Characterizing Visualization Insights from Quantified Selfers' Personal Data Presentations

Eun Kyoung Choe ▪ *Pennsylvania State University*

Bongshin Lee ▪ *Microsoft Research*

m.c. schraefel ▪ *University of Southampton*

Data visualization and analytics research has great potential to empower people to improve their lives. Aiming to design a visualization system to help nonexperts gain and communicate their personal data insights, the authors examined the insights people draw from their personal data and how they communicate them.

As a result of advances in self-monitoring technology and the prevalence of low-cost monitoring sensors, we are witnessing the rise of the practice of self-tracking—also known as the quantified self (QS) movement. Through collection and analysis of their own data, quantified selfers (Q-Selfers) develop an awareness of their behaviors, draw meaningful inferences, identify unrecognized problems, and change their behaviors.[1] At QS meet-ups and conferences, many Q-Selfers share both their knowledge about self-tracking as well as their personal findings drawn from their own data. To share these insights effectively, Q-Selfers often use visualizations they created with their data. Data visualization and analytics research has great potential to empower people to improve their lives by leveraging this data.[2] Many Q-Selfers, however, are not visualization experts or data scientists (see earlier work for detailed profiles of Q-Selfers[1]). Thus, it can be challenging for novices to rapidly construct visualizations, translating questions into data attributes, designing visual mappings, and interpreting the visualizations.[3] Consequently, their visualizations are often not ideal for conveying their insights (such as creating a line chart with two lines to convey a correlation), and they sometimes present insights that might not be scientifically valid (for example, stretching the meaning of a result without statistical evidence).

Part of the challenges in effectively conveying insights from personal data is that the concept of insight is poorly understood in the personal data context. Despite several research efforts attempting to understand and define insights,[4–7] the concept of insight as framed in these studies is too general to operationalize into personal data visualization functions for novice users. Moreover, prior research in characterizing insight[6] and understanding the processes of gaining insights[7] were explored in the context of helping scientists and visualization experts. If we are to develop domain-specific visualization systems, we need to understand domain-specific insights that are of particular interest to the users. With such understanding, we can better design visualization systems that can effectively convey insights from personal data. Thus, we examined nonexperts' insights in personal data contexts with an aim to develop an interactive visualization system that helps with identifying and communicating personal insights.

In this article, we present the results of a predesign empirical study[8] to inform the design of systems for individuals to develop visualizations in personal data contexts. We examined what

Published by the IEEE Computer Society

insights people gain specifically from their data, how they use visualizations to communicate the insight, and how these findings can inform the design of systems that support identifying and communicating these personal insights. We explored these topics through the analysis of personal data presentation by Q-Selfers, who believe in the notion of self-knowledge through numbers and aim to draw meaning out of their personal data. The insights Q-Selfers report during their presentations are particularly relevant for this work because they represent the gist of the personal data collection and exploration. In addition, visualizations and visual annotations Q-Selfers use in their presentations can directly inform us of their practices of storytelling with data and challenges associated with it.

In what follows, we begin by situating our work with respect to existing literature on visualization insights and QS presentations and personal informatics. We then report on findings from analyzing 30 video recordings of QS presentations focusing on insight types and visualization usages for communicating the insight. Lastly, we identify research opportunities in designing and developing visualization systems to effectively support exploration and presentation of personal data.

Visualization Insights

> "The purpose of visualization is insight, not pictures."[9]

Helping people find valuable insights has been considered one of the main goals of visualization.[9] What an insight actually is, therefore, has been the subject of considerable research; several researchers have attempted to understand, characterize, and define insights and to use insight for evaluating visualization systems.

Chris North argued that we can measure a visualization's effectiveness by measuring its ability to support the process of insight generation.[6] In contrast to the traditional evaluation method of using controlled experiments on benchmark tasks, North suggested an open-ended protocol, having domain experts be experiment participants and verbalize their findings (insights). Visualization researchers in collaboration with independent domain experts then analyze the insights by developing insight categories through rigorous coding. Although the open-ended protocol is more time consuming than traditional experiments, it can provide rich understanding on what insights visualization users gained. In our work, we adopted the open-ended protocol to identify data-driven visualization insights through iterative coding in personal data contexts.

As a part of insight-based evaluation, Purvi Saraiya and her colleagues defined *insight* as "an individual observation about the data by the participant, a unit of discovery."[10] Using this definition, they separated participants' open-ended remarks into distinct insight units and grouped them into four insight categories. In their research on the specific context of participants' remarks about biological and microarray data, these four categories turned out to be overview (overall gene expression), patterns (identification or comparison across data attributes), groups (identification or comparison of groups of genes), and details (focused information about specific genes).[10] Using these insight categories, researchers evaluated the insight generation capabilities of visualization tools,[10] whereas we developed insight categories to inform the design of visualization tools that support both data exploration and presentation. In addition, we identified distinctive characteristics of insights nonexperts derive from personal data.

Part of the challenges in effectively conveying insights from personal data is that the concept of insight is poorly understood in the personal data context.

Visual exploration combined with an individual's mental model can lead to discovery of insights, which researchers refer to as *facts*—patterns, relationships, or anomalies extracted from data under analysis.[5] Yang Chen and his colleagues developed a fact taxonomy from a deductive approach of reviewing visualization taxonomy literature combined with a user study of classifying comments in Many Eyes (IBM's free data visualization website) and expert interviews.[5] The fact taxonomy consists of 12 categories: trend, compound fact, outliers, difference, association, extreme, meta fact, value/derived value, categories, cluster, distribution, and rank. It was constructed for general usage on multidimensional data independent of any application domains. We find the taxonomy a helpful starting point to characterize visual insights with personal data visualization. Although visualizations and comments from Many Eyes are generated by the general public, they are not necessarily insights about people's own data; people often comment on the visualizations created by other people using publicly available datasets. Furthermore, the fact taxonomy is based only on multidimensional data types. Our analysis of Q-Selfers' presentation videos with personal data covering various data types can be a good complement to this general-purpose taxonomy.

Huahai Yang and his colleagues examined types of visual insights people derive from simple and composite information graphics—in particular, bar and line graphs.[11] As a part of their investigation, the authors developed a task-oriented visual insight taxonomy. They identified four types of basic insights (read value, identify extrema, characterize distribution, and describe correlation) and four types of comparative insights (compare values, compare extrema, compare distribution, and compare correlation). Although we share a similar goal of understanding visual insight, our work differs in two regards. First, our work sheds light on understanding insights derived by people with significantly higher engagement levels. In fact, earlier work noted *degree of engagement* as one of the important factors to help people gain insight.[7] Therefore, insights generated by people with high engagement levels might be different from those by one-time survey participants. Second, our video analysis method allowed us to analyze insights derived from various chart types beyond bar and line charts.

QS Presentations and Personal Informatics

The role of insights is particularly relevant in the QS community, which structures its meet-ups and conferences specifically to convey data-driven insights via the presentation of an individual's tracking experience. Previous work revealed challenges Q-Selfers face in connecting their data with the problem they wish to explore.[1] These challenges include lack of scientific rigor, missing important contextual information, and trying to track more data than necessary, leading to tracking fatigue and thus incomplete datasets for effective analysis. In this work, we have been particularly interested in how Q-Selfers think about the data itself with an aim to gain insight into how tools might need to align with these understandings to be useful to the community. The QS presentations offer a unique opportunity to witness what amounts to live think-aloud sessions, showing how both experts and nonexperts think about data, as well as what kinds of conclusions Q-Selfers would like to be able to draw from the data they track.

This focus on how Q-Selfers think about their data complements work in QS specifically and personal informatics more broadly for and can aid in thinking about how to understand and construct information. For instance, Ian Li and his colleagues argued that tools need to support reflection in terms of discovery and maintenance.[12] Rather than considering a tool that does a tracking job, John Rooksby and his colleagues pulled back from specific tool design to consider the narratives one wants to tell with their data across various trackers.[13] Such narratives range from *directive stories*, where data reflection is oriented toward progress with a goal, to *documentary*, where one tracks data simply as part of an account, to the intriguing story of fetishizing data and tracking itself as a way of being. Here we push in the other direction to investigate instead how Q-Selfers, especially non-data experts, understand the data they are collecting and how they translate that into information about themselves.

Analysis of Q-Selfers' Personal Data Presentations

At the QS meet-ups and conferences, Q-Selfers present their self-tracking experiences with respect to what they did, how they did it, and what they learned. In this work, we are particularly interested in the last topic—what they learned—in which Q-Selfers often discuss insights gained from their data.

Dataset

QS meet-up and conference talks are often video recorded and uploaded to the QS blog (quantifiedself.com). Between April 2013 and June 2014, 79 QS meet-up and conference talks were uploaded to the blog. To analyze the most up-to-date landscape of QS practice, we limited our dataset to these recent video posts, although the blog archives date back to 2008. Of those recent video posts, not all of the videos fit within our research focus. For example, some speakers presented a new tool they developed without describing their actual self-tracking practice and others presented their findings without any visualization. Some videos failed to capture the presenters' visual aids (such as a slide deck), so it was difficult to understand the context of tracking.

Thus, similar to earlier work,[1] we used the following two inclusion criteria to determine what was added to our dataset: the presenter presented his or her own QS practices, and video posts included data visualizations of some kind created with the presenter's own data. Of the 79 video posts we reviewed, 55 met the inclusion criteria. In the interest of time, we analyzed five randomly chosen video posts at a time until no new categories emerged. In the end, we included 30 video posts as the final dataset. The average length of the videos was 7 minutes 41 seconds, ranging from 4 minutes 32 seconds to 13 minutes 22 seconds (excluding Q&As). We transcribed this entire corpus of videos to aid with our analysis.

Data-Driven Insight

Insights do not necessarily come from data. For example, as reported in earlier work,[1] some Q-Selfers report high-level lessons about the self-tracking itself. For the purpose of this work, we only looked for insights drawn directly from data and presented with data visualization evidence. We defined our unit of analysis as an individual observation about the data accompanied by visual evidence (such as a chart, table, or image). Each data-driven insight therefore contains one visualization and one or more types of insight descriptions. For example, Figure 1 shows an example data-driven insight with a corresponding visualization; it contains several insight types such as identifying peaks, reading values, identifying an outlier, and self-reflecting on contextual information to help interpret the visualization.

To separate data-driven insights from the rest of a QS presentation (which might include motivations and the process of self-tracking), we captured screenshots of data visualizations and documented quotes that described the visualizations.

That said, some presenters described high-level lessons learned from their self-tracking practice without presenting any visual evidence or showed visualizations but did not explain what they learned. According to our definition, those cases are not data-driven insights and have been excluded from our analysis. From the 30 presentations, we identified 185 data-driven insights—mean = 6.17, standard deviation = 2.91, ranging from 2 to 12—each of which contained a segmented quote and an accompanying visualization.

Analysis Method

Two researchers separately extracted data-driven insights from five randomly chosen videos to agree upon the unit of analysis and to create an initial coding scheme. They then independently coded data-driven insights from the next five randomly chosen videos and convened to discuss any disagreements or modifications to the existing coding scheme. The two researchers repeated this process until they reached data saturation and developed the final coding scheme. After the end of each meeting, one researcher took a pass on the previously coded dataset with the revised coding scheme to update the code. One researcher also analyzed the screenshots of visualizations by categorizing the type of visualizations (such as line chart or bar chart) and annotations (such as labels) used to communicate the insight.

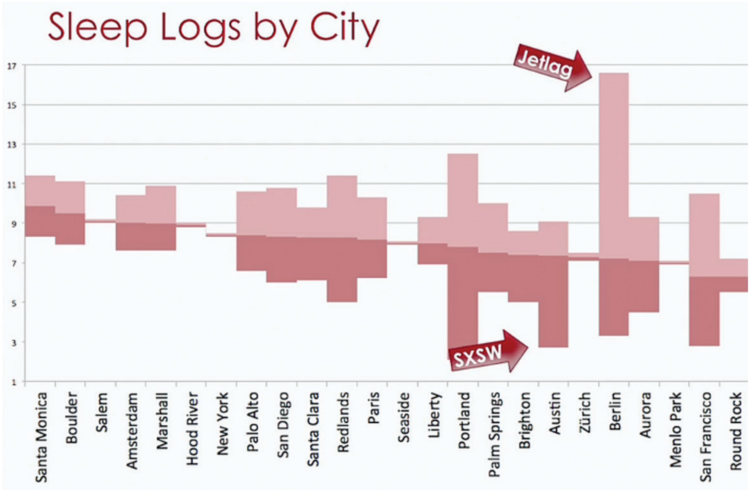


Figure 1. Example of data-driven insight with a corresponding visualization. Q-Selfer P51 explained the visualization with the following: "And the two peaks here. One of them was Berlin, where I was recovering from jetlag. That's almost 17 hours of sleep, which is a ton, and apparently it happened no other time. And in South by Southwest a couple of years ago in Austin, that was like basically a nap, two hours that night."

Types of Visualization Insight

From analyzing what people reported during their presentations, we identified eight data-driven insight types. Some types were further categorized into two to four subtypes. Table 1 summarizes the categories including type description, example quote, and the number of occurrences for each type. We present each type of insight in order of frequency.

Detail. Detail refers to specific data points that presenters find meaningful to report. They reported the measured value and reference usually with regard to extremes. These types are related to the *read value* and *find extremum* low-level tasks[14] and the *value/derived value* and *extreme categories*[5] noted by other researchers. In the P51's statements in Figure 1, for example, the presenter identified "17 hours" and "two hours" as "two peaks." Thus, we coded "17 hours" and "two hours" as identify values, whereas we coded the mention of "two peaks" as an identify extreme because it refers to maximum or minimum. Note that the presenter also pointed out Berlin and Austin, which we coded as identify references.

When communicating insights aided with visualizations, it is common for a presenter to point

Table 1. Types of visualization insights with description and example quotes.

Type	Frequency		Subtype	Description	Example quotes
Detail	137	62	Identify extreme	Explicitly state the identities of the data points possessing extreme values of the measure variable	"You can see that the first day and the last day were actually *the two hardest*." [P46]
		59	Identify value	Explicitly specify the measured value, its range for one or more clearly identified data points, or the difference between two measured values	"You'd see the values go from *about 77* up to *about 104*." [P9]
		16	Identify reference	Explicitly state the values of categorical variables, labels from the axes, or legends	"I had read books that were mostly by *men* and mostly by *white people*." [P32]
Self-reflection	95	66	External context	Uncaptured data provided by the presenter to understand and explain a phenomenon shown in the data	"For example, the day before my *vacation*, my heart rate was around 90 peaking at 110 at rest." [P36] "I didn't really feel any effects of the *calf injury* that I have been recovering from." [P55]
		15	Contradiction	Collected data contradicts existing knowledge	"I was also convinced before I looked at this data that I was reading equally between men and women, and the data actually *showed me that wasn't true at all*." [P32]
		7	Prediction	Predict the future based on the collected data	"If I stop going then I'll probably lose weight." [P61]
		7	Confirmation	Collected data confirms existing knowledge	"I see a bunch of things that are pretty consistent here and *that feel pretty right to me*." [P61]
Trend	67	67		Describe changes over time	"This is a plot showing just how much my triglycerides *have decreased*." [P37] "But when it comes to glucose, then the story was different, it was *impossible to see any trend*." [P17]
Comparison	64	37	By factor	Compare measured values by a factor (other than time)	"I was able to improve my *relaxation scores* every time I used this *focused breathing technique*." [P40]
		19	By time segmentation	Compare measured values segmented by time	"You can see on Thursdays and Fridays I actually tend to struggle a lot more." [P50]
		6	Against external data	Bringing in external data for comparison	"just to know where do I stand *in relation to other people*" [P50]
		2	Instances	Compare two specific instances	"You see actually two nights, one with little sleep, one with more sleep." [P6]
Correlation	21	21		Specify the direct relationship between two variables (but not as comparison)	"Anytime my sister's weight went down, her insulin levels went down. Anytime she gained weight, her insulin levels went up." [P17]
Data summary	17	17		Summary of collected data (such as number of data points and duration of tracking)	"I have completed about ***133*** books over *four years*." [P1] "This is *five years* of data." [P67]
Distribution	12	7	Variability	Explicitly state the variability of measured values	"Triglycerides move *all over the place*." [P37]
		5	By category	Explicitly describe the variation of measured values across all or most of the values of a categorical variable	"Overall, 2013 fall into work events, personal life, reading and writing, in approximately these percentages." [P1]
Outlier	4	4		Explicitly point out outliers or state the effect of outliers.	"Now, a quick look at this shows some pretty substantial outliers." [P10]

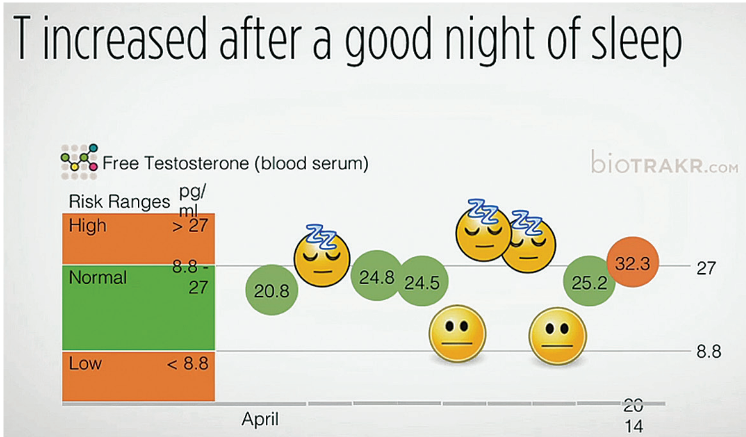
out the reference but not the associated value because the audience can read the value from the visualization as long as they know the reference. It is also common for presenters to specify extreme data points in their dataset, and as a result, we frequently observed combinations of identify value and identify extreme as well as combinations of identify reference and identify extreme. In the personal data context, identifying value can be applied broadly beyond conventional types of numerical or categorical data and visualizations. For example, both highlighting the trajectory of a running path on a map and accentuating one picture from a collection of pictures can be categorized as an identify value.

Self-reflection. Presenters shared insights gained from self-reflection on their own data. Self-reflection is a unique type of insight that is particularly relevant in the personal data context because it is based on highly personal information. Providing external contexts was the most frequent form of self-reflection. External contexts are uncaptured data, but they play a key role in understanding and explaining a phenomenon shown in the data. For example, P51 (in Figure 1) associated the longest sleep with jetlag and P36 (in Table 1) attributed the changes in his heart rate to a vacation. Presenters did not capture external contexts—such as jetlag and vacation—at the time of tracking main variables. Yet, they mentioned them during their talks and usually annotated the visualization with external contexts using text, shape, or other visual means to explain what was really going on at the time.

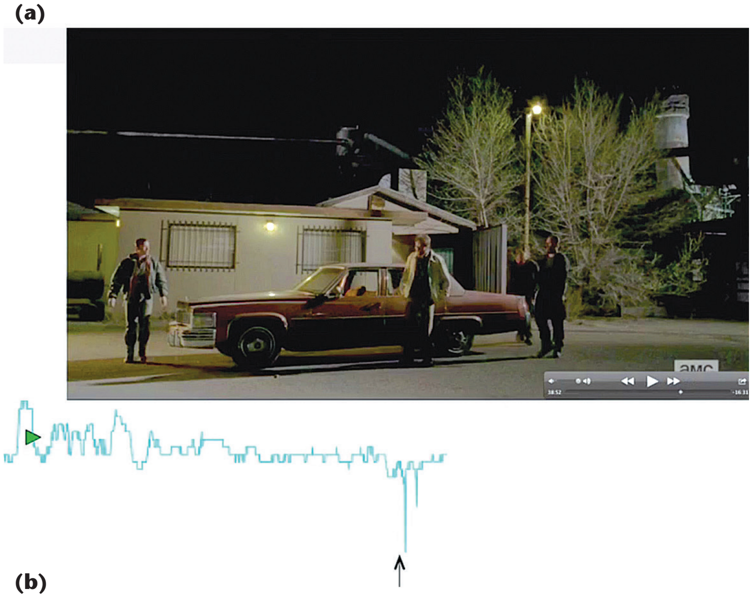
Other subcategories of self-reflection were confirmation of existing knowledge, contradiction (or disproof) of existing knowledge, and prediction of the future (at the time of self-reflection). For example, P42 described what seems normal (confirmation) and surprising (contradiction) to himself: "How much I stray from the moment is not a surprise to me; what was a surprise is the skew between how much I am lost in the future versus how much I am lost in the past."

Trend. Presenters frequently reported trends in terms of how data had changed over time. Given that most of Q-Selfers' personal data is time series, trend was a common insight type for Q-Selfers. Terminologies that presenters commonly used to describe trends include an increase or decrease and a fluctuation in the measured value over time. While less common, presenters also acknowledged having no clear changes over time, which we also considered as a trend. For example, P61 described what he found about his weight as follows: "I see a bunch of things that are pretty consistent here."

Comparison. Many Q-Selfers identified how various factors influenced their outcome behavior through comparison. For example, P40 (in Table 1) compared how various relaxation techniques affected his relaxation score. There are many ways to compare personal data: comparison by a factor



(a)



(b)

Figure 2. Sample visual annotations. (a) A scatterplot with color segmentation, value lines, and icons and (b) a line chart with an image and shapes.

of interest (such as a drug versus no drug), by time segmentation (for example, weekdays versus weekends), and against external data (such as my data versus my goal, social norm, or population data). Although time can be considered a factor, we created a separate category for comparison by time segmentation because all personal self-tracking data have time components.

Correlation. Comparison by a factor can sometimes be framed as correlation, and vice versa, but we identified correlation as a separate insight type when presenters specified clear associations between two variables. For example, P6 noted the correlation between sleep and testosterone: "What I found out was that the biggest differentiator, the biggest effect on my testosterone, was caused by sleeping. If I slept really well, if I slept more than eight hours, my testosterone the following morning was very high. The other way

Table 2. Types of visualizations ordered by their usage frequency.

Visualization	Frequency
Line chart	71
Bar chart	33
Scatterplot	23
Custom	21
Table	7
Dashboard	4
Pie chart	4
Stacked bar chart	4
Combined chart	3
Tally chart	3
Waterfall chart	3
Area chart	2
Calendar	2
Map	2
Matrix	2
Tag cloud	1

Table 3. Types of annotations ordered by their usage frequency.

Annotation	Frequency
Label	54
Shape	53
Trend line	32
Color/texture segmentation	27
Value line	17
Icon	12
Colored value	11
Image	9

around actually happen [sic] the same thing. So when I was really stressed, but I didn't really sleep much, my testosterone was low, indicated by these not-smiling smileys" (Figure 2a).

Other insight types include data summary, distribution, and outlier. Data summary refers to the number of data points and duration of tracking. For example, P1 stated, "I have completed about 133 books over four years." Distribution refers to the variability among the data or the categorical distribution of the data. Some presenters acknowledged the existence of high variability of measured values. They also went over the measured values in each level of the categorical variable. Lastly, presenters occasionally pointed out outliers in their data, as P25 commented, "As you can see, there are some interesting outliers here and there." In personal data contexts, some outliers are meaningful, and thus, presenters bring in external contexts to explain why they exist.

Visualization Usage

During their presentations, Q-Selfers frequently used visualizations to convey their data-driven insights. We analyzed types of visualizations with their usage frequency. We identified 16 visualization types (see Table 2). Line charts were by far the most frequently used, followed by bar charts, scatterplots, and custom visualizations (such as a grid of pictures or screen captures from a mobile app).

We also examined the visual annotations presenters used to convey their insight effectively. We identified eight annotation types (see Table 3). Labels and shapes (such as arrows and circles) were most commonly used, followed by trend line, color/texture segmentation, and value line (a line indicating a certain meaningful value). Sometimes, custom icons and images were used to enrich their presentations. For example, in Figure 2a, color segmentation and value lines were used to convey the normal range of testosterone level, and icons were used to represent sleep quality (Figure 2a). In Figure 2b, shapes (an arrow and triangle) and an image were used to explain a stress value drop.

Discussion

Our focus in this work has been to determine design requirements for how visualization systems can help non-data experts draw insights from their data. Based on our analysis of data-driven insights, we propose four areas for development of personal visualization systems: self-reflection, validity, communication, and annotation. We also reflect on the limitations of our method.

Self-Reflection as a Personal Insight

Previous research shows that even a simple form of self-monitoring feedback constitutes self-reflection and contributes to behavior change.[15] Q-Selfers' goal is not simply to quantify their behaviors but to reflect upon it, extract meaningful insights, and make positive changes.[1] From our analysis, we learned that self-reflection is one of the most common insight types that is often not shown in non-personal data contexts. Although self-reflection itself is an insight, it is often presented with other insights because self-reflection stems from one's attempt to understand why something interesting—a trend, comparison, or correlation—happened. For example, P45 tracked her running and reported the following data-driven insight: "While I was at Wharton my mileage was really low." She then verbally shared the external context that was not captured in her data to help the audience understand why her mileage was low: "part of that was because I was in the school, but part of it was my knee just wasn't getting any better." Although this type of self-reflection was an important insight, it was unclear when presenters self-reflected on their data—it might have happened during the data collection phase as well as the data analysis phase—or how they captured their self-reflection.

To fully leverage the benefits of self-reflection in behavior change, we see opportunities in designing personal visualization tools that encourage people's self-reflection. Personal data visualization systems can be designed to help people self-reflect on their data throughout various phases of self-tracking from data collection to data exploration. Furthermore, it would be helpful to support easy capture of these self-reflection insights—either qualitative or quantitative—against specific elements of a dataset that may be associated with the insight and access them both for exploration and for presentation purposes.

Gaining Valid Personal Insights

Although some Q-Selfers are technology savvy or have expertise in statistics, data analytics, or visualization, most of them do not have such data exploration and presentation skills. From the video analysis, we learned that Q-Selfers often report insights without conducting a statistical analysis. Among the 21 correlation insights we identified, statistical analysis was performed for only 10 cases (47.6 percent). Furthermore, we have no way of knowing whether even in these cases, the analyses were properly done. In addition, all 64 comparisons by a factor or by time segmentation were done "casually" (with eye measurement).

Critically, presenters did not seem to be aware that the differences they identified might not be statistically significant, hence technically showing no real effect. Furthermore, although many of them were essentially conducting a form of self-experimentation, the studies were not always properly designed. For example, the duration or the number of data points of the control condition did not match that of the test condition. Studies were not counterbalanced. Some presenters tended to point out things that they believed to be true without statistical evidence. Even though gaining scientifically valid insights might not be the main goal of self-tracking,[1] visualizations should not be used to reinforce bias. One way to reduce people's biases in interpreting their data is to incorporate basic statistical analysis for comparison and correlation into an interactive visualization system.

Communicating Personal Insights

Presenters mainly resorted to simple, standard chart types such as line and bar charts. Given that self-tracking data inherently is time-series data and that trend is one of the main insight types, it is not surprising that line chart is the most commonly used. However, more often than not, presenters created ineffective visualizations to convey their insights. For example, P26 used a line chart with dual y-axes—calories and temperatures—instead of a scatterplot to show the correlation between the two variables. Furthermore, we had to exclude a few potential data-driven insights because we could not comprehend the visualizations (even with their descriptions).

To help people choose the best visualization practice for a particular type of insight, we see opportunities in designing an automatic visualization suggestion engine that is similar to Show Me.[16] Although Show Me only accounts for properties of data such as data type (text, date, numeric, and so on), data role (measure or dimension), and data interpretation (discrete or continuous), insight type can also be another dimension for suggesting the best visualization for effectively communicating personal insights.

Visual Annotation for Highlighting Insights

The presentations in the video we analyzed were in-person, live presentations, where presenters could explain necessary external context during their presentation. Not surprisingly, presenters often used additional visual annotations such as a label with an arrow pointing to a specific visual object or region within a visualization (see Table 3). In addition, some presenters used custom icons (such as a smiley face or runner), images, and value lines to enrich their presentations. The basic visualization was usually created with standard tools (such as Microsoft Excel) and annotations were manually added to make the key point salient, meaningful, interpretable, and even enjoyable.

Visual annotations could play an even more significant role when people want to share their presentations asynchronously. Yet, despite the importance of visual annotations in conveying and highlighting insights, current visualization tools do not adequately support adding visual annotations. Therefore, visualization tools for creating basic charts equipped with an effective annotation capability could seamlessly support people's workflow in creating visualizations for the purpose of presentation.

We also observed that presenters mainly used static visualizations: they created charts with

Microsoft Excel or other statistical tools and then pasted them into a slide deck as images using standard presentation tools such as Microsoft PowerPoint and Apple Keynote. In the visualization community, telling stories with data has recently gained increasing attention, and some researchers specifically explored how to support more engaging, animated storytelling with data.[17] One promising avenue for future research is to explore ways to help presenters leverage interactivity, animations, and visual annotations to perform engaging presentations to attract and keep an audience's attention.

Limitations

We note that not all the QS meet-up and conference talks are recorded and uploaded to the QS blog. Therefore, the small number of videos we analyzed might not have accurately reflected the entire set of insight types people report. Another constraint with respect to our video analysis method was that we were unable to examine the process of gaining insights and creating visualizations because they were not portrayed in the videos. Moreover, because our unit of analysis was the data-driven insight that was communicated during the presentation phase (which occurs after the data exploration phase), it is possible that we might have missed some insight categories, especially the high-level insights detached from data or insights arisen during the data capture and data exploration phases.

Using other methods such as observations or interviews could help us better understand how and when people gain insights and how they create visualizations. Furthermore, presenters by no means represent the whole community of Q-Selfers or the general public. However, as a group of people who collect, analyze, and present data about themselves, Q-Selfers are in the vanguard of personal visualization and personal visual analytics, thereby providing invaluable insights toward understanding visualization insights in the personal data context.

As a first step toward building visualization systems to effectively support exploration and presentation of personal data, we studied what insights people gain with personal data and what visual annotations they use to communicate insights. Our work contributes to the fields of personal informatics and personal visualization research by identifying research opportunities in designing and developing visualization systems to support non-data experts to explore and present their data.

Specific areas for future research include supporting self-reflection process, helping nonexperts create and communicate with visualizations, and easing the process of annotation. To this end, we are creating a personal visualization platform that supports non-data experts to explore and create visualizations with their own data. We are examining how such a system can help people gain and communicate personal data insights we identified in this work.

References

1. E.K. Choe et al., "Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2014, pp. 1143–1152.
2. D. Huang et al., "Personal Visualization and Personal Visual Analytics," *IEEE Trans. Visualization and Computer Graphics*, vol. 21, no. 3, 2015, pp. 420–433.
3. L. Grammel, M. Tory, and M. Storey, "How Information Visualization Novices Construct Visualizations," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 6, 2010, pp. 943–952.
4. R. Chang et al., "Defining Insight for Visual Analytics," *IEEE Computer Graphics and Applications*, vol. 29, no. 2, 2009, pp. 14–17.
5. Y. Chen, J. Yang, and W. Ribarsky, "Toward Effective Insight Management in Visual Analytics Systems," *Proc. IEEE Pacific Visualization Symp.*, 2009, pp. 49–56.
6. C. North, "Toward Measuring Visualization Insight," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, 2006, pp. 6–9.
7. J.S. Yi et al., "Understanding and Characterizing Insights: How Do People Gain Insights Using Information Visualization?" *Proc. ACM Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (BELIV), 2008, article no. 4.
8. M. Brehmer et al., "Pre-design Empiricism for Information Visualization: Scenarios, Methods, and Challenges," *Proc. 5th ACM Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (BELIV), 2014, pp. 147–151.
9. S.K. Card, J. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, 1999.
10. P. Saraiya, C. North, and C. Duca, "An Insight-Based Methodology for Evaluating Bioinformatics Visualizations," *IEEE Trans. Visualization and Computer Graphics*, vol. 11, no. 4, 2005, pp. 443–456.
11. H. Yang, Y. Li, and M.X. Zhou, "Understand Users' Comprehension and Preferences for Composing Information Visualizations," *ACM Trans. Computer-Human Interaction*, vol. 21, no. 1, 2014, article no. 6.
12. I. Li, A. Dey, and J. Forlizzi, "Understanding My Data,

Myself: Supporting Self-Reflection with Ubicomp Technologies," *Proc. 13th Int'l ACM Conf. Ubiquitous Computing* (UbiComp), 2011, pp. 405–414.
13. J. Rooksby et al., "Personal Tracking as Lived Informatics," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2014, pp. 1163–1172.
14. R. Amar, J. Eagan, and J. Stasko, "Low-Level Components of Analytic Activity in Information Visualization," *Proc. IEEE Symp. Information Visualization* (INFOVIS), 2005, pp. 111–117.
15. A.E. Kazdin, "Reactive Self-Monitoring: The Effects of Response Desirability, Goal Setting, and Feedback," *J. Consulting and Clinical Psychology*, vol. 42, no. 5, 1974, pp. 704–716.
16. J. Mackinlay, P. Hanrahan, and C. Stolte, "Show Me: Automatic Presentation for Visual Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, 2007, pp. 1137–1144.
17. B. Lee, R.H. Kazi, and G. Smith, "SketchStory: Telling More Engaging Stories with Data through Freeform Sketching," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2416–2425.

Eun Kyoung Choe *is an assistant professor in Pennsylvania State University's College of Information Sciences and Technology. Her research interests include human-computer interaction, health informatics, and personal informatics. Choe has a PhD in information science from the University of Washington. Contact her at echoe@ist.psu.edu.*

Bongshin Lee *is a researcher in the neXus group at Microsoft Research. Her research interests include human-computer interaction, information visualization, and user interfaces and interaction techniques. Lee has a PhD in computer science from the University of Maryland at College Park. Contact her at bongshin@microsoft.com.*

m.c. schraefel *is a professor of computer science and human performance at the University of Southampton UK, where she holds a Royal Academy of Engineering Research Chair to model and evaluate interactive systems to support creativity, innovation, and discovery in science. She leads the human systems interaction design lab, exploring "wellth creation," looking at how to design interactive systems to best support health and well being as personal and social aspiration, rather than illness prevention. Contact her at @mcphoo, www.ecs.soton.ac.uk/~mc.*