, we overlaid the reported stressors in a calendar view. Figure 3b shows an example. It showed the days and times when specific stressors were reported and the color indicated the likelihood of the event determined to be stressful by the app. By hovering over or tapping on each segment within the visualization, participants could access additional details, including an expanded form of the abbreviation, the corresponding stress likelihood, and the average minutes reported for that hour. To facilitate interpretation, a color bar is situated on the right side of the visualization, indicating the stress likelihood values associated with different segments.
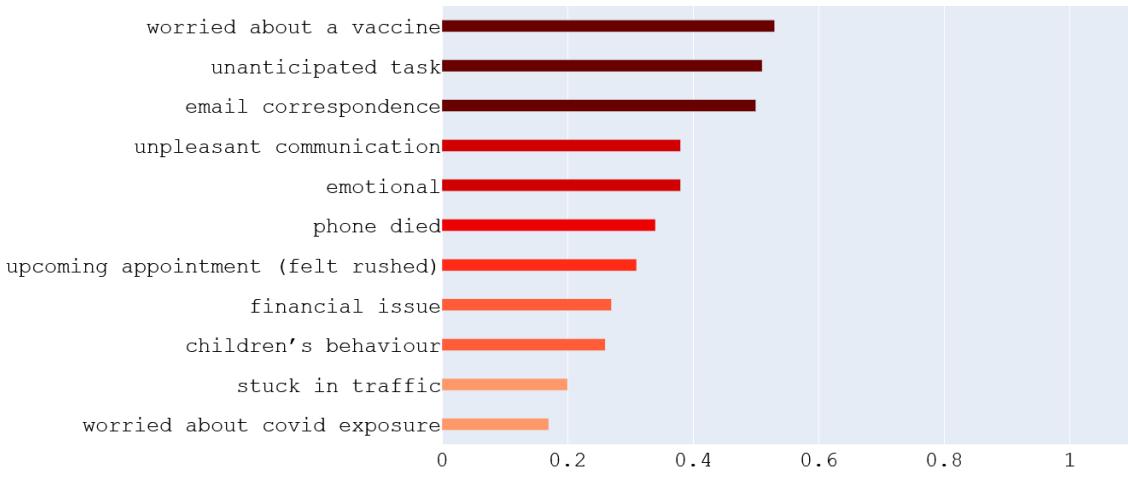
Fig. 4. Stressors associated with the highest stress likelihoods (Week 6)

6 WEEK 6: STRESSOR RANKING

Each event presented to participants for stress rating and annotation has associated with it a likelihood with which the app determined this event to be indicative of stress. If the participant chose to denote a presented event as stressful and provided a stressor, this stressor becomes associated with the stress likelihood computed by the app for this event. Different stressors differ in the average stress likelihood associated with them. We used these to rank the stressors and showed the average stress likelihood associated with the highest-ranked stressors. Figure 4 shows an example of such a visualization. The stress likelihood displayed here was computed from the app and not from the stress ratings to allow for the possibility that in future apps, the burden of logging can be reduced further by having participants simply choose a stressor and not have to provide a stress rating each time they are prompted.

7 WEEK 7: WEEKLY TREND IN SPECIFIC STRESSORS REPORTED PER WEEK

Different weeks may differ in which stressors are reported and how frequently. The frequency of reports for some stressors may also change if the participants make some changes to their behaviors and are successful in mitigating them. To give them the ability to see such trends in their stressor logs, we provided them with a way to visually see the frequency of different stressors they reported. Figure 5 shows an example. Participants had the option to toggle stressors on and off in order to selectively view their representation on the graph.

8 WEEK 8: WEEKLY PREVALENCE

We then provided an alternative way to view the frequency of stressors reported each week. The goal was to allow them to compare the frequency of different stressors on the same graph so they could compare the relative frequency with which specific stressors were reported each. We wanted the graph to allow a comparison of frequency for different stressors for a given week and see how this changed over time. Figure 6 shows an example. The size of the dots represents the relative prevalence. Each segment on the graph corresponds to a specific stressor, and hovering or tapping over a segment provides additional information, including the expanded form of the abbreviation, the week number, and the corresponding number of occurrences for that stressor.
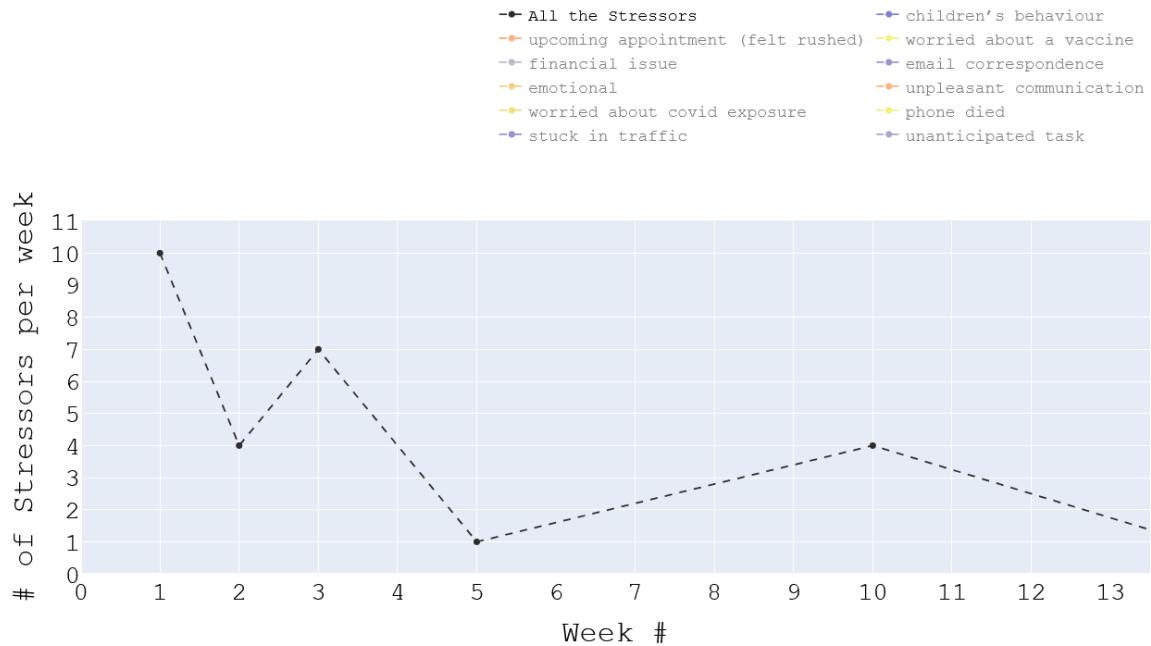
Fig. 5. Weekly trend in specific stressors reported per week over time (Week 7)
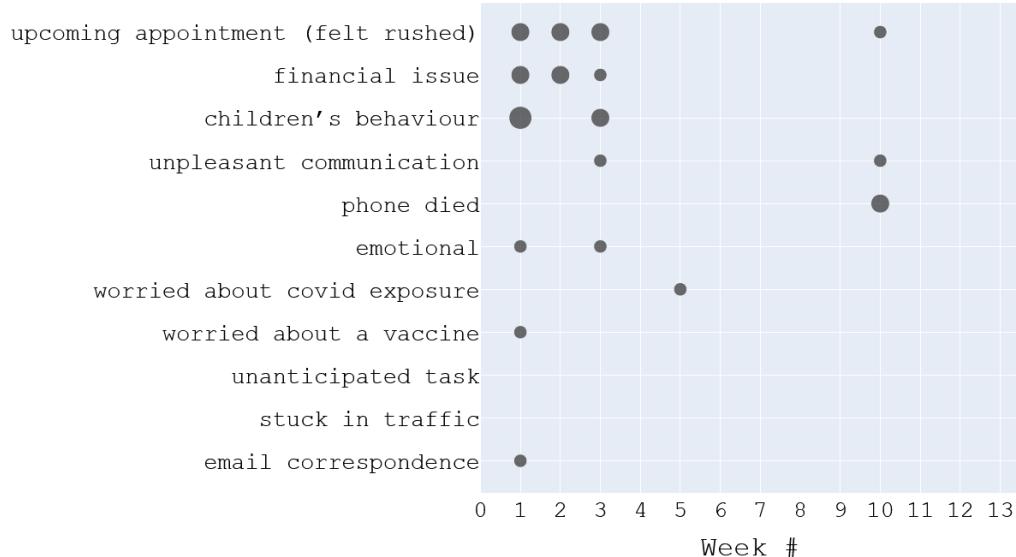

Fig. 6. Relative frequency with which specific stressors occur each week over time (Week 8)
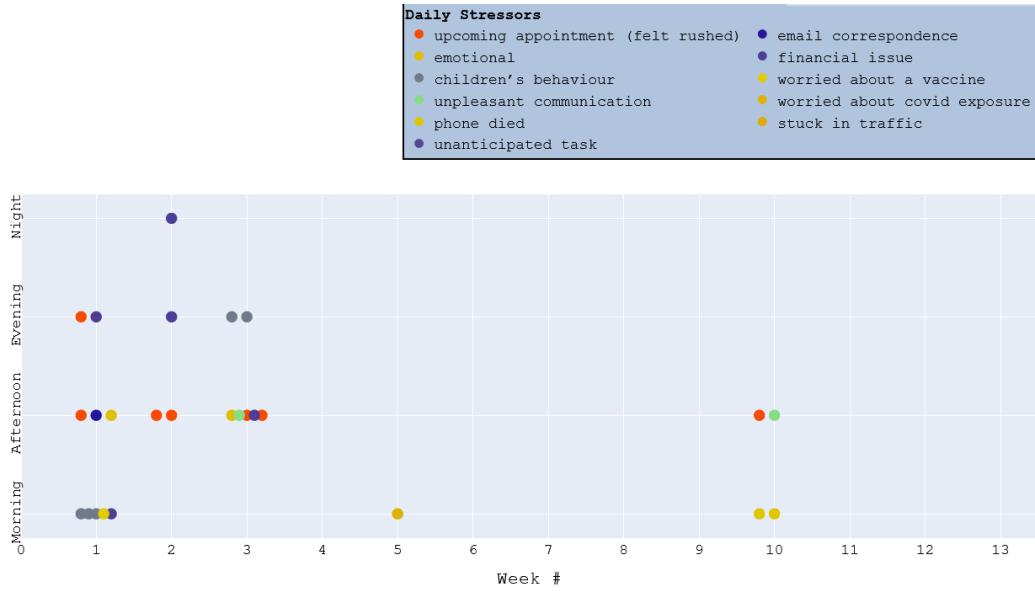
Fig. 7. Temporal patterns in stressor logs (Week 9)

9 WEEK 9: TIME OF DAY TREND

We wanted to provide a temporal context for different stressors so that participants could determine if there were specific temporal dynamics for some stressors. For this purpose, we provided them with a visualization to see the specific blocks of the day when different stressors occur. Figure 7 shows an example. It shows the blocks of time when specific stressors occur each week. The goal was to also allow them the ability to see any evolution in the temporal patterns over time. By interacting with the visualization, participants could select individual stressors by clicking or tapping on the corresponding legend item. This feature enabled participants to focus on specific stressors and gain insights into their temporal distribution throughout the MOODS study period.

10 WEEK 10: LOCATION TREND

Similar to facilitating the discovery of any temporal patterns in their data, we also wanted to facilitate the discovery of any spatial patterns in stressors. For this purpose, we revised the visualization from Week 9 by replacing the vertical axis to show the reported semantic locations instead of blocks of time. Figure 8 shows an example of such a scatter plot. To focus on specific stressors, participants could interact with the graph by clicking or tapping the corresponding items in the legend. This interactive functionality allowed the participants to isolate and explore individual stressors, gaining insights into their occurrence patterns across various locations over the duration of the MOODS study.

11 WEEK 11: DAY OF THE WEEK PATTERNS (DAILY TREND)

Some participants may have specific days of the week that may be more stressful than others depending on the nature of their work or personal life. Figure 9 shows an example of a visualization that represents the frequency of reported stressors across different days of the week, spanning from the commencement of the study to the present date. In addition to showing which days of the week had the most stressors reported, it also allowed the ability to see trends in
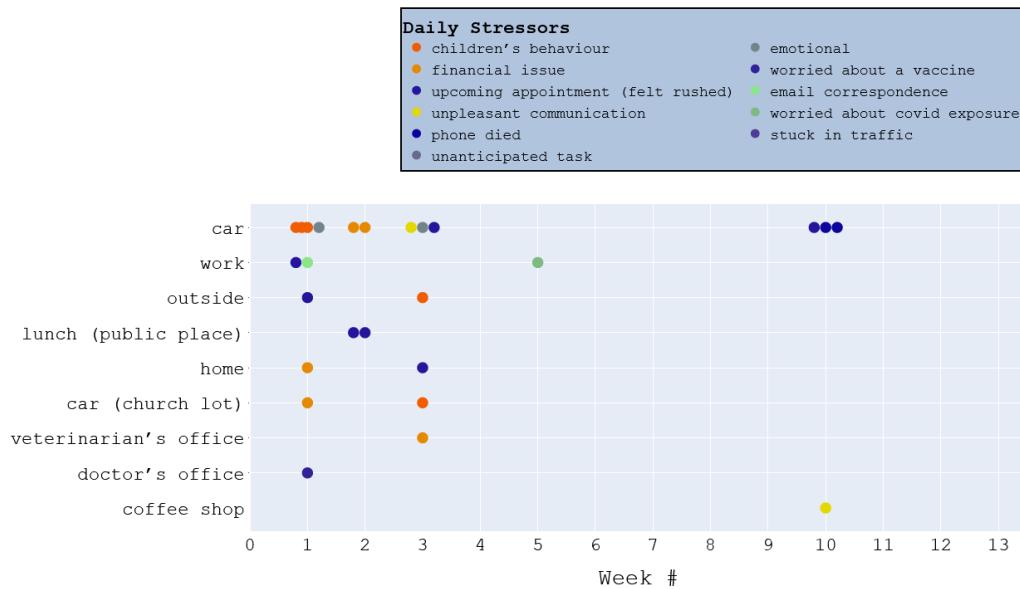
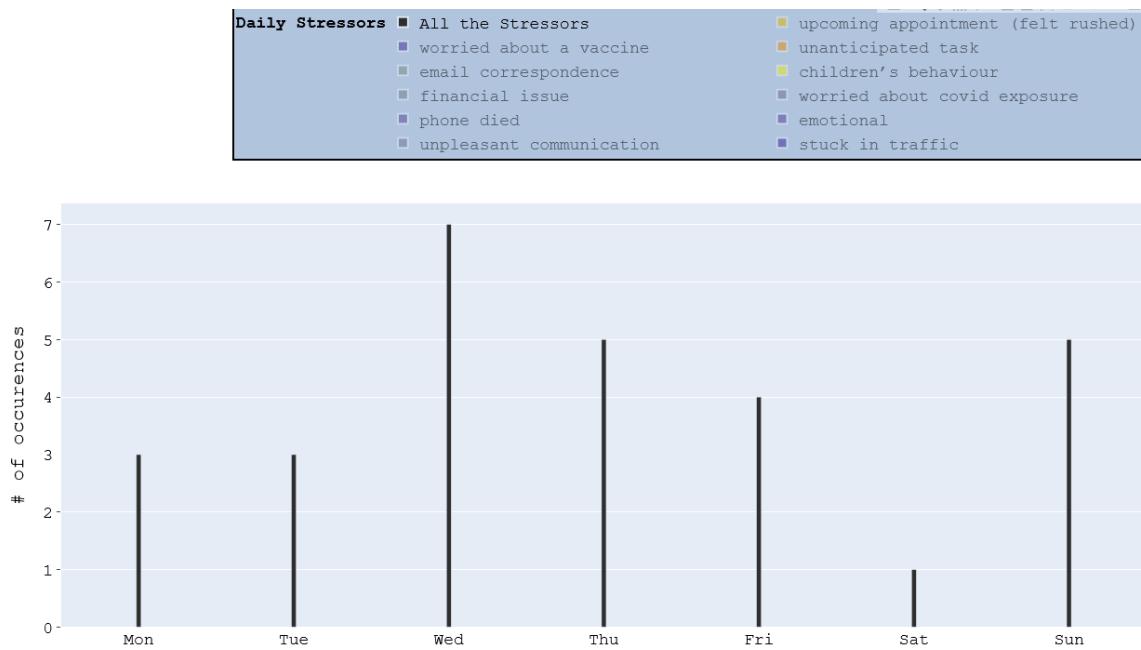Fig. 8. Spatial patterns in stressor logs (Week 10)

Fig. 9. Day of week patterns in stressor logs (Week 11)

specific stressors. By interacting with the graph and clicking or tapping on the corresponding items in the legend, the participants could isolate individual stressors for closer examination. This interactive feature enabled participants to
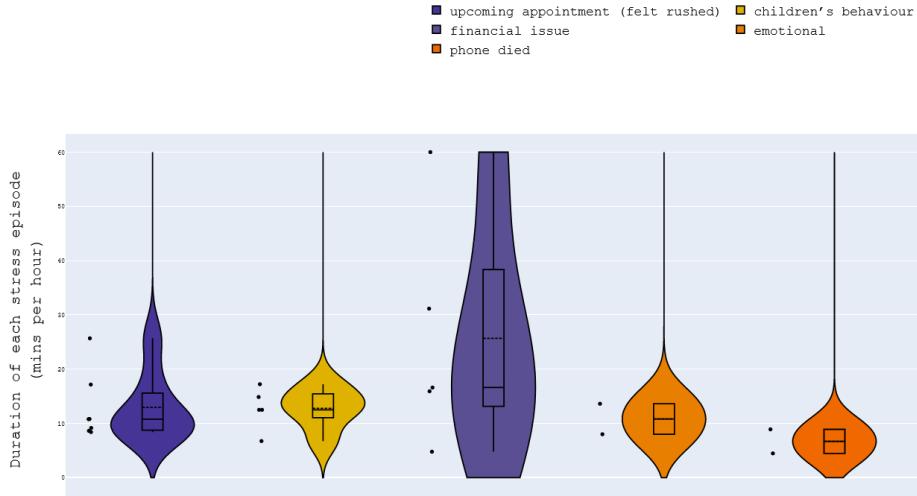
Fig. 10. Distribution of the duration of stress episodes associated with the five most prevalent stressors (Week 12)

explore how specific stressors had been reported and distributed over various days of the week throughout the MOODS study.

12 WEEK 12: STRESSOR DURATION

The app estimated the duration of each bodily event that was presented to participants for stress rating and stressor annotation. Wa wanted to allow the participants to see if stressors differed in terms of the duration of each episode attributed to them. For this purpose, we presented a series of violin plots illustrating the duration distribution of the five most frequently reported stressors. Figure 10 shows an example. Each violin plot comprises box plots at their centers, providing insight into the central tendency and variability of the data. Additionally, the broader areas of the violins depict the probability density distribution across various values, which is smoothed using a kernel density estimator.

13 WEEK 13: PREVALENT STRESSOR DURATION

To dig deeper into the duration patterns, we next presented weekly patterns in the average duration per week for the top five reported stressors throughout the weeks of study participation. Figure 11 shows an example. The visualization features bars, with red bars indicating the current week and gray bars representing previous weeks. This presentation allows participants to compare the duration of stressor periods over time, highlighting any variations or trends in stressed periods across different weeks.

14 WEEK 14: WORD CLOUDS

Finally, when we have the most amount of data in the stressor logs collected, we created Figure 12 illustrates word clouds to show patterns in the stressors reported and their associated reported locations. In this visualization, different stressors and locations are represented by words, with font sizes reflecting their frequency of occurrence. Larger fonts indicate more frequently reported stressors or locations. This word cloud offers a quick visual insight into the relative prevalence of different stressors and the locations associated with reported stressors.
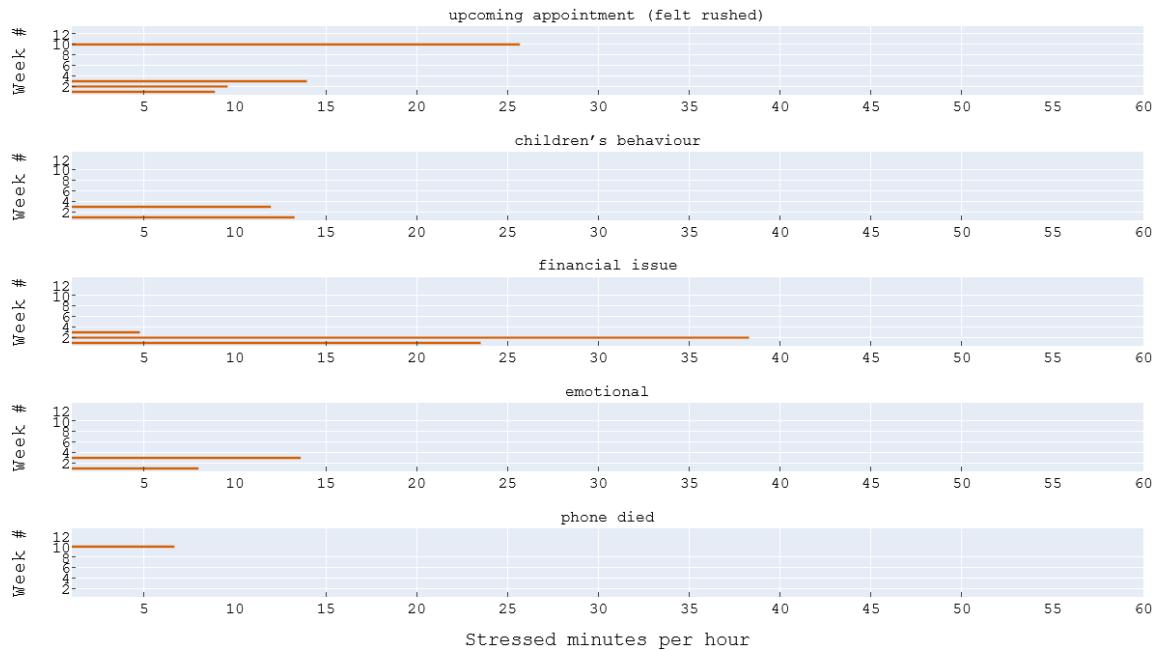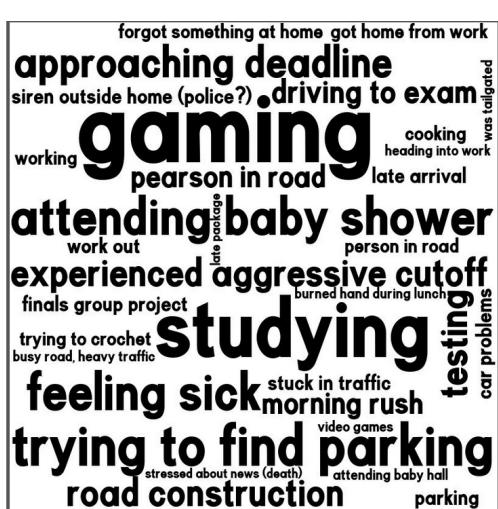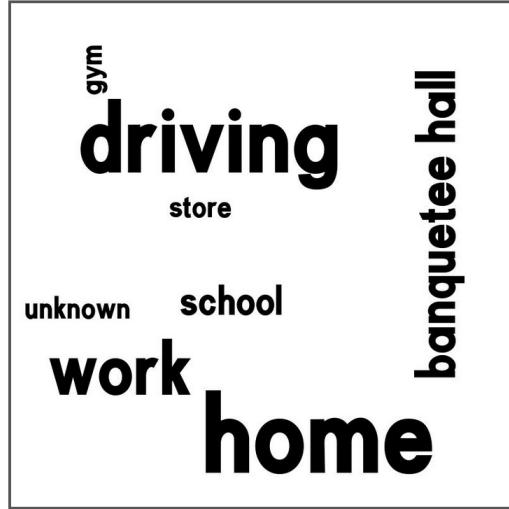
Fig. 11. Average stress episode duration associated with the five most prevalent stressors each week (Week 13)

(a) Word clouds from the stressors log

(b) Word clouds of locations reported with stressors

Fig. 12. Visualizations for Week 14