Design and Evaluation of Health Visualizations for Older Adults

Thai D. Le

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

George Demiris, Chair

Hilaire Thompson

David McDonald

Program Authorized to Offer Degree:

School of Medicine

Biomedical and Health Informatics

University of Washington

Abstract

Design and Evaluation of Health Visualizations for Older Adults

Thai D. Le

Chair of the Supervisory Committee:
George Demiris, PhD, FACMI, Alumni Endowed Professor
Department of Biomedical Informatics and Medical Education

Older adults, those 60 years and above, represent the quickest growing demographic group in the United States. Additionally, health related changes associated with aging make this population one of the primary consumers of health care resources. Innovative informatics solutions can assist older adults in maintaining health and independence. One such approach is through smart home technologies, residences with technology embedded within the infrastructure of the home to unobtrusively monitor and assist older adults with activities of daily living. To present data collected from home based monitoring including smart homes, and other informatics tools such as telehealth in a meaningful manner, I describe work in the development of health visualizations for older adults. Though a body of work has shown that older adults find utility in technology to support their health and wellness, there has been limited research examining how this would translate to data visualizations. I start by looking at potential differences in how older adults process graphical information compared to the general population through a set psychophysics experiments. I then apply a user-centered

design approach to iterate on health visualizations from early mockups to fully interactive prototypes. I describe different approaches for evaluating visualizations with older adults, and report on the findings of the evaluations. Finally, I thematically analyze the evaluation sessions to extract themes associated with how older adults utilize health visualizations. Based on these themes, I provide a set of recommendations to assist other researchers and designers in this domain as they develop older adult focused visualizations. This work represents an end-to-end process from initially identifying older adult visualization needs through to the design and evaluation of interactive visualizations. The three primary contributions in this research are: 1) comparing graphical perceptual needs of older adults with that of the general population, 2) comparing different approaches towards evaluating health visualizations, and 3) providing a set of guidelines to inform the design of health visualizations for older adults.

Abstract

Older adults, those 60 years and above, represent the quickest growing demographic group in the United States. Additionally, health related changes associated with aging make this population one of the primary consumers of health care resources. Innovative informatics solutions can assist older adults in maintaining health and independence. One such approach is through smart home technologies, residences with technology embedded within the infrastructure of the home to unobtrusively monitor and assist older adults with activities of daily living. To present data collected from home based monitoring including smart homes, and other informatics tools such as telehealth in a meaningful manner, I describe work in the development of health visualizations for older adults. Though a body of work has shown that older adults find utility in technology to support their health and wellness, there has been limited research examining how this would translate to data visualizations. I start by looking at potential differences in how older adults process graphical information compared to the general population through a set psychophysics experiments. I then apply a user-centered design approach to iterate on health visualizations from early mockups to fully interactive prototypes. I describe different approaches for evaluating visualizations with older adults, and report on the findings of the evaluations. Finally, I thematically analyze the evaluation sessions to extract themes associated with how older adults utilize health visualizations. Based on these themes, I provide a set of recommendations to assist other researchers and designers in this domain as they develop older adult focused visualizations. This work represents an end-to-end process from initially identifying older adult visualization needs through to the design and evaluation of interactive visualizations. The three primary contributions in this research are: 1) comparing graphical perceptual needs of older adults with that of the general population, 2) comparing different approaches towards evaluating health visualizations, and 3) providing a set of guidelines to inform the design of health visualizations for older adults.

Introduction

Aging has an impact on many components of health, including physiological, social, and cognitive changes. Innovative informatics solutions can be applied to effectively support successful aging for older adults. One informatics approach involves integrating data from health monitoring technologies to provide near real-time collection of health information. However, though this approach makes it possible to collect a diverse range of data, this information can be challenging to synthesize. Data visualizations are powerful resources that leverage the human visual system to abstract data into representations that afford detection of trends and patterns. Though a body of work exists describing different approaches towards visualizing data collected from monitoring technologies within the home of older adults [1–3]; these are often designed for researchers or clinicians. The literature is sparse on the design of health visualizations for older adults as consumers of the information. Representing data to an older adult stakeholder group through appropriately designed visualizations provides an important resource that could empower them to engage with family members and clinicians in maintaining their health and wellness. In this research, I provide an end-to-end description of the design and evaluation of health visualizations for the older adults. This work contributes to the literature by first identifying graphical needs of older adults and comparing them to those of the general population. The results from the graphical studies informed the design of two high fidelity interactive visualizations. I then provide a comparison of different methods for evaluating health visualizations. Based off of the evaluations, I extracted key themes associated with how older adults would utilize health visualizations. I provide a set guidelines for designing health visualizations that has been missing from the design literature for older adults.

As part of the design process, I first examined the graphical perceptual needs of older adults. This is presented through a paper published in *Perception* where I describe a set of experiments on basic graphical elements and compare performance between two population groups: older adults and a general population[4]. Though graphical perceptual studies have been conducted previously, none have looked at how the findings translate to an older adult population. These findings are important given that aging is associated with changes in visual acuity and cognitive tasks such as processing speed, impacting the ability to perceive graphical displays. The results of the study also provide a set of building blocks to better understand how graphical elements should be composed to effectively represent information to older adults.

I use the results of the study to iterate on designs of health visualizations with an older adult consumer focus before developing two high-fidelity prototypes. In the second paper, I describe a study comparing different visualization evaluation methods. Evaluation of information visualizations is an essential component of the design process, allowing researchers to quantify differences across visualizations and reiterate as part of the design cycle. However empirical approaches to evaluation vary. I compare two evaluation methods: benchmark evaluations and insight evaluations. In a benchmark evaluation participants are given a set of well-defined tasks and asked to complete the tasks through exploration of the visualization. Metrics such as task completion and task accuracy are used to compare visualizations. In an insight evaluation, participants are asked to openly explore the visualization while verbalize insights generated during the exploration process. This is analyzed later to identify the number and types of insights generated and compared between visualizations. I present a summary of findings from both evaluation approaches and identify the tradeoffs between the methodologies.

In the third paper, I synthesize the results from the evaluation sessions to present a set of themes associated with how older adults utilize health visualizations. Though older adult specific guidelines exist, they are often within the context of websites. From these themes, I provide a set of recommendations for other researchers and designers developing health visualizations for older adults. These recommendations are independent of the specific visualizations that I evaluated and are generalizable for others to consider when developing health visualizations. This represents a separate set of information needs compared to health visualizations.

From the three papers, I present an end-to-end flow for the design and evaluation of health visualizations for older adults. As technology is increasingly used as a tool to support aging, we have more ways to capture data and more data sets than before. However, older adults should not be forgotten as primary stakeholders amidst this abundance of data. Improving care involves including older adults within the decision-making process. To do this, we need effective visualization approaches for older adults that facilitate effective processing of health data and promote understanding of complex wellness information.

Paper 1: Elementary Graphical Perception for Older Adults: A Comparison with the General Population

Authors: Thai Le (BS), Cecilia Aragon (PhD), Hilaire J. Thompson (PhD, RN), George Demiris (PhD)

Abstract

We identified the graphical perceptual information needs of older adults ($\geq$ 60 years of age) through a set of psychophysical experiments on bar, stacked, and pie charts. The results are compared with those of a general population ($\leq$ 60 years of age). We conducted the experiments as online remote studies with 202 total participants across two experimental types: 1) comparison judgments of graphs (50 older adults, 50 general population) and 2) proportion judgments of graphs (52 older adults, 50 general population). Older adults took longer than the general population to complete tasks across both comparison (4.09 seconds) and proportion judgments (3.66 seconds). However, this translated to an approximately equal level of perceptual accuracy. Bar charts were the most effective graphical display when considering both speed and accuracy. Older adults were more accurate using pie charts compared to the general population in the comparison task.

Key Terms: Aged, Visual Perception, Psychophysics, Data Display

1. Introduction

Graphical visualizations, when designed appropriately, are valuable resources that support data analysis and dissemination of information. Graphs represent visual abstractions of underlying data that can amplify cognition, detect trends, and promote insight. One area of research in graphical visualization focuses on empirical experiments to compare the efficiency of different elementary graphical elements. Understanding how different graphical displays impact accuracy and speed of perception can inform the appropriate design of data visualizations.

Macdonald-Ross (1977) and Lewandowksy and Spence (1989) provide extensive reviews of empirical research on the perception of graphical visualizations. However, participants within these studies were often recruited from a college or high school population, reflecting a narrow subset of the general audience. Though Cleveland and McGill showed that level of technical training did not impact the accuracy of graphical perception (Cleveland & McGill, 1986), there has been limited research on how the results generated by psychophysical experiments translate to an older adult population group. We addressed this gap of knowledge by providing a comparison of graphical perceptual tasks between older adults (at least 60 years old) and the general population (less than 60 years old). The tasks were informed by Simkin and Hastie's original studies comparing bar, stacked bar, and pie charts across two experiments: comparison and proportion judgments (Simkin & Hastie, 1987).

2. Background

2.1 The Older Adult Information Consumer

Older adults are a growing demographic group with a broad spectrum of information needs. However, aging related changes impact perceptual, cognitive, and psychomotor skills. A human-centered design approach is required to present information to older adults in an effective manner. Charness, Demiris, and Krupinski provide a framework for understanding how user capabilities interact with technological devices within a task (Charness, Demiris, & Krupinski, 2011). This results in a degree of fit that is characterized by satisfaction and comfort when capability meets demand, as opposed to frustration and discomfort when demand is greater than capability (Charness et al., 2011). Though this balance between capability and demand is described within the context of health technology, the framework is a valuable resource for understanding the visualization needs of older adults.

There is great diversity, even within the older adult population. Differences in educational background, experience with technology, and attitudes shape a broad user profile. Data visualization, by its nature, involves visual perception. Physical changes such as reduced visual acuity, color discrimination, and contrast discrimination, along with overall vision loss impacts the design of data visualizations (Brabyn, Schneck, Haegerstrom-Portnoy, & Lott, 2001). Cognitive changes such as decline in inductive reasoning and short term memory are also design considerations (Singh-Manoux et al., 2012). Data visualizations are powerful tools to help reduce the cognitive load of information. However, aging-associated changes inform the need to design visualizations that reduce clutter, have clear contrast, use discriminatory colors, and limit complexity.

One of the major challenges with designing for the older adult populations is a lack of awareness of user needs along with unclear guidelines on how to address these needs (Czaja & Lee, 2006). Czaja and Lee provide an overview of cognitive abilities that may be impacted with increasing age and their implications for performance on technology-based tasks. These include perceptual speed, visualization ability, working memory, psychomotor speed, spatial memory and reasoning, and visuo-spatial abilities (Czaja & Lee, 2006). Fisk et al. provide guidelines for the design of interfaces for older adults based on age-related changes in cognition (Fisk, Rogers, Charness, Czaja, & Sharit, 2012). The recommendations are a valuable reference for designers and include heuristics such as presenting information in consistent locations, adhering to principles of perceptual organization, highlighting important information, avoiding technical jargon, and minimizing demands on working memory (Fisk et al., 2012). Further heuristics also exist for the design of older adult focused websites (D. E. Chisnell, Redish, & Lee, 2006; D. Chisnell & Redish, 2004). However, these heuristics are only broadly applicable within a data visualization context. We address this limitation by focusing on older adult perceptual needs for data visualization at the fundamental level of graphical elements. These are building blocks to inform the design of graphical displays that convey information to older adults.

2.2 Value of Data Visualizations

Data visualization and information visualization are two terms frequently used to describe the process of abstracting data into visual representations that support analysis. For the purposes of this research, we adopt the convention used by Few in which data visualization is an umbrella term encompassing information visualization (Few, 2009). There are two key aspects of data visualizations: they are visual representations that abstract data and amplify cognition. Information visualizations provide an added component of computer-supported interactions for exploration of the data (Card, Mackinlay, & Shneiderman, 1999). Bertin, in early work during the 1960's, defined three key goals of data visualization: provide a recording of information, communicate the information across an audience, and allow for processing of information (Bertin & Berg, 2011).

It can often be challenging to effectively analyze raw data despite the wealth of information available underneath. One reason for this may be the limitation of working memory, the short-term system that allows us to store and manipulate information transiently. Miller found that this capacity spans seven chunks of information (Miller, 1956), though further research has shown that this is flexible and impacted by type of information (Cowan, 2005). Research has shown that working memory is negatively correlated with age (Dobbs & Rule, 1989; Salthouse & Babcock, 1991), a potential factor to consider when evaluating the effectiveness of visualizations for older adults. Baddeley proposed of the visuo-spatial sketchpad component of working memory. This system constructs and manipulates visual information, though it is limited to, at most, four chunks of information (Alvarez & Cavanagh, 2004; Baddeley, 1992; Luck & Vogel, 1997). Despite this limitation of visual working memory, data visualizations are effective representations of information because they are able to chunk large amounts of data into meaningful visual components and augment our working memory with an external storage (Ware, 2005). This allows us to comprehend large amounts of data, identify properties and trends not initially inherent, check for unusual outliers and generate hypotheses (Ware, 2004). Scaife and Rogers point out that despite the breadth of work in graphical perception, there is still limited and fragmented understanding of how graphical representations work (Scaife & Rogers, 1996).

2.3 The Basic Vocabulary of Visualization

Bertin frames data visualization within the context of a language system for the eye, describing both the elementary components of visualizations and providing rules for how they are effectively composed (Bertin & Berg, 2011). The building blocks of this language include shape, size, texture, intensity/value, color/hue, orientation, and position (Bertin & Berg, 2011). However, he also notes that there is a close association between understanding different types of information and the appropriate visual element used to represent them, in essence, a relationship between form and function (Bertin & Berg, 2011). Ware expands on the work of Bertin and highlights the value of pre-attentive attributes within visual perception (Ware, 2004). Pre-attentive attributes are those that are quickly and unconsciously perceived, allowing certain components to stand out from a visual representation. These attributes are grouped into form (orientation, shape, line length, line width, size, curvature, added marks, enclosure), color (hue, intensity), 2-D position, and motion (Ware, 2004). Within the context of data visualization, Ware's work provides a valuable foundation for understanding how data should be encoded into visual elements that are quickly perceived. Mackinlay further provides a ranking of graphical components based on accuracy of perception, classified by quantitative, ordinal, and nominal data types (Mackinlay, 1986). These guidelines are grounded by human perception; choosing the appropriate visual components to encode data allows us to quickly and effectively extract out meaning from a visualization.

2.4 Psychophysical Research on Graphical Displays

Psychophysics focuses on the relationship between a perceived stimuli and its true magnitude. Within the data visualization literature, psychophysical experiments provide valuable information on the effectiveness (both speed and accuracy) of visual encodings for differing graphical display. Cleveland and McGill presented a set of experiments to assess accuracy of perceptual judgments on graphical displays (Cleveland & McGill, 1986). For each graphical display, participants were asked to assess the magnitude of a stimuli compared to its standard reference. The experiments found that the graphical judgments from most accurate to least accurate were: position along a common scale > position along identical but non-aligned scale > length > angle > slope > area (Cleveland & McGill, 1986). Spence focused on the use of graphical displays for comparisons as opposed to absolute judgments of magnitude in his psychophysical experiments (Spence, 1990). Out of eight display types (horizontal and vertical lines, bars, pie and disk slices, cylinders, boxes, and table entries), Spence found that the bar, box, and cylinder displays allowed for both quick and accurate comparisons of numerical quantities (Spence, 1990). Simkin and Hastie conducted a set of experiments that looked at the impact of graphical displays on accuracy and speed for two judgment tasks: comparing between absolute magnitudes and assessing proportion of a segment to the whole (Simkin & Hastie, 1987). The chart types included a simple bar, stacked bar, and pie chart. The authors found that, for comparison judgments, the simple bar chart was most accurate followed by stacked bar and pie charts. For proportion judgments, the pie chart and bar chart were equally as effective for accuracy, more so than the stacked bar chart (Simkin & Hastie, 1987).

More modern research in the field of psychophysics includes work by Heer and Bostock who demonstrate the viability of using crowd-sourced online techniques for experiments in graphical perception (Heer & Bostock, 2010). Using Amazon's Mechanical Turk, the authors successfully replicated Cleveland and McGill's findings on proportional judgments, though with slightly higher variability due to a wider population sample (Heer & Bostock, 2010). The benefits of a web-based, crowd-sourced approach include greater economy of price, reduced data collection time, and increased scalability (Heer & Bostock, 2010). Heer and Bostock further demonstrate that it is possible to conduct web-based, crowd-sourced graphical perception experiments to generate new insights, in particular related to the judgment of rectangular area based on aspect ratios (Heer & Bostock, 2010).

Hullman, Adar, and Shah summarize an alternative model of information visualization as a balance between cognitive efficiency and desirable difficulties that induce deep learning (Hullman, Adar, & Shah, 2011). Cognitive efficiency refers to the traditional model of information visualization as a set of components designed to augment external memory and information storage while highlighting trends within the data (Hullman et al., 2011). Though cognitive efficiency is important towards designing data visualizations, the authors propose that introducing desirable difficulties can promote active processing by the user (Hullman et al., 2011). This can result in greater long-term retention and comprehension of the information while increasing engagement with the data visualization (Mayer, Hegarty, Mayer, & Campbell, 2005).

A limitation of the existing psychophysical research lies in the sampling of participants from a primarily university or general population. However, given that older adults differ from the general population due to changes in visual acuity and perception, it is unclear if the psychophysical research findings would translate appropriately. To truly understand the design needs of older adults for data visualization, we address the gap in psychophysical research by modifying Simkin and Hastie's study to provide a comparison of graphical perception between an older adult and general population group. We select Simkin and Hastie's study as a reference due to its flexibility at addressing both common use cases of data visualizations: comparison of graphs, and proportional judgments within a graph.

3. Methods

3.1 Study Design

We developed two experimental arms for the study. In the first experimental group, participants were given a pair of graphs within the same display type and asked to identify which graph out of the pair was smaller and to provide an estimate of the percentage with which the smaller one is of the larger one (comparison). In the second experimental group, participants were shown only a single graphical display type and asked to judge the percentage that the division represents of the whole bar or pie (proportion). The three types of visual displays included a simple bar chart, stacked bar chart, and pie chart [Figure 1]. There were 45 trials total for each experiment. These were created from 15 randomly generated numbers triplicated across the graph types. We presented participants with a random order of the 45 trials. This design was a replication of Simkin and Hastie's study, using the same visual display types. We reduced the number of total trials from 90 to 45 due to concerns of experimental fatigue for older adults. Simkin and Hastie's study was conducted in person on a controlled computer interface with gratuity. We translated the experiments into an online survey to allow for broader coverage and left the survey open for uncompensated voluntary participation.

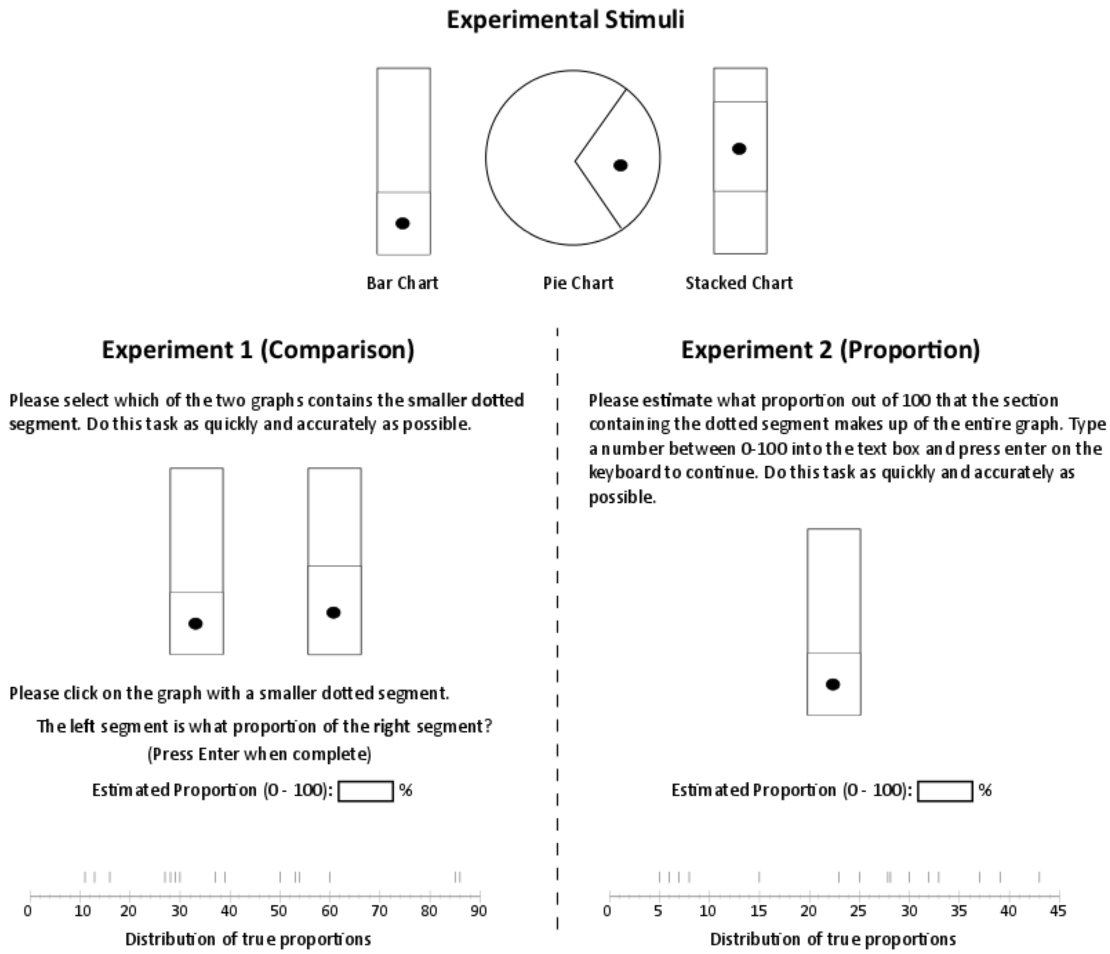

Figure 1: The three experimental stimuli types include a bar chart (position judgments), pie chart (angle and area judgments), and stacked chart (length judgments). In the first experiment type, participants make an estimate comparing how much the smaller dotted segment makes up of the larger one. In the second experiment type, participants estimate what proportion of the larger chart is composed of the dotted segment. The true proportion values for each of the 15 experimental tasks are shown below experimental descriptions.

3.2 Display Stimuli

We conducted the experiments through an unmoderated online survey. We controlled for different monitor sizes by fixing the content panel in which tasks were displayed at a 12’’x12’’ dimension.

We generated both experimental stimuli datasets following the design of Simkin and Hastie (Simkin & Hastie, 1987). In the comparison experimental group, we presented participants with pairs of graphical displays within the same graphical display type (simple bar, stacked bar, or pie chart). All bars and pie charts had fixed dimensions across the trials such that they represented equivalent area. A dot was placed on a segment of each chart to indicate the desired regions for comparison. The length of the bar segment or sector of the pie chart was determined by randomly selecting two integers between 3 and 47 inclusive. A pair of integers (X, Y) was determined valid if the ratio of the smaller to larger integer was between 0.05 and 0.95 inclusive. In addition, a randomly generated offset value was created for the stacked bar and pie chart, representing the distance from baseline (the bottom of the stacked bar, or at the twelve o’clock position of the pie chart) that each segment was drawn. We generated 15 pairs of integers along with their offsets. We used these values to generate the paired simple bar, stacked bar, and pie charts. As a result there were 45 total trials, with every 15 pairs of integers within each graphical display type [Figure 1].

For the proportion experimental group, we presented participants with a single graph per trial drawn from one of the three graphical families. The simple bar, stacked, and pie charts were generated using the same technique as the comparison experimental group. In this case, pairs of integers were not needed, only 15 randomly generated integers with their offsets. These were used to create the 45 trials where each of the 15 values is repeated across the three graphical display types.

3.3 Participants

We recruited across two population groups: older adults (at least 60 years old) and general population (less than 60 years old). Participants were recruited through electronic mailing lists of nursing and medicine-affiliated faculty, recruitment fliers on campus at the University of Washington, older adult retirement communities and senior centers, online forums, and snowball sampling. We restricted the survey to English speaking participants. We stopped recruitment when at least 50 participants had completed the full survey for each population group x experimental type combination. We chose a sample size of 50 participants for each experiment type, as that would allow us to identify a Cohen's effect size of 0.29 at $\alpha = .05$ significance with 90% power from a one-way ANOVA (Cohen, 1988). We conducted surveys between October 2013 and April 2014. The university's Institutional Review Board approved all study procedures.

3.4 Data Analysis

We assessed differences in demographics between the two population groups using ANOVA or Chi-Squared as appropriate. To assess accuracy of judgments, we calculated the absolute difference between a participant's judgment and the true value for each experimental task. We made an estimate of the distribution of the reaction times and absolute error values using the midmean, a robust estimate of location. We calculated the mean of the midmeans of the 15 trials in each display type, experimental group combination (Simkin & Hastie, 1987). We tested for differences in means of midmeans across display type using ANOVA. Comparisons between the means of midmeans of display types were made using the Newman-Keuls test as part of post-hoc analysis. To quantify the effect of age group on accuracy and response time adjusted by display type, we applied an OLS linear model to fit the mean of midmeans for each experimental stimulus as a function of display type, age group, and true value of the stimulus.

Because each experimental stimulus is presented once across each display type, we also compared performance (absolute error of estimate and completion time) using the

Wilcoxon Signed Rank Test for matched pairs. We applied the Wilcoxon matched pair tests as an extension of the methodology by Simkin & Hastie to remove potential variation at a participant level. We used the results of the tests to determine a ranking of display types by accuracy and completion time.

4. Results

4.1 Population Demographics

A total of 202 participants completed the full surveys, distributed across the experimental groups as: Experiment 1/Older Adults (N = 50), Experiment 1/General Population (N = 50), Experiment 2/Older Adults (N = 52), Experiment 2/General Population (N = 50). A comparison of demographics between the two population groups is found in Table 1. We found race differences between the two population groups (X^2 = 27.1, df = 3), however gender (X^2 = 2.45, df = 1) and education level (X^2 = 4.31, df = 4) were not significantly different between the two groups.

	Older Adults (N = 102)	General Population (N = 100)	F-Value/X^2 (DF)
Age (Mean, SD)***	69.9 (7.2)	35.2 (14.6)	462 (1)
Gender (Female)	74.5%	63.4%	2.45 (1)
Race***			27.1 (3)
Asian or Asian American	1.0%	25.0%	
Black or African American	1.0%	1.0%	
Hispanic or Latino	2.0%	4.0%	
Non-Hispanic White	96.1%	71.0%	
Education Level			4.31 (4)
High School/GED	6.9%	5.0%	
Some College	14.7%	24.0%	
Associate's Degree	6.9%	6.0%	
Bachelor's Degree	28.4%	33.0%	

Graduate Degree	43.1%	33.0%

Table 1: Demographic comparisons between population groups (* p<.001, ** p<.01, * p<.05).**

4.2 Experiment 1 (Comparison Task)

4.2.1 Older adults

We did not find a difference in completion time when comparing mean of midmeans across graphical display type (p=0.24, F-value=1.5, df=2) for older adults using ANOVA. However matched pair comparisons using the Wilcoxon Signed Rank Test identified the following order in completion time (quickest to slowest) as: bar > stacked > pie (effect sizes: 0.147 – 0.396). We found differences in mean of midmeans absolute error by display type (p<.001, F-value=12, df=2). Further pair-wise comparisons using the Neuman-Keuls test found that the bar assessments (mean=4.11, SE=0.145) were more accurate than the stacked bar (mean=5.30, SE=0.201) chart (p<.001) and similarly the pie chart assessments (mean=4.533, SE=0.175) were more accurate than stacked bar (p<.05), though of lesser magnitude. The matched pair comparisons identified a similar order of most to least accurate display type: bar > pie > stacked (effect sizes: 0.108 – 0.236) [Figure 2].

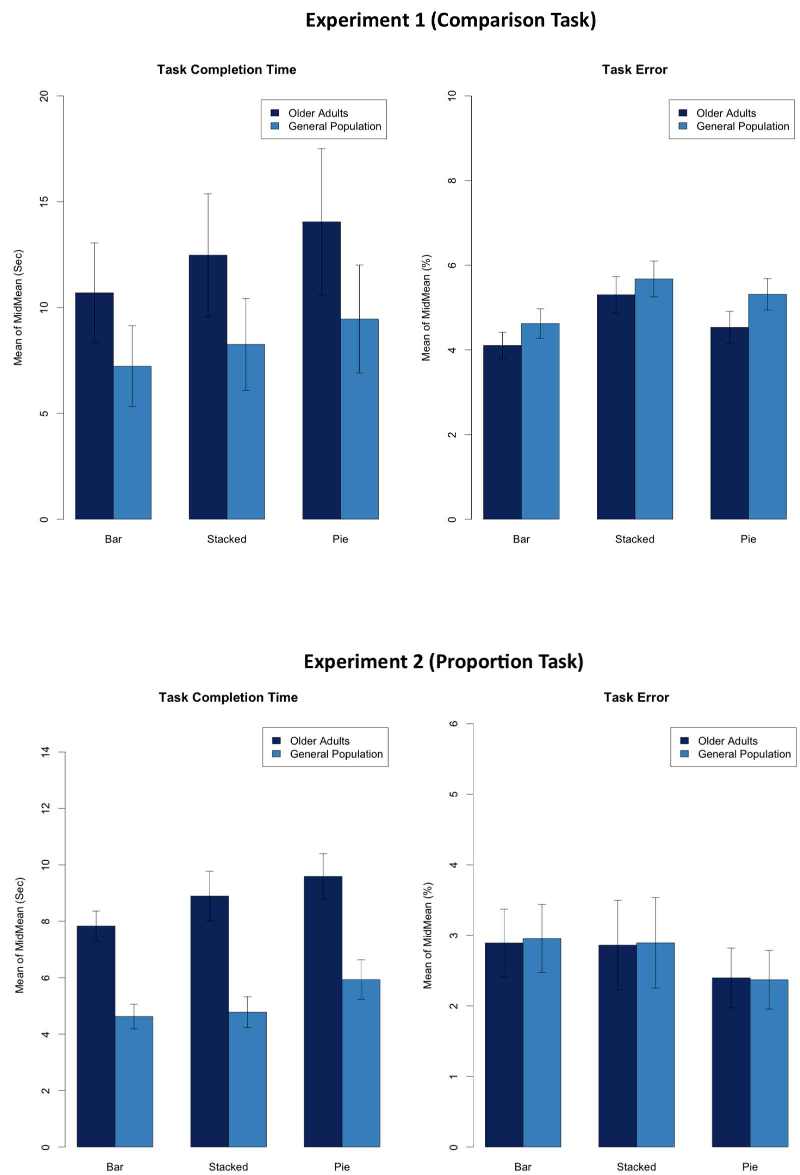

Figure 2: Mean of midmeans across all stimuli by display type for both experimental tasks.

4.2.2. General population

For the general population, we also did not find any differences in mean of midmeans completion time by graphical display type (p=.32, F-value=1.16, df=2) using ANOVA.

However, matched pair comparisons using the Wilcoxon Signed Rank Test identified the following order in completion time (quickest to slowest) as: bar > stacked > pie (effect sizes: 0.149 – 0.340). We found that the mean of midmeans absolute errors differed ($p<.001$, F-value=9.05, df=2) with the bar plot assessments (mean=4.62, SE=0.163) being more accurate than both the stacked bar (mean=5.68, SE=0.196), ($p<.001$) and pie chart assessments (mean=5.31, SE=0.173), ($p<.01$). The matched pair comparisons found a similar ranking from most accurate to least accurate display type: bar > stacked, pie (effect sizes: 0.095 – 0.155) [Figure 2].

4.2.3 Comparison

An OLS linear model of response time as a function of display type, age group, and true stimulus value found that older adults took 4.09 (SE=0.977) seconds longer than the general population ($p<.001$). In addition, the pie chart took on average 2.79 (SE=1.20) seconds longer than the bar chart comparisons ($p<.05$). With accuracy as a response variable, the linear model identified: older adults made estimates that were 0.56 (SE=0.142) percentage points more accurate than the general population ($p<.001$), pie charts were 0.56 (SE=0.174) percentage points less accurate than bar charts ($p<.01$), and stacked charts were 1.13 (SE=0.174) percentage points less accurate than bar charts ($p<.001$) [Table 2].

	Experiment 1 (Comparison)		Experiment 2 (Proportion)	
	Response Time	Accuracy	Response Time	Accuracy
Bar Chart (reference)				
Stacked Chart	1.40	***1.13	*0.61	-0.05
Pie Chart	*2.79	**0.56	***1.53	**-0.54
Older Adults (reference)				
General Population	***-4.09	***0.56	***-3.66	0.02

True Stimulus Value	-0.03	0.00	*** 0.03	***0.04

Table 2: Linear models of response time and accuracy as functions of display type, age group, and true stimulus value. Point estimates are provided with significance values (* p<.001, ** p<.01, * p<.05).**

4.3 Experiment 2 (Proportion)

4.3.1 Older adults

For the older adult population group, we found a difference (p<.01, F-value=6.35, df=2) in mean of midmeans completion time across display types based on the ANOVA test. Further pair-wise comparisons identified a difference between the bar and pie charts, with bar chart comparisons being made 1.76 (SE=0.003) seconds quicker (p<.01) compared to pie charts. Bar chart assessments were also made 1.06 (SE=0.038) seconds quicker (p<.05) than stacked charts. The matched pair comparisons using the Wilcoxon identified bar chart assessments as the quickest followed by stacked and pie charts (effect sizes: 0.056 – 0.193)). We did not find a difference in accuracy across display type based on the ANOVA (p=0.28, F-value=1.31, df=2). However, matched pair Wilcoxon tests found that the pie chart assessments were more accurate than the stacked chart and bar chart (effect sizes: 0.066 – 0.099). There was not enough evidence to indicate a difference in distribution of accuracy for the bar chart compared to the stacked chart [Figure 2].

4.3.2 General population

For the general population, we also found differences in mean of midmeans completion time by graphical display type (p<.01, F-value=7.14, df=2) using ANOVA. Pair-wise comparisons found that the bar chart assessments were 1.31 (SE=0.004) seconds quicker than the pie chart (p<.01) while the stacked chart assessments were 1.15 (SE=0.004) seconds quicker than the pie (p<.01). The Wilcoxon matched pair tests identified the following order from quickest to slowest as: bar > stacked > pie (effect sizes: 0.065 – 0.268). We did not find a difference in accuracy across display types using the ANOVA ($p=0.19$, F-value=1.74, df=2). However, when we conducted matched pair comparisons, we found pie charts were more accurate than both bar and stacked charts (effect sizes: 0.115 – 0.127) [Figure 2].

4.3.4 Comparison

An OLS linear model of response time as a function of display type, age group, and true stimulus value found that older adults took 3.66 (SE=0.239) seconds longer than the general population ($p<.001$). In addition, the pie chart assessments took 1.53 (SE=0.293) seconds longer than the bar chart assessments ($p<.001$) to make comparisons while the stacked chart assessments took 0.61 (SE=0.293) seconds longer than bar chart assessments ($p<.05$). With accuracy as a response variable, the linear model did not identify differences between age groups though we found that pie charts were slightly more accurate than bar chart comparisons with estimates on average 0.54 (SE=0.198) percentage points closer to the true value ($p<.01$) [Table 2].

5. Discussion

5.1 Performance Differences

The three graph types in this study highlight different perceptual elements of judgment. The bar graph uses position to perform comparisons. The stacked chart uses differences in length, and the pie chart uses a mixture of area and angle to assess values. We found that older adults were consistent in taking longer amounts of time to make assessments across both comparison and proportion tasks compared to the general population. However, this is offset by matching or even slightly better accuracy on the tasks. An explanation for this could be that aging related differences in visual perception and cognitive processing are compensated for with longer response times, though the visualizations themselves are able to display the information at the same level of precision.

By graph type, we found that the bar chart facilitated quickest response time followed by stacked and pie charts. These findings were consistent across age group and experimental trials. The results indicate that perceptual judgments based on position for the bar chart are quickly perceived compared to length or angular based assessments.

For experiment 1, comparing two graphs within the same display type, we found that the bar displays were most accurate for both population groups. For older adults this was followed by pie charts. In modeling the process of making these comparison judgments, Simkin and Hastie proposed that participants superimpose the smaller segment onto the larger one in the stacked and pie charts (Simkin & Hastie, 1987). An anchor is identified via the larger segment (often of 0%, 50%, or 100%), and then participants scan the remaining difference of the superimposed image relative to the anchor to make an estimate of the proportion (Simkin & Hastie, 1987). In a bar chart, rather than superimposing the smaller segment onto the larger one, it is possible to project horizontally the ray from the smaller segment to the larger segment before performing the anchoring and scanning steps (Simkin & Hastie, 1987). This is a less cognitively challenging process and provides an explanation for the increased accuracy of bar charts. Older adults were more accurate in making comparisons using the pie chart compared to the stacked chart, indicating that angular judgments were more effective for older adults compared to realigning lengths within the stacked display.

For experiment 2, we found age related differences in completion time with older adults taking longer than the general population; however accuracy rates were similar across age groups. By display type, we found that there was a slight improvement in accuracy for the pie chart compared to either bar or stacked charts. Simkin and Hastie proposed that the proportion judgments are made through an anchoring and scanning process (Simkin & Hastie, 1987). Participants first anchor from the overall container to references of 0%, 25%, 50%, 75% or 100% and then scan the difference of the anchor from the segment to make an assessment of proportions (Simkin & Hastie, 1987). The improved accuracy of pie charts is primarily due to the task, a part-to-whole comparison. Participants did not need to cognitively isolate segments of two pie charts and superimpose them to make a comparison as in experiment 1. In addition, the anchoring values within a pie chart may be of greater familiarity for participants, corresponding to distinct 90 degree angles or straight lines. However, the accuracy of pie charts for proportion estimates comes at the cost of processing speed, as indicated by the slow response times. Processing angles is more time consuming and cognitively challenging than position or length (Cleveland & McGill, 1984; Simkin & Hastie, 1987), but once processed, the distinct reference angles of 0, 90, and 180 degrees may help participants anchor towards a more accurate estimate of proportions.

5.2 Comparison with Simkin and Hastie

We modeled the experimental procedures based on Simkin and Hastie's original study (Simkin & Hastie, 1987). However, a primary difference was that we implemented the experiment as an online survey as opposed to in-person sessions in front of a computer screen. Simkin and Hastie also did not report measures of variance for their quantitative findings. Given these two key limitations, we were unable to provide a quantitative statistical comparison of our findings with those of the original paper. However, we are able to provide a qualitative discussion of the differences in findings.

For experiment 1 (the comparison assessment), Simkin and Hastie found that the judgments made with the bar charts were quicker than those of either the stacked or pie chart. This difference in response time was also true for experiment 2 (the proportion assessment). The results from our study showed a similar relationship, though a further difference in response time was identified with stacked charts performing better than pie charts.

For accuracy, Simkin and Hastie found that bar charts were more accurate than stacked and pie charts in experiment 1. These results were validated for the general population group within our study, though older adults had a further distinction with pie charts having higher accuracy then stacked charts. In experiment 2, Simkin and Hastie found that bar and pie charts performed equally well in accuracy, with stacked charts being poorer than both. Our findings were similar in that pie charts were more accurate in the proportion task compared to stacked charts. The accuracy of bar charts was indeterminate compared to the other two display types.

5.3 Limitations

We recognize that limitations exist within the study. The online distribution of the survey biases potential respondents, especially for older adult populations who may not have computer access or are not as comfortable with performing the online tasks. Further expanding on this research with paper printouts of the displays and in person sessions would allow for the inclusion of a more representative sample. We also were unable to control for the setup of the experimental environment (for example, how far apart participants sat from the monitor, whether a laptop or personal computer was used, if a mouse was used). This contributes to variability across participants. However, we accounted for some of this variability by conducting matched pair comparisons, focusing on differences between display types by experimental stimuli as opposed to aggregate population group comparisons. Participants may also differ in their appropriation of time for answering questions; some tending to spend longer for more accurate responses. To control for this difference, an extension of this work will limit the time that the graphs are displayed before asking participants to make judgments.

5.4 Conclusion

Though graphical perception has been studied frequently within the field of psychophysics, it is limited to a student or general population. We extend the work on graphical perception to an older adult population. Given the aging associated changes in visual acuity and cognitive processing speed, it is important to understand the effectiveness of fundamental graphical elements on perception for older adults. We found that, overall, older adults took longer to process the graphical displays compared to the general population, though this added delay is reflected through a similar level of accuracy. The bar chart remains the more effective graphical display for making comparisons in terms of both response time and accuracy. However, for older adults, we found the pie chart to be a slight improvement over stacked chart in accuracy for proportion tasks. We also demonstrated that, depending on the functional goal of the display (to make comparisons or proportion judgments), participants differed in performance across bar, stacked, and pie charts.

Funding: This work was supported by the National Library of Medicine (Training Grant T15LM007442).

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. doi:10.1126/science.1736359

Bertin, J., & Berg, W. J. (2011). *Semiology of graphics: diagrams, networks, maps*. Redlands, Calif.: ESRI Press : Distributed by Ingram Publisher Services.

Brabyn, J., Schneck, M., Haegerstrom-Portnoy, G., & Lott, L. (2001). The Smith-Kettlewell Institute (SKI) longitudinal study of vision function and its impact among the elderly: an overview. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *78*(5), 264–269.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. San Francisco, Calif.: Morgan Kaufmann Publishers.

Charness, N., Demiris, G., & Krupinski, E. (2011). *Designing Telehealth for an Aging Population: A Human Factors Perspective*. CRC Press.

Chisnell, D. E., Redish, J. C. (Ginny), & Lee, A. (2006). New Heuristics for Understanding Older Adults as Web Users. *Technical Communication*, *53*(1), 39–59.

Chisnell, D., & Redish, J. (2004). Designing Web Sites for Older Adults: Expert Review of Usability for Older Adults at 50 Web Sites. AARP.

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531–554. doi:10.1080/01621459.1984.10478080

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man-Machine Studies*, *25*(5), 491–500.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition.). Hillsdale, N.J: Routledge.

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Czaja, S. J., & Lee, C. C. (2006). The impact of aging on access to technology. *Universal Access in the Information Society*, *5*(4), 341–349. doi:10.1007/s10209-006-0060-x

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, *4*(4), 500–503. doi:10.1037/0882-7974.4.4.500

Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis*. Oakland, Calif.: Analytics Press.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2012). *Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition*. CRC Press.

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203–212). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1753357

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting infovis with visual difficulties. *Visualization and Computer Graphics, IEEE Transactions on*, *17*(12), 2213–2222.

Lewandowsky, S., & Spence, I. (1989). The Perception of Statistical Graphs. *Sociological Methods & Research*, *18*(2-3), 200–242. doi:10.1177/0049124189018002002

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi:10.1038/36846

MacDonald-Ross, M. (1977). How Numbers Are Shown: A Review of Research on the Presentation of Quantitative Data in Texts. *AV Communication Review*, *25*(4), 359–409.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, *5*(2), 110–141.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology. Applied*, *11*(4), 256–265. doi:10.1037/1076-898X.11.4.256

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. doi:10.1037/0012-1649.27.5.763

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185–213.

Simkin, D., & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, *82*(398), 454–465. doi:10.2307/2289447

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., … Dugravot, A. (2012). Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. *BMJ*, *344*(jan04 4), d7622–d7622. doi:10.1136/bmj.d7622

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 683.

Ware, C. (2004). *Information visualization: perception for design*. San Francisco, CA: Morgan Kaufman.

Ware, C. (2005). Visual Queries: The Foundation of Visual Thinking. In S.-O. Tergan & T. Keller (Eds.), *Knowledge and Information Visualization* (pp. 27–35). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11510154_2

Paper 2: A Comparison of Health Visualization Evaluation Techniques with Older Adults

Authors: Thai Le (BS), Hilaire J. Thompson (PhD, RN), George Demiris (PhD)

Abstract

We applied three different approaches towards evaluating interactive health visualizations with older adults. Health visualizations are a valuable resource for older adults to monitor trends in wellness and to engage in care with family members and health care providers. However, there has been limited work in comparing different visualization evaluation approaches with older adults. We evaluated the visualizations through a benchmark evaluation, insight-based evaluation and summative perceived usability questionnaire. In a benchmark evaluation participants are presented with defined tasks to complete. Metrics of completion time and accuracy are used to compare across visualizations. An insight evaluation asks participants to explore the visualization in an open-ended manner. The researcher goes back and codes for the generation of insights during the exploration process. We used a System Usability Scale to assess perceived usability after interacting with each visualization design. We were unable to identify statistically significant differences between visualizations using the benchmark evaluation, moderate differences with the perceived usability scale, and more granular differences through the insight evaluation. Our study identified differences amongst the three evaluation techniques while also confirming the value of insight evaluations in comparing visualizations.

Key Terms: Data and Knowledge Visualization, Evaluation/Methodology, Interactive Data Exploration and Discovery

1. Introduction

Health visualizations can be valuable resources to represent complex data collected from different sources. The value of visualizations lays in their ability to represent trends and patterns that are otherwise less apparent in raw format. For older adults, health visualizations provide a resource to monitor health and wellness over time, while also encouraging greater engagement in care. However, developing appropriate visualizations is a non-trivial task due to the limited guidelines that exist for designing and evaluating health visualizations with older adult consumers. In this manuscript we address the challenges of evaluating health visualizations with older adults by providing a comparison of three different evaluation methods across two interactive visualizations of health data.

The evaluation methods included: 1) benchmark evaluation, 2) insight evaluation, and 3) subjective usability questionnaire. In a benchmark evaluation, participants are given directed tasks to complete. Metrics of completion time and accuracy can then be used to compare across visualizations. An insight evaluation allows participants to explore the visualizations in an open-ended task while thinking aloud. Insights are identified and coded. The number of insights generated and the quality of the insights are compared across visualizations. For subjective usability questionnaire we used the System Usability Scale as a holistic measure of the participants' experiences interacting with the visualization (1). Each approach emphasizes one of three areas in visualization evaluation as described by Santos et al: effectiveness, efficacy, and satisfaction (2). We summarize the key similarities and differences across each method and provide lessons learned from applying each approach with older adults.

2. Background

2.1 Health Visualizations Older Adults

There is a body of work describing the impact of data visualizations on medical decision-making from both health care provider and consumer perspectives (3–6). Elting et al. found that both the method of display and the framing of data influenced how physicians interpreted clinical trial data (4). From the patient perspective, Feldman-Stewart et al found that a participant's perception of treatment risks and benefits was influenced by the type of graphical display (5). This is not surprising given the impact data visualizations can play in illuminating trends within the data. Older adults recognize the utility of health visualizations for personal use, often having favorable impressions of visualization mockups (7–9). However, health visualizations designed specifically for older adults are currently limited within scientific literature. Reeder et al found that older adults differed in their information needs compared with health care providers (8). Older adults emphasized a higher-level overview of data primarily to identify longitudinal changes in health and wellness. In contrast, health care providers described using visualizations at a finer level of granularity for diagnostic purposes (10). Given these different information needs, it is important to align design guidelines with the intended audience.

Though design guidelines and cognitive theories on information visualization exist, they are often understudied for use in older adults. Chisnell et al provide heuristics for designing web sites with older adult users (11,12), Czaja et al present guidelines for interface design of computer systems (13), and the National Institute of Aging provides a checklist for older adult friendly websites (14). However, these existing guidelines are focused on promoting navigation and clarity within a general web structure. Information visualizations differ in promoting insight and understanding of a dataset rather than to broadly navigate through information. Gong and Chandra describe the design of a dashboard to monitor aging in place within the context of a smart home environment (15). The design process applied a human-centered framework that included interviews with stakeholders and development of use case scenarios. However evaluation of the design was still an open-ended issue to address as future work.

2.2 Evaluating Visualizations with Older Adults

Evaluation of information visualizations is an essential component of the design process, allowing researchers to quantify differences across visualizations and reiterate as part of the design cycle. However, there is a lack of defined approaches towards evaluating and comparing visualizations with older adults. Reeder et al conducted semi-structured interviews with older adults to gather high level insights on 3 different visualizations of sensor data collected from a smart home environment (9). The interviews provided subjective feedback on design preferences and generated recommendations for iterative redesign. This approach helped highlight key design considerations across different versions of the visualization (9). Le et al evaluated health visualizations through focus groups. They presented three static visualizations to older adults and asked questions related to how participants analyzed the data and how they might use the visualizations. Though not a primary objective, the focus groups also generated design guidelines based on feedback from the participants (16). These included reducing metaphoric comparisons within the visualization, limiting differences in color gradient, and reducing the number of visual cues. In both studies general high-level differences were identified amongst the visualizations. However a systematic evaluation was not conducted to help designers compare the visualizations based on quantifiable metrics.

2.3 Methods to Evaluate Information Visualizations

Current approaches to evaluating visualizations are drawn from the information visualization literature. These span across different spectrums depending on the stage of the visualization design (17) though there is also strong discussion over which techniques are appropriate to apply when evaluating visualizations (2,17,18). For the purposes of this discussion, we focus on summative evaluations to compare performance of near finished designs as opposed to formative evaluations to inform further iterations of designs.

A standard approach to evaluating visualizations focuses primarily on assessing effectiveness through empirical task-based methods (benchmark evaluation). Participants are given defined tasks to complete on a visualization and metrics of accuracy and completion time are recorded. These quantitative measures allow for comparisons across visualization designs. A primary concern with empirical approaches is ecological validity. By having the researcher define select tasks for participants to complete, a functional bias is created against the visualization. The evaluation then focuses on how effective a visualization is at facilitating specific tasks. Though this can be a useful form of comparison, it limits the robust nature of health visualizations as an exploration tool and may be misaligned with how participants would use the visualization.

Given the goal of information visualizations as described by Card, Mackinlay, and Shneiderman is to provide insights into the data (19), both Saraiya and North propose an alternative open ended evaluation approach termed insight evaluations (20,21). Defined by Saraiya et al, an insight is an individual observation of the data leading to a unit of discovery in knowledge (20). For an insight-based evaluation, participants are asked to openly explore the information visualization in a think aloud manner. With minimal guidance from moderators, participants verbalize insights generated through interaction with the visualization. Researchers later code these insights. The number of insights can be quantified, categorized, and compared across visualizations.

A third approach to evaluation involves subjective questionnaires asking participants to rate their experience interacting with the visualization. This approach addresses the perceived usability and learnability of information visualizations. The questionnaire is given after participants have had an opportunity to fully explore and interact with the visualization. Perceived usability with an experience, though subjective in nature, has been shown to influence satisfaction and repeated use of products (22). The System Usability Scale (SUS) is a long-standing industry standard evaluation questionnaire consisting of 10 questions aggregated on a scale of 0-100 (1). The SUS is a robust questionnaire validated across domains addressing issues of usability and learnability of systems (23,24). Given that scales have yet to be developed specifically to evaluate information visualizations, the SUS serves as a viable alternative.

The three evaluation approaches described previously highlight different focuses of information visualizations. The benchmark evaluation technique emphasizes cognitive efficiency, whether or not a design facilitates effective analysis of the data. Insight evaluations focus on efficacy of visualizations and their ability to support the generation of naturalistic insights. The summative questionnaire assesses subjective satisfaction of a system. All three areas of effectiveness, efficacy, and satisfaction are essential towards quantifying the user experience when interacting with information visualizations (17). In this paper we provide a comparison of all three approaches within the context of evaluating health visualizations for older adults. The findings inform how the techniques may overlap as evaluation mechanisms and we provide a discussion on which technique would be more appropriate for evaluations with older adult stakeholders.

2.4 Preliminary Work

As part of a participatory design framework, we have developed and evaluated static visualization mockups of health data (16). Data were simulated based on a pilot study in which older adults would interact with a health kiosk to collect a wide range of data across domains of physiological, social, cognitive, and spiritual health. The data collection method varied and included physiological measurements, self-reported questionnaires, and cognitive games (25). Preliminary evaluation of mockups with both older adult and health care providers led to an iterative redesign and higher fidelity interactive visualizations (10,16). The visualizations were developed in Tableau, a data analysis tool that supports interactive visualization design (26).

We developed two visualizations based on the same underlying data set [Figure 1 and Figure 2]. The data consist of measurements aggregated into three levels of granularity. At the top level is overall wellness, composed of an average of four different components (physiological, social, spiritual, and cognitive domains). The components are further broken down into subcomponents, defined as specific measurements such as activities of daily living, processing speed, memory, or social engagement. The principles used to develop the visualizations are described fully in previous work (16). At a high level the visualizations aim to present wellness information as a holistic integration of different domains of health, allow older adults to track wellness longitudinally while also diving into specific domains as needed. The visualizations were also limited to familiar graphical displays (bar or line graphs) as recommended by users through focus groups. An interactive slider allows users to weight each component of wellness relative to the total overall wellness score. Though the visualizations are based off of the same data set, their interactions and layouts differ. We conducted an evaluation of the visualizations across three different approaches: a benchmark evaluation, insight-based evaluation, and perceived usability questionnaire. We present a comparison of the findings across evaluation techniques along with lessons learned from applying each method.

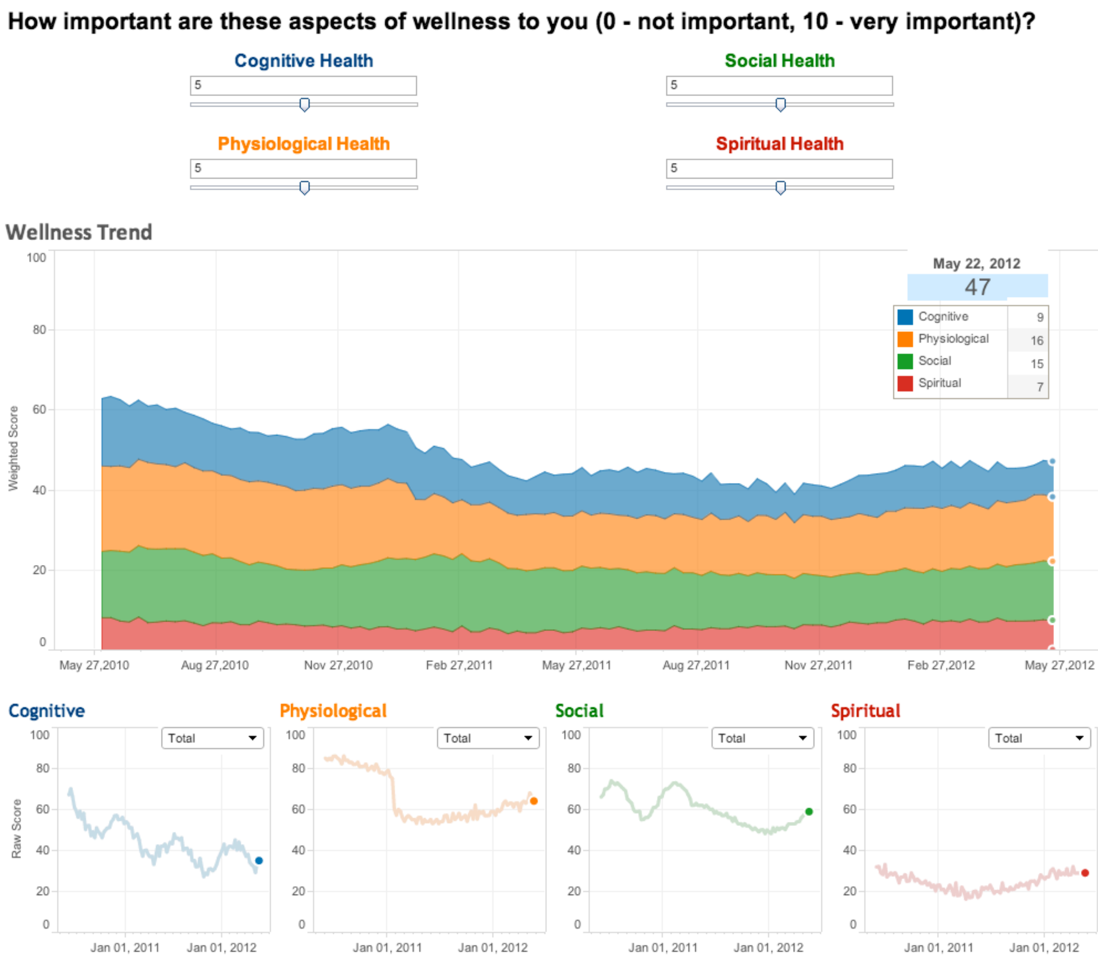

Figure 1: Visualization A in which each component of wellness is stacked as layers to form the overall wellness. Subcomponents of each component of wellness (cognitive, physiological, social, and spiritual) can be viewed by using the dropdown menu.

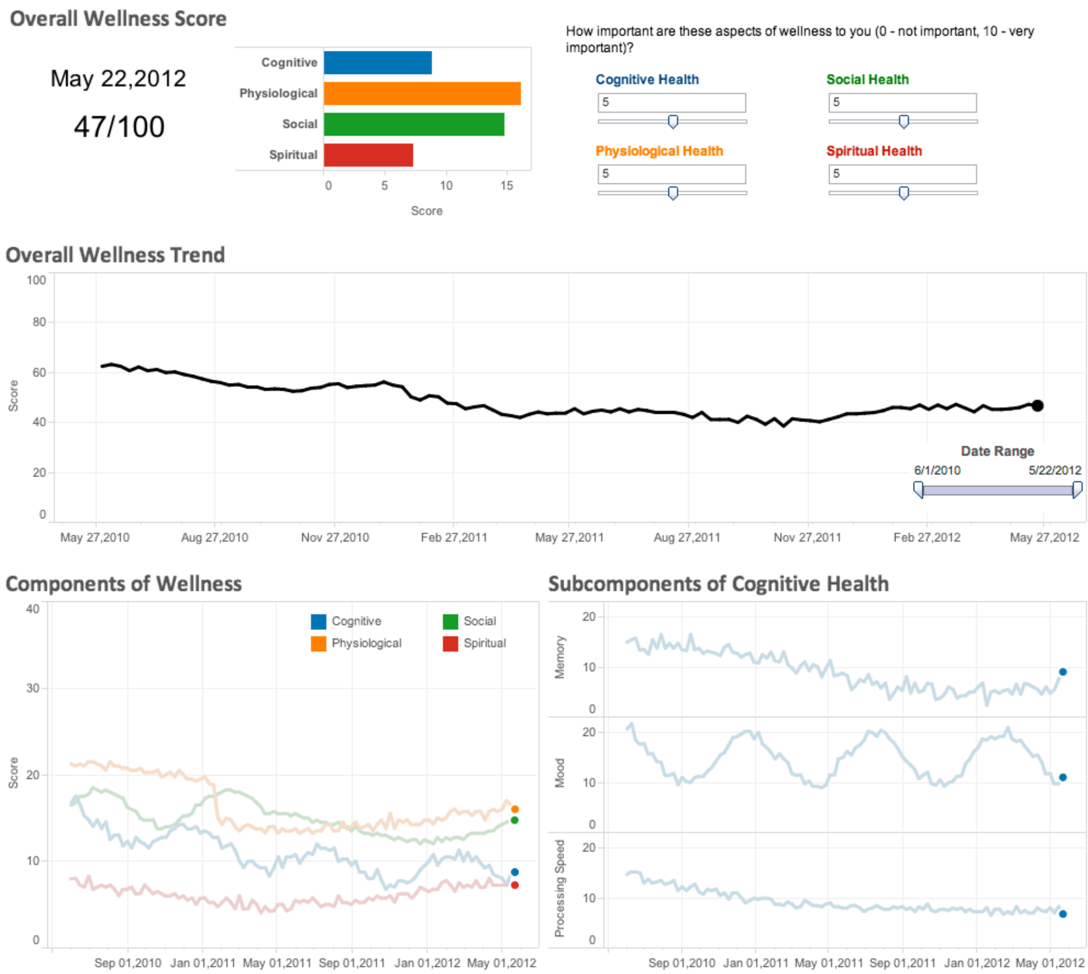

Figure 2: Visualization B in which overall wellness is isolated as a single line graph on top. The components of wellness are shown simultaneously. When a component is selected, the subcomponent graphs are updated to represent subcomponents for that area of wellness.

3. Materials and Methods

3.1 Participants and Setting

We recruited participants through older adult independent living facilities or community centers throughout the Puget Sound area. Recruitment methods varied depending on the facility and consisted of informational flyers or announcements through the community newsletter. We restricted participation to members of the community at least 60 years old, English speaking, and willing to participate through informed consent.

All procedures were reviewed and approved by the University of Washington Institutional Review Board.

3.2 Study Design

We conducted one-on-one sessions with older adults through a within subject 2 (visualization type) x 3 (evaluation method) design. Participants were first given a presentation describing the context of the study, data collection methods, and construct of wellness. We then presented the first of two visualizations, emphasizing that the data came from a hypothetical participant. We provided an overview describing the different interactions available within the visualization. Participants then completed both a benchmark and insight-based evaluation on the visualization. This was followed by a post session assessment of usability through the System Usability Scale. The same procedure was repeated for the second version of the visualization. We counterbalanced the order of visualizations presented, the order of the evaluation method (insight or benchmark), the task sets used for benchmark evaluation, and the order within the task sets. The sessions lasted at most 1.5 hours and participants received a $15 gift card for participation. We audio- and screen- recorded the sessions for later analysis.

3.3 Benchmark Evaluation

We presented participants with five tasks spanning across the breadth of interactions available within each visualization [Table 1]. The tasks were designed to incorporate Shneiderman's task taxonomy of overview, zoom, details-on-demand, relate, and extract (27). For example, a sample task involving overview would ask participants to identify if the overall wellness trend has increased, decreased or stayed the same. All tasks had definitive answers, allowing an assessment of task accuracy. We created two task sets requiring the same analytical approach to solve, only differing in minor detail such as time frame of analysis or type of wellness component to analyze. We counterbalanced task set shown to participants across visualizations and randomized the order of tasks.

Task Set
Task 1: From the period of Oct 2011 – Dec 2011, how has Jeffrey's overall wellness changed? (increased, decreased, stayed the same)
Task 2: By approximately how much has Jeffrey's overall wellness changed from the start of 2011 to the end of 2011?
Task 3: Which component was the largest contributor to Jeffrey's overall wellness in February 2011? (cognitive, physiological, social, spiritual)
Task 4: In the month of March, 2011, what subcomponent of spiritual health was the lowest? (spiritual engagement, spiritual strength)
Task 5: Suppose Jeffrey values cognitive health more as he ages. By how much does his overall wellness change if he rates cognitive health at 10 (very important)?

Table 1: Example task set shown to participants during the benchmark evaluation. Two task sets were generated with the same task objects, only differing in small details such as time frame or type of component.

3.4 Insight Evaluation

For the insight evaluation, we asked participants to openly explore the visualization while thinking aloud. We limited moderator involvement during the think-a-loud, though if necessary prompts were provided to encourage exploration. The prompts included: "Are there areas of the visualization you find interesting?", "What made you examine that part of the visualization?" and "How would you find out the answer to your question?". These prompts encouraged further exploration without biasing a specific area of focus. There was no defined time limit on the insight evaluation; we moved on when participants saturated on insight generation, indicated they were finished exploring, or ran into a hard limit on the session duration.

3.5 Summative Questionnaire

We administered the SUS when both the benchmark and insight evaluations were complete. This gave participants an opportunity to fully explore the visualization before providing an assessment of perceived usability. The SUS was administered on paper as we asked participants to rate each statement on the 5 point Likert scale. We administered the SUS for each visualization block.

3.6 Analysis

For the benchmark evaluation, we recorded task completion time along with task accuracy. We conducted matched pair t-tests to identify potential differences in mean task completion time between the two visualization types for each task type. To compare accuracy, we conducted McNemar's test for binary matched-pair data with the null hypothesis being no difference in accuracy proportion for a given task between the two visualization types.

For the insight evaluation, we first transcribed the sessions and then coded the transcripts to identify insights. We identified insights as observations about the data with the following characteristics described by North: *complex, deep, qualitative, unexpected,* and *relevant* (21). We also included a time stamp for each insight generated during the evaluation process. We categorized each insight based on the following properties:

- Uniqueness: is the insight unique from others generated during the evaluation session?
- Hypothesis: does the insight generate further questions and exploration about the data?
- Breadth or Depth: is the insight about a broad level trend in the data or is it a focused finding involving specific components of the visualization?

- Low-level analytic task: which analytical processes were used to arrive at the insight? These processes are derived from Amar et al.'s taxonomy of retrieve value, filter, compute derived value, find extremum, sort, determine range, characterize distribution, find anomalies, cluster, and correlate (28).

We compared mean number of unique insights generated between visualization types using matched pair t-tests. We also applied matched pair t-tests to compare the different properties of insights generated during the evaluation across visualization types.

We scored the SUS scale based on the methodology described by Brooke in which each question was added up with Likert values coded from 0-4. The scale was inverted for questions that were worded in the negative. This sum was normalized to a range of 0-100 by a scale multiple of 2.5 (1). We compared mean SUS score between visualization types using the matched pair t-test.

Given that the dependent variables for each of the three evaluation methods differed, we were unable to directly compare results across methods. However, we provided a descriptive comparison based on the overall findings and conclusions derived from each evaluation method. In addition, we highlighted the lessons learned in applying different visualization evaluation approaches with older adult participants.

4. Results

4.1 Demographics

We conducted sessions with 21 older adult participants (15 female, 6 male). Participants were predominantly white/Caucasian (19 participants) with mean age 70.5 (SD: 5.0) years. Participants' highest education was predominantly some college or an advanced degree [Table 2].

	Older Adults (N = 21)
Age (Mean, SD)	70.5 (5.0)
Gender (Female)	15
Race	
Black/African American	1
White/Caucasian	19
Other	1
Marital Status	
Divorced	9
Married/Partnered	5
Single	7
Highest Education	
High School/GED	1
Associates Degree	2
Some College	6
Bachelors Degree	6
Graduate Degree	6

Table 2: Demographic characteristics of the participant sample. All participants were at least 60 years old recruited from independent living facilities or older adult community centers.

4.2 Benchmark Evaluation

In comparing task completion times between visualizations we did not find statistical significance for any of the tasks. Task completion time was highly variable with participants differing in motivation and persistence in completing tasks. Between tasks, participants took longest on Task 2, while Task 1 and Task 3 were completed relatively quickly across both visualizations. When we restricted completion time to only participants who correctly completed tasks, we found a moderate difference in completion time for Task 3, with participants taking 16.6 (95% CI: [-6.3, 39.5]) seconds longer with visualization B compared to visualization A (p=.15).

	Visualization A	Visualization B
Task 1	53.1 (31.7, 74.4)	41.0 (28.6, 53.3)
Task 2	72.0 (47.4, 96.5)	86.1 (48.1, 124.1)
Task 3	52.2 (29.4, 75.0)	63.5 (43.7, 83.2)
Task 4	61.4 (44.2, 78.6)	83.9 (61.4, 106.4)
Task 5	77.3 (60.7, 93.9)	63.6 (49.0, 78.1)

Table 3: Mean time to complete task for the visualizations. A 95% confidence interval is provided on mean estimates. Matched pair tests did not identify statistically significant differences in completion time between visualizations.

Participants were able to complete the tasks across a broad range of accuracy levels. From most accurate to least were Task 3 (35/40 completed correctly), Task 1 (31/40 completed correctly), Task 4 (25/40 completed correctly), Task 2 (10/40 completed correctly), and Task 5 (6/40 completed correctly). We developed tasks to range in difficulty; this was reflected appropriately through completion rate. McNemar's matched pair test on accuracy of tasks between visualizations did not show statistical significance for task type. As a result, across visualization types there was strong concordance on accuracy. We did not find evidence to conclude that visualization type impacted proportion of participants who are able to correctly complete tasks.

Task 1 (31/40)	Vis B (incorrect)	Vis B (correct)
Vis A (incorrect)	0	7
Vis A (correct)	2	11
Task 2 (10/40)	Vis B (incorrect)	Vis B (correct)
Vis A (incorrect)	12	4

Vis A (correct)	2	2
Task 3 (35/40)	Vis B (incorrect)	Vis B (correct)
Vis A (incorrect)	1	1
Vis A (correct)	2	16
Task 4 (25/40)	Vis B (incorrect)	Vis B (correct)
Vis A (incorrect)	4	4
Vis A (correct)	3	9
Task 5 (6/40)	Vis B (incorrect)	Vis B (correct)
Vis A (incorrect)	15	2
Vis A (correct)	2	1

Table 4: 3-way contingency table of task type x visualization display x task accuracy. McNemar's matched pair test for proportions did not identify differences in accuracy between visualizations for each task type.

4.3 Insight Evaluation

Participants generated a mean of 13.8 (95% CI: [10.8, 16.7]) insights during the evaluation with visualization A. For visualization B, participants generated a mean of 20.0 (95% CI: [14.9, 25.1]) insights. A paired t-test found that participants generated on average 6.2 (95% CI: [0.4, 12.1]) more insights with visualization B compared to visualization A ($p<.05$).

Examining insights by breadth/depth across visualizations, we found that visualization A generated an approximately equal distribution of high level breadth first insights (46%) compared to in-depth insights (54%). Visualization B generated more in-depth insights (59%) compared to breadth (41%), though the difference in proportions was not considered statistically significant ($p=0.16$).

Breadth/Depth of Insight	Visualization A	Visualization B
Breadth	134	171
Depth	155	249

Table 5: Categorization of insights by breadth/depth for each visualization type. We did not find differences in proportion between displays.

We found 22.2% of insights in Visualization A generated hypotheses to drive further exploration or to provide an explanation for the trend in data. For example, a typical hypothesis driving insight would be "When I saw it going up and down I thought for some reason it would correspond to the seasons or something… Oh yea December it goes down until April, so maybe it is seasonal." In Visualization B, 30.7% of insights were hypotheses driving. This difference of 8.5% (95% CI: [1.8%, 15.4%]) was statistically significant ($p<.05$).

In categorizing insights through Amar et al's functional taxonomy of low-level tasks (28), we found that a majority of insights represented a characterization of distribution (56.7% in visualization A, 55.5% in visualization B). These insights often described trends in the data, for example if a trend was increasing, decreasing, or highly variable. The next most common role of insights was to identify a correlation (25.6% in visualization A, 31.9% in visualization B). Comparing between visualizations, we found that there was a 6.3% (95% CI: [.2%, 12.6%]) increase in insights that involved retrieving values in Visualization B compared to Visualization A ($p<.05$). In contrast, Visualization A had a 5.9% (95% CI: [.3%, 11.5%]) increase in insights that identified anomalies compared to Visualization B ($p<.05$).

Functional Categorization	Visualization A (Total Insights = 289)	Visualization B (Total Insights = 420)

Retrieve Value *	17.0%	23.3%
Filter	14.2%	12.4%
Compute Derived Value	2.4%	2.6%
Find Extremum	9.4%	10.7%
Sort	5.5%	3.6%
Determine Range	6.9%	6.9%
Characterize Distribution	56.7%	55.5%
Find Anomalies *	17.3%	11.4%
Cluster	10.0%	8.1%
Correlate	25.6%	31.9%

Table 6: Percentage of insights that fit each functional categorization. The categorizations are not mutually exclusive. T-tests were conducted to compare proportions between visualizations for each functional category. (* p<.05, ** p<.01, ***p<.001)

4.4 Summative Questionnaire

Participants rated visualization A on the SUS with mean score 66.1 (95% CI: [54.6-77.7]) while visualization B was rated with mean score 73.3 (95% CI: [64.0-82.5]). A score of 68 on the SUS scale is considered average at the 50^{th} percentile (29). This indicated that participants perceived visualization B as slightly above average on usability, while visualization A was on par with the median. A matched pair t-test comparing SUS scores showed that visualization B was rated higher than A by a mean score of 5.9 (95% CI: [-1.6, 13.5]), though this was not considered statistically significant (p=0.12).

5. Limitations

There are limitations to consider within the study. Our population was a convenience sample of older adults. As such, participation was voluntary and we had a fairly homogenous sample in ethnicity. Our participants also skewed towards higher educational levels with 57% having at least a bachelor's degree. There were also concerns about fatigue during the evaluation sessions, lasting approximately 90 minutes. To alleviate these, we gave participants breaks between visualizations and evaluation tasks. However as the session stretched, motivation to complete the tasks could have decreased. We alleviated some of these issues by counterbalancing the tasks and visualizations. There was also variability in participants' approaches towards the insight evaluation. Some considered certain insights too obvious to verbalize; others felt uncomfortable with the open-ended nature of the tasks. We controlled for these individual differences by conducting matched pair comparisons focusing on differences within a session as opposed to aggregate group level comparisons.

6. Discussion

6.1 Individual Evaluation Techniques

We found inconclusive results when comparing visualizations through the benchmark evaluation. Task completion time was highly variable, primarily because participants had differing levels of persistence in their approach. Given this variability, it was challenging to identify statistically significant differences in completion time between visualizations. We did expect to find a difference in task completion rate between visualizations. Though Visualization A and B represented the same data, there were significant differences in how the information was organized and presented. Visualization A emphasized individual components of wellness and their change in time. Visualization B, in contrast, supported comparisons of components by showing them aligned within the same graph. We would expect that performance on the tasks would reflect this difference in design, however the results did not show significant differences. There was a strong degree of concordance in task completion between the two visualizations. Instead, the primary factor impacting task completion was the task itself. Certain tasks participants struggled consistently with regardless of visualization design, similarly other tasks were completed easily by participants.

The insight evaluation indicated differences in efficacy of the visualizations. Participants generated more insights through Visualization B compared to Visualization A. One of the primary reasons was due to the overlay of components of wellness on the same graph in Visualization B. This supported comparisons between components much more effectively. Similarly all subcomponents for a given component of wellness were shown simultaneously. As a result, participants were more easily able to identify correlations between trends and generate hypothesis-driving insights. This was reflected in the increase of hypothesis and correlation insights generated by Visualization B compared to Visualization A. In contrast, Visualization A focused on individual elements and their change over time. For example, each of the components of wellness was shown on an individual graph. Dropdowns allowed users to shift focus to individual subcomponents of wellness. By emphasizing individual graphs, participants focused on trends for each element over time. This would allow them to better identify anomalies in the data (such as sudden drops or increases), consistent with the increase in number of insights related to anomalies for Visualization A.

Findings from the SUS did show a modest, though not statistically significant improvement in perceived usability for Visualization B compared to A. This was consistent with subjective feedback from participants who indicated a preference for Visualization B more frequently than Visualization A. Participants found the overall wellness trend of Visualization B simpler to use and to identify scores on specific dates compared to Visualization A. In contrast, the stacked layers on visualization A made it difficult to compare individual components, though participants did find the colors more vibrant and visually appealing. A common issue to both visualizations was the association of different graphs within the overall dashboard. Selecting a given point on a graph highlights the same point on different graphs within the dashboard. However, participants often did not notice this difference since the changes were within the periphery of focus.

6.2 Comparison of Evaluation Techniques

The three evaluation techniques had different response variables; as a result direct comparisons were not valid. However we are able to discuss differences in overall findings. One challenge with benchmark evaluations is that they are highly dependent on the tasks selected (30). This was reflected in our study where a predominant factor in task completion was the task itself, not the visualization type. We had trouble identifying differences between visualization types using the benchmark evaluation technique. The insight approach showed a difference in total number of insights generated. When looking at the insights in detail, we were able to find differences in types of insights generated that were validated by observations of how participants interacted with the visualizations. This provided a level of detail that we were unable to infer from the benchmark evaluation. The SUS evaluated the visualizations from the dimension of perceived usability. We found that the SUS scores validated some of the findings from the insight evaluation. Perceived usability has been found to correlate with improved engagement (22), leading to further exploration of the visualization and generation of more insights.

North, Saraiya, and Duncan provide a comparison of benchmark and insight evaluations for microarray visualizations (31). They were able to identify differences in task completion time amongst three visualizations for 2/7 tasks. There was a difference in accuracy for 1/7 tasks (31). The authors found a closer association between the insight evaluation and the benchmark evaluation, with findings from the insight evaluation validating the benchmark evaluation. We did not find a similar relationship and instead found high variability in task completion time across participants. The insights reported by North, Saraiya, and Duncan were further classified into domain specific categories, allowing granularity into which types of insights were generated more frequently by visualization (31). Our approach generalized the categorization of insights based on low-level analytical tasks proposed by Amar et al (28). This provides information on what functional tasks are supported by the visualizations through insight generation. We found that this approach does not require domain specific knowledge and allows for extension to other visualizations.

7. Conclusion

We approached evaluation of visualizations from the perspective of efficiency, efficacy, and perceived usability as assessed through benchmark, insight, and summative questionnaire evaluation methods. We found that the methods differed in their conclusions with a slight overlap of the SUS questionnaire and the insight evaluation. However, we found that the benchmark based evaluation approach was highly dependent on the tasks. As a result it did not necessarily reflect how participants would explore visualizations. Participants also commented on the discordance between certain tasks and their natural pattern of behavior, indicating that they would rarely look for such specific information. Though the insight evaluation provided a more naturalistic assessment, it also came at the cost of analysis time. We transcribed and qualitatively coded each session for insights. This was a time intensive analysis approach compared to analysis of the benchmark evaluation or SUS. However despite this increase in time commitment, we found that the insight evaluation was able to successfully distinguish visualizations and also identify functional tasks supported within the visualization. This was an added level of information consistent with observational behaviors when interacting with the visualizations. Our study identified differences amongst the three evaluation techniques while also confirming the value of insight evaluations in comparing visualizations.

References

1. Brooke J. SUS: A “quick and dirty” usability scale. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996. p. 189–94.

2. Santos BS. Evaluating Visualization techniques and tools: what are the main issues? Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization [Internet]. 2006 [cited 2014 Sep 25]. Available from: http://www.dis.uniroma1.it/beliv08/pospap/santos.pdf

3. Hoeke JO, Bonke B, van Strik R, Gelsema ES. Evaluation of techniques for the presentation of laboratory data: support of pattern recognition. Methods Inf Med. 2000 Mar;39(1):88–92.

4. Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators’ decisions to stop clinical trials: prospective trial with repeated measures. BMJ. 1999 Jun 5;318(7197):1527–31.

5. Feldman-Stewart D, Kocovski N, McConnell BA, Brundage MD, Mackillop WJ. Perception of quantitative information for treatment decisions. Med Decis Mak Int J Soc Med Decis Mak. 2000 Jun;20(2):228–38.

6. Morrow DG, Leirer VO, Andrassy JM. Using icons to convey medication schedule information. Appl Ergon. 1996 Aug;27(4):267–75.

7. Huh J, Le T, Reeder B, Thompson HJ, Demiris G. Perspectives on wellness self-monitoring tools for older adults. Int J Med Inf. 2013 Sep 4;

8. Reeder B, Le T, Thompson HJ, Demiris G. Comparing information needs of health care providers and older adults: findings from a wellness study. Stud Health Technol Inform. 2013;192:18–22.

9. Reeder B, Chung J, Le T, Thompson H, Demiris G. Assessing Older Adults' Perceptions of Sensor Data and Designing Visual Displays for Ambient Environments. An Exploratory Study. Methods Inf Med. 2014 Apr 14;53(3).

10. Le T, Reeder B, Thompson H, Demiris G. Health Providers' Perceptions of Novel Approaches to Visualizing Integrated Health Information. Methods Inf Med. 2013 May 7;52(3):250–8.

11. Chisnell DE, Redish JCG, Lee A. New heuristics for understanding older adults as web users. Tech Commun. 2006;53(1):39–59.

12. Chisnell D, Redish J. Designing Web Sites for Older Adults: Expert Review of Usability for Older Adults at 50 Web Sites. AARP; 2004.

13. Czaja SJ, Lee CC. The impact of aging on access to technology. Univers Access Inf Soc. 2006 Dec 8;5(4):341–9.

14. Making Your Web Site Senior Friendly: A Checklist. National Institute on Aging; 2002.

15. Gong Y, Chandra A. Design of a holistic dashboard for technologies enhancing aging in place. 2011 13th IEEE International Conference on e-Health Networking Applications and Services (Healthcom). 2011. p. 254–7.

16. Le T, Reeder B, Yoo D, Aziz R, Thompson H, Demiris G. An Evaluation of Wellness Assessment Visualizations for Older Adults. Telemed E-Health. In Press;

17. Carpendale S. Evaluating information visualizations. Information Visualization [Internet]. Springer; 2008 [cited 2013 Oct 23]. p. 19–45. Available from: http://link.springer.com/chapter/10.1007/978-3-540-70956-5_2

18. Bertini E. Evaluation Comes in Many Guises. Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization

[Internet]. ACM; 2006 [cited 2014 Sep 25]. Available from: http://dl.acm.org/citation.cfm?id=1168149

19. Card SK, Mackinlay JD, Shneiderman B. Readings in information visualization: using vision to think. San Francisco, Calif.: Morgan Kaufmann Publishers; 1999.

20. Saraiya P, North C, Duca K. An insight-based methodology for evaluating bioinformatics visualizations. Vis Comput Graph IEEE Trans On. 2005;11(4):443–56.

21. North C. Toward measuring visualization insight. Comput Graph Appl IEEE. 2006;26(3):6–9.

22. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Inf Manage. 2006 Jan;43(1):1–14.

23. Lewis JR, Sauro J. The Factor Structure of the System Usability Scale. In: Kurosu M, editor. Human Centered Design [Internet]. Springer Berlin Heidelberg; 2009 [cited 2014 Oct 26]. p. 94–103. Available from: http://link.springer.com/chapter/10.1007/978-3-642-02806-9_12

24. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Int J Hum-Comput Interact. 2008 Jul 29;24(6):574–94.

25. Thompson HJ, Demiris G, Rue T, Shatil E, Wilamowska K, Zaslavsky O, et al. A Holistic Approach to Assess Older Adults' Wellness Using e-Health Technologies. Telemed E-Health. 2011 Dec;17(10):794–800.

26. Business Intelligence and Analytics [Internet]. Tableau Software. [cited 2014 Oct 26]. Available from: http://www.tableausoftware.com/

27. Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. Visual Languages, 1996 Proceedings, IEEE Symposium on [Internet].

1996 [cited 2013 Oct 23]. p. 336–43. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=545307

28. Amar R, Eagan J, Stasko J. Low-level components of analytic activity in information visualization. Information Visualization, 2005 INFOVIS 2005 IEEE Symposium on [Internet]. IEEE; 2005 [cited 2014 Sep 22]. p. 111–7. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1532136

29. Sauro J. Measuring Usability with the System Usability Scale (SUS) [Internet]. 2011 [cited 2014 Oct 29]. Available from: http://www.measuringu.com/sus.php

30. Saraiya P, North C, Duca K. Comparing benchmark task and insight evaluation methods on timeseries graph visualizations. Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization [Internet]. 2010 [cited 2013 Oct 23]. p. 55–62. Available from: http://dl.acm.org/citation.cfm?id=2110201

31. North C, Saraiya P, Duca K. A comparison of benchmark task and insight evaluation methods for information visualization. Inf Vis. 2011 Jul 1;10(3):162–81.

Paper 3: Understanding Older Adult Use of Data Visualizations as a Resource for Maintaining Health and Wellness

Authors: Thai Le (BS), Nai-Ching Chi (BS), Shomir Chaudhuri (BS), Hilaire J. Thompson (PhD, RN), George Demiris (PhD)

Abstract

It is important to develop efficient strategies and tools for promoting healthy aging as the older adult population continues to grow. An approach to this is through the development of data visualizations derived from monitoring technologies. A well-designed visualization that translates data from health monitoring technologies into consumable information may empower older adults within their care. Presently, there is limited research on how older adults use data visualizations and the potential barriers that impact utility. We conducted semi-structured interviews with 21 older adults to get their perspectives on use of health visualizations. Through an affinity mapping exercise, we extracted key themes associated with how older adults utilize health visualizations. Based on these themes, we provided a set of recommendations as points of consideration for other designers developing older adult focused health visualizations.

Key Terms: Data and Knowledge Visualization, Older Adults, Interactive Data Exploration and Discovery

1. Introduction

The older adult population is one of the fastest growing demographic groups in the United States (1). Associated with the aging US population are changes in health and wellness, which present unique challenges in balancing the need for independence with considerations for safety. One approach towards alleviating concerns over health and wellness of older adults while maintaining independence is through the integration of technology within the home. E-health and smart home technologies allow for a holistic assessment of wellness across dimensions of cognitive, physiological, social, and spiritual health (2). However, the data collected from these resources represent potentially complex concepts of wellness. Data visualizations help translate this information into a more consumable format. For older adults, this has the potential of bridging the gap between the abstract collection of data and the tangible representation of integrated wellness. Though a body of work has shown that older adults find utility in technology to support their health and wellness (3–6), there has been limited research examining how this would translate to data visualizations (7). Perceived utility has been found to play a key role in adoption and use of systems (8–10). Visualizations are intended to serve a need, however if this does not align with those of the older adult consumer it becomes a resource of limited impact. In this paper, we describe findings from evaluations of interactive visualizations with 21 older adults focusing on how the visualizations may be used as a resource for maintaining health and wellness. Our work highlights key points of consideration for other researchers and designers as they develop visualizations for older adults. We complement the findings with recommendations to support health visualization development for the older adult consumer.

2. Background

2.1 Informatics as a Resource to Support Aging

With the improvement of medical care and preventive health measures, there is an increase in life expectancy within the United States. The older adult demographic group, those at least 65 years old, is growing at an unprecedented rate and by 2040, this group is expected to represent 20% of the U.S. population (1). This growth is due to both an increase in life expectancy and the aging of the baby boomer population. Due to the increased prevalence of chronic diseases and conditions associated with the aging process, the cost of providing health care to an older adult is three to five times higher than the general population (1). The 2007 CDC report on *The State of Aging and Health in America* identified a key public health goal of supporting dignity during the aging process and maintaining quality of life for older adults (11). Within this context, research in biomedical informatics and gerontological care can provide valuable contributions towards developing tools to help older adults maintain well-being and independence.

One area of work for health assessment involves integrating data from health monitoring technologies. Efforts in this area have the potential to provide near real-time and continuous collection of health data with greater accessibility compared to episodic assessments at clinic visits. These in home health monitoring tools include wireless physiological assessment devices, motion-sensors, embedded infrastructure sensors, and cognitive assessments embedded in computer software. Though there are some potential concerns raised over the obtrusiveness of in home health monitoring technologies, prior research by Wild et al. and Demiris et al. have found that older adults recognize the potential applications of health monitoring tools and accept them when the perceived utility of the monitoring technology exceeds concerns over privacy (3).

2.2 Data Visualizations for Older Adults

Health monitoring technologies make it possible to track health and wellness continuously, unobtrusively, and reliably while providing stakeholders with quantifiable feedback. A significant challenge towards promoting these value propositions lies in demonstrating tangible insights gathered from health monitoring data. It is not enough to have the infrastructure and technology in place to collect data, especially since older adults may not be familiar with the detailed datasets generated, which also often requires clinical knowledge to interpret. Appropriately designed visualizations can bridge this gap between data and information. A well-designed visualization that translates data from health monitoring technologies into a consumable medium for older adults can promote active engagement in healthcare and furthers communication amongst members of the care team.

There is a limited body of literature on data visualization research with a specific older adult consumer focus. Mynatt et al highlight the importance of maintaining an awareness of long-term health and well-being for older adults (12); this is operationalized through the development of a Digital Family Portrait (12,13). The portrait is a visualization of sensor activity collected from a smart home environment with different icons bordering the digital picture frame representing activities of daily life (13). As part of the design process, Mynatt et al conducted a needs analysis with family members to identify what information should be represented, how family members assess well-being of older adults, and what social values contribute to the use of the visualization tool (13). The authors evaluated the visualization through a field trial with an older adult/family member dyad, finding that the initial designs were too complex and ambitious in conveying ten levels of information. However, the authors did find that changes in the digital portrait initiated between the older adult and family members (13). A more extended one year case study found differences in how older adults and family members used the visualization, concerns over privacy and security, and value of providing awareness and connectedness through the Digital Family Portrait (14).

Reeder et al describe the design of sensor visualizations from a six-month pilot study with older adults (7). The authors conducted semi-structured interviews with older adults during the course of the pilot study, elucidating issues related to perceived usefulness of sensor data. The authors found that visual displays of sensor data were useful for caregivers of older adults experiencing cognitive decline. From the older adult perspective, visualization of sensor data was useful for consultations with their health care providers about activity levels. Reeder et al integrated the feedback from the semi-structured interviews in conjunction with design principles in data visualization to develop three visual displays of sensor data (7).

Le et al focus on design of data visualizations for integrated health and wellness, applying Dunn's wellness model to categorize data collected into cognitive, physiological, social, and spiritual wellbeing (15,16). This visualization was evaluated with both older adults and health care providers in separate focus groups (17,18). The authors focused on how older adults approached the visualization process and provided a set of design recommendations based on feedback from participants (18). The authors also found that at a high level, older adults identified with the value of having visualizations as a health assessment tool (18). However the focus of the authors' work was on understanding the processes involved in analyzing visualizations as opposed to the utility the visualizations.

Our work complements the existing literature by examining the perceived utility of an integrated wellness visualization tool. We extended visualizations outside of the sensor focus described by Reeder et al (7) and towards integrated wellness collected from multiple data sources. We focused on interactive graphical visualizations as opposed to metaphorical representations as described by Mynatt et al (13) to allow for more exploration within the visualization. The exploration process allowed older adult consumers to better conceptualize how they would utilize the visualizations over the course of the evaluation session.

2.3 Older Adult Design Guidelines

Data visualizations can support consumers for both health assessment and shared decision-making when designed appropriately. However, the older adult population represents a unique demographic group with different design considerations than the general population due to aging associated differences in vision, cognition, and motor control (19). Redish and Chisnell provide a set of heuristics for website design with older adults as primary consumers based on a literature review (20). Evaluations using the heuristics are demonstrated across various websites for different older adult profiles taking into account age, ability, aptitude, and attitude (21). Demiris, Finkelstein, and Speedie provide a set of guidelines for the design of web systems related to system interface, training and support of users, and page content (22). The National Institutes of Aging also has a set of guidelines for older adult friendly web sites which includes organizing web information appropriately, using readable text, writing appropriate text for the audience, making information easier to find, and providing alternative media sources for the information (23). Much of the design recommendations focus on a usability perspective, optimizing a user's ability to navigate, find, and interact with information. Translated to data visualizations, these guidelines can be used to support effective comparisons and identification of trends. However, usability is only half of the challenge in designing data visualizations. We address the second half associated with utility of health visualizations through a qualitative analysis of interviews conducted with older adults. Having this focus on utility within the design framework is a valuable resource towards encouraging use of a system.

2.4 Prior Work

We iterated on designs of health visualizations with older adults using a user-centered approach (24). We presented early visualization mockups to older adults through focus groups (18). We used focus groups as a methodological approach to encourage discussion on the design prototypes, allowing participants to generate recommendations in a collaborative environment. Based on the findings from the focus groups, we refined the visualizations into higher fidelity interactive prototypes. The visualizations consisted of bar and line charts representing simulated wellness data (25). Interactions allowed users to view wellness at differing levels of granularity while principles of brushing and linking were applied to connect multiple views of data (26). For this work, we focused less on the design and evaluation of visualizations and instead we report findings related to use of health visualizations from an older adult consumer perspective. In particular, we wanted to understand: 1) do older adults identify utility in visualizing their wellness data, 2) what are different use cases for the visualizations, and 3) what are barriers limiting use of the visualizations. From these findings, we also provide guidelines that support other researchers and designers in the development of older adult health visualizations.

3. Materials and Methods

3.1 Semi-Structured Interview and Evaluation Sessions

We conducted one-on-one semi-structured interviews with 21 older adults to obtain their perspectives on use of health visualizations. During this interview, we presented participants with two interactive visualizations developed over the course of iterative user-centered design (see Appendix). Part of the interviews consisted of evaluation components where we asked participants to complete benchmark tasks and openly explore the visualization. We described the design and evaluation of the visualizations as part of prior work in Le et al (25). We prompted participants during the open exploration phase of the sessions with questions such as: Are there areas of the visualization you find interesting? What made you examine that part of the visualization? Do you have any questions that you would like to answer in the visualization? At the completion of the session we asked participants about their overall perspectives on the visualizations and any potential use cases they may find for such a resource.

3.2 Participation and Setting

We recruited participants for the sessions through contact with independent living facilities and older adult community centers throughout the Seattle, WA area. Depending on facility preferences, recruitment included distribution of flyers on community boards, postings within newsletters, and snowball sampling. Sessions took place at the community center or a common room in the independent living facility, lasting at most 90 minutes. We presented participants with a $15 gift card at the completion of the session. All participants had to be at least 60 years old, English speaking, and willing to participate through informed consent. Participation also included audio- and screen- recording of the sessions. The University of Washington Institutional Review Board approved all procedures involving human subjects.

3.3 Qualitative Analysis

We transcribed the sessions verbatim, removing any potential personal identifiers. TL conducted an initial high-level review of the transcripts, coding for any reference to use and utility of the visualizations. TL, NC and SC then completed an affinity mapping exercise to aggregate coded excerpts into themes. Affinity mapping is an inductive process that organizes ideas and concepts into themes (27). In this process, we printed excerpts generated from the initial high-level review onto separate notecards. We displayed these notecards on a table and collaboratively reorganized the notecards based on content, grouping together similar excerpts. The process came to a conclusion when all excerpts had been grouped. We then labeled the groups based on their thematic content. Over the course of the affinity mapping, the initial high-level codes were also validated collaboratively (this resulted in the exclusion of certain transcript excerpts that did not fit the code of utility). We reviewed the categorizations generated from affinity mapping separately at the conclusion of the exercise and reconciled differences as a group. Affinity mapping allowed us to generate themes associated with use and utility of health visualizations through a bottom-up, inductive approach emergent from the raw data of participant interviews. These results were presented through a qualitative description of the themes (28,29).

4. Results

We conducted semi-structured interviews with 21 older adult participants with average age 70.5 years old (sd=5.0 years). A majority of participants were female (n=15) and Caucasian (n=19). Participants varied in highest education level completed across high school/GED degree (n=1), associate degree (n=2), some college (n=6), bachelor's degree (n=6), and graduate degree (n=6). Participants also varied in marital status across divorced (n=6), married/partnered (n=5), and single (n=7). Health literacy as assessed through the Rapid Estimate of Adult Literacy in Medicine Short Form found that two participants had a 7th to 8th grade reading level while the remaining participants had at least a 9th grade reading level (30).

Qualitative analysis identified 5 themes associated with use and utility of health visualizations for older adults. An overview of these themes is presented in Table 1. We report on the themes in detail below.

Themes	Summary	Recommendations
Use of Visualization as an Intervention	Participants believed these visualizations could be useful in detecting trends in a person's life as long as it included guidance for facilitating improving the trend.	Integrate health promotion recommendations within the visualization.
Use of Contextual Information to Improve Utility	Participants stated a desire for contextual information associated with the visualizations (i.e life events, community events etc.)	Allow annotations within the visualizations to contextualize changes in the visualization.
Perceived	Many participants expressed concerns with	Provide multiple

Limitations Due to Computer Literacy	not being able to access or manipulate the visualizations due to a lack of understanding in using computers.	mediums through which the participants can access the data (i.e. paper printouts)
Timeframe for Representing Data	Participants saw the value in the data over a monthly or yearly time frame in order to more clearly detect patterns. They also expressed the limitations in the data as a daily measure of health.	Focus on designing visualizations to detect longitudinal changes over time.
Sharing of Health Information	Participants were open to sharing their health information with health care providers. They were less open to sharing the same information with family members or friends.	Allow for custom sharing of information. This includes who should be able to view the data and what level of information they should have access.

Table 7: Overview of primary themes impacting utility of health visualizations. Based on the themes, we provided recommendations that should be considered when designing other health visualizations for older adults.

4.1 Use of Visualization as an Intervention

Participants found that the visualizations could be used primarily as an intervention tool. The visualizations supported the identification of trends, in particular gradual differences that would not otherwise be noted day-to-day. As an example, participants commented that sudden changes in wellness measures such as social or spiritual health could be correlated with activities such as a visit from relatives or a trip out of town. These differences the participants already had knowledge of and the visualization, though it indicated a sudden change, would not provide personal insight into the data. However, detecting changes such as cognitive decline was an important use case for participants. These changes would not be as noticeable, therefore the visualizations allowed participants to step back and identify patterns over time. Detecting differences was only half the story for participants; the visualization provided greater utility if there was some guidance to redirect the course of decline in health:

> *"I think I want to see more information like these are the things you can do to improve your cognitive abilities, these are the things you can do to improve your physiological, each one of these things I would like to see that on a separate page somewhere. That would be more helpful to me rather than just looking at this and not knowing what's causing the things to go up or down or to change."* – P18

There was a sharp dichotomy of how participants perceived visualizations could be used as an intervention. Participants were generally positive about sharing the visualization with family members and health care providers. The visualizations could then be used for health assessment and external intervention:

> *"There were other times where it went down significantly. And those would be if the reports were sent to a caregiver or to a family member they would know that it might be appropriate to intervene."* – P01

However, for personal use of the visualization as an intervention, participants were mixed in response. There was a strong perception that, either individually participants already had a strong recognition of their health and wellness, or that if there was a significant decline, especially in cognitive health, there would be challenges in interpreting the visualization. In either case, personal use of the visualization as an intervention tool would not be as helpful for participants.

> *"It [the visualization] wouldn't be relevant to me because I kind of know. I know how my social thing is. I'm going. I know spiritual, I know my psychological, and I know myself."* – P07

"Well, from the looks of it, Jeffrey has deteriorated in all areas so how well he would be able to understand this graphical concept, I just don't think it would be really there for him." – P03

4.2 Contextual Information Supports Utility

Participants found that a significant challenge of using the visualizations was the lack of contextual information. They were able to identify differences in patterns within the visualization such as longitudinal changes or certain drop-offs in wellness. This prompted discussions hypothesizing potential causes associated with the changes. However, participants noted that this was all based on inference and they lacked the contextual information to understand what prompted the differences within the graph.

"You know, well it's hard, without knowing more about his history, it's really hard for me to say because I just kind of look at graphs and stuff, it's not the whole picture. To me, I would need, besides the graph, the visual and all of that, I would want to have the interaction between myself and him also and talk, find out what's going on as much as he would want to share with me." – P03

"I think once you get used to it, you can pick up a lot of information just glancing at it. And maybe a few clicks. Every now and then I think it would be useful to take to a deeper level to see if you can see patterns. Of course you're not just mashing what's there, you're mashing that with what you know about your life. I think it's a very large part of how it would be useful." – P09

As a result, participants found that the visualizations provided a limited understanding the data. The contextual information, either annotated within the visualization or through conversations with the participant, helped answer questions about why certain patterns appeared.

4.3 Perceived Limitations Due to Computer Literacy

Though participants successfully interacted with the visualizations while exploring and completing usability tasks, they expressed uncertainty about making full use of the system due to perceived limitations in computer literacy. For example, participants caveated their exploration process by indicating they used computers for limited tasks such as browsing the Internet or checking their email. Though this did not impede their ability to interact with the visualization, the perception of computer literacy impacted potential use of the visualization. As a participant noted:

> *"Imagine if most people would learn to use this very quickly? Well, most people are more computer literate than I am, I think… I would say people with computer skills would learn to use this of course. And the more competent they are the more quickly they would learn it."* – P13

Participants also commented that a lack of access to computers would limit potential use of the visualizations. As a result, alternative mediums to present the information would be valuable complements to the visualizations.

> *"I think almost everybody that I know, educated or not, a printout would do just fine. There's no way… most of us aren't computer savvy anyways, including myself… It would be more for a year period of time, maybe a half year period of time or maybe periodically, get a printout like this."* – P07

4.4 Timeframe for Representing Data

Participants indicated that visualizations of health data had optimal timeframes for representation. Day-to-day visualizations had limited use, primarily because of the high variability of the measures. As one participant noted:

> *"Day-to-day, well, I mean I looked at it day-to-day I might go a couple of days to see, because you go day-to-day you're going to have days when you're down and days when you're up. So I don't know if I would want to study it that much day-to-day."* – P14

Instead, participants indicated that visualizations were most useful at monthly or yearly intervals. This aligned with their use of the visualizations as a resource to monitor longitudinal changes in health and wellness. It was important for participants to view the visualizations over a long enough period of time that they were able to make an assessment on what constituted a normative pattern of wellness. Shorter time interviews limited the contextual view of the wellness, making it difficult for participants to assess patterns.

> *"… from certain date to date why would anybody want to know that anyways, particularly if you've got a graph that's showing you the dates from 2011 to 2012. You've got this to this, that's one year. Those are important and you can just use that… I mean a year's time would be plenty. In other words looking at these graphs would be plenty. You wouldn't need to do anything. You could note that okay I went down, it went down in any of them except social and that went up…"* – P07

4.5 Sharing of Health Information

Participants were receptive towards sharing the visualizations with other stakeholders, in particular health care providers. Participants envisioned the visualizations as a supplement to their conversation with health care providers.

"I think it could help them [health care provider] perhaps come to an understanding of what really is underlying the problems. Because so many medical problems, as you know and I know, have an emotional for want of another word, an emotional background, cause. Many times it's, why do you really have a headache and why is this really bothering you. And sometimes it'd be good to understand that you're worried, they could see that their … patient [is] suffering with problems with their cognitive skills they could act, whereas they may not have noticed it. I think a good practitioner would have known it already but possibly didn't and, so I think they could use it." – P10

However, in contrast to the open communication with health care providers, participants were less willing to share information with family members and expressed concerns with privacy. Though participants recognized the utility of sharing the information with family members, for example as an alert to monitor changes in health, participants wanted to maintain more control over what components of the visualization their family members were able to see.

5. Limitations

We recruited through a convenience sample of older adult independent living and community centers in the Seattle, WA area. We found that a majority of participants who chose to participate were of higher educational background and more homogenous in race than the general older adult population. Given that participants self-selected to take part in the study, there was also potential for bias towards those who had an interest in the study topic of interactive health visualizations.

6. Discussion

Based on the themes identified through semi-structured interviews with older adults, we provide a set of design recommendations. We have aligned the recommendations with the findings on use and utility of visualizations from this study. The recommendations are generalizable towards development of future health visualizations for older adults.

Emphasize High Level Visualization of Health Data

Participants indicated that there was greater utility in visualizing health data over extended periods of time. Visualizations should support this objective by providing overview of health information over a long enough period that participants would be able to assess baseline levels and variations from the norm. This approach towards visualization is not new; Shneiderman has emphasized this in his mantra of information visualization "Overview first, zoom and filter, then details-on-demand (31)." However, we have found that though providing an overview is essential, there was less utility in providing zoomed views of the data. This was because, within the context of wellness assessment, participants found day-to-day variability subject to multiple factors that limited any conclusions drawn from the visualization.

Provide Annotations within the Visualization

Though participants recognized that there was variability day-to-day, they indicated that contextual information would be a valuable resource in explaining trends in the data. Allowing annotations within the visualizations would help contextualize the data, serving as a personal reminder for participants while also informing family members or health care providers. The visualizations allowed participants to extract information related to trends and patterns in wellness; however the values were still an abstraction. Participations commented on how this provided little information as to why there was an increase or decrease in the data. Different approaches to this could include automatically populating contextual information from external sources such as electronic health records or digital diaries in addition to manual annotations.

Integrate Health Promotion Recommendations

Participants found that though the visualizations showed changes in health and wellness, it left them unclear about next actionable steps, for example to reverse a declining trend in health. For family members and health care providers, this information was useful to instigate a conversation and intervene if changes were noted. However, participants had trouble translating the information from the visualizations into action. One approach to alleviate this is by providing recommendations to improve wellness based on trends within the visualization. The recommendations allow participants to take a proactive approach; allowing them to ground the abstract visualizations with actionable items. This concept of self-efficacy within health promotion has been shown to be an essential predictor of both behavior and health outcome (32–34). Integrating health promotion recommendations shifts the paradigm of health visualizations towards a data-to-information-to-action approach that older adults find more impactful and of greater value.

Allow Print Media Access to Visualizations

Though older adults are steadily growing in their adoption of technology, as a demographic group they are still far below the national average in the United States (35). Internet adoption is at a current high of 59% for older adults; however within this population are two distinct profiles. Younger, higher-income, and more educated older adults are comfortable with technology and have positive views towards its benefits. The older and less affluent demographic group is more disconnected with technology and generally has a skeptical attitude towards the benefits of technology (35). Participants within our sessions expressed concerns about access to the visualizations, especially for themselves or others whom they observed as less comfortable with using computers. Developing visualizations that can be distributed as a printed medium can address some of the challenges associated with technology adoption while also opening up access to a broader audience.

Support Custom Sharing of the Visualizations

We found that participants identified different levels of comfort when sharing the visualizations with others. There was a strong use case for sharing the visualizations with both family members and health care providers. Given the sensitive nature of health data, it was important that participants were allowed to control who was able to view the visualization and what level of detail to present the visualization. This differed depending on the stakeholders. The level of sharing could also be adjusted over time depending on the changing needs of the older adult.

7. Conclusion

We conducted semi-structured interviews with older adults while evaluating two different visualizations. Through an affinity mapping exercise, we extracted key themes associated with older adults' utilization of health visualizations. Based on these themes, we provide a set of recommendations as points of consideration for other designers developing older adult focused health visualizations. As the opportunities to embed technology into the home increase, health visualizations make it possible to present data back to the older adult consumer in an effective manner. By examining how older adults perceive of the utility of health visualizations, we lay the groundwork for design choices that impact eventual use and adoption of systems that generate data for such visualizations. A further area of work includes extending the visualizations to a longitudinal study in which older adults interact with live versions of the visualization grounded in their data. Direct investment in the visualizations may alter how older adults interact with and use the visualizations outside of reflections on perceived utility. In addition, older adult specific visualization guidelines are still limited. Information needs of older adults differ from those of health care providers; as such there is work to be done in defining recommendations on how graphs should be designed and incorporated into systems to fit the needs of older adults.

References

1. A Profile of Older Americans: 2013 [Internet]. Administration on Aging; [cited 2014 Nov 7]. Available from: http://www.aoa.acl.gov/Aging_Statistics/Profile/2013/docs/2013_Profile.pdf

2. Thompson HJ, Demiris G, Rue T, Shatil E, Wilamowska K, Zaslavsky O, et al. A Holistic Approach to Assess Older Adults' Wellness Using e-Health Technologies. Telemed E-Health. 2011 Dec;17(10):794–800.

3. Wild K, Boise L, Lundell J, Foucek A. Unobtrusive In-Home Monitoring of Cognitive and Physical Health: Reactions and Perceptions of Older Adults. J Appl Gerontol. 2008 Apr 1;27(2):181–200.

4. Coughlin JF, D'Ambrosio LA, Reimer B, Pratt MR. Older adult perceptions of smart home technologies: implications for research, policy & market innovations in healthcare. Engineering in Medicine and Biology Society, 2007 EMBS 2007 29th Annual International Conference of the IEEE [Internet]. 2007 [cited 2013 Oct 28]. p. 1810–5. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4352665

5. Melenhorst A-S, Rogers WA, Caylor EC. The Use of Communication Technologies by Older Adults: Exploring the Benefits from the User's Perspective. Proc Hum Factors Ergon Soc Annu Meet. 2001 Oct 1;45(3):221–5.

6. Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M, et al. Older adults' attitudes towards and perceptions of "smart home" technologies: a pilot study. Inform Health Soc Care. 2004 Jan;29(2):87–94.

7. Reeder B, Chung J, Le T, Thompson H, Demiris G. Assessing Older Adults' Perceptions of Sensor Data and Designing Visual Displays for Ambient Environments. An Exploratory Study. Methods Inf Med. 2014 Apr 14;53(3).

8. Nielsen J. Usability engineering. San Francisco, Calif.: Morgan Kaufmann Publishers; 1994.

9. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q. 1989 Sep;13(3):319–40.

10. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man-Mach Stud. 1993 Mar;38(3):475–87.

11. Centers for Disease Control and Prevention and The Merck Company Foundation. The State of Aging and Health in America 2007. The Merck Company Foundation; 2007.

12. Mynatt ED, Essa I, Rogers W. Increasing the opportunities for aging in place. Proceedings on the 2000 conference on Universal Usability [Internet]. ACM; 2000 [cited 2014 Nov 8]. p. 65–71. Available from: http://dl.acm.org/citation.cfm?id=355475

13. Mynatt ED, Rowan J, Craighill S, Jacobs A. Digital family portraits: supporting peace of mind for extended family members. Proceedings of the SIGCHI conference on Human factors in computing systems [Internet]. 2001 [cited 2013 Oct 28]. p. 333–40. Available from: http://dl.acm.org/citation.cfm?id=365126

14. Rowan J, Mynatt ED. Digital family portrait field trial: Support for aging in place. Proceedings of the SIGCHI conference on Human factors in computing systems [Internet]. ACM; 2005 [cited 2014 Nov 9]. p. 521–30. Available from: http://dl.acm.org/citation.cfm?id=1055044

15. Le T WK. Integrated data visualisation: an approach to capture older adults' wellness. Int J Electron Healthc. 2012;7(2):89–104.

16. Dunn HL. High-level wellness for man and society. Am J Public Health Nations Health. 1959;49(6):786–92.

17. Le T, Reeder B, Thompson H, Demiris G. Health Providers' Perceptions of Novel Approaches to Visualizing Integrated Health Information. Methods Inf Med. 2013 May 7;52(3):250–8.

18. Le T, Reeder B, Yoo D, Aziz R, Thompson H, Demiris G. An Evaluation of Wellness Assessment Visualizations for Older Adults. Telemed E-Health. In Press;

19. Charness N, Demiris G, Krupinski E. Designing Telehealth for an Aging Population: A Human Factors Perspective. CRC Press; 2011. 114 p.

20. Redish J, Chisnell D. Designing web sites for older adults: A review of recent research. Retrieved June. 2004;9:2008.

21. Chisnell D, Redish J. Designing Web Sites for Older Adults: Expert Review of Usability for Older Adults at 50 Web Sites. AARP; 2004.

22. Demiris G, Finkelstein SM, Speedie SM. Considerations for the design of a Web-based clinical monitoring and educational system for elderly patients. J Am Med Inform Assoc JAMIA. 2001 Oct;8(5):468–72.

23. National Institute of Aging, National Library of Medicine. Making Your Web Site Senior Friendly. 2002.

24. Norman DA, Draper SW. User Centered System Design; New Perspectives on Human-Computer Interaction. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.; 1986.

25. Le T, Thompson HJ, Demiris G. A comparison of health visualization evaluation techniques with older adults. (Under review);

26. Buja A, McDonald JA, Michalak J, Stuetzle W. Interactive data visualization using focusing and linking. , IEEE Conference on Visualization, 1991 Visualization '91, Proceedings. 1991. p. 156–63, 419.

27. Martin B. Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Rockport Publishers; 2012. 208 p.

28. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health. 2010;33(1):77–84.

29. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health. 2000 Aug;23(4):334–40.

30. Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, et al. Development and Validation of a Short-Form, Rapid Estimate of Adult Literacy in Medicine: Med Care. 2007 Nov;45(11):1026–33.

31. Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. Visual Languages, 1996 Proceedings, IEEE Symposium on [Internet]. 1996 [cited 2013 Oct 23]. p. 336–43. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=545307

32. Strecher VJ, Others A. The Role of Self-Efficacy in Achieving Health Behavior Change. Health Educ Q. 1986 Jan;13(1):73–92.

33. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychol Rev. 1977;84(2):191–215.

34. Holloway A, Watson HE. Role of self-efficacy and behaviour change. Int J Nurs Pract. 2002 Apr;8(2):106–15.

35. Older Adults and Technology Use [Internet]. Pew Research Center; 2014 [cited 2014 Nov 10]. Available from: http://www.pewinternet.org/files/2014/04/PIP_Seniors-and-Tech-Use_040314.pdf

Appendix

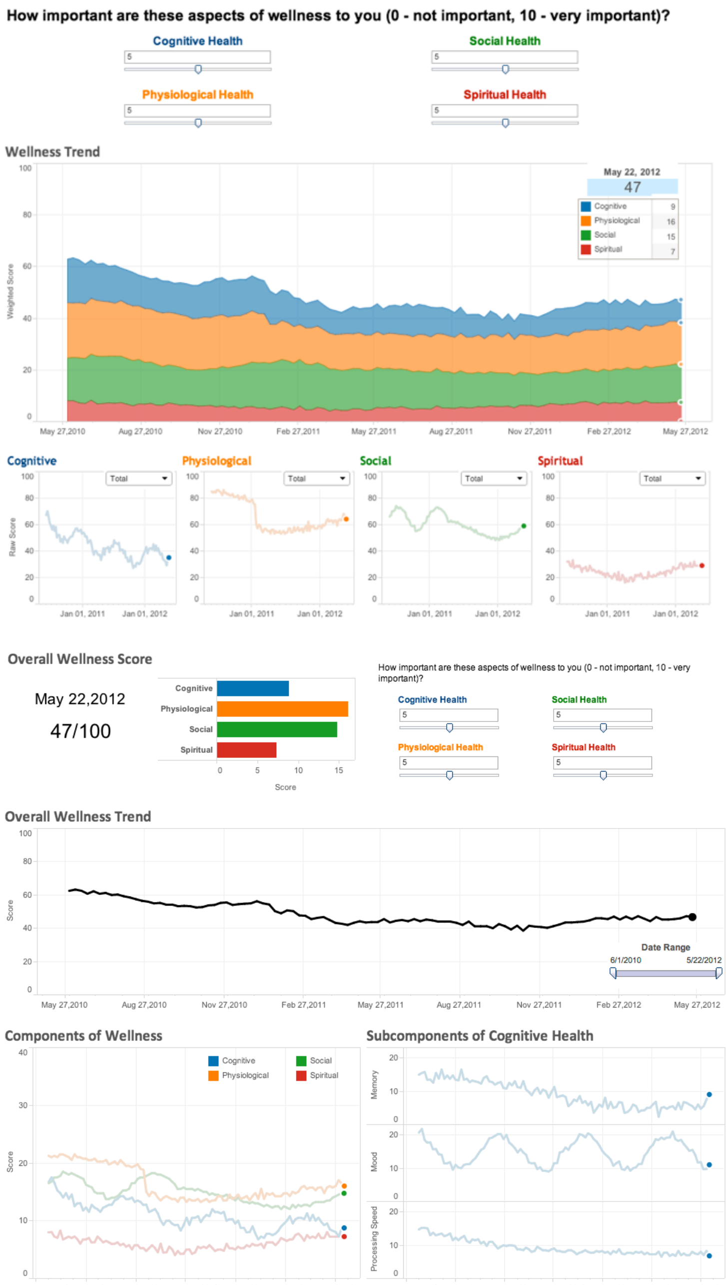

Figure 3: Visualization designs representing wellness information. Data are synthetically generated from a pilot study that collected data across dimensions of cognitive, physiological, social, and spiritual health using e-health technologies (2). The two interfaces represent the same data set though differ in organization, layout, and emphasis of information. Different interactions include zooming, brushing, linking, and filtering.

Conclusion

Through the course of the three papers, I have provided a holistic approach towards the design and evaluation of health visualizations for older adult stakeholders. This is an area of research that has been understudied within the existing literature. Though information technologies provide great potential to continuously collect and monitor health; the older adult perspective is often missing in the design of visualizations. As a result we are missing out on a potential opportunity to assist older adults in using this information to improve their health and maintain independence. Appropriately designed visualizations help translate the often abstract and dense data collected from health monitoring technologies into more relevant representations. For older adults, these visualizations can be empowering tools to not only manage their wellness but also to promote engagement within care across stakeholder groups such as family members and health care providers. Over the course of this research, I addressed three gaps in knowledge: 1) limited understanding of older adult graphical and health information needs, 2) unclear understanding of how visualization evaluation approaches compare to each other, and 3) lack of older adult specific health visualization guidelines.

In the first paper I addressed the limited research on graphical perception with older adults through a set of psychophysics experiments [4]. Though the findings in graphical perception have laid the foundation for how to build graphs effectively, much of this research has been done with a college or general population. Given that aging is associated with changes in visual acuity and processing speed, it is unclear if the same psychophysics findings would apply to an older adult population. To evaluate this I extended a study by Simkin and Hastie to allow for the comparison of results between older adults (at least 60 years old) and the general population [5]. The studies evaluated the effectiveness of three graphical display types (bar, stacked, and pie charts) on perception. Through online surveys, I created two experimental types: a set of comparison judgments and a set of proportion judgments. I sampled 50 participants across each age group (older adult, general population) and each experimental type (comparison judgment, proportion judgment). A comparison between age groups showed that older adults took longer to process graphical displays though this translated to an approximately equal level of accuracy. This indicates that the effectiveness of graphical elements to display information accurately is independent of age; however aging related changes in processing speed impact the amount of time required to make these judgments. Consistent with both Simkin and Hastie's original study [5] and Cleveland and McGill's work on graphical perception [6], I found that the bar chart remains the most effective taking into account accuracy and speed. This work in graphical perception provided me with a foundation for developing interactive visualizations. Leveraging these findings along with feedback from older adult and health care provider focus groups [7,8], I created visualizations that consisted predominantly of bar and line graphs.

In the second paper, I provided a comparison of different approaches towards evaluating health visualizations. In particular I evaluated the visualizations using a benchmark, insight, and summative questionnaire approach. In the benchmark evaluation, I gave participants a set of 5 tasks and measured the time and accuracy of completion. I found that there was high variability in task completion time (even for a within participant study design), limiting my ability to detect differences between visualizations. I found that selection of the tasks was the primary factor influencing accuracy, masking any potential differences between visualizations. Given the strong dependence on task type, a benchmark evaluation would be more appropriate for visualizations where specific optimizations were required to facilitate a comparison. For the insight evaluation, I asked participants to openly explore the visualizations while verbalizing their experience. I later reviewed the sessions and coded for the generation of insights. Comparing between visualizations, I found differences in the number of insights generated and also specific characteristics of the insights. This approach provided a more naturalistic evaluation of the visualizations; better aligning with the objectives of the visualizations as exploratory tools. By categorizing and comparing characteristics of insights, I also gained a finer understanding of differences between visualizations. However, the insight evaluation approach is balanced by the heavy investment in time required by the researcher to review sessions and to code for insights. For the summative questionnaire, I administered the System Usability Scale to measure perceived usability of the visualization [9]. This detected moderate differences between visualizations that aligned with qualitative feedback from participants over the course of the evaluation. The SUS scale was quick and simple to administer though the raw numerical values provided limited understanding as to what contributed to the score. Overall the three methods focused on different aspects of the evaluation: the benchmark looks at efficiency of visualizations, the insight looks at efficacy, and the summative questionnaire looks at satisfaction. Each of these dimensions can be important for comparing visualizations; selecting the best approach depends on aligning it with the intended objectives of the visualization.

In the third paper, I took the results from the evaluation sessions and synthesized them to identify factors that contributed to the utility of health visualizations. This applied a qualitative analysis approach where I first coded at a high level for references to use of health visualizations. I then conducted an affinity mapping exercise with the assistance of two other researchers. This was a bottom up approach that grouped the excerpts into higher-level thematic categories. From this analysis I found five key factors that impacted use of health visualizations: 1) ability to intervene based on trends in the visualization, 2) need for contextual information to complement the visualization, 3) perceived limitations due to computer literacy, 4) preference for longitudinal visualizations, and 5) differences in how the visualizations would be shared. Based on these thematic findings derived from the evaluation sessions, I also provided a set of recommendations generalizable outside of my specific visualizations. These are points of consideration to assist other researchers as they embark on the design of health visualizations for older adults.

The three papers presented here described the design and evaluation of health visualizations for older adults. There are certain areas in which this work can be further expanded. As currently constructed, the visualizations were created through a simulated data set based on a pilot study. The next step would involve conducting a field study with older adults in which live smart home and health monitoring data are collected and visualized. This is an important extension that personalizes the visualization work within the participant's natural environment. Having a live and personalized context for the visualization will shed further insight into how older adults use the visualizations. This also naturally extends to a study that examines the older adult, family caregiver, and health care provider interactions when the visualization is shared. In particular, how does the visualization impact communication amongst these stakeholders and in what ways does it support older adult independence? Though the visualization work described here was within the context of health monitoring and smart home data, this could be extended to other data sources. Integrating the visualizations with electronic health record data or connected health devices can provide a holistic view of the older adult's health and wellness. Alternatively, the visualizations generated from the older adult's home environment could be embedded within the electronic health record, serving as an integrated bridge between the health care provider and older adult. The information from the visualizations would a continuous resource to remain updated on the older adult's wellness while serving as a common frame of reference for older adult - provider communication during clinical visits.

From this thesis work, I have presented an end-to-end process flow for the design and evaluation of health visualizations. I focused on the older adult as consumers of the information, representing a different set of needs than other stakeholders. Over the course of this research, I contributed key findings to the literature. I presented the first work in graphical perception that examined older adults as a specific population group, showing that differences existed compared to the general population. I further provided a comparison of different techniques for evaluating health visualizations. I then identified key themes associated with how older adults would use health visualizations. From those themes I generated a set of recommendations for future researchers in this domain. There has been limited research on visualizing health information to older adults and my work sheds insights into this domain. Appropriately designed health visualizations serve as powerful resources to empower older adults within their care while also encouraging communication with other stakeholders. This work provides extensive contributions to the health visualization literature for older adults while also opening up further avenues of exploration for future research.

References

1. Wang S, Skubic M. Density map visualization from motion sensors for monitoring activity level. 2008 IET 4th International Conference on Intelligent Environments. 2008. p. 1–8.

2. Mynatt ED, Rowan J, Craighill S, Jacobs A. Digital family portraits: supporting peace of mind for extended family members. Proceedings of the SIGCHI conference on Human factors in computing systems [Internet]. 2001 [cited 2013 Oct 28]. p. 333–40. Available from: http://dl.acm.org/citation.cfm?id=365126

3. Jakkula V, Cook DJ. Mining Sensor Data in Smart Environment for Temporal Activity Prediction. Poster Sess ACM SIGKDD San Jose CA [Internet]. 2007 [cited 2013 Nov 27]; Available from: http://eecs.wsu.edu/~vjakkula/ACMKDSD.pdf

4. Le T, Aragon C, Thompson HJ, Demiris G. Elementary graphical perception for older adults: A comparison with the general population. Perception. 2014;43(11):1249–60.

5. Simkin D, Hastie R. An Information-Processing Analysis of Graph Perception. J Am Stat Assoc. 1987 Jun 1;82(398):454–65.

6. Cleveland WS, McGill R. An experiment in graphical perception. Int J Man-Mach Stud. 1986;25(5):491–500.

7. Le T, Reeder B, Yoo D, Aziz R, Thompson HJ, Demiris G. An Evaluation of Wellness Assessment Visualizations for Older Adults. Telemed J E-Health Off J Am Telemed Assoc. 2014 Nov 17;

8. Le T, Reeder B, Thompson H, Demiris G. Health Providers' Perceptions of Novel Approaches to Visualizing Integrated Health Information. Methods Inf Med. 2013 May 7;52(3):250–8.

9. Sauro J. Measuring Usability with the System Usability Scale (SUS) [Internet]. 2011 [cited 2014 Oct 29]. Available from: http://www.measuringu.com/sus.php