A FRAMEWORK TO DESIGN AND EVALUATE WEARABLE INTERACTIVE SYSTEMS FOR HEALTH

by

REEM ALBAGHLI

B.A., Kuwait University, 2000

M.A., Kuwait University, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Computer Science

2018

This thesis entitled:

A Framework to Design and Evaluate Wearable Interactive Systems for Health

written by Reem Albaghli

has been approved for the Department of Computer Science

____

Prof. Kenneth M Anderson

____

Prof. Leysia Palen

____

Prof. *Shaun Kane*

Date____________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol # 16-0331 and 16-0652

ABSTRACT

Albaghli, Reem (Ph.D., Computer Science)

A Framework to Design and Evaluate Wearable Interactive Systems for Health

Thesis directed by Professor Kenneth M. Anderson

Interactive wearable devices—or wearables—are increasingly being used to track the health-related data of their users. The design of such systems is a considerable challenge that stresses the analysis and design techniques of software engineering and human-centered computing and places considerable demands on user experience designers to produce systems that are reliable, accurate, and understandable while also remaining customizable by end users. Through my work, I aim to address this situation by developing a design framework that can be used to organize the activities during the development life cycle of a health-related wearable system. The WISE framework proposes a set of heuristics and activities by which health-related wearable systems can be both designed and evaluated. To demonstrate the power of my framework, I conducted five studies to help design the user interface of a wearable system that helps users visualize their heart rate data and monitor that data for anomalies that may require the attention of a medical professional. Over the course of my five studies, I identified core elements that embodied the heuristics of the WISE framework. This helped in defining important activities and principles when designing visualizations that are easy to understand by users and provide them with value in terms of being able to communicate with their doctors about the health of their heart. To illustrate the generic nature of my

framework, I have also demonstrated how my framework could be used to design health-related wearable systems that monitor other medical conditions such as depression and Parkinson's disease.

CONTENTS

Table of Contents

TABLES

FIGURES

INTRODUCTION

Interactive wearable devices—or wearables—are increasingly being used to track the health-related data of their users. The design of such systems is a considerable challenge that stresses the analysis and design techniques of software engineering and human-centered computing and places considerable demands on UX designers to produce systems that are reliable, accurate, and understandable while also remaining customizable by end users. In my work, I develop a system that focuses on the intersection of systems for consumer wearables and heart health. Through my research I identify a framework that can be used to organize the activities during the development life cycle of a health-related wearable system. The framework is organized around four desirable heuristics of interactive systems: *contextualization*, *customization*, *interactivity*, and *simplicity*.

These heuristics are useful both in terms of evaluating existing health-related wearable systems but also as targets that a designer is trying to achieve when creating such systems. The framework offers a variety of activities that can be used during analysis and design to understand the needs of the users of the system and to elicit the information needed to create systems that exhibit these heuristics. The ultimate goal of the framework is to help designers produce health-related wearable systems that are understandable and valuable to their users. By "understandable" I mean that the users of these systems can easily monitor some aspect of their health, know the current state of their medical condition, and be able to communicate that information to their doctor. By "valuable" I mean that the system makes it straightforward to perform these tasks and provides a user with confidence that they have some aspect of control over their health.

I design a health-related wearable system that monitors the heart rate of a user and provides the ability to identify anomalous readings that may require the attention of a medical professional. My goal with this project was not to implement a completely finished product deployed on a modern wearable device but rather to illustrate how improved user understandability and system perceived value can be achieved. Through this process I inductively identified the types of insights that shaped my framework, and later show what these insights can provide during such a design process and the types of activities it recommends to elicit such insights.

Given this context, my research questions center on the feasibility and utility of creating design frameworks for understandability within this application domain as well as on what is achievable with respect to monitoring heart rate data to provide insight into heart health. To wit:

- Are the design heuristics of contextualization, customization, interactivity, and simplicity beneficial to producing health-related wearable systems that are understandable to a wide range of users?
- What design activities are needed to elicit information that is useful to designing understandable health-related wearable systems?
- What are the relationships between these design heuristics and understandability and how do these elements contribute to the perceived value of a health-related wearable system by its users?
- What health-related insights are possible by tracking heart rate data from the sensors of modern wearable devices?

- What elements are needed in the display of heart rate data to make them understandable and provide high value to the users of health-related wearable systems?

This dissertation is organized as follows. I first present an overview of my framework and show how it can be used to evaluate the understandability and value of health-related wearable systems. I then demonstrate the process through which I design a health-related wearable system for monitoring heart rate data such that it provides users with insights into their heart health. To create the system, I designed and implemented five separate user studies. I provide a high-level overview of all five studies and then present each study in detail. Then, at the end, I present the final interface of the system and evaluate it with respect to how well it achieves the heuristics proscribed by the framework. Taken together, the five studies combined represent a comprehensive evaluation of the benefits achieved once the framework heuristics are present in a system. However, before I get into the details of all five studies, I will situate my work with respect to related work in. To illustrate the generic nature of my framework, I will also present some initial thoughts on how my framework can be used to design health-related wearable systems that monitor other medical conditions such as depression and Parkinson's disease. Finally, I conclude with a discussion on the benefits of my framework and its implications for the future design of health-related interactive wearable systems.

THE WISE FRAMEWORK

I refer to my composition of interactive system heuristics and activities for designing and evaluating health-related wearable systems as the WISE framework. WISE stands for Wearable Interactive Systems Enhanced; it is a framework that aims to increase the successful adoption of health-related systems designed for wearable technology. Such adoption can be achieved if the user perceives the system as valuable. It can be perceived as valuable if it makes it easier for users to monitor their health and/or medical conditions and convey information about their conditions to their doctors and other medical professionals. The other important goal of the WISE framework is producing understandable systems; a user will not adopt a system unless it provides valuable information that is easy to understand especially when first interacting with the system. The framework is organized around four main heuristics of interactive systems that are critical to maximizing the value and understandability of health-related interactive wearable systems: *contextualization*, *customization*, *interactivity*, and *simplicity*. I will describe each of these heuristics and provide examples of how each one plays a role in increasing the value and understandability of a system.

My framework also consists of a set of analysis and design activities that are critical to eliciting knowledge and requirements from the potential users of the wearable system being designed to help imbue the four main heuristics of the framework into that system thereby increasing its understandability and value to those users. The activities include both qualitative and quantitative methods and are applied in a highly iterative fashion as the design of the target system evolves. The activities include:

- detailed interviews with medical professionals and potential users of the target system
- user studies designed to elicit feedback on visualizations of user data that can be generated via sensors on modern wearable devices
- experiments that have participants record the user data of interest by wearing devices that have the required sensors

- surveys of participants that participate in user studies and experiments to capture as much qualitative data as possible to help identify factors that contribute to the understandability of the visualizations

- a commitment to performing multiple iterations of the visualizations of the captured user data with a goal of boosting understandability each round

- use of the "think aloud" technique to have users explain the meaning, value, and potential benefits of the designed visualizations and to provide insight into their understanding of the information contained in those visualizations

- use of intercoder reliability measures when analyzing user feedback and open-ended survey responses to gain high confidence in the interpretation of results

- frequent consultations with medical doctors to verify the utility of the visualizations with respect to enabling conversations with patients about their tracked health-related data

None of these activities are necessarily unique to this design framework; rather it is the combination of these activities and the goal of using them to gain insight into making health-related interactive wearable systems understandable that is novel to my work. I should also note that my framework is dependent on the concepts and techniques provided by software engineering life cycles to produce functional and reliable software systems. The WISE framework limits its contributions to providing insight into how to make wearable systems more valuable and understandable to their end users. Those benefits can only be fully realized if fundamental software engineering techniques are followed to produce an actual functioning prototype of the system that operates in a reliable fashion.

I will now discuss the four heuristics of the WISE framework in more detail.

Customization (a.k.a. Personalization)

A crucial property that health-related wearable systems can provide is the ability to customize or personalize the information they provide about the data they track such that it reflects knowledge about their users. For example, the definition for an anomalous heart rate reading means something entirely different for a middle-aged, sedentary adult than it does for a fit young athlete. When a system customizes its display to reflect a user's own attributes and conditions, it can lead to boosts in the understandability of the displays since the user can naturally anchor off of attributes in the display that reflect themselves. Increased understandability then boosts the value of the system to the user. Furthermore, when a system provides customized information to its user that information will hold a higher value for that user; it will feel as though the system is speaking personally to them. As a result, it is important to understand for a given medical domain—such as the health of a user's heart—in what ways can a wearable system customize its displays and its algorithms for its users. Thus, designers must understand the implications that various user characteristics—such as age, gender, height, weight, athletic ability, and medical conditions—have on how the analysis of a particular individual is carried out which might then, in turn, have an impact on when the system provides alerts to the user about anomalous conditions. We have found that one way to elicit this information that is accessible to UX designers is through iterative interviews with medical professionals as well as with users in the target population for the system. It is important to iterate as one will learn more about a domain over time allowing for more sophisticated questions to be asked in later interviews; in turn, this elicits more information on how a system can be customized for different types of users.

Contextualization

A second key property for health-related wearable systems is contextualization. I define contextualization as the ability of the system to help its users correlate the data its tracking with the user's actions and situations. A certain heart rate value while a user is active can be completely harmless, while the exact same value can be alarming when recorded at rest. When the system is able to identify or record the state a user was in—including both physical and emotional states—then it is able to contextualize the readings that it presents later when the data is being reviewed by a user's doctor or analyzed for trends over time. Achieving contextualization is one of the most challenging heuristics to add to a system. While sensors on wearable devices are becoming more sophisticated, they are still a long way from providing accurate contextualization. There are different methods that can be used to help achieve contextualization:

- Relying on the User: One way to achieve contextualization is to rely on a user to log their their major actions and activities throughout the day. Such an approach is not desirable since users are often reluctant to do this type of work after the fact and the main goal with wearable technology is to make things easier for a user.

- Partial Reliance on the User: A second way to add contextualization is to find ways to partially automate the capture of activities and other attributes that can later be used to contextualize the readings being tracked by the system. For instance, a system that captures the speed that its user is traveling can make inferences about whether that user is sitting, standing, walking, running, or driving. Using machine learning, such inferences can be done in an automated fashion and used to contextualize information displays at a later point. Advanced instances of this method might offer users the ability to override the automated classifications or have code that detects when it might be good to pop up a dialog to ask a user a simple multiple choice question to capture additional information about observed events in the data. For instance, a spike in heart rate without a corresponding increase in the activity of a user might be explained by the consumption of a caffeinated drink. Once enough of this data has been collected, the app can further contextualize data by recognizing patterns in the data and use those observations to help a user understand the impact their behaviors are having on their health.

- Third-Party Apps: A third method for contextualizing the data captured by a health-related wearable application is by utilizing the data contained in other apps on the wearable device or on another device that is carried by the user, especially data that by its nature would contain important contextual information, e.g., the calendar on a user's phone. As the user schedules events—such as a doctor's appointment, a workout, a set of important meetings, or a party—this information can be brought into an information display of the user's tracked health data. For instance, it is known by doctors that a visit to the doctor's office can cause stress in their patients and so an anomalous heart rate reading taken at the start of such a visit can likely be attributed to increased stress levels rather than a medical condition. They would need multiple examples of the anomalous readings across many days and different types of events before such readings would become a cause for concern. There are many types of applications out there that could provide important contextual information: workout apps, meditation apps, games, etc.

- Sensors: Finally, a fourth method for providing contextualization in wearable apps are all of the contextual clues that can be inferred from the various sensors on modern devices. Heart rate sensors are common and can provide insight into the activity level of a user; as mentioned above, accelerometers are useful sensors to have to compare and contrast the data coming from the heart rate sensors. Other sensors that may be used in the future are gyroscopes, oximeters, and thermometers. All of the data coming from these sensors can help to provide additional insight into what a user is doing.

These methods can be used by themselves or in combination to help provide context around the health data being tracked by a given health-related wearable system. With respect to our framework, contextualization is critical to help a user understand the data being presented to them and to place additional value on what the system is helping them to monitor.

Simplicity

Simplicity is a key property to increasing the understandability of a system for its users. If a system does not attempt to keep the information that it conveys to users in a form that is straightforward to understand, then the system is at risk for not being adopted or used. This is similar to the concept of progressive disclosure (Nielson, 2006). Progressive disclosure is a design philosophy that recommends showcasing only the most important features of a website or a desktop application at first launch. Additional features are then revealed when they are needed so that the overall user experience keeps a nice balance between functionality and simplicity. Nielson's work on progressive disclosure showed that this design approach was proven to boost the usability of web and desktop systems. In my own work, health-related wearable systems can generate a lot of complex information about a user very quickly. As such, striving for simplicity in the visualizations that these systems produce is essential to ensure users can understand the information being displayed. As I will show later, this property proved effective in my own user studies; once balanced simplicity was achieved I saw a boost in understanding by my participants especially those that did not have much experience with science given their educational background or with statistical data visualizations in general.

Health-related wearable systems can be quite sophisticated and, as a result, there can be pressure on designers to create interactions and interfaces that reflect the underlying complexity. What is needed is balanced simplicity; that is, an attempt to keep the information displays of a system simple and easy to understand relative to the complexity of the application domain or the functionality of the underlying system. In the user studies that I will discuss below, I found that users can be easily intimidated by a system at first glance. My initial visualizations would cause an intimidated user to declare that they did not understand anything from the display. In that instance, I would engage the user in think aloud techniques asking them to simply describe what they were seeing. Such engagement would eventually lead them to get past their initial fear and begin to understand the information being presented but one cannot rely on such interactions occurring for all users. As such, I had to iterate multiple times to find a set of elements that balanced the complexity of the information I was displaying with easy to understand notations and contextual clues to help them to want to engage with that information on their own.

These issues are especially important to tackle in the domain of health-related wearable systems; the reason for this is nicely illustrated by a quote from James Heckman, a Nobel Laureate in Economics who observed “There's a strong relationship between education and health, that the better educated people are healthier people." This observation provides strong motivation to prioritize the need for simplicity in the design of wearable health systems (Conti, et al., 2015). According to Heckman's work, it is possible that less educated people will find it more challenging to grasp the concepts conveyed in the visualizations of such systems when simplicity is not achieved. Yet that very same population might have a stronger need for these systems given the potential benefit they have to enhance their health knowledge and thus their health and the overall quality of their lives.

As an example of these issues in modern systems, consider the information made available in Apple's Health app on its iOS platform. The interface is simple and clean but certain design choices limit its utility for users interested in heart health. For instance, all of the important information on heart rate behavior that occurs throughout a day is gone since the display simply blends all the variance that occurs in an hour into a set of two numbers that show a user their minimum and maximum readings for that hour. A user can bring up a display that provides a lot of information about heart rate on a single graph but I observed during the studies I describe below that some of my participants with less scientific education found that information to be overwhelming and not understandable. As a result, a designer must strive for balanced simplicity in their displays to make the system more understandable and thereby increase the value of the system to its users.

One way to mitigate the complexity of information displays in wearable systems is to increase the interactivity of these displays which brings us to the last desirable property of health-related wearable systems identified by the WISE framework.

Interactivity

The final critical property of health-related wearable systems is interactivity. It is important for a user of these systems to be able to interact with and explore the data that is being presented to them via the system's displays. The benefits of interactivity are similar to the ones provided by active learning techniques in education (Brame, 2016) in which hands-on experiences have been shown to provide boosts in student learning and retention of concept inventories and thus student understanding of the domain being taught. Similarly, when a user is able to explore their tracked and analyzed data in a wearable system—e.g. changing the level of detail that they can see or jumping from one view of the data to a related view—these direct experiences increase engagement and understanding of the underlying information. As mentioned above, simplicity and interactivity go hand in hand and together can make complex data accessible and understandable to users who might otherwise struggle to understand it. An interaction may be able to teach a user on the fly about the type of information they can find and how to access it but only if simple explanations appear as the user is browsing the information in a visualization. Such a balance can add clarity, boost understandability, and thus the overall value of the system.

Evaluation of Existing Applications

One benefit of the WISE framework is that existing health-related wearable applications can be evaluated in terms of their ability to support the four identified heuristics and thus make inferences about the understandability of these systems and the value they provide to their users. Since I have made use of the WISE framework to design a health-related wearable application that monitors heart rate data, I have chosen to demonstrate the evaluative capabilities of the WISE framework on a set of existing applications that also track heart rate data.

I surveyed consumer applications on the market that track health data and observed that most of these applications centered on fitness behaviors. The main focus of these applications would be variables such as step count, calories burnt, workout zones to achieve maximum calorie burn, etc. Some of these applications would also include features to perform sleep analysis. Of the applications that also examine heart rate, they do so within the fitness frame as well. I identified four applications that specifically focus on heart rate and I conducted preliminary analysis of these four applications with respect to the WISE framework heuristics. I wanted to look at how each application fared in terms of customization, contextualization, interactivity, and simplicity. Each of the applications performed differently on the WISE framework, faring better on certain measures than others. Overall, I found that none of these applications achieved traction with all four heuristics of the WISE framework. As such, I believe there is room for improvement in health-related wearable applications that target heart rate data and heart health.

I now present the four heart-rate applications that I evaluated and provide context about how each one fared with respect to the four heuristics of the WISE framework.

FITIV Pulse Heart Rate Monitor

FITIV is an app that gives in-depth daily fitness stats such as steps taken, average heart rate, daily high and low heart rate readings, hourly average heart rate readings, and Apple's activity rings (FITIV, 2017). Features most related to heart health are its hourly average heart rate readings and a notification that warns users if they are overtraining.

In terms of the WISE framework, the FITIV application offers little in the way of contextualization other than the colored bands with labels such as “red line” and “warm up” (see Fig. 1). The application also offered almost no functionality to customize its displays to a given user and there was no interactivity in the app except to allow a user to tab between the various displays of the application. The displays themselves were not interactive. However, this application does do well with respect to simplicity. A user can fairly quickly gain a good grasp of the different workout zones and the percentage of their workout spent in that zone. The use of colors and numbers on the display are not overwhelming either. These choices increase user understandability of the data analysis.

Figure 1. FITIV User Interface

Heart Watch. Heart and Activity

The HeartWatch application (see Fig. 2) offers the ability to configure notifications for the following scenarios (HeartWatch, 2015):

- when heart rate exceeds a bpm value.
- when heart rate drops below a bpm value.
- a daily reminder to check your overall progress

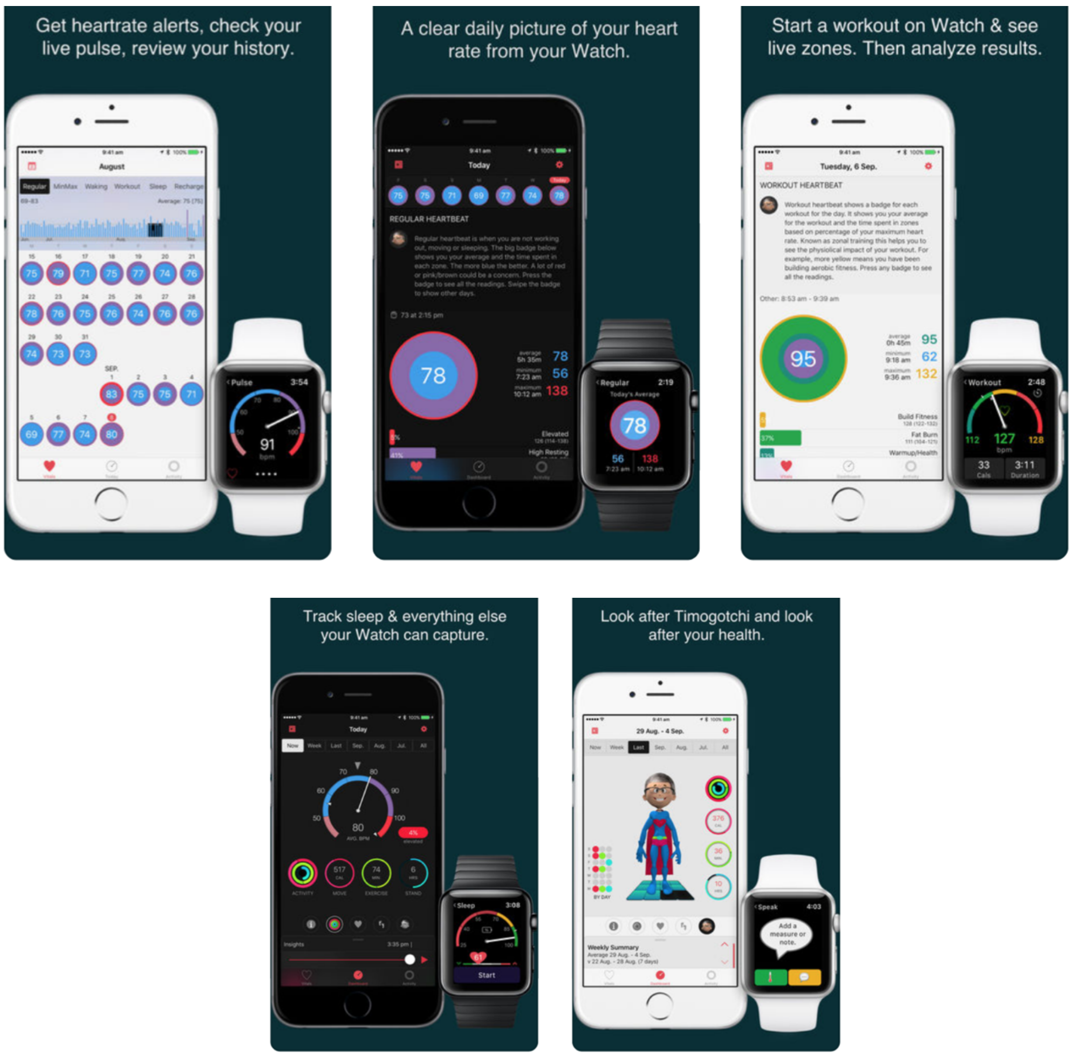


Figure 2. Heart Watch User Interface

Contextualization: There is no contextualization offered by this application and therefore no personalization. This lack of contextualization decreases the value a user sees in the analysis the application provides in terms of learning about their general heart health and behavior.

Customization: The application provides some customization. A user is allowed to select a specific maximum and minimum heart rate of their own choice. The app notifies a user when these boundaries are crossed. This feature is geared towards fitness and relies on the user to know his/her desired maximum and minimum heart rates. This selection is incredibly difficult to get right as these values not only vary according to age and gender but also by activity and the athletic level (i.e. fitness) of a given user.

Interactivity: This application is not very interactive. Like FITIV, it provides basic navigation among screens but not much beyond that.

Simplicity: The visuals of this application are focused primarily on one form of representation, namely—the activity rings popularized by the Apple Watch—and some profile infomation visible on some screens. A lot of the important analysis on heart rate performance is abstracted behind the rings and hence lost to the end user. This representation is too simple to allow its user to gain an understanding of their heart health.

Heart Graph

Heart Graph (Heart Graph, 2014) is an app that provides a representation of heart rate data similar to an ECG format (see Fig. 3). It provides different workout zones in coded colors and displays maximum and minimum heart rate.

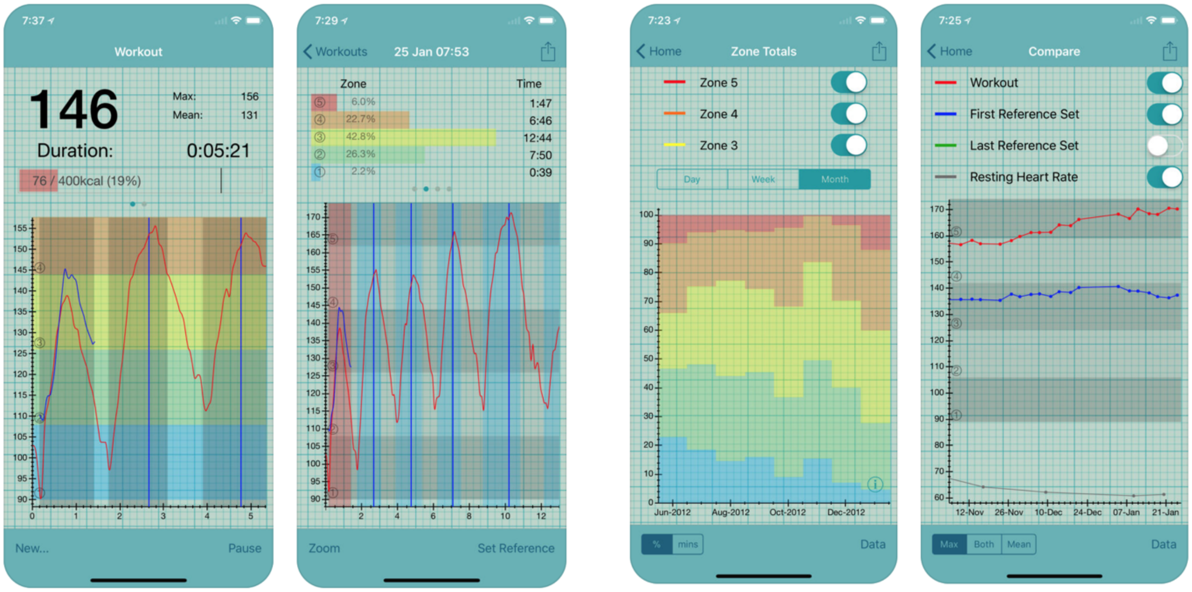


Figure 3. Heart Graph User Interface

Contextualization: There is no contextualization in these charts. This only speaks to workouts, which makes the information less valuable to the general user.

Customization: There is no health-related customization to allow the user to personalize the metrics of the analysis.

Interactivity: There is no interactivity, making it harder for the user to understand the presented information or to obtain additional information.

Simplicity: The data on the graphs here are too simple; little information can be understood at first glance. The workout zones are color coded for noting to allow users to understand when each zone was reached during a workout. However the lines on the graph are not easily understood.

Cardiogram

The Cardiogram app (Cardiogram, 2016) offers a timeline of a user's heart rate but with tabs, added by the user, that show your heart rate each day across events (see Fig. 4). This interface allows the user to track spikes related to stress, diet, or exercise if they log that data regularly. Also, this application will provide information on workouts that shows how many minutes a user spent in each heart rate zone.

Figure 4. Cardiogram User Interface

Contextualization: Visually the application developers did a good job in achieving contextualization. The tabs added on the chart label each data collection period with the action associated with it. They also utilize color coding to help the user distinguish these activities easily. However, this is achieved by heavy reliance on the user to log most of the information. They popup questions for the user to choose from a list of possibilities. Below, we show the contextualization support of Cardiogram in action.

First, a user is prompted (see Fig. 5) with symbols that can be used to categorize a particular reading.

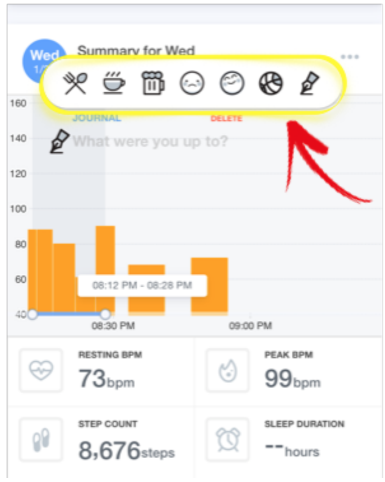


Figure 5. Cardiogram Contextualization Prompt.

Second, a user can add a note to explain the reading (see Fig. 6).

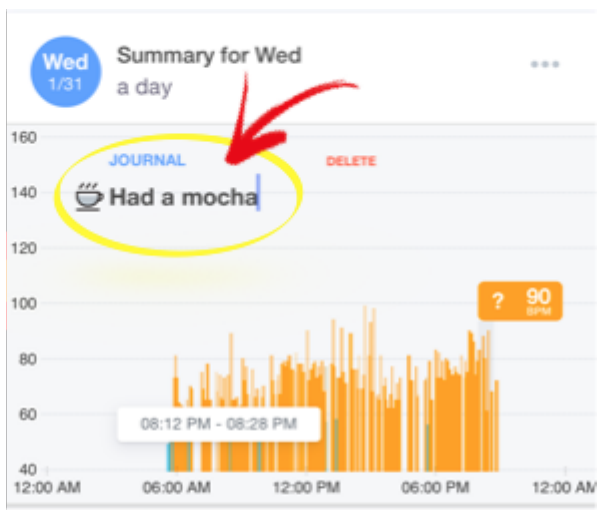


Figure 6. Adding a contextual note in Cardiogram.

Finally, the user's selections annotate all subsequent displays of that heart rate reading (see Fig. 7).

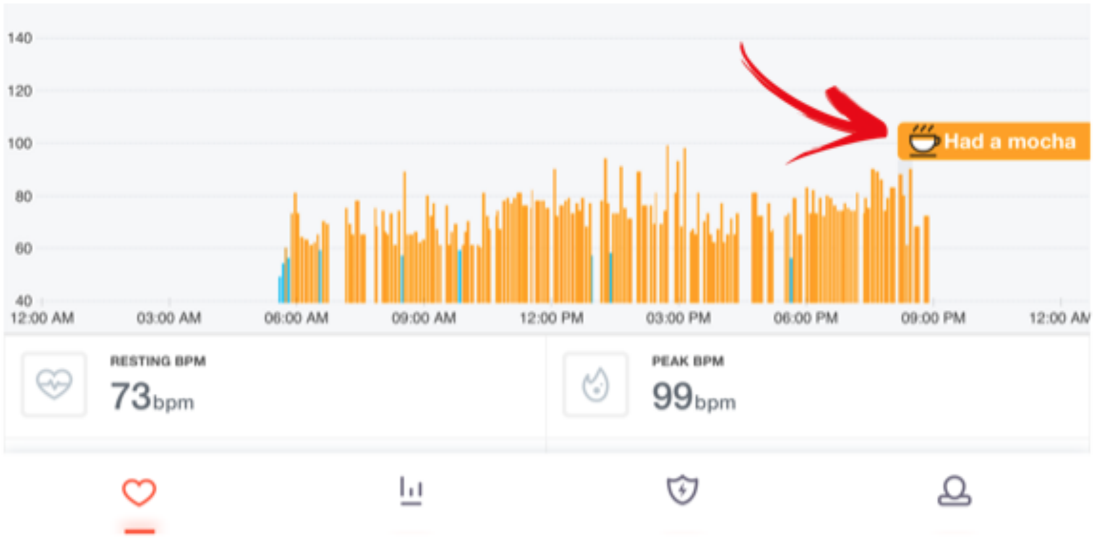


Figure 7. Contextualization in Cardiogram

Customization: The application provides a fair amount of customization; however the analysis provided by the application was not personalizable (e.g. the user is not able to learn their own health limits according to the information in their profile).

Interactivity: The graphs in this application are interactive. However the graphs did not offer a lot of screen real estate to give more flexibility for the user to explore with ease.

Simplicity: The design of the application and its visuals achieved a nice balanced level of simplicity. This makes it visually interesting and easy to grasp many concepts at first glance. However not much screen real estate is used to allow the user to view their data at the full limits of the screen of their device.

Discussion

The review of these four applications have demonstrated that the WISE framework can help to draw our attention to the support wearable applications provide to their users with respect to the four main heuristics I have identified that contribute to the understandability of a system and the value it provides to its users. In each case, I could abstract away details of all the features of an application and provide instead an overview that provides insights into the overall value of the application. I will now demonstrate how I can use the activities associated with the framework to iteratively elicit the knowledge and requirements needed to design health-related wearable applications that are understandable by their users and provide high value to their users with respect to tracking and managing their health.

RELATED WORK

As discussed above, my research involves producing a heartrate health wearable system that is understandable and is perceived as offering high value to their users. Through this process I identified the heuristics and activities of my framework, and then demonstrate its utility to design health-related wearable systems with high understandability and perceived value. As a result of this context, I situate my work in the fields of human-centered computing and health informatics. As a result, in this section, I talk about a variety of related work that touches on issues related to wearables as well as with respect to work that examines issues related to achieving particular design properties in interactive systems.

I situate my work within the tradition of other research that considers behavior change and wearable technology. Specifically, I examine issues related to user adoption and wearable device performance. Most of the literature on user adoption focuses on how to use wearable technology to monitor behavior change (Fritz, et al., 2014). However, the literature places less emphasis on how to further improve and make easier the experience of using self-tracking for the end-user of these systems.

For instance, Gouveia and colleagues conducted a study on behavioral change that tracked 256 users over a period of 10 months through a health tracking application called Habito (Gouveia, et al., 2015). They found that when researchers customized their recommendations to the technological ability of participants and then used certain gamification techniques, adoption of the tracking application increased anywhere from 20% to 50%. Another study—conducted by Glance and colleagues—sought to equip participants with Fitbit monitors over the course of four months (Glance, et al., 2016). Researches tracked step counts and gave participants a well-being survey, to see if there were any measurable changes before and after tracking began. They found that there was no measurable effect on a user's BMI. However, there were measurable reductions in non-HDL cholesterol as well as an improvement in the self-health and well-being of the participants. It is important to note that this study examined tracked steps and did not consider heart rate data. Nonetheless, it is an important study because it shows how a wearable device can lead to a measurable effect in self-rated well-being and how to induce gradual behavior change.

Other scholarly work—such as that conducted by Ye and others—examines the use of traditional medical devices to self-track medical data. With respect to Ye's work, researchers created a model and an algorithm to interpret data captured through an intelligent-tracking jacket which worked in sync with an ECG monitor. This study focused on a user population made up of five firefighters wearing these tracking jackets as they carried out their normal duties (Ye, et al., 2011). It offered a possible model that can be transferred to other tracking devices. In a second study, also focusing on portable ECG devices, Gay and colleagues monitored post cardiac surgery patients (Gay, et al., 2009). They equipped them with a ECG device and a personal digital assistant or PDA device. Then, the researchers gave participants tailored advice on their exercising patterns based on the tracked data. This research established a precedent in which tracked data could potentially help detect certain cardiac conditions. Researchers further demonstrated how to transmit reports to medical staff treating these respective conditions. In yet another study that deployed an ECG and a personal fitness tracker in parallel, researchers found that the personal fitness tracker was slightly less accurate than the ECG (Kroll, et al., 2016). Nonetheless such studies demonstrate the growing tendency of the medical community to investigate the potential of harnessing personal tracking capabilities. Other studies meanwhile have noted that there is a dearth of work that examines the accuracy and precision of wearable devices, particularly when it comes to medical applications (El-Amrawy Nounou, 2015). We also highlight the fact that in these studies an ECG device was used. We contend that an ECG monitor is not a practical device to carry out self-tracking with because it is cumbersome. It is also not easy for the layperson to understand and interpret an ECG reading intuitively.

Other user adoption studies, examine feedback design and device interaction. In this tradition, work by Gouveia and colleagues examined the importance of the glance (Gouveia, et al., 2016). They define this as a five second window that the user gets when he or she glimpses at their device to activate the screen (Gouveia, et al., 2015). The researchers believe that this glance is critical because it delivers important feedback in a short burst of time that then influences a user's engagement with the device. Other research led by Schmidt and colleagues designed a digital coach that helps highlight a user's respective strength and weakness (Schmidt, et al., 2015). Consequently, the application generates a personalized training plan to help users overcome motivational problems.

Performance studies make up a small cluster of user studies on wearables. For example, Guo and colleagues compared two major fitness trackers, the Fitbit and the NikePlus. They examined the performance of these trackers based on four indicators: how accurate the data was, the quality of the data provided, API availability, and user experience with the device (Guo, et al., 2013). Whereas Narasimha and authors compared sensor accuracy between an actual ECG and a mobile device ECG. They concluded that a smartphone was a much easier device to use (Narasimha, et al.,2016). Finally, there are studies such as those led by Goyal and colleagues (Goyal, et al., 2016) and Paul and authors (Paul and Irvine, 2014) that examine the security and user privacy issues arising from the pursuit of self-tracking. Overall these studies demonstrate the presence of real privacy issues attached to wearable devices that in turn operate across multiple levels and platforms, each requiring distinct privacy policies. Such studies indicate that a user should be made well-aware of these policies before they embark on self-tracking.

While this body of research does much to advance work on self-tracking and wearables, it seeks to improve fitness measures and does not necessarily seek to analyze medical conditions in a user-friendly manner. Much of this research also subselects on a sample of unique users. That is, the sample analyzed in these studies consists of specific members of a population with shared characteristics—for example, already diagnosed patients with medical conditions or, in one case, firefighting professionals. Hence, we know little with respect to how to target mainstream users who may not necessarily have any characteristics in common but nonetheless may have symptoms of concern. These wearable device users whom are unsuspecting of having a medical condition are a key segment that I am to support in my research. I do so because these individuals often do not profile themselves as patients-in waiting and these users require systems that are easy to understand to get them to use them and then provide significant value to keep them using the system long enough for the system to be able to identify anomalous conditions that convince the user to seek medical help.

With respect to creating visualizations for health-related wearable systems, an important distinction must be made based on the purpose of those visualizations. In general, good visualizations created for the purpose of analysis require different criteria than for ones created for the purpose of presentation. It is thus important to tell the difference. Well designed visualizations for the purpose of presentation require two important criteria: memorability and engagement (Robert Kosara, 2016). Presentation-oriented techniques need to be specific rather than general and compact rather than scalable to get an idea across to an audience. Visualizations for analysis would not benefit from such criteria as their use is allow a user to identify patterns across a large amount of data and to encourage them to ask plenty of questions about that data.

The numerous iterations that I performed in my user studies (discussed below) on the visual design of my health-related wearable system are geared to a general population. The main purpose of those visualizations is presentation. As I will present, the evolution of my heart rate visualizations was to make them more distinct with features such as colors and with fewer but more specific and clear labels. These changes help them to become more memorable to the user. Adding the element of interactivity to my visualizations made them more engaging and stronger in their ability to convey the desired information to the user.

Another important concept that I implemented as my visualizations evolved via the feedback from my participants was the overall way in which heart rate information was presented. These changes were in alignment with the lessons developed by Shneiderman (1996). A key set of rules to follow when designing information visualizations for user interfaces is to first offer an overview, then the ability to zoom and filter, and then provide details-on-demand. As is the case with my framework, we found that applying balanced simplicity to achieve better user understandability was key. In Shneiderman's approach to query information visualizations, it helps to present information quickly, while simultaneously allowing the user more control when exploring the data. This concept is especially useful in wearable environments where high resolution, large, color displays and fast data retrieval are not as available as they are in desktop environments.

It is important to note that the WISE framework is not meant to serve as a standalone software development life cycle. It very much focuses on guiding developers to creating wearable systems that are understandable and valuable to their users. This narrow focus means that designers and developers must still draw upon other techniques to produce a functional and usable system. One example of this type of work is an important set of heuristics developed by Jakob Nielsen to produce usable interactive systems (Nielsen, 1995). His heuristics involve everything from making the state of a software system visible to its users, focusing on recognition not recall, making use of aesthetic design elements and helping users recover from errors. All of his heuristics are are essential to designing a system with a high level of usability and should all be followed in parallel to any work being done in service to the goals of the WISE framework during a software life cycle. The WISE framework is not targeting usability; instead it's goals are to produce systems with high user understandability and high perceived system value. The WISE frameworks heuristics thus compliment the goals of Nielsen's usability heuristics, leading designers to create systems that have high potential in being adopted and used long term.

APPLICATION DOMAIN

I designed a health-related wearable system that tracks heart rate data with the goal of providing its users with insights into their heart health. In this section, I present an overview of this application domain and describe at a high level the five user studies I conducted to create this system; it was by doing these studies that I was able to identify the elements contained in my framework.

Members of the public who have heart-related medical conditions often have to deal with invasive methods for tracking the state of their health. Continuous heart-rate monitors are bulky. If a patient would like to conceal that they have a medical condition—as they might in the workplace—these awkward monitors have to be worn uncomfortably under clothing. I believe that wearable technology has great potential for helping members of the public monitor their basic cardiac health, and has the potential to alert users when a concerning condition is present. In my research, I am interested in exploring the design space of what medical conditions can be detected or monitored via the existing set of sensors that are deployed in wearable products today, such as fitness trackers or devices like the Apple Watch.

In my work, I am aiming to enhance the use of monitoring techniques that are considerably less invasive and demanding on our user population. My goal is to make an actual difference in how people perceive wearable technologies in terms of personal health awareness and management. Products like the Apple Watch or the Fitbit can give endless data on a user's heart rate, or the number of steps taken; however they do not have the ability to analyze and visualize that data in a way that would be understandable and of high value to their users.

Given the recent widespread deployment of wearables that contain heart-rate sensors (such as the Apple Watch), I am especially interested in the insights that can be provided by the monitoring of a person's heart rate and the benefits that can be achieved with the proper analysis of that data for a given user. I conducted a number of studies that had the following goals: a) learning the potential benefits of wearing a smartwatch with heart rate tracking, b) investigating the possible insights that these devices can offer physicians, c) experimenting with gathered sensor data and exploring the outcomes available from different types of analysis, and d) studying how the users perceived the representation of the analysis and to learn about their sentiments and perspectives on the results of my work. To achieve this, I conducted five user studies (see Fig. 8):

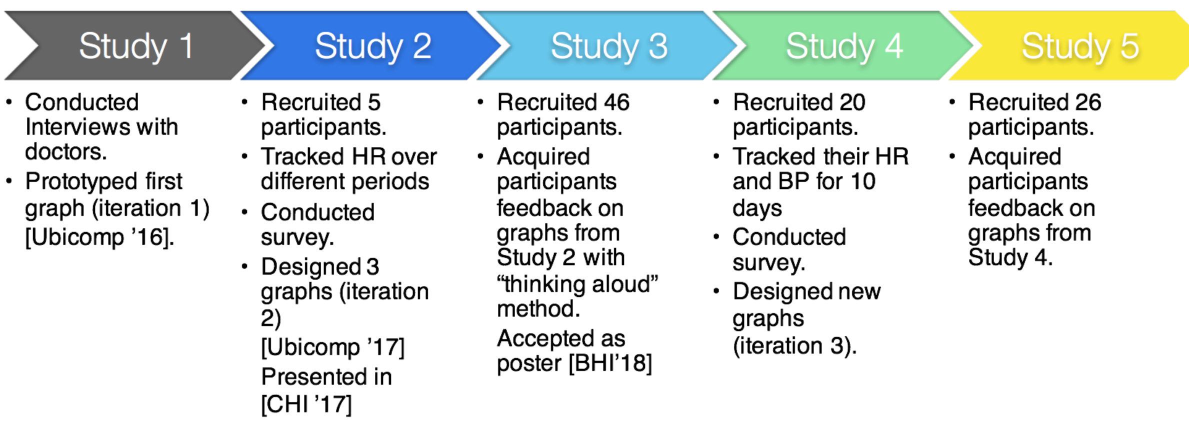


Figure 8. Overview of my five user studies including references to papers that were published as a result of each study.

High-Level Overview

In this research, I seek to gain insight into user experiences with a smart watch in the context of tracking heart rate data using a mixed methods approach (Creswell and Clark, 2006). Such research involves investigating the problem by blending insights gained from both qualitative and quantitative forms of inquiry (CDC, 2011). To aid my research objectives, in study 1, I interviewed cardiologists to gain their insights on the background of how such a wearable application could benefit an individual user or help to enhance doctor-patient interactions. I also designed the first iteration of a visual display for heart rate data. I then started my second study, guided by the information learned in Study 1.

For Study 2, I tracked five participants in a pilot user study and designed the second iteration of the heart rate displays. In Study 3, I collected the feedback of forty-six participants on the displays produced as a result of Study 2 (Albaghli, et al., 2017). I examined the perceptual statements made by our participants about the heart rate displays and conducted a thematic analysis of their feedback to help guide my work in the next set of studies.

In Study 4, I collected the smart watch data of twenty participants. These participants were also asked to manually use a medical device (an electronic blood pressure device) multiple times a day to measure their HR and blood pressure while simultaneously activating the heart rate sensor on their watch. I later conducted a survey instrument and drew on qualitative forms of inquiry to amass knowledge on the subjective experience of our respondents during Study 4. I once again used the results to create a new version of the heart rate displays. In Study 5, I worked with an undergraduate research student to create an interactive software prototype of the new version of the heart rate displays. I then recruited twenty-six participants to once again provide their feedback on the displays with respect to understandability of the information and perceived value to the end user. Sixteen of those participants had participated in prior studies allowing me to see how their perceptions changed as the visual design of the heart rate data evolved.

In summary, the qualitative part of my research occurred via the interviews in Study 1, and the user perception analysis that occurred in Studies 3 and 5. The quantitative part of my research occurred in Studies 2, 4, and 5 when I analyzed the tracked heart rate data of my participants. I also performed quantitative analysis of surveys administered at the end of each user study and conducted descriptive analysis of my datasets to graph these descriptives collapsed over different periods of time and themes, with a particular focus on outliers.

Study 1

The first study was a qualitative research study consisting of interviews with medical professionals, focusing, in particular, on cardiologists. The aim of the study was to explore the possible scenarios my proposed wearable system could support and identify the benefits that wearable technology with heart rate sensors could provide to its users independent of their medical condition. During that study, I developed a high-level software architecture for my system and developed an initial visual design for the way in which the system would report the results of its analysis to its users. As such, the primary activities that I made use of in this study were interviews and prototyping.

Study 2

The second study was a preliminary pilot study that tracked five participants. Each participant was asked to wear a fitness tracker—in this case, an Apple watch—to track their heart rate for ten days. Some participants already owned an Apple watch, and volunteered to provide us the data from before the ten days tracked in this particular study. We then asked each participant to fill in a survey once tracking was complete. Finally we designed a set of visual representations based on the tracked user data as well as the information gathered from the interviews of Study 1, and the survey information provided by the participants of Study 2. As such, the primary activities that I made use of in this study were analysis of interview data, surveys, user tracking, and prototyping of visual designs.

Study 3

In Study 3, I recruited forty-six participants to evaluate the visual designs produced by Study 2. Participants were asked to explain what they saw in the designs in terms of the meaning of the presented data, the possible benefits, and likes and dislikes. They were also asked to provide feedback on how the designs should be improved. During their evaluation, we asked the participants to think aloud while being voice recorded. They were also given the freedom to type their response if they preferred. All responses were recorded and analyzed using intercoder reliability. The primary activities were feedback elicitation and analysis. Through this study I started to identify the heuristics of the framework.

Study 4

For Study 4, I recruited twenty participants and tracked their heart rate data for ten days using two different types of devices: an Apple Watch and an ECG medical device. This was done to assess the accuracy of the Apple Watch by comparing its readings with the readings produced by the ECG while wearing both devices at the same time. Once the data tracking was complete, participants filled in a survey. I then produced a new version of the visualizations guided by the conclusions learned from Study 3 and Study 4. In this study I made use of activities such as user tracking, surveys, and prototype design. The goal was to ensure that our analysis based on data produced by wearable heart rate sensors was accurate enough to rely on. It was also to gain insight into the issues that arose while a user made use of a device for an extended period of time. The surveys provided insight useful to forming the framework more precisely, and achieving a subset of the WISE heuristics and touched on issues of understandability and user-perceived value.

Study 5

In Study 5, I recruited twenty-six participants. Each participant was presented with an app that implemented the new visualizations. They were asked to interact with the app's visualizations and explain what they saw in terms of the meaning of the presented data, the possible benefits, and personal likes and dislikes. They were also asked if they had any suggestions for improving the designs. I asked them to think aloud while being voice recorded during their evaluations. They were given the freedom to type their response if they preferred. The primary activities were feedback elicitation and analysis as well as verification that changes to the designs led to boosts in user understandability and user-perceived value of the final designs. The goal was to evaluate the presence of the key heuristics of the WISE framework in the visual designs.

Summary

Over the course of the five studies, I engaged a variety of stakeholders: medical professionals, people with medical conditions, healthy adults focused on fitness tracking, etc. and used their feedback to perform three iterations on the visual designs. Some participants were included in more than one of these studies. Figure 9 provides an overview of how participants were distributed across all five studies.

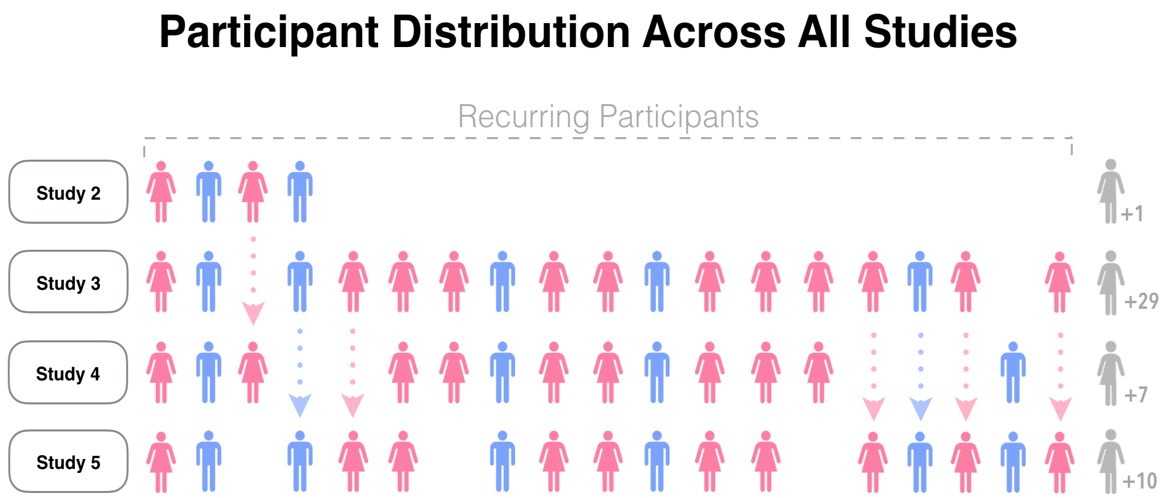


Figure 9. Some participants were included in multiple user studies. Each column represents one participant and each row shows which studies to which they contributed. On the right, participants in grey represent the remaining number of single-time participants per study.

I performed a user study to assess the accuracy of the heart rate sensors deployed in one commercial wearable device to gain confidence that the information my system generates would indeed be information that users could rely on when in conversation with medical professionals and to gain additional feedback on my visual designs. Finally, I worked with an undergraduate research assistant to design an interactive version of my visual designs to get data related on the interactivity property of the WISE framework and to prepare for the development of a more substantial software prototype of the entire system. As I will discuss below, over the course of these five studies, as the framework heuristics came into focus, I gathered evidence that the WISE framework helped me to create the design of a health-related wearable system that has high understandability with respect to its visual displays and has great potential to deliver high value to its users.

I will now present the method, analysis, and results of each of my five user studies. I will then present how the WISE framework could be used to develop designs of wearable systems for other medical conditions before presenting my conclusions.

STUDY 1

The first study was used to elicit information about the domain of heart rate data and the types of scenarios that could be supported by a wearable system that tracked this type of data.

Method

I conducted multiple interviews with seven doctors, with a focus on cardiologists, to understand what can be learned about the health of an individual via their heart rate. I believe that an area of great potential in this area is using a wearable tracker to detect easy-to-miss symptoms much earlier than usual and advising individuals to see their doctors when detected. Furthermore, the fact that such devices are worn on the wrist means that work in this area has great potential to be considerably less invasive than current methods and devices used by doctors today. In general, I was interested in how the information from these interviews can be used to design new applications that helps users understand their heart rate performance and behavior in different situations to potentially aid those individuals in monitoring (or detecting) existing heart conditions as well as helping them better convey information about their conditions to their doctors. My initial work in this area has been published in a UbiComp 2016 workshop called “Designing, Developing, and Evaluating the Internet of Personal Health” (Albaghli et al., 2016).

In my interviews, doctors were asked to share information about how they make use of heart-rate data in their own practice and to offer suggestions for the types of analysis that could be applied to continuously-monitored heart-rate data. They also provided insight into the types of visuals they would want to see generated by our proposed wearable system.

Analysis

The content of my interviews with doctors were analyzed in a deductive approach to find potential applications that utilize tracked heart rate data from wearable technology. I looked for applications that could be used to help users with or without diagnosed medical conditions or symptoms. Therefore my analysis considered different scenarios and personas to generate the applications discussed in the results.

Results

Potential Applications

My interviews identified four scenarios for a health-related wearable system that tracks, analyzes, and displays heart rate data. These scenarios provide insight into how developing monitoring apps for wearable devices can serve both patients and doctors.

Managing Conditions

One scenario is to develop an application that allows an individual to record whenever they felt they were experiencing a troubling symptom. These user events could then be correlated to the readings captured by their device and examined by a doctor to determine if the event, or readings, are related to a medical condition under investigation. The doctors indicated that there are devices that do this currently but they are more invasive than modern fitness trackers.

Highlighting Missed Symptoms

Cardiac arrhythmias (tachycardia, bradycardia, and atrial fibrillation) are disorders that can be detected by checking heart rate. These disorders involve slower or faster heartbeats as well as fluttering beats. Arrhythmia can occur in healthy people; however, as the frequency of the irregular beats increase, the chance that an individual has a disorder increases. The doctors said that these conditions can go unnoticed, especially with older populations. Younger people often self-report, but even then it would be helpful to have a non-invasive screening tool that can detect and track potential disorders. An earlier diagnosis would allow for earlier treatment and avoid complications. One concern here is that it is not yet known if trackers collect enough data to reliably detect cardiac arrhythmias.

Better Awareness of Self-Health

A third scenario involves motivating users to examine and track their heart rate regularly and to be on the lookout for alarming highs and lows given a user's gender, age, and activity level. A challenge here will be determining how many times anomalies must occur before a warning is generated. An additional challenge is getting a user to understand their readings and how those readings relate to what is expected for their individual situation. A final challenge is then motivating users to change their behavior to help improve their condition if anomalous behavior is detected. The work in my subsequent user studies focused mainly on exploring this particular scenario.

False Positives at the Clinic

The final scenario is situated around a visit to a doctor's office. Cardiologists are often faced with situations where a patient's data, collected by monitoring the heart through any number of tests, present concerning numbers but these patterns turn out to be false positives. One reason for this situation is where patients are nervous just to be in the cardiologist's office in the first place undergoing multiple tests. Having access to a user's longitudinal data could then help a doctor determine whether a patient's readings represent a real condition or were caused by anxiety.

Prototyping First Heart Rate Display and Initial Software Architecture

For this domain, I decided to make use of the Apple Watch as the primary device I would use to conduct my user studies. The Apple Watch is widely deployed and provides a well-developed software ecosystem for developing wearable applications. With respect to my application domain, the heart rate sensors on the Apple Watch measure the heart rate through the use of green LEDs and light-sensitive photodiodes. The heart rate is measured throughout the day. WatchOS then synchronizes the sensor data with the user's iPhone. This data gets automatically stored and displayed in Apple's Health app. With the use of HealthKit (Apple, 2014) my own heart-rate app can retrieve that sensor data and store it in a cloud-based database for long-term collection and analysis. Using this approach, my app can process this information and generate alerts as needed. This software architecture is shown in Fig. 10.

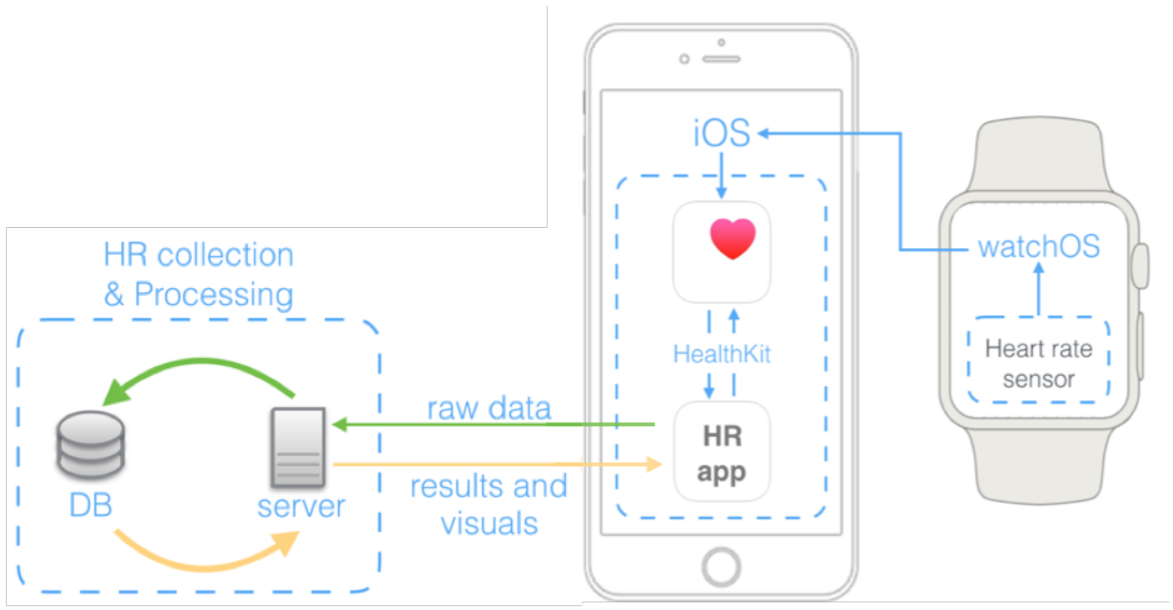


Figure 10. The proposed software architecture for my health-related wearable system.

To ensure user privacy, I would ensure that all user data would only be stored in the cloud after their consent. That data would be stored on our server and linked back to their device using techniques provided by Apple's HealthKit. If a hacker obtained my server's data, the information could not be linked back to actual users, since the anonymous user id only makes sense from within the context of my heart-rate application running on a specific device. If I were to develop a complete software prototype, my privacy policy would follow Apple's recommended practice and would explain that all user information would be stored on their devices while only their heart rate data and an anonymous user id is stored on my servers. Likewise, analysis results will be stored and viewable only from user devices.

I have designed a prototype visual that can help physicians better understand a certain aspect of a user's history (see Fig. 11). This visual can also provide users with a deeper knowledge about their health and cardiac behavior. This particular design represents iteration 1 of the visual designs my wearable app would use to increase user understandability and value. The visualization would graph the user's heart rate in beats per minute (bpm) across time. The background would highlight the safe zones for heart rate readings. The red area would highlight readings considered to be too high or alarming given the user's age, while grey would show readings lower than the healthy minimum expected of the user's age. These are in gradient colors in order to convey the higher alarming state as the gradient gets more intense in color.

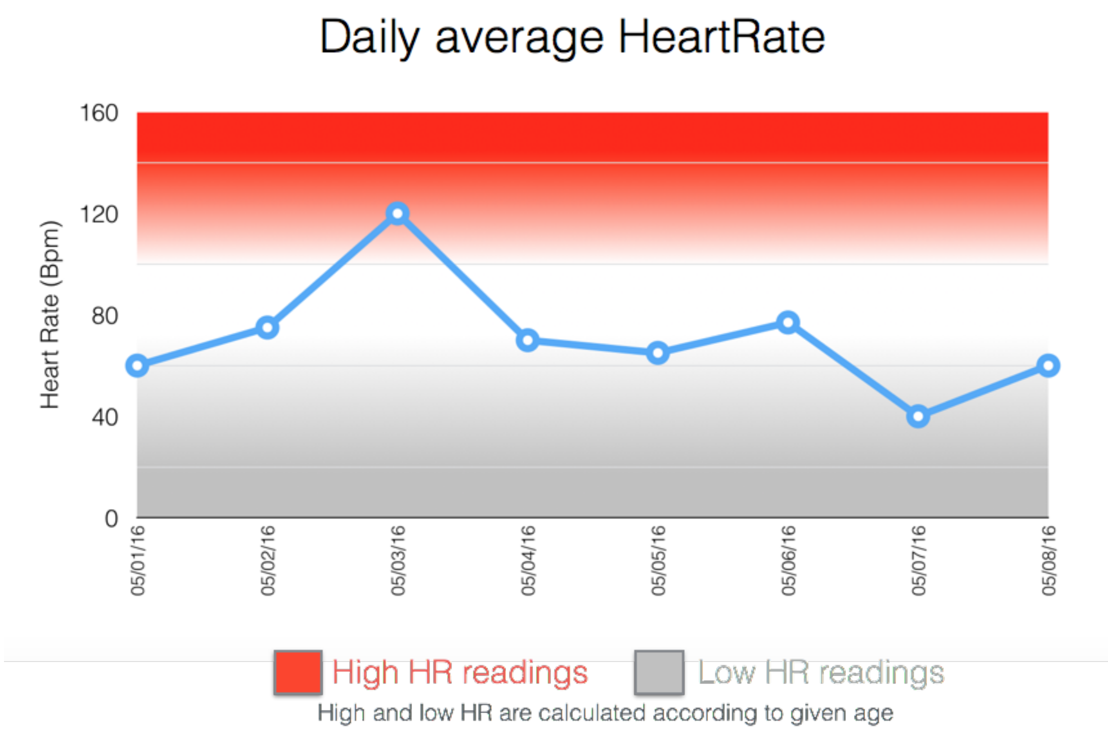


Figure 11. The initial design of a visual display of heart rate data.

Summary

At the end of Study 1, I had a better understanding of the needs of patients who have heart conditions and the types of scenarios that a wearable application could support around heart-rate data. I used that information to create an initial sketch of a visualization for heart rate data that could be used to elicit feedback from potential users within this domain. All of this work, set me up to begin Study 2.

STUDY 2

Study 2 was the first to gather tracked heart rate data from users with an Apple Watch. We generated a set of visual designs for presenting their data to them in a variety of formats. These visualizations were inspired by the initial visual design from Study 1 but were also influenced by the data received, information from the doctor interviews from Study 1, as well as survey data from the participants in Study 2.

Method

As mentioned above, I recruited five participants for this study who volunteered their heart rate data while wearing an Apple Watch. The data collected from the participants covered different periods, ranging from 10 to 500+ days. Snowball sampling was used to select participants while recruiting was done through word of mouth and acquaintance introductions. Some of my participants owned an Apple Watch and were using its data collection features already. I collapsed the collected data into different time periods of use to better compare new, intermediate, and long term use of the watch. The participants were all adults and included three male and two female participants with ages ranging from 32 to 63 years. The participants were asked to fill out a questionnaire after their heart rate readings were retrieved from their watches. Half of the survey questions were written with the aid of a cardiologist. The survey was designed to help me identify information related to underlying cardiac risk, general health, and environmental behaviors. The questions were designed to fulfill the following goals:

- Gain a preliminary understanding of the participant's health, activity, and dietary behavior.

- Learn about a participant's experience and what their self-rated health perception is, and how it might have changed with the consistent availability of this type of data.

- Understand how demographics and socioeconomic factors might play a role in the potential benefits of using data acquired through wearable technology.

- Get a deeper understanding of participant experiences and how the watch may have caused changes in routine or behavior.

I then designed the second iteration of heart rate visual representations. These were mainly designed according to the lessons learned from the doctor interviews from Study 1. We also considered the user input we gained through the survey results provided by the five participants of Study 2. The visualizations were designed using R, a language and environment for statistical computing and graphics. We chose R for the wide variety of statistical and graphical techniques it offers and the fact that it is highly extensible.

Analysis

For Study 2, I analyzed the survey data from the five participants and conducted descriptive analysis of their answers. I then proceeded to graph the tracked heart rate data from each participant collapsed over different periods of time with a particular focus on identifying outlier readings. This exercise allowed me to then design a range of visual representations for this data that I discuss below in the next section. I wanted to see which representation would give us the most insights without making it too complex for the average user. In the end, I found that the box-plot graph achieved this goal better than the rest.

Results

First, I looked at heart rate data over time for each participant. I wanted to visualize the data in a way that shows the following information in a simple snapshot:

- Where are the majority of heart readings clustered? This is important because a user can use it to quickly see their average resting heart rate measure, making it easy to notice deviations if/when they occur.

- Where do the max, min, and average heart rate values within the distribution lie?

- How do their readings compare to the normal healthy range of a resting heart rate?

- How many outliers are there? How frequent are they and what are their values?

Participants not satisfied with their health care providers have indicated a higher appreciation of the value a watch can offer. The participant with the longest period of use emphasized the fact that the watch is always on and continuously collection data as the most important aspect of why they wear the watch. Overall, the watch did encourage activity and offered motivation for use.

When participants were asked to rate the ease or difficulty of acquiring data from the watch, 60% of respondents reported that it was somewhat easy to access the data. When the same participants were asked if they would recommend the watch to a friend, 3 out of 5 participants said they would. 80% of participants also self-reported that their healthcare provider was probably aware of their medical history, but they were not certain. On this point, wearable tech data can inform the healthcare provider with specific information about the patient that would be helpful in understanding the patient's overall medical history. My data analysis showed that Participant 4 used the watch for 137 days, which for this study I coded as an intermediate user. This participant self-reported a mild cardiac condition, and mentioned that the watch helped them in understanding their symptoms better, and helped them understand their physical limits so they knew when to relax.

These insights were derived from analyzing the responses to the questionnaire. However, the sample size in this phase of my research was small and therefore my results are limited. In addition, none of the participants that contributed to this study were cardiac patients. A large amount of questions in the questionnaire were designed to help identify cardiac patients or people with any cardiac problems, even if mild. This further limited the utility of this study.

However, for my larger tracking study (Study 4), I address this limitation and draw deeper conclusions since I increased the number of cardiac patients in my participant population. Study 2 helped in learning how to improve the design of the questionnaire; in particular, to help make it more understandable to my participants and more useful for my analysis needs. In addition, as I went through the recruiting process and data gathering for Study 2, I encountered a number of obstacles—even though Study 2 had just five participants—that I had to overcome. As such, I learned valuable skills performing this study that made it easier when I got to the larger sample size of Study 4.

For instance, one participant in Study 2 had tracked data for one day only; I had to replace that participant with someone else. Most likely that participant did not charge the watch properly every night. A second obstacle that I handled was related to exporting the collected data from the watch; I went through several methods and attempts to successfully achieve that and, as a result, that aspect of Study 4 ran smoothly. In addition, upon implementing the questionnaire, I realized I needed more generic questions about lifestyle and diet, and had to add them to the questionnaire in a way that made sense to the user and that led me to make amendments to the study's approved research protocol. As a result, I am now experienced with that aspect of experimental design. I also learned that I will need to conduct a number of interviews with some of the participants, as well as with cardiologists, to better understand the results and I adjusted my plans accordingly to ensure that this issues were covered in Study 4.

As part of Study 2, I developed a number of visual representations to display a participant’s heart rate data after reviewing all of the collected data from my five participants. I designed the three visualizations described below to emphasize different aspects of the data to the user, satisfying different goals.

Graph 1 in Fig. 12 is basically a box plot. I made use of a box plot for the way it represents the heart rate data. It nicely abstracts away a large number of readings that reside within the box instead showing where most of the readings were clustered. It then highlights the maximum, minimum, and median of these readings and then displays those readings that are outliers as dots above and below the box. This representation delivers a balanced amount of tracked heart rate data that users will both understand and value.

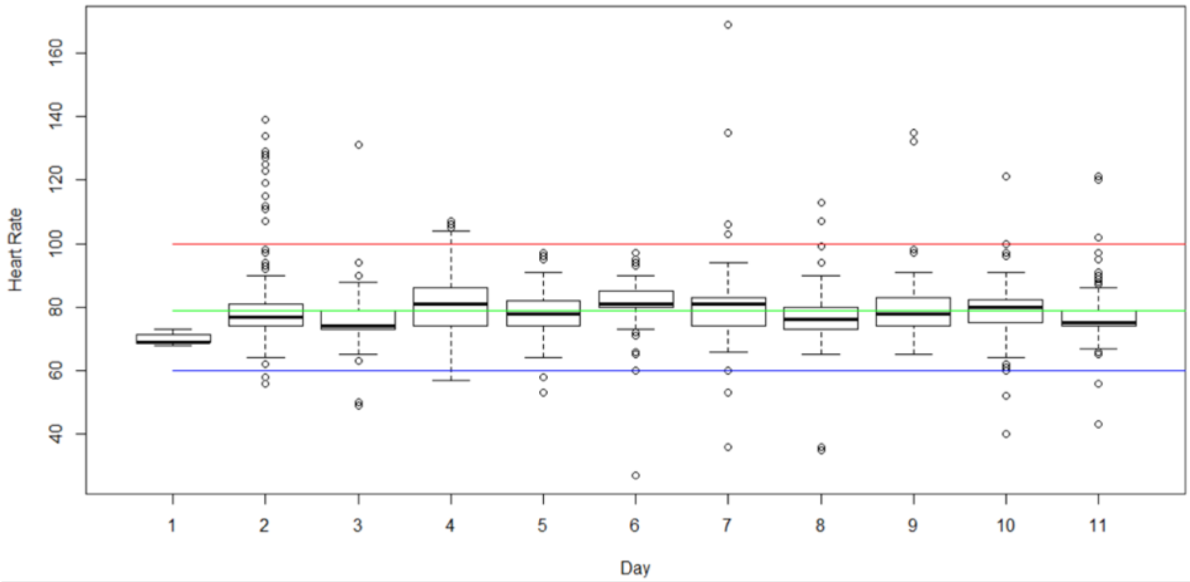


Figure 12. Graph 1 from Study 2: Snapshot of heart rate data over a short time period.

Graph 1 aims to display the variability of heart rate performance over a short time span. This graph is designed for a period between 5 to 12 days. There are three horizontal lines added to the graph to highlight the maximum and minimum normal heart rate limits in red and blue, and the overall average represented by the center green line. The median of each day is represented by the bold black line included in each box. Outliers are defined as data points that are distant from most of the other data points in a data set. Most data points lie within the boxes, hence the boxes present the range of the majority of the readings. As mentioned above, outliers are represented by the black circles lying beyond the limits of the box. Since this graph covers a relatively small period of time, it offers a user a snapshot of their heart rate data at a high level of granularity.

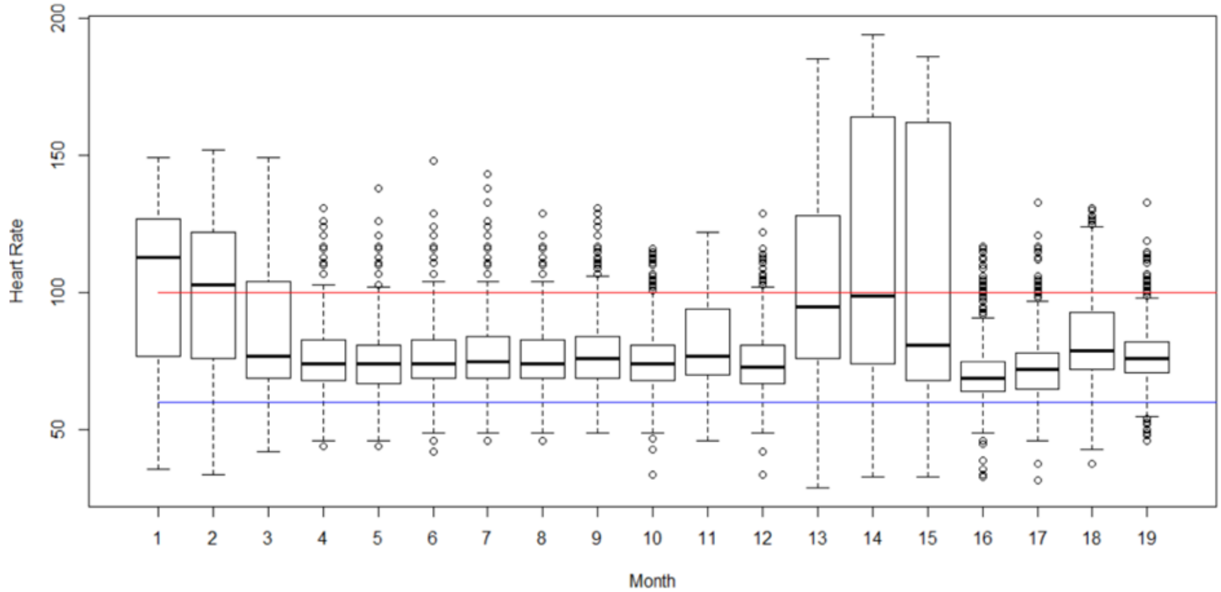


Figure 13. Graph 2 from Study 2: "Big Picture" view of heart rate data across months.

Graph 2 in Fig. 13 shows how the resting heart rate has been changing over a longer period of time. This graph is significantly less granular than Graph 1. It is designed to help a user identify behavior changes or events that occurred with their heart health on a broader scale. This graph again makes use of the box plot technique to visualize the distribution of the heart rate data. The period presented in Fig. 13 is 19 months. Healthy heart rate ranges are presented in the blue and red lines to show recommended maximum and minimum heart rate readings for a generic user. This generic standard, taken from the medical literature, is between 60 and 100 bpm (beats per minute). At a glance, we can see that for this user, their median heart rate readings lie mostly between the 60 to 100 threshold, which the doctors I interviewed in Study 1 told me would be interpreted as a good sign with respect to heart health. However, months 1, 2, 13 and 14, either surpassed the 100 bpm maximum threshold, or approached it closely. This particular graph would prompt a doctor to ask about those months to get more contextual information. For example, they might ask if the user was going through a particularly stressful situation in those months or if they changed how they exercised. Through this visual display, a user or doctor's attention is brought to these changes in readings and finding the reason(s) behind this change can lead to a better situation for the user and their health. This version of the visual design highlighted the need for both customization and contextualization since both would be useful in the interpretation of these data points.

Figure 14. Graph 3 from Study 2: Highlighting Alarming Readings

Graph 3 in Fig. 14 is called an out-of-range graph; it makes use of a scatter plot but only shows heart rate readings that are considered too high, too low, or are significant outliers based on other readings gathered on that day. The three different groups are shown with different colors and a legend provides information on the size of these alarming readings with respect to all of the heart rate data gathered for a user. That percentage will vary by user and its significance would have to be determined by a user's doctor. The idea behind this graph is to remove data related to "normal readings" and get a user to really focus on anomalous readings and if they are concerned by them to get them motivated to go talk with their doctor.

Summary

At the end of Study 2, I was well positioned to move from simple data gathering to more advanced user studies. I had gained experience in survey and experimental design and I had gathered real heart rate data that I could use to generate the next round of visual designs for my wearable applications. I generated three different types of graphs to present heart rate data and I was now ready to have those graphs evaluated with respect to their understandability.

STUDY 3

In Study 3, I recruited forty-six participants to evaluate the visual designs produced by Study 2.

Method

I generated three graphs that served different goals in Study 2 (Albaghli, et al., 2017). As discussed above, these graphs were designed for users of wearable devices with heart rate sensors. Each graph served a different high level goal, providing a user with the ability to review heart rate data over short and long time spans and to review readings that are considered outliers.

For Study 3, I recruited forty-six participants. My participants were recruited through word of mouth and verbal consent was obtained. Details on the demographics of my participants are shown in Table 1. The purpose of this study was to have the participants review the graphs generated in Study 2 and provide feedback on them. Participants were given the freedom to choose between providing feedback via text or audio. Thirty of the forty-six participants chose to deliver their feedback via audio; three provided a mixture of both, while the remaining thirteen texted their feedback. All data collection occurred through a messenger application that is available on both iOS and Android platforms. All participants used this application daily and were thus familiar with it; as such, no unnecessary inconvenience occurred on the part of the participant as I collected the data.

Table 1. Demographics for Study 1 Participants.

Gender	Female 28			Male 18	
Age	18 – 25 11	26 - 35 13	36 - 45 15	46 - 55 2	55 - 65 5
Highest degree earned	Highschool (HD) 8	Bachelor - non science (B) 13	Bachelor of Science (BS) 12	Master (M) 8	PhD 5

Participants received a brief study description and were also encouraged to be clear about their interpretation and opinion, as candor was key in making this experiment a success. This encouragement was necessary as a number of participants were not confident with their own knowledge on graph interpretation. They were assured that their genuine feedback was the sole purpose of this experiment and that it was important to reach a wide range of user backgrounds to enhance the quality of the study's outcomes. I provided information to each participant as to what constitutes an outlier in the visually presented data. Participants were then asked to interpret the three graphs with respect to how useful and readable the visuals were to them. Participants were further asked to explain what they liked or disliked about the graphs and to elaborate on the possible benefits of each graph. All participants were encouraged to think aloud.

Analysis

The audio data provided by our participants was transcribed. I stopped data collection after forty-six participants because I had reached the point of saturation. In qualitative methods, saturation is defined as the point where no new information comes to light (Seale, 1999).

To analyze this data systematically, we developed an emergent coding scheme. In other words, I inductively arrived at the themes that were used to categorize the information provided by my participants. I coded the data with a partner and used intercoder reliability (Lavrakas, 2008) to ensure we reached the same conclusion regarding how to assign a response to the same category.

The main categories surfaced by the analysis centered around the concept of understandability. As mentioned previously, I define understandability as: *did the participant interpret the data correctly and were they able to comprehend the important concepts in the patterns conveyed in the graphs.*

Results

A raw analysis of the text in the feedback of my participants highlighted the following frequently-used terms:

Table 2. Frequently-Used Terms in Describing the Heart Rate Visualizations

Terms	Frequency
Min, Max, Average, Median	139
Norm(s), Range	100
Reading(s), Point(s), Data	96
Clarity, Clear, Easy, Simple	69
Legend, Line, Grid, Square	62
Color(s), Blue, Red, Green, Orange, Grey	49

The use of these terms suggest that the user prioritizes these tags in order to understand the visuals and these priorities provided me with insight into what features of the visuals needed improvement. Since these frequencies provided me with some utility, I decided to generate word clouds of all the feedback provided for each of the three graphs (see Fig. 15). For Graphs 2 and 3, it is clear to see much fewer visible words that stood out in large size when compared to the word cloud of Graph 1. This means that the user terminology used to explain Graph 1 was more consistent among respondents and it reinforces the results I present next when analyzing the feedback manually (in particular, it reinforces the result that my participants all understood Graph 1 at a higher rate than Graphs 2 and 3).

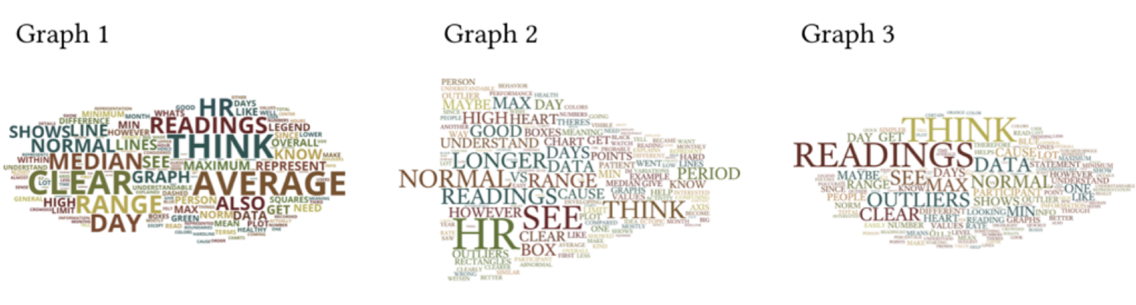


Figure 15. Word Cloud Representations of the feedback of the Graphs from Study 2.

As mentioned above, the purpose of this user study was to evaluate the understandability of the graphs designed in Study 3. If a participant closely identified the meaning or significance of the visual, we coded them as understanding the visual. 80% of participants understood Graph 1. 77.5% of participants understood Graph 2. 70% of participants understood Graph 3. Across the three graphs, 75.8% of participants overall understood the information presented in the graphs. Finally, out of all the participants, 69% of female participants understood the graphs overall, and 91% of males understood the graphs overall. These findings suggest that the graphs in their current form are clear, but need further improvement in terms of features, simplicity, contextualization, and customized user interpretation. Moreover, there is a gender divide in terms of understanding the graph. One recurring theme that emerged was a focus on the term “box,” referring to the box plots in Graphs 1 and 2. This representation has to be revisited in terms of the overall visual design as they caused a lot of confusion or hesitation. Many users initially would say “I don't get these boxes” or “what are these lines sticking out of the boxes.” Then most of them would actually explain the data represented in the boxes by saying “ok so this is the range of my HR readings" or "this is the highest reading and this is the maximum and that line is the average." This shows that although many of my participants stated that the boxes were confusing or intimidating, they were still able to comprehend the main concept they represented. As a result, the majority of the users were able to understand the patterns in the data in spite of this confusion.

Other feedback generated by my participants about the three graphs generically was the need for more color representations to be able to distinguish concepts in a clearer way. Another was the need for functionality that would allow a user to zoom in and pinpoint a certain reading while easily determining the value of that reading. Finally, a majority did not want to see too much detail on the graphs; they found the existing set of details either distracting, intimidating, or confusing. The full set of responses from my participants for Study 3 is located in the Appendix.

Summary

At the end of Study 3, I had strong motivation to gather more data to help get the information I needed to produce a new version of the visuals to attempt to boost the overall percentages of understandability in the users who reviewed the graphs.

STUDY 4

For Study 4, I recruited twenty participants to gather more quantitative and qualitative data to provide input into a redesign of the visuals created at the end of Study 2.

Method

After the completion of gathering user feedback on the graphs in Study 3, I wanted to obtain more tracked heart rate data and feedback from my participants to design new visuals that better speaks to a user's needs and perspectives based on what I learned from Study 3. I recruited twenty participants via word of mouth. All twenty participants volunteered to contribute their data to my study. I explained to each participant that no private or identifying information shall be collected. I also informed them that their data would be anonymized and secured. Each participant was provided with an Apple watch as well as a medical device for measuring heart rate with a high degree of accuracy. For the medical device, we chose to use an electronic blood pressure monitor. The instructions for using the device were explained to each participant. To wit: they should wear the watch throughout the day, but take it off to charge at night. They were also asked to measure their blood pressure at least three times a day, preferably with the first at the start of the day; the second during their daily routine; and the third right before they went to sleep. The Apple watch does not record heart rate continuously, nor does it take readings at regular fixed intervals. Because of the way the heart rate sensor operates in the watch—and to guarantee that the watch is actually recording the heart rate while the medical device is being used—all participants were asked to activate the heart rate app on the watch while the medical device was recording its data. Each user tracked their data across ten days. After completion, the participants were asked to complete an online survey.

The twenty participants recruited for this study were all at least 18 years old. We aimed to include a diverse demographic in our population pool. One important criteria we wanted to fulfill was to find participants that were diagnosed with cardiac related conditions (e.g. hypertension), or participants that were not necessarily diagnosed but were aware of having cardiac-related symptoms. A breakdown of the demographics for these participants is shown in Table 3.

Table 3. Demographics for Study II Participants

Gender	Female	Male			
	14	6			
Medical Condition	Diagnosed	Aware of it		Neither	
	3	3		14	
Age Group	18 - 29	30 - 39	40 - 49	50 - 60	>60
	4	7	6	2	1

Some of the participants were given a video or live demonstration on how to carry out the daily readings with the two devices while other were quickly familiarized with the process after a few attempts. This divide was due to several reasons as some participants were more familiar with mobile device and app use, while some participants pre-owned a blood pressure device and preferred to use their own device during the tracking experiment. This exception was allowed only when their device had high accuracy and quality. The method used to log the manual daily readings was done in two ways. Participants were given the choice of filling out a spreadsheet or to take a picture of the screens of both measuring devices (the blood pressure device and the Apple Watch) as soon as the readings appeared on both screens. I offered the latter option to lessen the burden on participants. Fifteen participants felt more comfortable taking pictures of their readings.

The users in Study 4 tracked their readings for ten days. After the completion of the tracking, all twenty participants received a survey. This survey asked about their overall experience for the duration of the tracking period. The survey consists of thirty-eight questions; a copy of the study is included in the Appendix. The questions on this survey include thirty multiple choice questions, five essay questions, and three short answer questions. The survey was designed based on insights gathered from our Study 2 (Albaghli, et al., 2017).

After the analysis of this survey was finished, I took the results of everything learned in Study 1 and 3 and started work with an undergraduate research assistant to create an interactive set of visualizations. These visualizations were based on the three graphs from Study 2 but with enhancements based on all that I learned. I used D3.js, a JavaScript library, as the main framework to generate these graphs. I selected D3.js for its capability of producing dynamic, interactive data visualizations in web browsers. These visualizations could then be embedded in a mobile app so we could deploy them to a new set of participants in Study 5.

Analysis

For each of our twenty participants, we exported the Apple Watch data. From these exports, we extracted the heart rate data in beats per minute as well as the time and date for each reading. These readings were taken as the participant wore the watch throughout the day. The average number of heart rate samples for participants was 2527 readings over a period of ten days. As some users pre-owned a watch and volunteered their tracked data prior to the experiment, we used this to generate graphs covering longer time spans. We used this data to design new visualizations governed by what we gleaned from the user feedback acquired in Study 3. We enhanced these diagrams with a richer representation of information through balanced simplification. We sought to improve user understandability. The results are presented below.

In addition, I was able to successfully collect thirty pairs of the manual readings that were taken from both the watch and the blood pressure device simultaneously. With this data, my aim is to increase the credibility of using an Apple Watch and similar data trackers to analyze heart rate. To assess the credibility, I first compared Apple Watch heart rate readings with those acquired from a medical device to evaluate its accuracy. Second, I examined the related literature to identify how the human heart behaves and what its quantitative limitations are. Next, I sought feedback from doctors to help me contextualize the heart readings. Finally, I synthesized the self-tracked data and participant observations to draw conclusions that are presented below.

With respect to the survey in Study 4, I designed it with the following goals:

- Gather participant demographic information
- Gain an idea about participants' sentiments on the experiment and the use of wearable technology in general, as well as using the Apple Watch.
- Understand self-rated perceptions of participant health and how much attention they pay to their own well-being.
- Gauge if any change occurs to a user's self-rated perception, after participating in the experiment.

I report on the results of this survey data below; the results from this survey were useful to refining the overall design of the WISE framework.

Results

Device Accuracy

To gauge the accuracy of the heart rate data measured on the watch, I compared it to the heart rate data my participants collected on their medical devices. I used blood pressure devices recommended by doctors and highly rated for their performance and quality. I considered the watch readings to be accurate when there was a difference of no more than five bpm or less with the heart rate reading from the device serving as our ground-truth data. Any difference between the two readings higher than five bpm were classified as inaccurate (on the part of the Apple Watch). I then calculated how accurate the Apple Watch was by dividing the number of accurate readings by the total number of readings measured. Fig. 16 shows an example of one of our participant's data sheet. I applied the method described here only after seeking the medical opinion of a doctor on how they would interpret patient heart rate readings.

date	Day	time	apple HR	M-decive HR	Sys	Dia	Off beats	Day No.	accuracy
12/21/2017	Thursday	7:20 AM	102	101	119	66	1	10	1
12/21/2017	Thursday	9:40 PM	106	106	137	71	0	10	1
12/21/2017	Thursday	11:30 PM	94	94	121	70	0	10	1
12/22/2017	Friday	8:42 AM	92	90	118	60	2	2	1
12/22/2017	Friday	4:58 PM	112	101	108	69	11	2	FALSE
12/22/2017	Friday	10:30 PM	95	92	108	64	3	2	1
12/23/2017	Saturday	9:45 AM	83	83	111	68	0	3	1
12/23/2017	Saturday	5:26 PM	87	89	105	64	2	3	1
12/23/2017	Saturday	11:11 PM	87	87	123	83	0	3	1
12/24/2017	Sunday	6:27 AM	83	83	122	83	0	4	1
12/24/2017	Sunday	5:29 PM	104	104	124	73	0	4	1
12/24/2017	Sunday	11:15 PM	89	89	124	61	0	4	1
12/25/2017	Monday	6:50 AM	86	86	128	85	0	5	1
12/25/2017	Monday	5:08 PM	87	86	117	74	1	5	1
12/25/2017	Monday	11:32 PM	92	92	113	55	0	5	1
12/26/2017	Tuesday	7:13 AM	86	86	132	85	0	6	1
12/26/2017	Tuesday	5:30 PM	83	83	136	87	0	6	1
12/26/2017	Tuesday	11:26 PM	94	95	131	85	1	6	1
12/27/2017	Wednesday	7:12 AM	83	78	126	78	5	7	1
12/27/2017	Wednesday	3:41 PM	88	87	119	82	1	7	1
12/27/2017	Wednesday	9:36 PM	86	84	120	76	2	7	1
12/28/2017	Thursday	6:53 AM	78	77	123	76	1	8	1
12/28/2017	Thursday	5:37 PM	94	94	125	91	0	8	1
12/28/2017	Thursday	11:45 PM	78	78	118	71	0	8	1
12/29/2017	Friday	8:11 AM	80	80	129	92	0	9	1
12/29/2017	Friday	6:04 PM	91	90	119	72	1	9	1
12/29/2017	Friday	10:00 PM	89	89	104	60	0	9	1
12/30/2017	Saturday	10:51 AM	83	83	140	95	0	10	1
12/30/2017	Saturday	6:50 PM	91	91	126	64	0	10	1
12/30/2017	Saturday	11:55 PM	90	90	145	95	0	10	1
Accuracy									96.67%

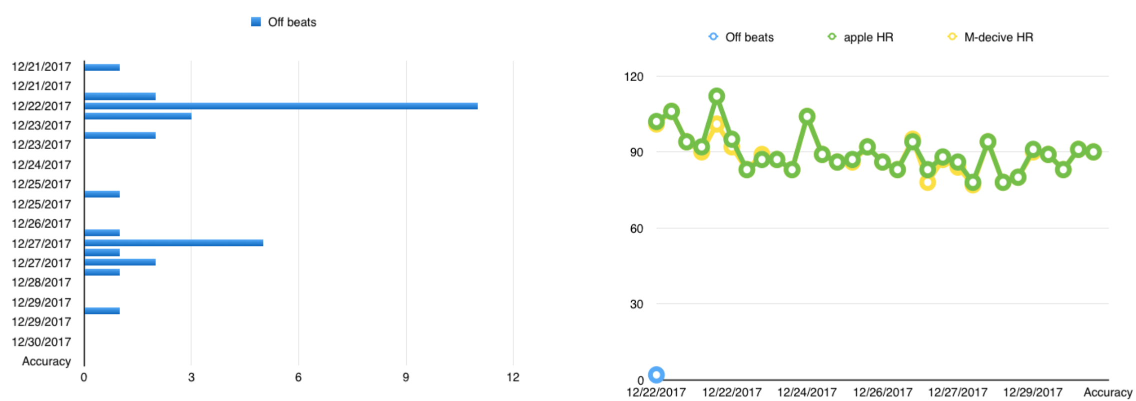


Figure 16. An example of the tracked readings from a Study 4 participant and an example of the calculation of the accuracy between the participant's watch and their device.

We found that the heart rate sensors on the Apple Watch were 85.33% accurate across all participants. All participants provided a total of three readings per day, for a total of thirty readings. The participants were asked to be at a stable resting position, preferably seated, when taking these measurements. We did this to ensure that the readings from both devices were accurate. The details of our device accuracy findings are in Table 4.

Table 4. Health data tracking details of Study 4

P NO.	WATCH READINGS COUNT	MEDIAN HR	MEAN	STANDARD DEV	MIN	MAX	ACCURACY
1	1415	87	88	10	37	160	100.00%
2	2835	79	83	18	28	191	100.00%
3	3292	84	85	10	45	142	83.33%
4	4001	76	77	12	43	160	63.33%
5	2243	65	67	13	35	164	96.67%
6	4037	64	70	21	39	167	100.00%
7	2389	81	83	15	34	178	80.00%
8	2632	89	92	21	36	194	96.67%
9	2902	88	88	15	35	200	90.00%
10	2797	97	97	14	29	158	53.33%
11	2162	73	74	13	36	159	96.67%
12	2215	74	76	13	39	185	70.00%
13	2175	74	75	11	40	202	90.00%
14	765	90	91	16	37	198	86.67%
15	2125	85	87	15	31	212	86.67%
16	2258	65	67	14	34	176	100.00%
17	2384	79	81	13	36	161	90.00%
18	1962	74	76	14	36	132	86.67%
19	3501	96	101	21.3	90	187	96.67%
20	2453	92.5	92	16	31	174	40.00%

The average age of my participants was 38.7, with a median age of 20. Of the participants, approximately 70% were female, and 30% male. The participants were highly educated with 96% having at least a bachelor's degree (see Table 5 for a complete set of demographic information). Seven participants indicated that they work out at least once a week, whereas eight participants stated they never work out. Only four out of the twenty participants indicated either high blood pressure or low blood pressure symptoms. Two out of the twenty participants indicated that they currently take medication that affect their heart rate and or blood pressure.

Table 5. Description of Demographics and Key Variables

Variable	Description	Mean	Median	SD (range)
Age	Respondent's age	38.70	20	11.11 (18 - 64)
Gender	Biological Gender as Identified by the Participant (Male = 6, 30%) (Female = 14, 70%)	NA	NA	NA
College degree	Respondent has earned a bachelor's degree or higher	95.90%	NA	NA
Relationship Status	Self-rated relationship status (Single = 30%) (Married = 70%)	NA	NA	NA
Self-Rated Health	How would you rate your attention to your own personal health? *(Ranges from I don't pay attention to my health to I pay great attention to my health)* *(I don't pay attention to my health = 10%)* *(I pay moderate attention to my health = 40%)* *(I pay adequate attention to my health = 15%)* *(I pay great attention to my health = 30%)* *(I pay great attention to my health = 5%)*	NA	NA	NA
Tracked Heart Rate	Beats per minute as captured by the Apple watch	82.50	80.63	14.77 (NA)
Accuracy of Apple watch to Medical Device	Watch readings was considered accurate when the difference with the medical device did not exceed 5 bpm	85.33%	90.00%	0.16 (40%-100%)

Note: Sample size = 20 Respondents, *Tracked Heart rate reflects the mean of all means, medians, and standard deviation, to compensate for a total of 50,543.

Twelve out of twenty participants (60%) indicated that after using the Apple Watch, their perception of health changed. One participant who answered yes to both change in their health perception and behavior, stated that the watch "attracted [his or her] attention to my habits… [such as] … walking, and standing… [it also] encouraged…deep breathing breaks." A majority—fourteen out of twenty participants (70%)—indicated that after using the Apple Watch, their behavior, routine, and or activity changed. One user remarked, "I exercised more often. I also spent more time walking outside to get in standing time." Another user expressed, "I started meditating in the middle of the study, I wanted to do something to be more aware, as I felt like I wasn't paying attention to my body enough. Also, seeing how I could add the watch/pressure measure in my routine encouraged me to try and add meditation as well." On the question of explicit behavior change, users conveyed in their remarks activity changes in breathing, waking up earlier, exercising, standing up more frequently, increased overall physical activity, a better sleep cycle, more attention and mindfulness of one's own behavior, a better commitment to tracking goals and commitment to lifestyle change as a result of the tracking, an interest in mediation, and a hint of self-competition with their past and future performance. All of this may suggest an improved ability for the user to self-actualize and attempt to reach their peak performance when engaging in self-tracking.

Seventeen out of twenty participants (85%) indicated that after using the Apple Watch, they would recommend using it for tracking heart rate to a friend, family member, or colleague. Of these, one person indicated that they would be "somewhat unlikely" to recommend the watch. Twelve out of twenty participants (60%) indicated that it was "extremely easy" for them to acquire their heart rate information from the Apple Watch. Two participants stated that it was "neither easy nor difficult." Overall, the participants found the process of self-tracking through the watch to be straightforward.

Boosting the Credibility of Heart Rate Readings

Another goal I wanted to achieve for Study 4 is to enhance the accuracy and validity of the heart rate data when they represent either outliers or might be considered alarming. This is important since outliers are usually ignored, yet we believe some readings can be valid and constitute an alarming reading. It is important to distinguish between what constitutes an outlier and what constitutes a reading that is alarming. An alarming reading is any one which lies outside the recommended healthy heart rate range given the active state of a user. An outlier is simply a reading that is distant from the majority of all other readings taken on a given day. In statistics, an outlier may be due to measurement variability or it may indicate experimental error; in those cases that data is then sometimes excluded from a final data set.

After establishing this distinction, we can identify conditions for when a reading can be considered alarming but is not considered an outlier. As an example, if a user of average good health is resting (and therefore his normal expected range for heart rate is between 60 and 100 bpm) and yet a large portion of his readings were around 110 bpm then all of these readings would be considered alarming. They are all above the expected maximum but not considered an outlier since all of those readings clustered together and thus their distance is not far enough away from the rest of the data to be considered an outlier. The reverse case is also true; a reading could be considered an outlier but not alarming; this might happen if, for instance, a reading under 100 bpm is taken and it just happens to be far from the rest of the readings taken that day. Since it is less than 100 bpm, it is considered safe but still an outlier. Finally, it is possible to have a reading that is both an outlier and is considered alarming.

Based on our own observations through the tracking studies, looking at how the devices performed with the users, and based on the information we learned after consulting with doctors on how they interpret readings, we identified three cases that can be used in the future to filter out data that is considered a true outlier and can thus be excluded from analysis and those which are not outliers and thus should be evaluated to see if they are also alarming:

Case A: If a user's heart rate data includes three on-the-surface outliers that were less than five beats apart and measured within two minutes then they are considered valid readings and should be included in the final data set. Mathematically, if X is an array of heart rate readings and t is an array of timestamps that correspond to the heart rate readings in X, then

if (| x[n] - x[n+1] | <=5) and (| x[n+1] - x[n+2] | <=5) and (t[n] - t[n+2] <=120sec)

then we keep these values in the data set and evaluate them to see if they can also be considered alarming.

Case B: If there are two outlier readings that are exactly the same or one beat apart, within 1 minute, then it's considered a valid reading. if (| x[n] - x[n+1] | <=1) and (t[n] - t[n+1] <=60sec) then we keep these values in the data set and evaluate them to see if they can also be considered alarming.

Case C: Drawing from knowledge about human physiology, it is known that a human heart can only increase its rate by 10 bpm within a single minute. Therefore, if a person is working out and their heart rate is 100 bpm, then we know that the highest bpm rate that the person can be in the next minute is 110 bpm. However, since an error of 5 bpm is considered acceptable for heart rate sensors, we can modify this statement and say that if we measure a rate of 100 bpm in one minute then we know that if the readings jump up to 115 bpm in the next minute we can consider that a valid increase. If it goes more than that, then we can exclude those readings from our data set. So, if a user has a reading of 70 bpm, then 105 bpm, and then 72 bpm, we can safely assume that the second reading is a false positive and throw it out since it is not physically possible for there to be such an increase. Therefore, Case C, would look for values in sequences that were different by more than 15 bpm and identify them as ones that we can safely remove from our final dataset. The other place where this fact can be used is to identify normal patterns of increase in bpm when someone has started exercising or walking and therefore classify those transitions for later contextualization.

Case C, then, can be used to “clean up” or filter data since it will identify outlier data that should be removed from the visualizations presented to the user and from any analysis that is then performed on that data. In Fig. 17, we should a set of heart rate ratings for a user and highlight in red those data points that would be filtered out using the rule imposed by Case C.

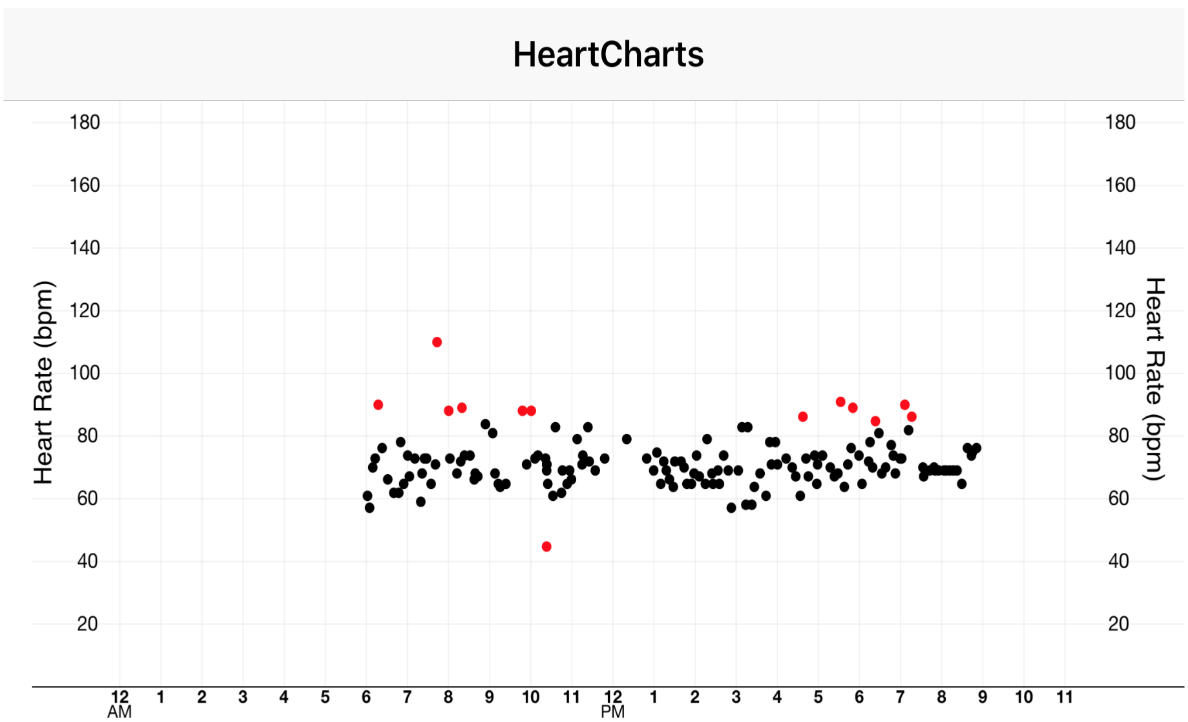


Figure 17. Results of applying Case C to a set of heart rate data. Data points highlighted in red would be removed from the final display seen by a user; those data points are also removed from any analysis that is then performed on that data.

These three conditions relate to one another and impose constraints. All data considered to be outliers should be filtered by Case C before detection of Cases A and B are performed. For instance, if someone has a sequence of readings that are outliers: “106, 108, 124, 105” all within a minute, then we would filter out the 124 knowing that it is not possible for it to change in that way. Once filtered, the sequence of “106, 108, 105” would then be considered an instance of case A and be retained in the final data set.

Applying these case rules can help decrease false positives and strengthen the credibility of anomalous situations that we would want to alert the user about if they occur repeatedly. These rules for filtering out spurious values (case C) and for identifying true positives hiding in the data that is otherwise considered an outlier or alarming. I have started to explore the use of these cases in the data we have collected from our participants. I can filter a participant’s heart rate data using Case C and then look for repeated instances of Cases A and B. I can then highlight those instances in the visualizations that my system would produce. Fig. 18 shows an example of valid instances of Case A and Case B being highlighted in this way.

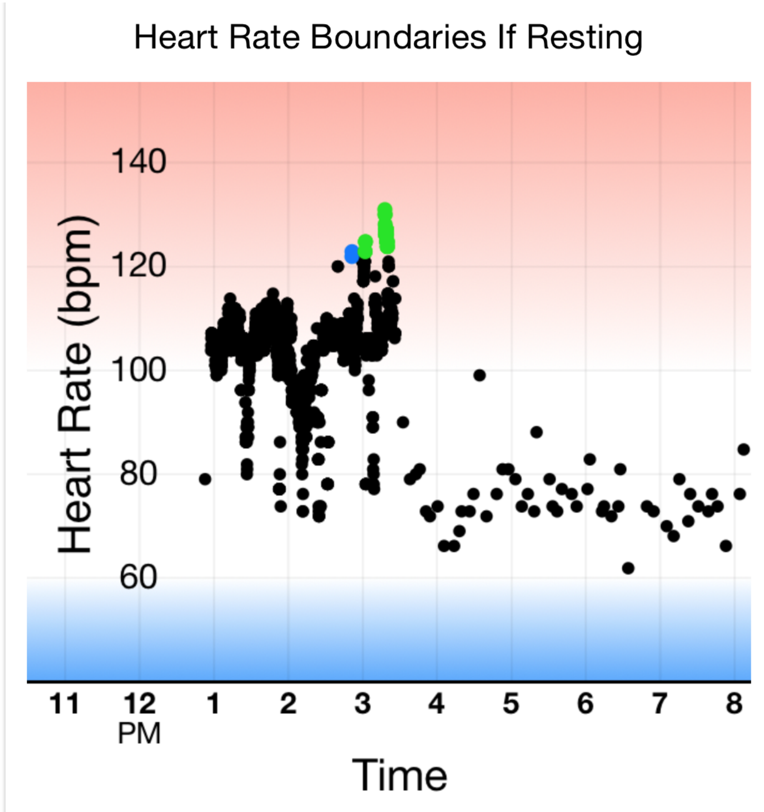


Figure 18. An example of highlighting Cases A (green) and Case B (blue) in heart rate data.

Heart Rate Visualizations

Based on the results of Study 1 and Study 3, I started working on a redesign of the visuals from Study 2. The key new component of the visual design is a set of three overlays that can be applied to the data in the graphs to enhance the way the data is displayed and feedback is presented to a user. A key component to this design is the assumption that a user has filled out a profile that contains information about their measurements and basic health. With this profile, the overlays can then customize the visuals based on that information. The profile is key to being able to provide data analysis within the app that varies according to a user's age, medical conditions, and fitness level.

As such, to add value to visualized heart rate data, the graphs should consider a user's profile. Moreover, subsequent analysis should visually explain to the user the range of their readings rather than simply glossing over minimum and maximum readings, as is the norm for similar applications.

To achieve this in our health-related wearable system, I and an undergraduate student researcher created visuals that utilized D3.js to provide interactive box plots and scatter plots. The box plots are used to represent data from larger time intervals, such as years, months, and days. They aid in spotting possible changes to a user's routine, or considerable changes occurring in their prevailing resting heart rate. The scatter plots are used for a more granular data representation spanning over shorter time period within hours. They can emphasize alarming readings recorded out of healthy heart rate boundaries. Being able to create plots for different time intervals allowed me to create a program that lets a user click to "zoom in and out" of their data by changing the plot type, time interval, and data subset. Along with the ability to change the interval of displayed data, the program analyzes each data point with respect to neighboring points to display a heat map of the concentration of values above the maximum heart rate. The readings are not idly displayed but analyzed according to the rules presented above (Cases A, B, and C) to decrease false positives and thereby enhancing the credibility of any highlighted alarming readings.

In addition to making the visualizations interactive, I have added the notion of an "overlay" that can be invoked on a display of heart rate data to customize and contextualize the display based on data in the user's profile as well as the results of an analysis applied to that data. The first overlay is known as the *Resting Heart Rate Analysis overlay*. The idea with this overlay is to update the visual to draw the user's eye to where their heart rate data should be when they are at rest (see Fig.

19). To do this, heart rate data is shown as circles plotted on a graph of bpm over time. The number of elements is greatly reduced with just a few lightly colored lines in the background to produce a grid against which the circles can be interpreted. Two gradients—similar to the ones in the original design proposed in Study 1—are added to the display to show the boundaries of when heart rate readings would be considered alarming. If the super majority of the displayed readings fall within the white band of the display then all is well. Furthermore, the boundaries of the gradients adjust according to data in the user's profile. While most people will have the standard range of 60 bpm to 100 bpm, the display will change appropriately if that range does not apply to the current user.

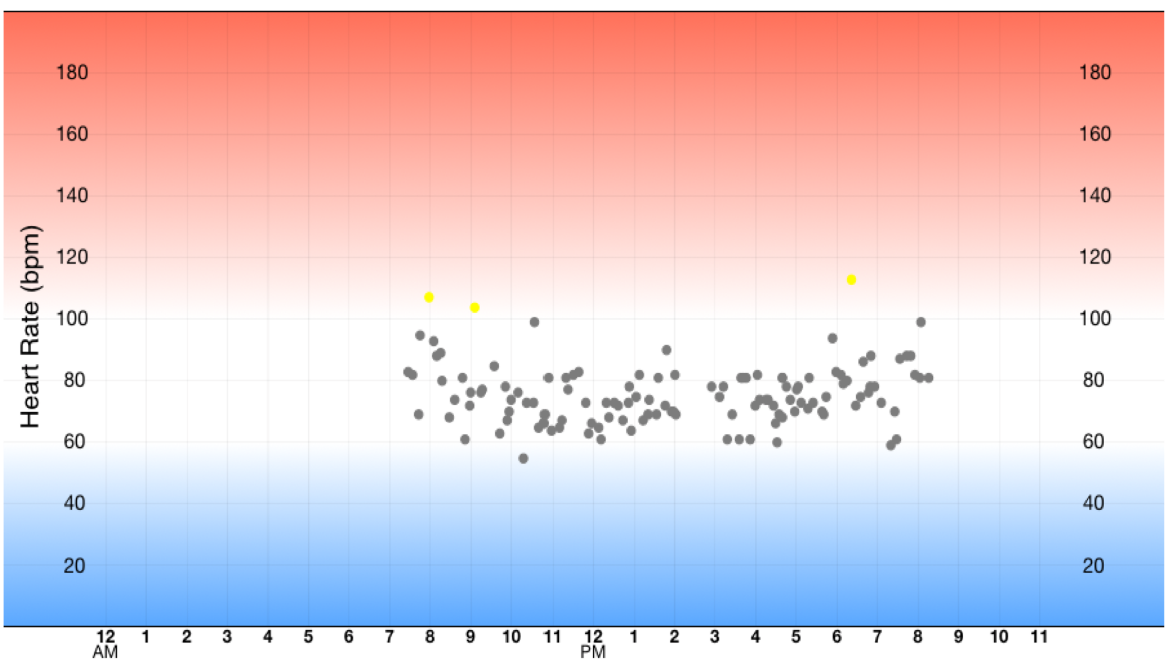


Figure 19. Updated Heart Rate Graph with the Resting Heart Rate Analysis Overlay

The second overlay I created is called the *Active State Heart Rate Analysis overlay*. This overlay again adapts to information in the user's profile and updates the display to show a greatly expanded range of legal values for the user when they are actively working out. In particular, the red gradient adjusts upward to show the safe maximum heart rate that a user should not exceed even during a high intensity workout (see Fig. 20). These numbers are drawn from recommendations published by the AHA (American Heart Association, 2015).

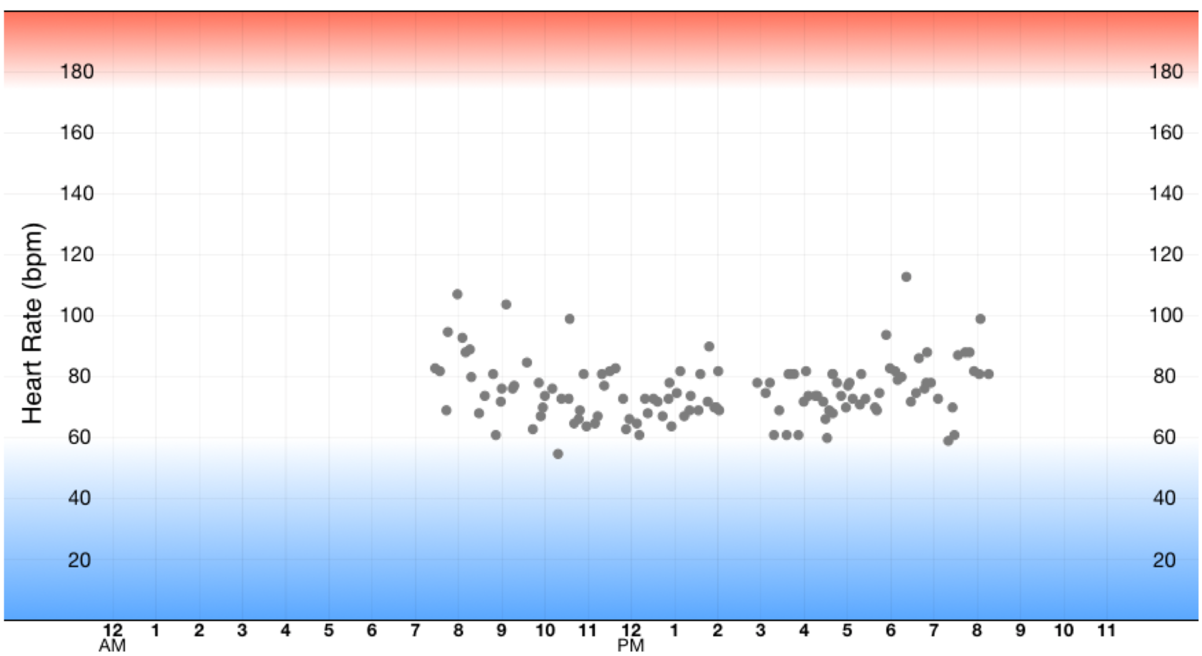


Figure 20. Updated Heart Rate Graph with the Active Heart Rate Analysis Overlay

The third overlay is called the *Alarming Readings Heart Rate Analysis overlay*. As with the previous two, the third overlay also makes use of information from a user's profile to deliver more customized outcomes. The third overlay shown in Fig. 21 highlights heart rate readings that are out of the normal range for the current user. High concentrations of alarming readings are displayed using a heat map technique. Fig. 21 spans over a period of 1 to 2 weeks and is especially helpful for medical practitioners and cardiologists that are initiating an investigation of a new patient (Albaghli and Anderson, 2016). This snapshot can provide a more telling timeline of heart rate performance in the recent past. It also provides insight into how a user's heart rate performs across their typical environment and routine. This information can familiarize doctors with the longer-term behavior of their patients as opposed to the false readings that may potentially occur within a medical environment that can be stressful to a new patient.

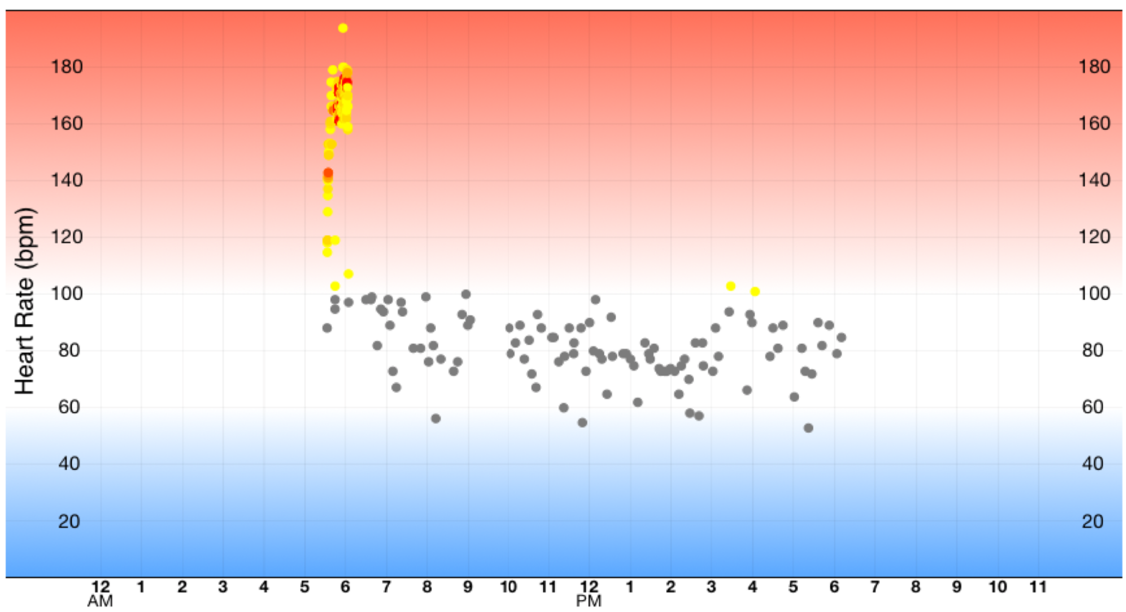


Figure 21. Updated Heart Rate Graph with the Alarming Readings Heart Rate Analysis Overlay

The visualizations in Figures 19, 20, and 21 are showing data for a short time span of just two days. This granular view allows for the display of individual heart rate readings contextualized according to the active overlay. The new interactive visualizations have a considerably expanded set of functionality. For instance, it is possible to view longer time spans and when the user does this, the visualization shifts to using box plots automatically. The first time a user sees a box plot, the app pops up a help screen to explain the representation to the user (see Fig. 22). When the user dismisses the help box, they can then click on one of the box plots to see the values associated with it and click it again to "zoom in" and see the heart rate readings for that particular day (see Fig. 23). The granular views will automatically show the selected overlay and the user can change the active overlay with a single click. With these changes, the updated visualizations address much of the feedback received from my previous user studies and represent an attempt to incorporate all four heuristics of the WISE framework.

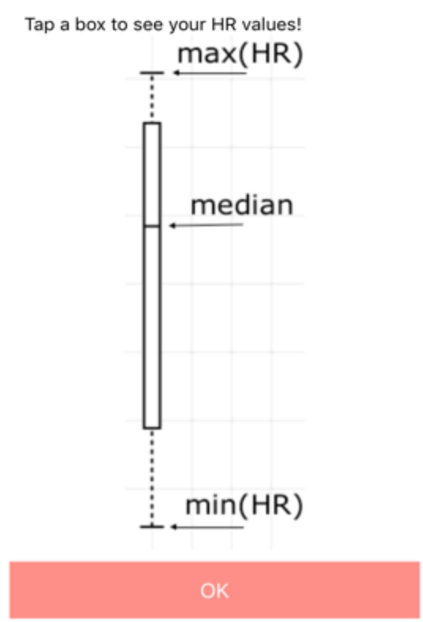


Figure 22. Pop-Up Help Display to Explain the Box Plot Notation. The help text reveals that box plots are interactive, and a user can tap them to see their associated values.

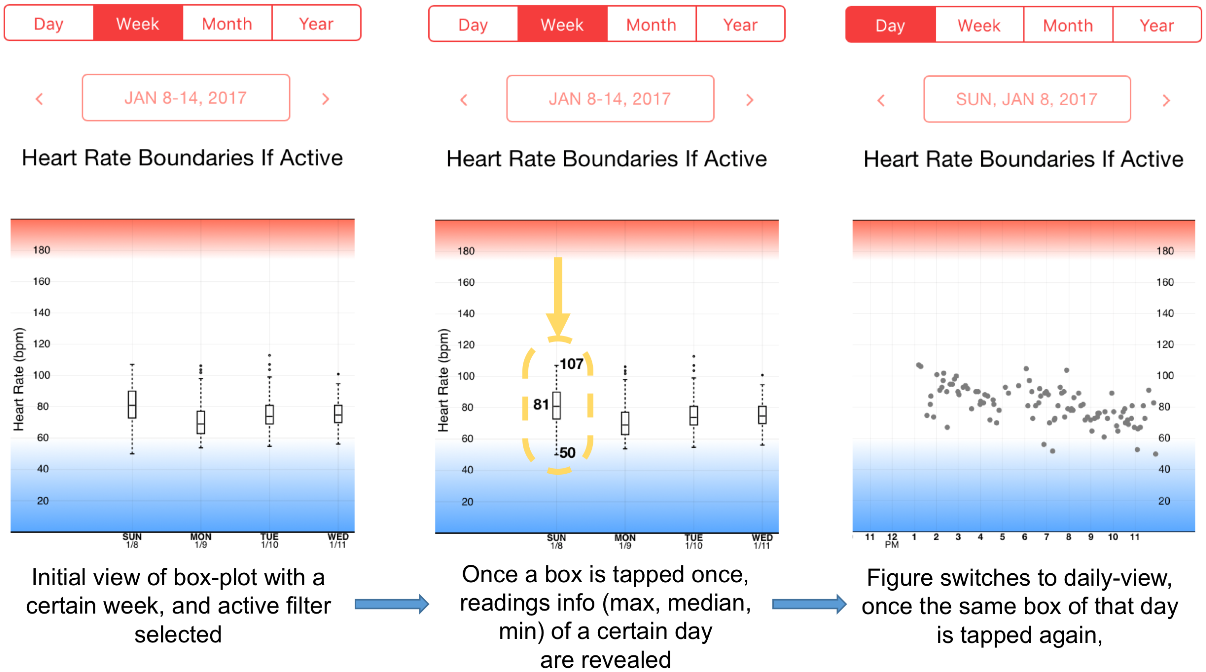


Figure 23. Example of interactive box plots in updated heart rate visualizations with the Active Heart Rate Analysis overlay selected.

Notifications

After reviewing the related medical literature, consulting with doctors, and making my own observation based on the data I collected, I developed a number of criteria with respect to the notifications a health-related wearable app should display to its users. These systems must have carefully timed notifications. This serves to alert the user of changes in their heart rate data only when it is relevant.

One important piece of information that each user should be familiar with is their median resting heart rate. The average resting heart rate should be determined by the end of each day. These measures should be averaged out week by week and compared against former weeks and months. When the resting heart rate shows a decrease or increase of more than 15 bmp, a notification should alert the user of this change, just to serve the purpose of familiarizing the user with their new resting heart rate. However, if the analysis of this change informed by the user's profile information determines that this change is concerning, then an alarm should be issued to the user recommending that the user make an appointment to see their doctor. This alert should be accompanied with a report showing the big-picture view graph from Study 2 but using the new visualizations set with a time span of months. Such a view will provide more information about behavior changes that reflect in changes to a user's heart rate over a longer period of time. This information can then be combined with a view of the heart rate data centered on when the change was detected with the alarming readings overlay selected.

Summary

At the end of Study 4, I ended with a wealth of data that was useful in confirming the accuracy of the heart rate sensors on the Apple Watch, rules to help "clean" heart rate data in a way to reduce false positives and boost the chance that alarming conditions represent "true positives" that a doctor should review, a set of insights from survey data of the Study 4 participants that greatly influenced the attributes of the WISE Framework, and a new set of interactive visualizations that addressed the concerns raised by the results of Study 1 and Study 3 and that exhibit all four of the heuristics recommended by the WISE Framework. In Study 5, these new visualizations are evaluated.

STUDY 5

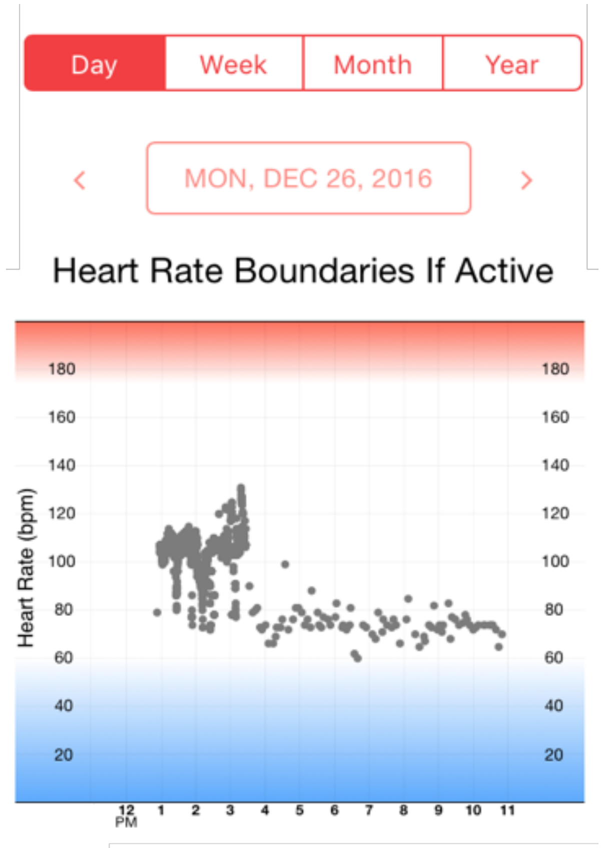
In Study 5, I recruited twenty-six participants to interact with the new visualizations and to gather data to assess if the new design impacted user understanding and perceived value.

Method

The objective of this final user study is to evaluate user understandability of the last iteration of my visualizations. The thinking aloud method was used in this study again to achieve this objective. For this study I recruited twenty-six participants. Participants were located all around the world. Some of them were assisted by a local recruiter since accessing the new interactive visualizations required downloading an app onto a phone or tablet device. In these situations, the recruiter would record their voice while they explored the new interactive visualizations on the device. The remaining participants were provided with instructions to help them download the app on their phones. All participants were then given the following instructions:

- Please think aloud while you explore the app. Share your thoughts and hesitations; describe what you think things mean or if they make no sense as well; share your likes and dislikes of what you see.
- Please take your time to read all tips or labels on the screen. There is no time limit and no need to rush.
- The data you will see is from a user wearing a watch that tracked their heart rate; you can pretend it is your data if you wish.

If participants were successful in finding all the features on the app, we asked them to look at a particular date that showed enough variance in the data to show the different cases of what constitutes healthy and not healthy heart rate data while switching between the resting and active overlays. The charts of that date with the resting and active overlays can be seen in Fig. 24.


Day
Week
Month
Year
MON, DEC 26, 2016
Heart Rate Boundaries If Active
Heart Rate (bpm)
180
160
140
120
100
80
60
40
20
12 PM
1
2
3
4
5
6
7
8
9
10
11

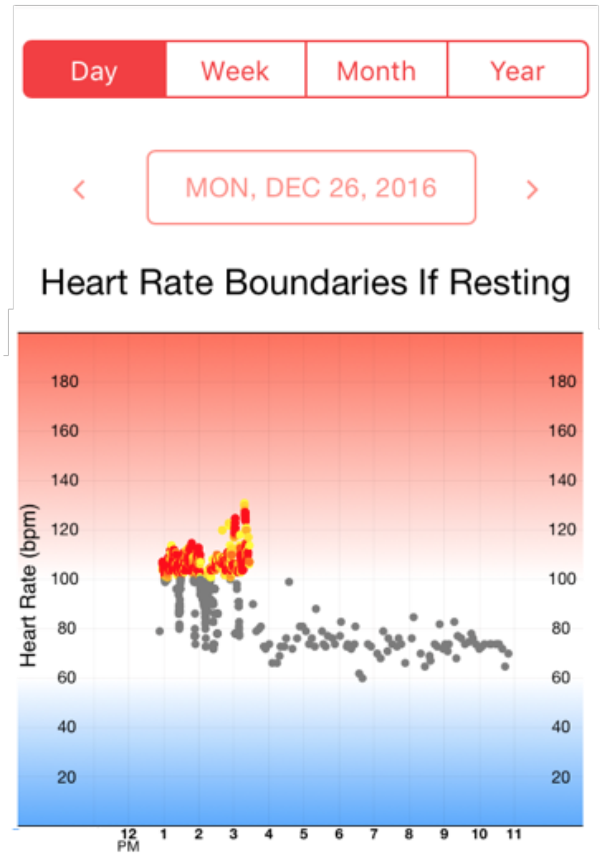


Figure 24. The difference in heart rate analysis once the active overlay (on previous page) is switched to the resting overlay (above). It is up to the user to select the overlay based on their memories of the day in question.

Only a few participants faced issues with understanding some of the terms shown in the interface since English was their second language. In such cases, the recruiter and I would translate the meaning of those words into their native language but provided no additional context or background information. This practice did not affect the experiment's outcome as I wanted to test the understandability of the visualized data independent of the language in use on the screens. Ideally, in the future, there would be versions of the app translated into multiple languages to avoid this issue. Unfortunately due to funding and time constraints, this was not possible for Study 5. The demographic information for the participants in Study 5 is shown in Table 6.

Table 6. Demographic Information for Study 5

Gender	Female		Male		
	16		10		
Age	less than 25	25 < age < 35	35 < age < 45	45 < age < 55	55 and higher
	3	8	11	0	4
Education	High school	Bachelor in arts and literature	Bachelor in Science	Masters	PhD
	5	8	7	3	3

Analysis

Study 5 is an evaluative study of the last version of our interactive visualization system. The heart rate data was displayed in interactive charts that allowed different views of the data as described above. The purpose of this study is very similar to Study 3 where I recruited forty-six participants to look at three different static charts and perform think aloud sessions while they looked at each and tried to make sense out of them. The goal of that study was to gauge user understandability given the state of the graphs at that time. Another goal was to have the opportunity to listen to the user and get their opinion on the data. Even when a user fully understood the graphs, I wanted to know what they did not like and what they preferred should be there and why. In Study 5, I wanted to look at the first goal again. After I took all the lessons learned from Study 3 in particular, the goal now is to gauge user understandability of the new graphs given their new interactive features.

In Study 5, some of the recruited participants also participated in the previous studies. Two participants persisted in contributing to Studies 2-5. Study 5 recruited participants through word of mouth. Participants were audio recorded during these sessions. All of the content of these audios were transcribed; these transcripts appear in the Appendix. One participant allowed us to video record their hands during their think aloud session. As mentioned above, I managed to recruit a number of participants that had also contributed their feedback in Study 3. I aimed to have at least 4 to 5 of those participants that did not achieve full understandability in Study 3. This goal was to help me compare their scores with this last study to affirm whether the final iterations, as guided by the heuristics of the WISE framework, were successful.

Results

Similar to Study 3, I used intercoder reliability again here to assure higher quality of my evaluation and how I and my assistant gauged a user's understandability. Participants were given a score from 1 to 3, where 3 means they understood all the concepts, including the difference in the analysis of the resting and active overlays, and what the heatmap colors and backgrounds expressed in the alarming overlay. A score of 2 meant that they understood all the basic information, the heart rate representations, and being able to navigate across different time intervals and what constitutes high and low heart rate readings, while a score of 1 is for participants who did not grasp these concepts. We were successful in recruiting some of the participants of Study 3 who scored lower than a 3 in that study. We wanted to look at those participants and see how they scored in this new study with the latest iteration of the visuals embedded in an interactive app. The results are shown in Table 7 below.

Table 7. Understandability Scores of Participants in Study 5

Participant	Gender	Age	Education	Understandability Score
1	male	66	High School	2
2	female	30	Bachelor of Arts	3
3	male	36	Bachelor of Arts	3
4	male	32	Bachelor of Arts	3
5	male	38	Bachelor of Arts	3
6	male	28	Bachelor of Sc.	3

7	female	63	Bachelor of Sc.	3
8	female	27	Bachelor of Arts	3
9	male	31	Bachelor of Sc.	3
10	female	44	PhD Sc.	3
11	female	35	Bachelor of Sc.	3
12	female	40	Ms Sc	3
13	female	44	Bachelor of Arts	3
14	male	36	Ms Sc	3
15	female	39	PhD Sc.	3
16	female	32	Bachelor of Arts	3

17	female	18	High School	2
18	male	43	Bachelor of Sc.	3
19	female	40	Bachelor of Sc.	3
20	male	30	PhD Sc.	3
21	female	41	Bachelor of Sc.	3
22	female	56	Bachelor of Arts	3
23	male	62	Ms Sc	3
24	female	22	High School	3
25	female	18	High School	3
26	female	36	High	3

			School	

The feedback from participants in Study 5 (available in the Appendix) has shown a very good understanding of the important concepts we wanted to convey to the user about their heart performance, their heart health, and its boundaries. The results show that 92.3% of all participants have shown full understandability of all the concepts, scoring at level 3. The remaining 2 participants (7.7%) scored at level 2. Most of our participants held bachelor degrees, but were almost equally divided between high school and higher education degrees (see Fig. 25). We do realize the limitation of our small sample size, however according to Nielsen, recruiting 5 participants is usually sufficient to reach saturation, and thirty is sufficient for quantitative studies aiming at statistics (Neilson, 2012). We also aimed to recruit participants from Study 3 who scored lower than a 3. We succeeded in recruiting seven such participants, six of which scored 2 in Study 3, and one which scored 1. We found that those seven participants in Study 5 boosted their performance: six of them scored 3, while one scored a 2. This is an excellent improvement in the understandability scores of these participants after the application of the heuristics of the WISE framework to generate the new version of these visualizations.

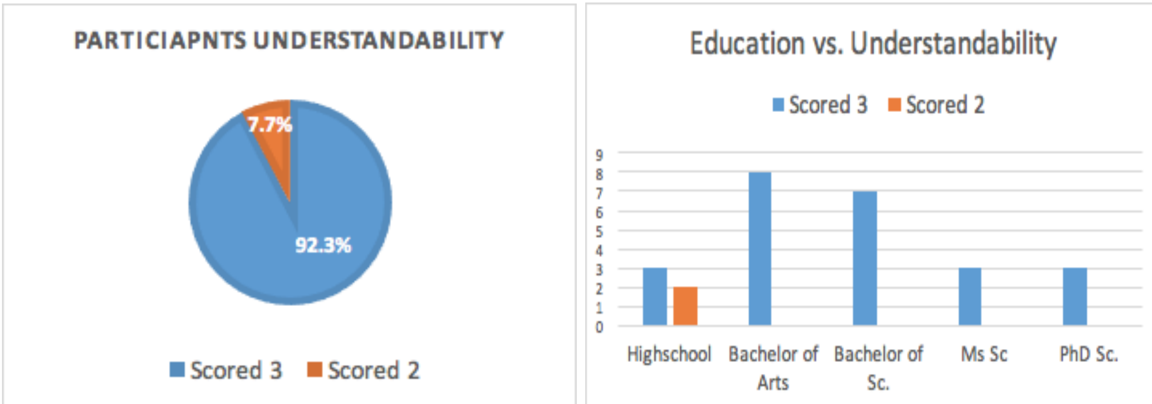


Figure 25. Study 5 Participant Understandability (left); compared with level of education (right).

Summary

The final user study presented good evidence that the changes made to the interactive visualizations of my health-related wearable system for tracking heart rate data led to increased user understandability. My hope is that if my current prototype was developed into a fully-implemented wearable application that it would similarly score well with respect to perceived user value. Certainly a system that is understandable has removed a considerable barrier to providing value to its users. Showing how the WISE framework can guide the design and development of a health-related wearable system for tracking heart rate data, I now discuss how the WISE framework can be used to help designers working on applications for other medical conditions.

BEYOND HEART HEALTH

A large number of human health conditions, or disorders, are directly connected with the measurement of biometric variables. The process of diagnosis, management, and treatment, often benefits from tracking these variables across time. This tracking is often carried out via expensive or stigmatizing medical devices, that are typically used for a limited amount of time. With the continuing progress of wearable technology, and its prevalence both socially and economically, there is no doubt that some of the earlier mentioned processes of managing health conditions can take advantage of this new technology. In spite of the limitations of these devices—such as their potential lower accuracy compared to medical devices—they offer important elements that are important. They are invisible, concealed in everyday devices such as watches that offer everyday functions like telling time. They also facilitate longevity of use and have a good chance of promoting better user health awareness when such systems are successfully executed. These goals can only be achieved when carefully considering not only what the user needs to know, but also what the user wants, how they want it, and how they can easily understand it. The WISE framework focuses on these outcomes. The framework aims to increase the user's perceived value of a system, while also increasing the users understandability of the data in these systems and its analysis. In Fig. 26, I highlight the key attributes of health-related wearable systems in this context. They are technologically advanced: small in size with fast processors & sensors and increasing data storage. They are invisible (non stigmatizing): embedded in typical, daily-use, consumer devices. As a result, they offer features that allow a user or patient to prevail in their use; they are affordable, trendy, socially accepted, and aesthetic. They also offer features that contribute to longevity of use: simultaneously offering non-health related functions that are needed daily (e.g. alarms, telling the time, etc.). These features align with existing campaigns to boost patient empowerment such as the Patient Empowerment Campaign (EPF, 2017).

Figure 26. The positive factors provided by wearables for health.

One example of a health condition that can benefit from having such an intersection is Parkinson's disease. Parkinson disease is a condition that involves symptoms (Parkinson's Foundation, 2017) of more than one category. These are symptoms affecting motor skills and movement such:

- Tremors
- Rigidity
- Slowness of movement
- Affected handwriting
- Soft or low voice
- Trouble sleeping

The other set of symptoms are more related to human behavior and mental disorders:

- Depression
- Anxiety
- Apathy
- Insomnia

Prevention and treatment of Parkinson's disease is currently managed through exercising, medications, and surgery. An important part of treatment and gauging the advancement of the disease is done through tracking the performance of these specific exercises and the recurrence of motor and mental symptoms.

Currently there is a research study that is ongoing. It is focusing on gathering data through the Apple Watch and iPhone by making use of a ResearchKit App developed by Sage Bionetworks and the University of Rochester. Their app utilizes the sensors to measure and track multiple variables: tremors, balance, gait, certain vocal characteristics, and memory. This is done through surveys and tasks that the user should follow in order to gauge their performance and symptoms. These tasks should be done before and after taking their medications, as well as by the end of the day to compare and evaluate the progression of the disease and eventually improve quality of life. Their main goal is to collect larger amounts of data to support their medical research advances in understanding and treating the disease.

This is one example of how one major disease can utilize wearable technology in allowing scientists to better understand the disease. However, improving the health and well-being of patients can also be achieved by giving a stronger role to the patient, allowing them to better understand and observe their own condition. We need to empower patient knowledge and understanding, as they are the ones present and observing their own conditions more than anyone else. Consequently, this will draw more importance to increasing understandability and system value held by the users in order not to miss what can be life changing benefits. With all this wealth of tracked data for a Parkinson's patient, seeing regularly updated visualizations that show their performance in a way easy to understand will help increase their awareness of their condition and perhaps adoption of the wearable system being developed for this study.

Parkinson patients using wearable technology can make good use of having the WISE framework applied to their wearable health systems. Contextualized representations can lead to identifying patterns of the activities, time of day, or other factors that more commonly surround the occurrence of their tremors. Once these representations are interactive, and have the right amount of simplicity, patients will be inclined to log useful, timed, notes about their mood, or symptoms. This can also encourage users to share data from other tracking apps to help achieve a broader, yet more accurate and informed picture of their symptoms progression, and their compliance to treatment.

One example of a visual representation would be to show the score of the patient's vocal recording. It would show the analysis of what an average score should be for the given patient's profile and progression level of their disease. It would allow the user to activate an overlay that contextualizes these scores by marking events such as exercising, or taking medication, or feeling happy. The patient with a tap on the screen can add another variable that shows the score of their motor-touch test survey. Another biometric variable to be used is to make use of the watches sensor that detects movement such as the accelerometer. An app can inform the user to activate a certain feature only when they feel their tremors occurring. This would activate the accelerometer. Tracking of such tremor data can again be represented in a qualified way to help identify progression. Hypothetically, a patient might notice that their voice and movement both improve right after a morning exercise session; however only motor skills show improvement after taking their medications. With elements of the WISE framework in use, the patient can better understand, remember, and correlate their systems and its analysis with their daily activities. This consequently increases the perceived value such a system would hold.

Mental illnesses, such as depression, is another disorder that can make use of the WISE framework for health apps. Depression patients are encouraged to exercise, eat well, and keep a journal and write about their feelings (WebMD, 2015). The main symptoms of depression can be common place. However, it is the intensity and duration of these symptoms that can help determine if a person has depression. Some of the main symptoms and feelings that can be present with depression are feeling:

- Sad, empty, or anxious.
- Helpless, worthless, or guilty.
- Hopeless.
- Irritable.
- Less interest in activities (hobbies, games, eating etc.)
- Less energetic.
- Trouble concentrating.
- Changes in the way you sleep.
- Changes in appetite.
- Aches and pains.

These symptoms can vary from one person to the next. Depression is a disorder that requires treatment. This includes going to regular therapy sessions with a psychologist. Treatment can also include administering medications on a regular basis that can last for long periods of time. There are also natural depression treatments that are related to changing a person's behavior (R. Morgan Griffin, 2015; Encyclopedia of Health, 2018). These changes can involve physical activity, lifestyle, or adopting different ways of looking at things. All of these are natural treatments that can be initiated by the patient and can enhance how a patient feels. These treatments can be tracked, better managed, and motivated with the support of wearable health systems that are understandable and valuable to a user. Below, I list activities that could help patients with depression that can take advantage of the potential benefits offered by wearable health systems:

- Getting in a routine: Setting a gentle daily schedule on your device, with reminders set by the user, can help patients feel a sense of control in their daily routine.
- Setting goals: Features to help set daily goals to help accomplish any task, starting with small realistic goals. The user can feel encouraged looking at a chart showing their progress on keeping up with their goals.
- Journaling: writing down thoughts and feelings to understand them more clearly. Keeping a journal can help gain control of emotions and improve mental health. Recording a journal, tracking the days they submitted journal recordings and showing the user a word cloud of any positive words they might have used are all examples.
- Exercising: Regular exercise provides many health benefits and there are a wide variety of applications that support a user track their exercise.

- Getting enough sleep: Tracking of sleep is available on many wearable fitness trackers. This can show how much hours of sleep a patient is getting and tracking progress.
- Challenging negative thoughts: A user can use a feature to help flag times they are feeling negative and down. This can track how often this happens and can also prompt activities that can help on the wearable device.
- Eating healthy: Plenty of apps enable journaling what one eats, and gives progress reports and reward badges to help motivate the user just for simply logging that data.
- Doing something new. (e.g. visiting a museum, volunteering, taking a language class). Based on location, a user can get alerts of interesting events happening nearby.

Other than the mentioned natural treatments, the formal medical treatment of depression patients mostly includes prescribed medications and therapy sessions. These sessions are enhanced when a patient can remember how they were feeling in the weeks prior to a session, and if any of it was affected or triggered by events or medication. It is clear again that empowering the patient with not only data, but data they understand, data they are willing to contextualize will encourage them to seek treatment or stick with their treatment. This can also lead to better, and more informed communication between the doctor and the patient, as the majority of these sessions rely on patient observations that are usually not recorded efficiently, and more likely relying on a patient's memory.

One major concern with such disorders is the will of the patient to take initiative. However, a system with such focus can also make use of audio recorded dictation, where a patient can record their journal with minimum effort, and those recordings can be logged not only with timestamps but also with data measuring the strength and pitch of a user's voice and their speed of speech and pauses. All of this can be further contextualized with events in their calendar, or labels of their phone alarm if present, along with the time medication was administered. This can be overwhelming for a patient to report clearly to their doctor.

The WISE framework would encourage the development of software that can help build a picture that would be much easier for the patient to narrate during their therapy sessions. It might motivate them to carry out a certain activity when they can evidently see in their contextualized data that it did enhance their condition with some consistency. A central application for depression can draw this information from multiple sensors and apps, creating a more rounded image of how the user is progressing and that will not only help report a better picture to their treating physician, but also empower, educate, and motivate the patient with systems that are custom to their personal attributes, contextual to their own schedule and behavior, simple and easy to understand, and offering them more control via interactivity.

The WISE framework can play a considerable role in other health conditions. It can similarly have a potential for the enhancement of people's understandability of their health and well-being for both patients and healthy users.

CONCLUSIONS

My research has centered on the potential that interactive wearable devices have with respect to helping people manage their health and improve their well-being. Given the wide deployment of wearable devices with sophisticated sensors in recent years, that potential is high and ever increasing. As I have shown, the design of such systems is a considerable challenge; designers and developers must commit to truly understanding both the medical condition they want to target as well as the needs of the people who will use their system to track that condition. These demands place considerable stress on the analysis and design techniques of software engineering and human-centered computing and the need to produce systems that are reliable, accurate, understandable, and customizable by end users. To help with this situation, I have designed and implemented a hearth-health wearable system with the aim to increase user understandability and user perceived value. Through this work, I inductively identified and developed the WISE framework to help designers and developers organize their work during the development life cycle of a health-related wearable system. The framework does this by stressing the need to produce systems that possess four desirable heuristics—*contextualization*, *customization*, *interactivity*, and *simplicity*—and providing a set of activities that are useful in producing the information needed to design these heuristics into a health-related wearable system. While these heuristics and activities are important, they are simply means to meeting the overall goal of the framework which is to help designers and developers produce wearable systems that are understandable and valuable to their users.

Initially I selected a medical condition—heart disease—that I wanted to study to see if I could design a wearable system that tracks heart rate data to provide insights into the health of a user's heart. My worked involved extensive interviews with doctors and a wide range of potential users to elicit requirements and information that drove the design of a set of interactive visualizations that later embodied the four heuristics of the WISE framework as they came to light. The resulting system and its interactive visualizations were understandable to a wide range of people with different educational backgrounds, medical conditions, and fitness levels. Over the course of five user studies, and the process of identifying the key heuristics and activities involved, I demonstrated how the framework could be used to generate a prototype of a health-related wearable system that provides a rich set of information for users to explore to gain an understanding of their heart health contextualized by a unique set of overlays that perform analysis and display results customized based on information contained in a user's profile.

I then performed a thought experiment to demonstrate that the WISE framework is not tied to a single application domain and that it could be used to guide the development of health-related wearable systems that target other medical conditions and track data other than heart rate. In the future, I intend to expand this work in several directions. I would like to complete the development of my software prototype to allow it to be deployed and used at a much larger scale than what I achieved in my user studies. I would like to further extend and refine the list of activities that the WISE framework recommends for guiding the development of a health-related wearable system and I would like to further explore the sets of metrics that can be used to clean and analyze heart rate data hopefully leading to more sophisticated notifications that can be generated for people using my system to manage the treatment of their condition. The potential to have positive impact on people's lives with this line of work is significant and I look forward to extending my work on the design and development of health-related wearable systems.

References

Albaghli, R. and Anderson, K. A vision for heart rate health through wearables. In Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM, New York, NY, USA, 1101–1105. 2016. < https://doi.org/10.1145/2968219. 2972715>

Albaghli, R., Raja, U., and Anderson, K. An innovative approach to better cardiac health through wearable technology. In Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. of the ACM International Symposium on Wearable Computers, pp. 877-882. 2017. <https://doi.org/10.1145/3123024.3125506>

American Chemical Society. Importance of Hands-On Laboratory Science. Public Policy Statement 2017-2020. 2017. <https://www.acs.org/content/dam/acsorg/policy/publicpolicies/education/computersimulations/hands-on-science.pdf>

American Heart Association. Target Heart Rates. 2015. <http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/Target-Heart-Rates_UCM_434341_Article.jsp#.Wk3PllQ-fR0>

Apple's WWDC 2014 - Session 203 - iOS. Introducing HealthKit. In Proc. The Apple Worldwide Developers Conference. 2015. <https:// developer.apple.com/videos/play/wwdc2014/203/>

Brame, C. Active learning. Vanderbilt University Center for Teaching. 2016. <https://cft.vanderbilt.edu/active-learning/>

Cardiogram for Apple Watch. Apple's App Store. 2016. <https://itunes.apple.com/us/app/cardiogram-for-apple-watch/id1000017994?mt=8>

CDC Center of Disease Control and Prevention. Million Hearts: Strategies to Reduce the Prevalence of Leading Cardiovascular Disease Risk Factors --- United States. 2011. <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6036a4.htm?s_cid=mm6036a4_w.>

Conti, G., Heckman, J., and Pinto, R. The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. In Proc. The Economic Journal. Econ J, 126: F28–F65. doi:10.1111/ecoj.12420. 2016

Creswell, J. W. and Clark, V. L. Designing and Conducting Mixed Methods Research. SAGE Publications. 2010. <https://us.sagepub.com/en-us/nam/designing-and-conducting-mixed-methods-research/book241842>

El-Amrawy, F. and Nounou, M. I. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? In Proc. of Healthcare informatics research 21(4):315–20. doi: 10.4258/hir.2015.21.4.315. 2015. <http://www.ncbi.nlm.nih.gov/pubmed/26618039>

EPF. Patient Empowerment Campaign. European Patients Forum and Patient Empowerment Campaign. 2017. <http://www.eu-patient.eu/campaign/PatientsprescribE/>

FITIV Pulse Heart Rate Monitor. Apple's App Store. 2016. <https://itunes.apple.com/us/app/motifit-workout-tracker-+/id942494517?mt=8>

Fritz, T., Huang, E. M., Murphy, G. C., and Zimmermann, T. Persuasive Technology in the Real World: a study of long-term use of activity sensing devices for fitness. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, 487–496. 2014. <https://doi.org/10.1145/2556288.2557383>

Gay, V., Leijdekkers, P., and Barin, E. A mobile rehabilitation application for the remote monitoring of cardiac patients after a heart attack or a coronary bypass surgery. In Proc. of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments. Article 21, 7 pages. 2009. <http://dx.doi.org/10.1145/1579114.1579135>

Glance, D. G., Ooi, E., Berman, Y., Glance, C. F., and Barrett, H. R. Impact of a Digital Activity Tracker-Based Workplace Activity Program on Health and Wellbeing. In Proc. of the 6th International Conference on Digital Health Conference. pp. 37–41. 2016. <http://dx.doi.org/10.1145/2896338.2896345>

Gouveia R., Karapanos, E., and Hassenzahl, M. How do we engage with activity trackers?: a longitudinal study of Habito. In Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 1305–1316. 2015. <http://dx.doi.org/10.1145/2750858.2804290>

Gouveia, R., Pereira, F., Caraban, A., Munson, S. A., and Karapanos E. You have 5 seconds: designing glanceable feedback for physical activity trackers. In Adjunct Pro. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and in Proc. of the 2015 ACM International Symposium on Wearable Computers. pp. 643–647. 2015. <http://dx.doi.org/10.1145/2800835.2809437>

Gouveia, R., Pereira, F., Karapanos, E., Munson, S. A., and Hassenzahl, M. Exploring the design space of glanceable feedback for physical activity trackers. In Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 144–155. 2016. <http://dx.doi.org/10.1145/2971648.2971754>

Goyal, R., Dragoni, N., and Spognardi, A. Mind the tracker you wear: a security analysis of wearable health trackers. In Proc. of the 31st Annual ACM Symposium on Applied Computing. pp. 131–136. 2016. <http://dx.doi.org/10.1145/2851613.2851685>

Griffin. R. M. 10 Natural Depression Treatments. WebMD. 2015. <https://www.webmd.com/depression/features/natural-treatments#1>

Guo, F., Li, Y., Kankanhalli, M. S., and Brown, M.S. An evaluation of wearable activity monitoring devices. In Proc. of the 1st ACM international workshop on Personal data meets distributed multimedia. (PDM '13). ACM, New York, NY, USA, pp. 31-34. 2013. <http://dx.doi.org/10.1145/2509352.2512882>

Health Encyclopedia. Journaling for Mental Health. University of Rochester Medical Center Rochester, NY. 2018. <https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentID=4552&ContentTypeID=1>

Heart Graph. Apple's App Store. 2014. <https://itunes.apple.com/us/app/heart-graph/id591655999?mt=8>

HeartWatch. Heart & Activity. Apple's App Store. 2015. <https://itunes.apple.com/us/app/heartwatch.-heart-activity/id1062745479?mt=8>

Kosara, R. Presentation-oriented visualization techniques. In Proc. of IEEE computer graphics and applications 36. pp. 80-85. 2016. < http://ieeexplore.ieee.org/abstract/document/7383166/>

Kroll, R. R., Boyd, J. G., and Maslove, D. M. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. In Proc. of the Journal of Medical Internet Research 18(9):e253. 2016. <http://www.ncbi.nlm.nih.gov/pubmed/27651304>

Lavrakas, P. J. Intercoder Reliability. In Encyclopedia of Survey Research Methods. Sage Publications, Inc. 2008. <http://dx.doi.org/10. 4135/9781412963947>

Narasimha, D., Hanna, N., Beck, H., Chaskes, M., Glover, R., Gatewood, R., Gudleski, G., Danzer, S., and Curtis, A. B. A smart phone-based ECG recorder is non-inferior to an ambulatory event monitor for diagnosis of palpitations. In Proc. Heart Rhythm Society Annual Scientific Sessions , S149. 2016. <https://www.researchgate.net/publication/303149896_Smartphone_based_ECG_recorder_versus_an_event_monitor_for_the_evaluation_of_palpitations>

Nielsen, J. Progressive Disclosure. Nielsen Norman Group 1.1. 2006. <https://www.nngroup.com/articles/progressive-disclosure/>

Nielsen, J. 10 Usability Heuristics for User Interface Design. Nielsen Norman Group 1.1. 1995. <https://www.nngroup.com/articles/ten-usability-heuristics/>

Nielsen, J. How many test users in a usability study. Nielsen Norman Group 1.1. 2012. <https://www.nngroup.com/articles/how-many-test-users/>

Parkinson's Foundation. 10 Early Signs of Parkinson's Disease. 2018. <http://parkinson.org/understanding-parkinsons/10-early-warning-signs?gclid=EAIaIQobChMI0IviruLb2QIVglh-Ch09agIAEAAYAyAAEgJB_fD_BwE>

Paul, P., and Irvine, J. Privacy Implications of Wearable Health Devices. In Proc. of the 7th International Conference on Security of Information and Networks. 5 pages. 2014. <http://dx.doi.org/10.1145/2659651.2659683>

Schmidt, B., Benchea, S., Eichin, R., and Meurisch, C. Fitness tracker or digital personal coach: how to personalize training. In Adjunct Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. of the 2015 ACM International Symposium on Wearable Computers. pp. 1063–1067. 2015. <http://dx.doi.org/10.1145/2800835.2800961>

Seale., C. Grounding theory. In The Quality of Qualitative Research. SAGE Publications Ltd, 87–105. 1999.

Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proc. Visual Languages. Proc., IEEE Symposium on. DOI: 10.1109/VL.1996.545307. 1996. < http://ieeexplore.ieee.org/document/545307/>

Web MD. 10 Natural Depression Treatments. 2015. <https://www.webmd.com/depression/features/natural-treatments#1>

Ye, C., Kumar, B. V. K., and Coimbra, M. T. Human identification based on ECG signals from wearable health monitoring devices. In Proc. of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies. Article 25. 5 pages. 2011. <http://dx.doi.org/10.1145/2093698.2093723>

APPENDIX

Study 3: Table X. transcribed user feedback

Participant	Gender	Age	Education	Understand-ability	Feedback
1	F	56	Bachelor degree science	2	I think the user can understand things clearly from the legend. The colors are clear, and the legends are clear in conveying the meaning of the user.
2	M	27	Senior college student	3	Chart should clarify what is this box. This box, that might be represented in terms of days or months, has 3-types Symmetric, left scroll, right scroll. He then explained these different types.
3	M	36	Master degree in science	3	First and second graphs are very clear. The third graph, I understood, but I didn’t like the big gap in the center. I didn't think its very well represented. For the general public that didn't study basic stats may not find them easy. I think G1 can be helpful in getting more information out of it. interns of categorization in terms of time of day eg. (morning, afternoon, evening), or in terms of day of the week as well (Monday, or weekend). Overall representation of Graph1 is good, a person can deduce more information or come to a conclusion. If you have info in terms of demographics can add relevancy to the data.

4	F	56	Bachelor degree art & Lit.	3	G1 what i understood there is variation in HR according to some days. In general readings are mostly within the normal. G2 Its easy to eat and understand. You can see the variations in the HR throughout the year. It's very understandable. I can see that clearly when the participant excited himself, the black hard lines are the medians. However the boxes are confusing a bit, I don't know what they represent. G3 The readings here to me are not clear, I couldn't read them. I don't know where the heart rate readings are. This shows that all heart rate readings seem they are starting off from one level, and this is confusing, cause I can find the rest of his readings in between the max and min range.
5	F	39	Bachelor degree science	3	I like the first graph more, it is clearer and I think that is due to the fewer number of days, hence I can see the center line. However, overall it is all understandable, but I didn't get the meaning of the boxes, what do you mean by them. The third graph is clear too. its a representation of the higher than norm and lower than norm represented in colors.
6	M	49	Medical degree	3	Graph 1 with minimum and maximum is clear and excellent. I think this will also be clear to general users. It shows when readings are outside of the normal range. The line that represents the average is also clear. But what do you mean by the outliers? I think they are correct readings that are ectopic. Overall I think this is a clear graph. Graph 2, I think it should mention or highlight when a reading is ectopic. it would help a surgeon to know when they need to electrocute the heart during the surgery. Most important thing to us is

					that if I have an ectopic value, that it wouldn't effect the blood pressure. If it did not, then we disregard this reading as an incorrect one. Its good, but less in clarity compared to Graph1. The rectangles are a little confusing cause they are exceeding the normal range and become longer in shape. Graph 3 is my favorite, its the simplest clearest. Its draw back that it doesn't have a lot of details. Compared to graph 1, you cant tell much if you were looking at it from a far. while graph 1, you can interpret a lot of info even when looking at it from a farther distance.
7	M	33	Bachelor degree art & Lit.	3	First 1 s 11 days. Each graph is covering different durations. The legends are clear. The max and min norm limits are clear in meaning. I don't get the rectangles. I get the hard black line is the median, but not the meaning of the box. I think If I'm working out I would most be interested in the outliers, the far values. Graph 2 is the exact same idea of G1 but in a longer period. But how does is portray 19 days and the title says months? The average is not included in Graph 2. Graph 3 Is very clear. It covers 140 days. the label "%11 of total readings". I prefer if the graphs has some bullet points to convey what I should get from this graph. some very brief explanation to make more understandable. In the beginning I though its complex, but no as I saw them again they are general clear. However, the outlier i think is very important. I think the hard black line means that my average HR during the day lied in that number. But the

					squares are not clear, I think they need to be explained.
8	F	25	Bachelor degree art & Lit.	3	Graph1 its clear. its HR vs Days. for example, day 2 was the day where most outliers occurred. The colored lines representing max, min, and average, are explained in the legend, I think they help you in classifying the readings. Graph 2 I don't think I got it. I don't get what do the boxes represent. The lines are clear. I understand the median representation in the black hardline. for example, does this mean that month 14 was so high and over the maximum norm? I'm not sure I understand it. Graph 3 is very clear, shows the days with really high occurrences of outliers easily and clearly. The last chart is easy to get. Its a representation of HR vs Days.
9	F	39	Bachelor degree science	3	G1 I can see its HR vs days. The boxes represent readings per day. Day 2 seems intense, maybe cardio, cause theres a lot of outliers. Day 7 also a lot of high and low outliers, maybe he was emotionally unstable. I had to remind myself what box plotting means, but yes overall I think its easy to understand. What to conclude from this Im not sure. G2 I can see that you changed it to HR vs months. that explains the longer boxes. it also explains

					why many days doesn't have outlier points. Another point, even the days with outlier points have become more condensed. Maybe there should be another box to combine them. I think this graph can give more accurate data/idea about someones health. G3 reminds me of scattered diagrams. I like it more cause Its visually more pleasant, and doesn't require a BG in statistics to understand it, and helps more in noticing trends. Maybe this will be enough to have a quick judgement on a case quickly, I think this would be valid though only when you have large scale data. It states “11.2% of total” does it mean total data collected out of 1 participant, or from all participants. Therefore I couldn't make out what this exactly meant. Overall, it easily shows the over the norm and below it, and whats trending. I like it more. while G1 and G2 I think they are good for someone who is more interested in detailed info. suggestions: I think the visuals, to serve a larger audience and gain benefit. the graphs should be simpler representation, and abstract the info. If the user is interested in knowing certain numbers, then G1 and G2 helps more, for trends G3 helps more.
10	M	33	bachelor degree science	3	All graphs are very straight to the point and clear. I think the second one, since its months instead of days, might be difficult to fully understand all information for some people since the data its a lot of data compressed in one graph. but other then that all are clear.
11	M	22	High	1	I think that the graphs are 2 complex, would rather have a simple representation that would

			school		show me for example only alarming readings represented in color like red for too high and blue for too low.
12	F	33	Bachelor degree art & Lit.	1	I didn't understand anything. (They just took a glance and then said that).
13	F	30	Bachelor degree art & Lit.	1	I didn't understand anything. (They just took a glance and then said that).
14	F	25	Bachelor degree art & Lit.	1	I didn't understand anything. (They just took a glance and then said that).
15	F	43	PHD of science	3	G1 HR vs day. blue and red represent min and maximum range. the green is the average of HR in each day. black hardline is the median of each day. Id like to know whats the difference between median and average, I think normal users need that explained. I think its interesting to see outliers WITH-IN the normal range. outliers are unexpected values far from the rest. G1 looks good and clear, but I would want to learn more about the outliers and the reason behind them. G2 its Hr vs months. so I have 19 months. I wish I can understand the meaning of the box lengths, as it is hear surpassing the normal range lines. i can see there is a big variance in median, but still mostly within the normal range. Most of the outliers are outside the normal range here, the max and min. I think this graph helps in giving me a bigger image of the heart rate performance. I can tell there were 3 months the HR values were so high, which might tell a fact about the HR in

					these 3 months. G3 shows Hr vs days. meaning from 0 to 140 for one participant. It shows only outliers above max and below min. and outliers in a different color that can be more extreme values. “11.2%” I didn't get this statement. I cant tell what I can get from this graph. It only focuses on values that are not normal. blue means over max, orange below minimum. (the participant went and zoomed in) If I could zoom in I would see the values that show frequent very high readings for example. The more readings I get on both sides, the more outliers I have in that day, such as this day, its crazy. Zooming will help me view details more clearly in concerning days, and will help exclude error readings if they turn out to be distant from any other value. If theres a lot of outliers, maybe i did a lot of activity that day or theres something wrong.
16	M	60	Master of science	3	G1 from a first quick look doesn't seem to be clear, it might need some more clarification in the legend when it comes to the squares. The hardline is the median, which is understandable, but the dashed lines coming out of the box are also not clear in meaning. I think the boundaries of the recorded data is either represented by the box, or by the dashed line edges that are coming out of either side of the box. For example day4 had the most activity, did he actually go over the 100-max, or does the box itself represent the limit of what his HR reached that day. Also, the average, the middle green line, does it represent the participant’s average, or the general average of people. Finally, the watch did give me info on my range, steps, and so forth, but it does not tell me if

					I'm doing ok or not?? there are no comparators. Also G1 and G2 I think the order of the items in the legend would be better if they followed the order they show up on the graph from top to bottom, i think it would make more sense I think. G2 is a longer period, over a year, its months vs HR. the Hr is mostly within the range. Again I wish to understand the users max and min are they the dashed lines, or the box edges. Also, the participant if he saw the chart would see if he saw this and wanted to know if he was doing good, the chart should give him that kind of reference to gauge their health. However, I can clearly compare info about my activity performed among months and motivate to do better if it showed I was slacking compared to prior months. I also, I can see the goal from this chart, cause when I wear the watch I wont see this chart, I will only see the raw readings from the watch. I imagine these charts are only for the person who will analyze the data, t=and the analyst would give his results from these graphs to the user. G3 is simpler than the others. The max and min are clear from the color coding. Its HR vs Days. However the blue is starting from a single level, I don't know how that is possible, maybe his HR is level? "11.2 total readings". Again is there a reference that I can compare my HR to the proper healthy HR rate? However its a simpler clearer graph, but cause its has less info, or maybe cause it serves a different goal from the other 2.
17	M	27	Associate's degree	3	Its pretty straight forward.. But I would do something about the outliers.. Get their average, and have them contribute to the graphs data. The

					comparisons work very well. Also, the months graph doesnt have an average line maybe that would help with giving a better perspective.
18	M	30	Bachelor degree art & Lit.	2	G1 it shows there an increase then a decrease in the data. the separator green line shows the center of the normal range. I think the graph with grid lines would make it more clear. G2 showed it clearer since the rectangles will longer and could show the variations in HR readings in a more visible way. G3 the number of points and the colors being overlapping, impairs my ability to distinguish between how high or high low the values are.
19	F	27	Bachelor degree art & Lit.	2	G1 I can see that the user had a HR that were mostly good and within the normal range, didn't surpass the max or min limit much. G2 shows that at first the HR did go over the max limit, after that it went back to normal, then at a certain day it was again close to the max but did not surpass it. G3 I didn't understand this one alot. I think the blue and the orange should be like the grey (I think she means that they are ll outliers) but what it seems that most of the readings are between the high and low.
20	F	56	Master of science	2	Graph 3 is the easier one to understand I'm getting the maximum, minimum, and outliers quickly and easily. Then G1 you have to concentrate little to get the important values.

					Graph 2 is complicated and you need to look more than one time in order to get some information about the heart rate.
21	F	32	Master of science	3	G1 what i know is that a box plot, the upper and lower lines are the max and min readings, therefore I got confused how the outliers were out of the whiskers? also, there are outliers within the norms! So, again, why those out of the norm are outliers, and why are they out of the box plot, cause the whiskers of the box should cover all data since they represent the minimum and maximum. G2 I have the same issue, but its HR vs monthly. Is there going to be a supplement to each graph to help explain whats is being expressed here? However for both graphs i think its a good representation for HR. a box plot is a good choice i think. G3 i think i have the same issue. How are we distinguishing between the outlier and the over max and over min. They are all outside the normal range, and I thought only the outliers are the ones outside the normal range. The "11.2%" statement, and see the graph, I don't what is this statement based on, and how did you choose that 11.2 of all the data per day ? I got that you chose the HR you chose per day? I don't know if it reflects the actual readings or not! Personally, since its HR and I want to see all the data, I would not want point-per-reading data, I would rather have an accumulative representation like the box plot to see a better overall picture. I can see max, min, median and the range of my data in one go.

22	F	34	Bachelor degree art & Lit.	2	G1 was the hardest to read, the Graph 2, then Graph three was the easiest in comparison. G3 was the closest to the normal graph representation of the HR we are used to, its almost giving a zigzag form through the overall shape of the recorded readings, hence it gave me a sense of whats happening in the performance of the HR. The legends of G3 is simple and clear, the rest had a lot more details. I kind of got a sense of whats happening in G2.
23	F	41	Master of science	3	G1 reminds me of quality control charts used in project management. you have an upper limit and lower limit, and the HR for a person throughout the day. I can see that you have median and average, but its not clear what the difference between them. I don't know whats the significance of maximum norm and minimum norm when there are outliers. you almost have outliers in each day. I can understand its the HR taken several times through the day. I can understand the bar charts are not the same length. G2 Im assuming that you are taking the data over several months period. It shows how is the HR of the patient is developing over the months, or how is the ranges developing over the months. For example for months 1,13,14,15 theres big variances and that can be indicator that theres abnormal activity in the heart. Maybe the patient is doing some hard work or exercises or whatever. Sorry for using the phrase patient, maybe the person is the better word. I can see that there is lots of outlier data that is appearing in this graph. G3 I can see that it is something in between the

					first and second in terms of interval covered. I can see that there is some vagueness to me. I understand that you left the area of the normal heart for most people empty, and you are checking here not the normal heart rate, but the people who are out side those limits and how they are trending. It seems that people with readings over the maximum norm are more than the ones lower than the minimum. I can see that the outliers towards the end of the graph towards the 120 mark become higher. For me it not that clear, though I can see the goal or purpose of this graph is more clear, it to highlight the values that are out of the safe normal range.
24	M	58	PHD of science	3	Chart 1: Heart rate vs days -The red line represents the maximum norm -Blue shows the minimum norm - green is the average - black is medium per day circles shows the abnormal ones (outliers) Graph 1: Minimum norm is almost equal to 60 bpm Maximum norm is almost equal to 100 bpm Average is almost equal to 80 bpm Chart 2: Heart rate vs months -The red line represents the maximum norm

					-Blue shows the minimum norm - green is the average - black is medium per day circles shoes the abnormal ones (outliers) Graph 2: Minimum norm is almost equal to 60 bpm Maximum norm is almost equal to 100 bpm Chart 3: Gives the outliers , high and low heart rate -Grey shows the outliers -Blue is over the maxim norm - Red is below the minimum norm The unit of the heart rate is given by BPM (beats per minute)
25	F	30	Bachelor degree science	3	G1 I get that the median is my own reading. Min and max are clear, they represent the minimum heart rate a person can get and the maximum heart rate a person can get. I can see that there were some days I was above and some I was below, but I never reached the average, bit I don't get what is the outlier. G2 looks like that this person has huge problems cause he doesn't have an average. Why are there boxes? and why are there lines? but graph 2 doesn't not have average. G3 its clear its not one of those plotted graph things. Im usually not good at scattered graphs. Again i don't get what it outliers, but theres a lot

					of them. Below norm makes sense, the high ones are between 100 and 120. I can see the below norm are between 60 to 50, which seems stable. But what is this person doing? are they running?
26	F	41	Bachelor degree science	3	G1 of course you need a minute or so to be able to read the graph to get what its about. So, HR variability by day for this participant. So its for 1 participant. The legend is a little unorganized. The minimum norm and the maximum norm and average, are kind of static or constants, they should have come up as in their order in the graph. The median since it has variability, it should have showed up in the legend before after the static values, its not very clear. I took 30 seconds to take everything in and what its about. The outliers was explained to me that are distant values. The boxes, I cant understand much, I can assume its the rate of the normal HR and thats their corresponding median. and its over 11 days of tracking for this participant. I get that the participants readings are curving up and then down. Went up till day 5 then down, then up again at day 8. I get its meaning, but maybe not its usability. The dashed lines coming out of the boxes i don't get that. And about the rate, for example day 7 is really up cause he had a lot of activity. In day 2 there are a lot of outliers but the median is under the green line. Definitely the lines coming out of the boxes i don't get. I suggest improving the graph, and adding some introductory text before seeing the graph. G2 Its a third P, so its not the same person. Its over 19 months, so its over a year and a half. You removed the average here I guess since its months. Still i don't know what the width of the box, I'll follow the median to see the HR of the

					person. The median helps me to conclude things. I can tell that there are outliers here, as well, but WHY are they not considered within the data. So I guess for me I need to know what the box represent, and their lines. Overall, if he's a normal person in the healthy range, then he/she should have HR wishing the normal range. I can see the P in month 1 and 2 showed readings way over the max, but does that mean all the readings in that month were really high? or the average of that month was high? and again the outliers are discarded from being considered with the data that makes the median. G3 HR bpm over 160 days. Im sure its still about 1 person. but it doesn't say which participant is it, or maybe I'm supposed to guess which participant is this from Graphs 1 and 2. I think this is P3 I think, since he has jumps in the same spots. Oh no this is a different duration, 160 days, so its totally different. and here all the normal readings are omitted here. So we are not looking at normal readings, we are only looking at abnormal information to check it out, the ones below the min, over the max, and the outliers. So this is what this graph represents, and no focus on the average. So here this helps us to see the abnormalities. Though it doesn't have a lot of information, but its the clearest one for me. "11.2% " Oh I get it! so all the readings we see is the abnormal readings, all normal ones are ignored here. I make more sense of this graph, but maybe because I saw the first 2 graphs first.
27	F	25	Bachelor degree art & Lit.	3	G1 is for a person wearing a watch. I understood its a watch that tracks the HR. G1 is for P1 by days. G2 s for P3 by months. The readings i can

					see are high, but the squares i don't get. The first graph shows that most readings are mostly stable within the normal range. However graph 2 shows inconsistent heart performance, and some of the squares/readings are going way over the max, and fewer times going lower.
28	F	19	High school science	2	G1 all was clear except the boxes. The colors helped me a lot and the legend and the numbers, made it clear to understand. But I felt it was crowded. The squares are not explained even in the legend and i feel they are making the graph unnecessarily complex. G2 is just like G1 but its in month, therefore the square became longer rectangles. G3 is so simple and will be understandable by everyone. What made it really clear is that its not crowded with information, and the colors. I think the graph without the added horizontal lines are better.
29	F	41	Bachelor degree science	2	G3 is best. data more accurate and can easily be read and seen. but doesn't show the average and median. The legend shows whats over max and min, but adding a highlight to the really concerning readings on the chart on the point itself in the plot, not only the legend.
30	F	18	High school science	2	G1 I felt was the easiest one, that is because the X axis had less numbers and hence made it less crowded and easier to read since theres more space between each grid. G2 I felt it to be more difficult to understand since

					the X axis had way more data squished together. G3 I think is the most challenging to read, since the points are heavily overlapping, therefore its inaccurate when you would try to get the value of a reading for a certain point, the resulting number will be inaccurate due to this overlap.
31	F	32	Bachelor degree science	3	G1 shows HR of P1. The outliers is the reading of the heart pulses. The blue line is the maximum normal healthy range, and similarly the red line. The median and average are just statistical measures. However, I didn't get the boxes, and why are they there, and what does the dashed lines represent? Also, the time of the readings are not mentioned. As a user for an app, I think they need to know these things. G2 is by months not days. Otherwise its all similar. The limit of the Y axis is different between G1 and G2, how come its not like G1's HR limit on the Y axis. also the timing, I don't get when the HR readings occur. G3 the grey are the outliers of the HR. the blue is the over maximum norm , does this mean the number of people that were over the max norm? I was told the the outlier was the pulses that are far from most data, does this mean this is for one person. "11.2" is this the percentage of people you measures? it also says bpm, what is bmp? I also don't know if the outlier is a single reading? or number of occurrences of readings at each data point? finally I suggest that the user would want to see the time of when a reading was taken.
32	M	46	Bachelor degree	3	Its all understandable, but statistically I think its better to highlight the alarming highs and lows to

			science		help the user in recognizing concerning readings. There are smart sensors that accommodate to changing normal ranges. This is a challenging software problem since it has to be accommodating to those changing limits.
33	F	30	Bachelor degree art & Lit.	3	G1 shows that the P has somewhat settle HR performance. on average it didn't show a lot of variance or extreme high or lows. It usually went between 65 and 85, except for the outliers, they are high, and almost always high, I don't get why is that. G2 Is over a period of months. OMG the performance is very irregular, goes up and down in a non gradual way. It went over the over the maximum and never went under the minimum. But I wanna know are these the readings of a normal person, or is there something wrong with him? G3 is it a grouping of the other 2 graphs? I think it is useful to show me heart performance over time. However the boxes I didn't get what they represent.
34	F	36	Bachelor degree art & Lit.	2	G1 I didn't get the squares. The lines are of max and min are clear. The average green line is also clear as well as the median. However I didn't get whats the difference between the average and the median. Why are there squares? it should be all lines. G2 is the same G3 It was the most that visually made sense, cause it didn't have a lot of details. I also easily concluded that the values in the middle are the intermediate values / average. Visually again it

					looked more obvious to the eyes, it made more sense and was easier to understand, it was self explanatory, I didn't need to loop up the legend much.
35	F	36	High school	1	G3 I understood a lot better than the other 2. G1 and G2 had too many lines and boxes.
36	F	39	PHD of science	3	G1 the graph is the clearest I think cause it gives the closest picture, like hour by hour, or no like shows all the details. It shows all the details u need to see and give the max min healthy range and the average. But the box, on what basis the range of the box keeps changing ? I'm not sure maybe I'm missing something. I think the box range represents the readings recorded across the hours. If that's true then it's actually clear. The outliers I think they are plenty for one of the hours, but overall it's understandable. G2 its the same as the graph elements in G1, only its graphed over a monthly period. I think it's a clear good graph. G3 it's nice for when you know what's wrong. Shoes the abnormal behavior. I like it cause it's for people who are not interested in the normal behavior of the heart, but they want to see the abnormal behavior , which is good. It clearer for people who only want to catch what's wrong. Overall all the graphs are understandable and clear.

37	F	39	Master of science	3	G1 & 2 they kind of have the same idea, only one of them is by day, and the other is showing the data by month. In general these 2 figure I consider clear. However, I think adding or emphasizing colors and highlighted data would probably make it more attractive to the eye and easier to read. For example if there were really high readings HR points, they would be in red to make them more visible, and then the high readings but less concerning would be in orange, and so on. Also, if there were numbers on the plot that would probably also help, for example the min and max and average would be displayed somewhere on the plot as numbers. G3 from the quick look its the most that caught my eye, maybe cause of the colors. But later when I took a closer look at it, I think its very unclear if I wanted specific information that I can use cause the dots are so close to each other that I wouldn't be able to make out clear numbers of a given corresponding reading. I suggest the data be less crowded here to allow more space.
38	F	42	Master of science	3	G1 and G2 I felt they are the same, not much difference in their visual representation. The third seemed to be mix between the first and the second, except that it covers a longer period of time. For example if I zoomed in G3 then I think I would get the something similar to graphs G1 and G2. As for benefits I think they can help me understand the status of my heart, or my HR performance during a workout for example, and if it was complying with the normal ranges and so on. I don't use an Apple watch and not big on looking my measures in numbers, but I assume that would be the benefit.

39	F	18	High school science	2	At first look I thought that the first 2 graphs specially were scary and complex. However, right after I took a closer look, I thought that of course if a legend is there, then it will always be clear. The legend explains everything, I don't think there would be any problem in interpreting the data given the legend being available.
40	M	37	Bachelor degree science	3	G1 shows a plot of HR readings against days. It covers the period of 11 days. From the legend I can read that the plot will be showing the boundaries of the minimum health HR and the maximum, and it also shows the average. Whats the difference between the median and the average however, to me they seem similar meanings. I can see that the second day have lots of outliers, this can mean that the participant had a lot of activity in that day. To make more sense of this, I would like to know the ratio of the outliers in comparison to the total number readings because since outliers in day 2 are very high, and were 5% of total readings, then I wouldn't be very concerned, unlike if they were 25% for example. G2 is similar to G1 but it is in months. I can see the first and second months are out of the normal range, but the boxes became much longer, why is that? This graph to me however, would cause me to seek a doctors advice if my median was going out of the normal range case it automatically implies a longer more consistent period of alarming readings. G3 shows a period of about 5 months, and only focuses on readings out of the normal range. I think its pretty vague, but the “11%” statement tells me the alarming readings represent only 11% of my entire readings therefore I wouldn't be that

					concerned since its a small percentage.
41	M	21	High school	2	Chart 1: what i understand its like Im studying someone by his HR and this is Patient 1 first day his HR was low. second day lots happened, don't get the dots yet. whats the red and blure, ah ok, its the high and low limits. fifth day mostly normal, 7th day too high, many dots go up. this person is either having a disease or you are studying without him knowing if he has a problem and you are trying to help him. Chart 2: P2 and HR by month not days. theres 3 months the dots are not there, it was between the normal i think, maybe those 3 months are the summer vacation. I think he's a student and had a vacation in the summer. Chart3: I can see the high and low, why is there nothing in the normal. Are you only measuring the high and low readings. Im not sure if this is for multiple people or one. I think this is for one person and its a test thats trying to figure out whats occurring more the high or low, and trying to find a solution for his problem.
42	M	20	High school	1	Chart 1: we have 2 problem first the dots in the chart i don't know what they are exactly. Is it talking about Heart Beat rate. Chart 2: I understood if the HR 100, the readings in months 40 to 12, the HR are not ok or normal. Chart 3: small percentage of outliers are less than 50, and some are higher than 50. Theres a big

					percentage are over 100. Theres a lot of outliers specially above 100 and less than 50.
43	M	22	BS	3	chart 1: Box plot for a patient that describes Heart beats per minute. Its across 11 days, mostly centered around 80 bpm. The box and whisker plots are in between 2 values which are 100 and 60. The only distinct thing noticeable that day 2 has a lot of outliers, meaning abnormal behavior in that day. Same case in day 11. I think most of abnormal behavior would be in day 2 and 7, other than that its mostly around the average centered in 80 beats per minute. chart 2: This is another box plot, its for a patient showing monthly variation for his heart beat. Judging by this graph the patient has a slight problem in months 1 2 and 3, basically because the box plot is quite lon and hence a lot of variation. There is Specially a concern in months 1 and 2 cause the median s above 100. so 50% of the reading at 100 or less, indicating a slight problem. It appears he has slight problems 13 1415 there is a serious heart problem that the patient is suffering from. 14 is the most concerning as the maximum is almost reaching 200, and these are not even outliers! these are readings close to the average. He is definitely having a serious heart problem. chart 3: Seems that this is a distribution of the high and low, the extreme Heart beats per minutes for several days. looks like in general there is an average between how many high and low readings that are being obtained. even though its 11.2% of the readings, its an indication that its an average between the 2. theres a gap that shows theres no low readings. there is a gap, this can

					indicate in HR may refer to a problem, also most outliers are considered high. Tho they don't represent most data, the outliers here tend to be in the upper section. meaning most outliers are a raise in bpm.
44	M	39	Bachelor degree science	2	chart 1: Heart rate variability for days. first day seems nothing recorded. second day HR was more than 100 alot, and a good part that was less then 100, and a fewer readings around 60 bpm. the following days revolve around 80. 3rd most are around 110. the fifth day most under 100. 6ith are most in the range of 60 to 100. 7th is going way high almost 160, …. in the 11th most are within range. I guess day one nothing was recorded. second day maybe it was a weekend and they exerted effort. the 7th day shows a lot of variance in comparison, either they have a problem or they weren't wearing the watch. Not sure if they were wearing the watch consistently or sporadically through out the day. i didn't get the rectangles. chart 2: this chart is by months. over 19 months wow! .. it shows most are over 100, but 456 and tis 9 are all similar. until month12 the readings get higher than usual, but then they return to the norm. the outliers. is this showing the max and min? the 60 and 100? but there are reading out of those ranges, so why are there really high readings or lower ones. chart 3: everything represented in dots. you maybe want to focus on the readings out of the

					normal range.
45	M	20	High school science	3	chart 1: HR is measured over 10+ days. the highest point and lowest points are shown. the black middle line is the average of each day. chart 2: similar to chart but the time is represented in months. but also, the dots represent the uncommon/gay outliers. chart 3: this chart shows all the readings, but the instances of high or low or outliers are the ones shown here. all together these particular points represented 11.2% of the readings, which means 88.8% of the times the HR was normal within the normal range.
46	M	43	PhD of science	2	chart 1: Things I believe I did understand: - The blue and the red lines present the heart rate range for the participant's age group. - The box presents the participant's heart rate during the day, where the lower part shows the participant's lowest heart rate, the upper part shows the highest rate and the black line in the middles is for the average rate. Things I didn't get L - The small circles in the chart. - The dotted vertical and horizontal lines above and below the boxes. Chart 2 " Heart rate variability by month for participant 3" Same as chart 1 Chart 3 "outliers, high, and low HR"

					I find this chart confusing for the following: - The color coding for low and high is different than the one in chart 1 and 2 - Data presented by those blue and red dots need to be zoomed out. It is hard to tell if the heart rate values belong to different or same day. - I don't find this graph useful. It doesn't give meaningful information that chart 1 and 2 couldn't provide.

Noteworthy feedback on Graphs 1, 2, and 3 from Study 3:

On G1 and G2

"the boxes are confusing a bit, I don't know what they represent."

"the time of the readings are not mentioned. As a user for an app, I think they need to know these things."

"The colors helped me a lot and the legend and the numbers made it clear to understand."

On G3

"This shows that all heart rate readings seem they are starting off from one level, and this is confusing"

"G3 I think is the most challenging to read, since the points are heavily overlapping"

“Visually again it looked more obvious to the eyes, it made more sense and was easier to understand, it was self explanatory, I didn't need to loop up the legend much.”

“G3 from the quick look its the most that caught my eye, maybe cause of the colors.”

“I can see the high and low, why is there nothing in the normal. Are you only measuring the high and low readings.”

“The third graph, I understood, but I didn’t like the big gap in the center.”

General:

“I think its better to highlight the alarming highs and lows to help the user in recognizing concerning readings.”

“Whats the difference between the median and the average however, to me they seem similar meanings.”

“I think If I'm working out I would most be interested in the outliers”

“adding a highlight to the really concerning readings on the chart on the point itself in the plot, not only the legend, would be better”.

“ I would like to know the ratio of the outliers in comparison to the total number readings because since outliers in day 2 are very high, and were 5% of total readings, then I wouldn't be very concerned, unlike if they were 25% for example.”

“I would not want point-per-reading data, I would rather have an accumulative representation like the box plot to see a better overall picture. I can see max, min, median and the range of my data in one go.”

Study 4: Survey Questions and tables

Q1. Please enter your initials:

Q2. Please enter your age:

Q3. What is your biological gender?

- Male
- Female

Q4. Do you workout regularly?

- No I don't
- Once a week
- 2 - 3 times a week
- 4 - 6 times a week

Q5. If yes how many hours per workout?

Q6. How often do you consume caffeine? [e.g. caffeinated tea/coffee, soda/soft energy drinks, chocolate]

- none
- 1 - 2 times a week or less
- 3 - 5 times a week
- once a day
- multiple times a day

Q7. How much water do you drink per day?

- less than 1 cup
- 1 - 2 cups
- 2 - 4 cups
- 5 - 7 cups
- 8 - 10 cups
- more than 10 cups

Q8. Do you currently smoke cigarettes, if so how often?

- non smoker
- rarely
- occasionally
- frequently

Q9. If yes, about how many cigarettes do you smoke in a typical day?

Q10. Do you have any of the following symptoms, and how often?

Palpitations or racing heart (i.e. rapid/strong or irregular beats due to agitation, exertion, or	Never	Rarely	Occasionally	Regulary

illness)

Chest pain	Never Rarely Occasionally Regulary
Light headedness or dizziness	Never Rarely Occasionally Regulary
Loss of consciousness	Never Rarely Occasionally Regulary
Fatigue or lack of energy	Never Rarely Occasionally Regulary
High blood pressure	Never Rarely Occasionally Regulary
Low blood pressure	Never Rarely Occasionally Regulary

Q11. For the symptoms selected in the previous question, what were they accompanied with?

Palpitations or racing heart (i.e. rapid/strong or irregular beats due to agitation, exertion, or illness)	Fast heartbeat Low heartbeat both don't know
Chest pain	Fast heartbeat Low heartbeat both don't know
Light headedness or dizziness	Fast heartbeat Low heartbeat both don't know
Loss of consciousness	Fast heartbeat Low heartbeat both don't know
Fatigue or lack of energy	Fast heartbeat Low heartbeat both don't know
High blood pressure	Fast heartbeat Low heartbeat both don't know
Low blood pressure	Fast heartbeat Low heartbeat both don't know

Q12. Do you take any medications that affect your heart rate or blood pressure?

- yes
- no
- I don't know

Q13. Are you diagnosed with any medical conditions that affect the heart rate or blood pressure?

- No
- High blood pressure (Hypertension)
- Low blood pressure (Hypotension)
- Thyroid disease
- Sleep apnea
- other (non cardiac diseases)
- cardiac medical conditions (please indicate):

Q14. How many times (per day) did you measure your readings with the medical device?

- 1
- 2
- 3
- 4

- 5
- more

Q15. What did you feel about the experiment? tell me about the use of the device as well as the watch for this period?

Q16. After using the watch, did your perception of your health change? Explain if yes

- No
- Yes

Q17. After using the watch, did your behavior or routine activity change? Explain if yes

- No
- Yes

Q18. What do you like the most about using the apple watch to track your heart rate?

Q19. What do you like the least about using the apple watch to track your heart rate?

Q20. When you are considering new products in this area, what attribute do you consider most often?

- price
- value
- brand
- innovation
- quality
- Other (Please specify):

Q21. Overall, how difficult or easy was it to get your heart rate information from the apple watch?

- Extremely easy
- Somewhat easy
- Neither easy nor difficult
- Somewhat difficult
- Extremely difficult

Q22. How likely would you recommend using the Apple watch for tracking your heart rate, to a friend, family member or colleague?

- Extremely likely
- Somewhat likely
- Neither likely nor unlikely

- Somewhat unlikely
- Extremely unlikely

Q23. What is the highest level of school you have completed?

Less than high school

- High school graduate
- Some college but no degree
- Associate degree in college (2-year)
- Bachelor's degree in college (4-year)
- Graduate degree (masters, PHD)
- Professional degree (JD, MD)

Q24. Please select the status that best describes you?

- Single
- Married
- Divorced
- Widowed
- Prefer not to say

Q25. Please describe your daily routine.

Q47. How busy does your daily routine get?

- 1 (not busy)
- 2
- 3
- 4
- 5 (very busy)

Q26. Are you a caregiver to any member in your family, including minors (e.g. your children) or elderly parents? If so how many? (choose 0 if not applicable).

- 0
- 1
- 2
- 3
- 4 or more

Q48. Are you financially responsible of anyone other than yourself? if so how many?

- 0
- 1
- 2
- 3
- 4 or more

Q27. How would you rate your attention to your own personal health?

- I dont pay attention to my health
- I pay moderate attention to my health
- I pay adequate attention to my health
- I pay good attention to my health
- I pay great attention to my health

Q28. When mild symptoms do occur of any health concern (eg. fatigue, dizziness, high pulse with moderate activity), when would you be motivated to seek medical advice?

- I ignore it until its serious
- If it occasionally repeats over months
- If it occasionally repeats over 2 to 3 weeks
- After occurring 3 to 4 times within a week
- After occurring 1 to 2 times within a week

Q29. Do you think other people see you differently when you are wearing the watch? if so why? and would that affect your choice of wearing it regularly?

Q30. Which statement best describes your current employment status?

- Paid employee
- Not working (looking for employment)

- Retired
- Self-employed
- Prefer not to say

Q31. During your most recent medical visit, how would you rate your satisfaction with the amount of time that your healthcare provider spent with you?

- Extremely satisfied
- Somewhat satisfied
- Neither satisfied nor dissatisfied
- Somewhat dissatisfied
- Extremely dissatisfied

Q32. During your most recent medical visit, was your healthcare provider aware of your medical history?

- Definitely yes
- Probably yes
- Might or might not
- Probably not
- Definitely not

Q33. How long have you been using your current healthcare provider?

- Less than 6 months

- 7 months to 1 year
- 1-2 years
- 3 or more years

Q34. How likely are you to continue using your current healthcare provider in the future?

- Extremely likely
- Somewhat likely
- Neither likely nor unlikely
- Somewhat unlikely
- Extremely unlikely

Q35. How likely is it that you would recommend your healthcare provider to a friend, family member or colleague?

- Extremely likely
- Somewhat likely
- Neither likely nor unlikely
- Somewhat unlikely
- Extremely unlikely

Q36. How positive or negative was your first reaction to using the apple watch to track your heart rate?

- Extremely positive
- Somewhat positive
- Neither positive nor negative
- Somewhat negative
- Extremely negative

Q37. Overall, how difficult or easy was it to get your heart rate information from the apple watch?

- Extremely easy
- Somewhat easy
- Neither easy nor difficult
- Somewhat difficult
- Extremely difficult

Q38. If you're interested in getting feedback on your analysis, please provide your email:

Study 5: Transcribed User Feedback

P no.	Gender	Age	Education	Score	User Feedback
1	M	66	High School	2	I can see on the graph a group of points, and I see an X and Y axis, I see a black line. I see day week month year. I see a date of a specific day selected. The menu gives me resting , and active HR information. After clicking resting. It shows me healthy heart rate when resting. It's the same drawing and points as before, but now there are added colors as an extra layer over the points.
2	F	30	B.A.	3	I see the numbers first, and the labels for them. Is this the time? Ok so then this tells me my HR according to time it occurred. I want to see when my HR reached the maximum. I can see when it rized, what it was when I first woke up for example. I can also see my HR across the day, week and month. It is easy to understand. On the menu on top, it give me the option to see my healthy rate when resting. The resting HR option show me the healthy HR in grey color. I see The legend is below. So the red and blue shows the HR that are over the maximum and lower than the minimum.

3	M	36	B.A.	3	This seems like a plotter that gives a chart for the HR. But why is the plot this way? The plot shows me the average of the HR. and the time frame of the HR across day, month or year. The X axis is the time, the Y axis is the HR readings. The weekly looks different, it represents data as a candle stick. The daily readings are starting at 6am, maybe the user wore it starting at that time. The menu, when I click on resting HR option, this shows me when the danger zones of the HR occur. So the 2 colors blue and red show me when the HR is overly stressed at dangerous rates, or when its too low at again risky low rates. HOwever I think this won't be easy for the elderly to easily maneuver this application. It needs to be also available in Android to reach more people.
4	M	32	B.A.	3	I can see the average is 80. the application shows my HR over 24 hours. I can see the HR in days, week, months and years. It navigates like a calendar. The top menu options give me 3 choices, resting active and none. After clicking resting, I can see the healthy HR is shown in grey. So this is a filter that shows me the state of my HR when I'm active or resting. The red colors shows me readings that are not good, over the maximum, and blue for the minimum. Its clear and easy to understand.
5	M	38	B.A.	3	These are points that are supposed to be my HR. The top options show me HR by days, on Sunday for example there was more activity, and then I can see by months. The top menu, gives me options, when I choose resting, the blue color shows the low range and the red shows me the danger zone.

6	M	28	B.S.	3	So at 6:00 HR was 60bmp. From 6 to 7 it was fluctuating between 60 and 70. I can see the HR by days months and years. On the top menu, when I choose "resting" for example, I can see at 8pm the HR was giving readings that the application is saying they are problematic. The blue color shown means that the HR is lower than needed, and the red shows the range that the HR is higher than the safe range, its very high and problematic.
7	F	63	B.S.	3	This is showing me my HR from 6am till night. Mostly the HR was normal, but at a certain time for 2 hours the HR was very high, I want to know why not shooted up like this. I can also see the readings in week, months, it shows me my HR in ranges, and similarly in years. When I click on the resting heart rate option from the menu, it shows the same chart. It tells me what's my average which is 89 which is normal. It shows me the HR at any given time. These colors highlight all the alarming HR values, or abnormal values that need to be checked. This is really nice! I like it, it shows when my HR is too high or too low.
8	F	27	B.A.	3	I can see points on the chart, and the time either by day month or year. I can see at 6am the HR was 80 and so on. When I click the interval of months, it shows me the average across the days, and across the year I can see which month was the highest rate recorded for example. The menu option on the top when I choose resting for example, it shows me the colors red and blue on the chart. The readings in red mean my HR was overly high, and the readings in blue show the readings that were

					too low.
9	M	31	B.S.	3	I can see the HR across time. The graph its clear the average s between 60 and 100 but there was a spike at a certain time, maybe they were running. But it decreased so suddenly. I can also see the average across different times from days, week, months, years.
10	F	44	PHD Sc.	3	I can see the HR in bpm across hours. I wish to zoom in to see each one dot. When I touch the dots I can see the numbers. The touch is not that accurate. I can see the date, nice. Now I click on weeks, it shows the median HR of one week. It gives me my summary over week, per day. It shows the outliers, median, max and min. This is an outliers which shows something might be wrong. When I click on the eye button, I see resting and active boundaries. But how would I know if I was resting or active. (she reads the legend). So the red gradient is if I am above the maximum healthy HR. Blue shows if its lower than the minimum. Is this the summary of my measurement? Oh! Ah! Ok I can choose both or just resting or just active. I went to Dec. 26 2016. On this date the chart shows alarming readings, because I got high HR and I am not in the active mode, that why the chart looks like this. If I click active, they all become grey, alarming readings becomes less.

11	F	35	Bachelor of Sc.	3	(she read the tutorial aloud). I see a chart with tabs for day, week, month, and year. The chart shows the HR in bpm from 8am to 3pm for this day. The monthly charts shows the ranges of the HR per day. The yearly chart I can only see 1 month. When I click profile I can fill the name and age. There is more info when I click on the eye button. It showed the HR in resting or active. If I click one of them, it colors the chart. For the active the range changes. So based on what kind of activity I was doing I can see my HR if its suitable or safe for resting or active. The interface is very simple, it doesn't require a lot of time to learn. It can be enhanced with features that would prompt the user to move if they were inactive for a long time.
12	F	40	Ms Sc	3	I can see the HR across the day in hours. I can also choose weeks, months, years. The weekly view shows the HR in bars, I think they indicate the majority of the HR resided within this box. I will go check now the resting HR boundaries under the eye buttons. I want to see if there is a difference between resting and active HR boundaries. (she's reading the info page). I think the app is pretty understandable. I'll go to the specific date. The HR is the same on both charts, but the boundaries change between the resting and the active. The resting chart shows that the HR at 140 and 150 is alarming to go this high high since they are not being active. I remember I read this can be an indicator that this is a problem. but those yellow/red dots make sense that they go grey when I click active, the heart is supposed to be higher, but they don't make

					sense if the person is resting.
13	F	44	Bachelor of Arts	3	(reading tutorial). There's a whole bunch of dots there. So this is by day, somebody's data starting from 7:30am in the morning. I switch to week, it shows me maximum and minimum. The screen from the start doesn't show me the label on the left stating this is HR in bpm. On average the HR is not too awful. I click months, its the same thing here, but I really like the default view would start on the left to see the label. I click the "eye" it gives resting active and none. I click resting, now I can see the HR boundaries if resting, cool! Now the chart has all these colors. Funny enough the chart with the colors doesn't move but looks nice.With active, i see info that yellow HR is alarming, reaching dangerous levels. So during a run, if there's peaks and high's that are immediately noticeable. The dude button gives me user info. I'm going to the date Dec 26 2016. I goto active, it looks crazy. When resting, Oh WOW, much more interesting. that makes me wonder about my heart if I’m having issues. That looks very very clear.

14	M	36	Ms Sc	3	I can browse HR across days in hours. On week, I can see the outliers and ranges. Going to month it shows me 30 days within the month. I click on the date, and I can explicitly goto a specific date. the person button gives me the profile info. There is the eye button. Show HR boundaries if active or resting. I click the question mark it gives me more info. Clicking on HR boundaries if active. It seems like from the first hours of the AM the HR is in the 120 range but then stabilizes between 80 and 60. When I goto resting, it seems higher in the beginning. Someone is resting here, when I click resting, resting is less, ok makes sense, the scale in the background is different. When I goto Dec 26 2016. When I click on HR boundaries, the red gradient and blue are predefined on to what is considered healthy or not. Readings in the white range are healthy, but in blue or red are not. When resting the user had alarming readings around pm, but when I chose active all readings are ok. So blue and red depend on if you are active or not, So if you are resting your HR should not go high like 120, but if active this is considered normal.

15	F	39	PHD Sc.	3	(reading the tutorial). I see the chart, it shows the HR on the Y axis and time in the X axis. The user's HR is between 60 and 100. Wasn't the user wearing the watch at sleep time? Anyway, it's showing the HR in dots. I switched to the following day there's no data, so he didn't wear the watch on this day. I don't like the dots a lot, maybe bars or lines to represent the HR, however I can clearly get the info. I click on "week" and I can see the max min and median on the boxes, it's clear I think. the dots outside the box are outliers. I can similarly go through months and years with same format. The outlier is either a wrong reading or a problem, Im guessing its wrong, otherwise it would have persisted enough to be included to in the max-min range. Im clicking on the other buttons. When I click on resting, and it shows me alarming readings in yellow. Since my condition is resting, my HR should not be high, and so when it's in the red range its considered alarming. If my condition is active, those same readings are considered safe, since it should reach higher values and so its not alarming to be high.
16	F	32	Bachelor of Arts	3	The axis on the left is the HR, the bottom line is the time. HR readings started at 10am and they go up and down. Across the week I see HR through the days. Each day gives me three numbers to show the HR. Same thing for months and year. But the year shows the user wore the watch for one month of that year. When I click the eye button, the HR when its in grey means its safe HR. When they are yellow/read means it's alarming and dangerous. The red gradient shows the overly high readings and the blue are the lower than

					healthy readings. Now I have resting and active options. When resting, if the dots are red/yellow means its abnormal, that there's something wrong. But if I'm active all the readings I see here are grey.
17	F	18	High school	2	This is a chart for different times (day week month year). If I choose day, this shows the date. The graph is HR vs. hours. These dots are the HR data. When I touch the dots, it shows the bpm, at what time during the day. Most readings are between 60 and 90, I guess that's healthy. When I press week, I see the days of the week, and the lowest HR recorded was 60, the highest 97. Same representation applies when I try months and year. When I click on the eye, I can see the HR boundaries for resting or active. When I click resting, most readings are in white range. When I click on "active" the boundaries change to a higher number. Going back to the day view, I see a scatter graph. When I clicked active most of the readings are in the white range. if they are in the red range, it starts at a certain number, after the HR passes this the number, they become alarming being yellow colored. So when I see the difference between resting and active, the boundaries that change. But whats the white range considered, it didt say, it just explained red and blue ranges.

18	M	43	Bachelor of Sc.	3	I am starting the app. I like that it has a tutorial. I can sort the data over days, weeks and so on. Main screen is clear. The icon of the "eye" I think is for the filters, it might be confusing. The alarming heart rate, can be in yellow. I think the yellow color can disappear in the graph, maybe another color is better. I am going to the week view, I can see outliers. How come you never mentioned to the user in the tutorial, but you did explain where to find the maximum , median and maximum. The tutorial looks a little different from the actual app. I think the app is clear, the colors of the alarming heart rate I'd rather be more visible. I like that I can double click on a box from the year chart, and it would take me to the months data, and then week and so on. When I go to a certain day and click on the active filter, the chart shows that my readings have no problems, but when I click on the resting filter, many readings are considered alarming, I'll panic and call my doctor. I think the app should find a way to figure out if I was active or resting, cause what if I saw these alarming readings after a month had passed and I can't remember if I was exercising or just sitting on my couch, Im sure I wont remember.

19	F	40	Bachelor of Sc.	3	I see instructions on how to interact and navigate the app. Tap once to get info, twice to zoom. Ok so this is my data hypothetically. The x axis is time, and y axis is Heart rate in beats per minute. I see the recordings starting from 7am till 2pm, so that's when I was wearing the watch. Many readings overlapping. When I touch the reading point, the number value of that reading appears. I navigate to the following day, no data recorded. I can click on week to see my data. I see rectangle blocks, they show my my readings per day. Some dots are out of the box, these are outliers I think. I can tap to see the max and min values, very understandable. Below there are 2 buttons, one with a person's icon to enter my information name and age. The other button with the eye shows me my heartrate boundaries in 3 states at resting or active. I click resting, it gives me a tip with information on the meanings of the colors. So I want to go to a day to see the dots. On Monday Jan 23 I have only 3 yellow dots, meaning it's not that risky, on the lower levels of red over-max range. Clicking on active all the values are considered safe. I think I saw all the app feature. Its very easy and very straight forward.

20	M	30	PHD Sc.	3	The tutorial is starting. "participant read all text aloud". I see a chart, Y axis is HR in bpm, the X axis is hours of day, I see the date, and flat data point, basically my HR is not doing much all day. I can access the calendar to select a specific date. If Im assuming its my data, then my resting HR is a bit high. Im looking at "week" HR, and I can scroll through the week. I can touch the box to see the values of min max and average. I'm back on day now, I feel the watch measures a lot more readings than what I see here, like only 10 readings within an hour. This user's readings are very stable. There's 2 icons in the bottom of the screen. I press the "eye" button, it shows my the HR boundaries, I can select this option to see my HR boundaries at active Its predicting what my healthy boundaries are for my HR at resting. There's blue and red gradients, is this where I want to be? I'm going to click the button again and choose resting. Now I see a larger red gradient, so if Im not active and some HR is in yellow and above a certain threshold then it's not ok. The gradients background are still there when I switch to weeks or months. The little person button gives me profile info where I can put in my info name and age. I think I covered the functionality.

21	F	41	Bachelor of Sc.	3	I think the tutorial is too long but useful. I can see the date, the Y axis Is HR beats in minute. I would like to see what's considered the normal rate. The user was wearing the watch from 8am till 11pm. I can switch to seeing HR by weeks. Ok the tip shows where I can find the max min and average readings. I can tap a box to see my HR value. If this is the max and min, where are there some dots beyond the max and min? But I like the fact that I can zoon through double tapping on the boxes. I can explore the readings from days to year. When I click on month, I can't truly tell the data is for which box, its jumping around, maybe the font should be smaller to fit. I can see some months with higher medians, so something different is going on with this user. I see two buttons, the one with little person shows user info. The eye button, when I click on more info it gives gives too much text. So the app doesnt see if I'm actually resting or active. Ok, I click resting, so I should use this if the user himself knew what state he's in. So if the user is resting on this day, the user clicks on the resting button, if dots are in yellow/red dots within the red range, that means he has a problem. Similarly with the blue range for readings too low. When I click active, this helps me compare in general how my heart is evaluated if I was either resting or active right? When active, the red boundary range went so far up, I think it's ridiculously high, but that means no readings will be considered alarming, which means my heart rate can take in higher values without being alarming.

22	F	56	Bachelor of Arts	3	The first chart shows the heart rate state through the day. He started wearing the watch from 8am until 2am, and I see the HR are regular and all between 60 and 100 and hence normal. I can see HR in days, weeks etc. I can see the date, the user started recording his HR from 6am until 11pm on this day. When I click on weeks, it shows the HR through the days of the week. The boxes are within the normal range meaning his HR is safe. I don't get the lines though. Wow t I can see how it is across a whole year. I click on the eye button, it says active and resting, and I see colors showing what healthy HR and alarming. If I click active, the colors mean the same thing. It shows me what's considered normal HR, and I can still switch from year to days. Now I'm seeing the date you wanted me to look at Dec 26, 2016. If resting the alarming range starts from 120 or 140 , most of the users heart rate are within the healthy range, except for the time between 1 to 3, This means he definitely was active her and did something. When I look at at the same day when hes active, the user's HR is all considered healthy. Wait I looked at them again. Going back to the graph when resting I think I was wrong. If this user was resting the whole time and he had those higher readings within the red range, then he has a problem. However in the active graph the red range starts at 180, which means the HR can safely go up to that level, but not beyond, cause it'll be too much for his heart.

23	M	62	Ms Sc	3	The chart shows my heart rate boundaries if resting.I can see the HR mostly in the normal range from day 1 to day 11. They are all normal between 60 and 100, its neither in the red high range or the too low blue range. The readings that are in yellow/red colors are high, and are residing in the abnormal range. If a person was looking at his rate, he will be able to know when his readings were alarming though he was at resting. Maybe he was upset or ate a lot that caused him to give such high readings though he was resting. If there was no cause to explain this, he must check a doctor. This definitely helps users to monitor their HR wellbeing with high accuracy. When clicking on HR boundaries if active, for the same day, my first observations is that the dots are the HR readings, I see they are identical to the ones from the previous resting chart. How is that possible. I noticed that the resting chart seems that the user was active more while the active chart all the readings were in grey and considered normal. It shows that the HR is in the normal range. I feel that the app is important and useful to monitor my HR and keeps record of my history and might make use of more features for added potential benefits. in total, all readings in white are healthy and safe. all alarming readings are either in red range being too high, or in blue which are too low.

24	F	22	High shool	3	The app shows me the HR over different times: days months and year. I can also enter my information in the account info button. When I click on the eye button, it says see HR boundaries for active or resting or none. When I click resting, ok this is explaining the colors. They grey color represents healthy HR, the colored yellow red is alarming, the more its red the more dangerous it is. The red background these are basically showing when you crossed the healthy limit. The higher you go up its worse. Basically the chart now after choosing resting, if he is resting, then he is not doing anything, the supposed healthy maximum is larger than the one of active filter. So when he is resting, so the readings entered in the dangerous zone since he was resting. But when active is chosen those dots, the HR, is considered normal since the HR should go high once active.
25	F	18	High school	3	The app is very simple has a simple structure and easy on the eye. I can easily see the HR in the chart. When I goto the eye button, it says it shows me the healthy boundaries if I am resting or active, grey dots are healthy HR and yellow is alarming ok! now if I choose active .. hmm ok , its clear. Ok now I'll see "resting" and .. Woaw!! the yellow readings are alarming well that's alarming HR then. Cause this means Im just sitting down maybe but my HR is shooting up like that.

26	F	36	highschool	3	(The participant was looking at data of Dec 26 2016) The chart shows me dots for the heart rate. The chart that specifically shows the resting heart rate boundaries it is showing me the same data, but I can see the higher heart rate beats when they reach high concerning levels. Since they are resting the reason might that they are very stressed, concerned, or having issues with their blood pressure. The colored in yellow readings are not healthy for a person. If this repeats this can mean or lead to having an illness. In the other chart option, for the active, It shows me that if I'm exercising I can have my HR that high but would be considered healthy and actually good for me.