The Future of Data Visualization in Personal Informatics Tools

Chloe Fan
MetroMile, Inc.
1870 Broadway St
Redwood City, CA 94063 USA
chloe@chloefan.com

CHI'13, April 27 – May 2, 2013, Paris, France.
ACM 978-1-XXXX-XXXX-X/XX/XX.

Abstract
In this position paper, I discuss data visualization trends in personal informatics tools using examples from research, consumer tools, my experiences in the Quantified Self (QS) community and companies, and challenges of visualizing data at MetroMile.

Author Keywords
Data visualization; personal informatics; challenges.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design.

Introduction
Data visualization is one of the main ways by which users of personal informatics tools make sense of their data in the reflection stage [3]. Academic research on visualizing personal data has focused on displaying the data in a way that can be understood quickly while engaging users (e.g. [1][2][4]). The challenge is finding meaningful and beautiful ways to visualize the data to capture user engagement and facilitate understanding of their behavioral patterns.

Quantified Self (QS) is a worldwide community of people who are interested in using personal informatics tools to track and make sense of personal data. Through interacting with members of the QS community since its first conference in 2010, talking to QS startups and projects such as BodyTrack (www.bodytrack.org), and working on data visualization solutions at MetroMile, I recognized a new common challenge in using data intelligence to find actionable insights from the raw data and communicating them in a meaningful way.

A Brief History

When personal informatics and QS were still in their early stages, users collected one stream of data, including any related metrics, for each tool they used. For example, a pedometer was used to track how many steps users took, but often, it would also track how many miles were covered, and even how many calories were burned, both derived from step count. The visualizations for this type of data were usually a number or a graphic showing the total over a specific time frame, usually a day. Academic research in this field produced visualizations showing these numbers in an engaging way, aiming to motivate people to walk more or adopt the tool for long-term use (e.g. [1][2][4]).

As self-tracking tools become outfitted with more advanced sensors and features, users are able to track multiple streams of data. Take, for example, BodyTrack's environmental base station (www.bodytrack.org). In addition to tracking multiple indoor air quality and environmental variables such as temperature and humidity, the BodyTrack website also allows users to combine streams of data into one view. For instance, users can see their physical activity alongside the weather, their sleep quality, and heart rate if they choose to. One example insight users could gain from using this tool is that their physical activity performance is affected by their sleep quality, which is affected by their bedroom's air quality. Users could use this information to improve the air quality in their room in order to sleep and subsequently, exercise better. This contextual data paints a richer picture of external factors that affect their behavior or other internal factors that are affected by their behavior, bringing users one step closer to changing it.

However, it is not easy to discover this information just by visualizing streams of raw data. Visualizing these multiple streams of data in an engaging way is not enough to highlight the important relationships and change behavior. With hundreds of data points interacting with each other, users need help finding the patterns that matter, to know what to pay attention to and what to change. Even before visualizing the data, data intelligence is needed to find the relevant insights to show, which then need to be visualized in a way that effectively communicates what the user should do with the information.

Data Visualization at MetroMile

I will describe my data visualization work at MetroMile (www.metromile.com) to illustrate the process of thinking through these challenges. To help car owners save money, MetroMile provides per-mile car insurance and data transparency regarding how much, where, and how users are driving. The data is gathered using a GPS sensor plugged into the car's on board diagnostic (OBD) port, which tracks time, location, and mileage information.

Miles driven for the month compared to budgeted miles for the month.

Daily miles driven, with average, highest, and lowest day.

Trip maps including date, start and end time, distance, and duration.

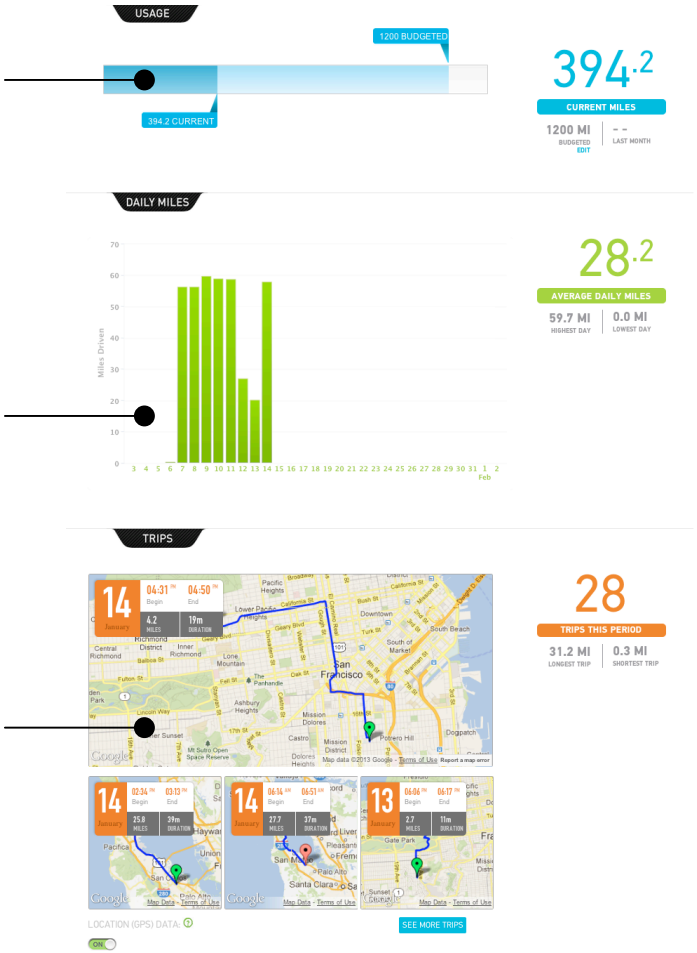

Figure 1. Example of what users see on their online MetroMile dashboards.

Figure 1 shows how we visualize this data with actual data from a sample user. The dashboard shows the most relevant information to users using simple graphs and large numbers to highlight how many miles users have driven that month, how many miles they drove on average per day, and how many trips they made.

Each of these three sections is effectively one stream of data with relevant data surrounding them. As we collect other types of data, the initial impulse is to ask, "How can we show more data in the same space?" In other words, how can we optimize the display of data? Figures 2 and 3 illustrate my redesign of the monthly mileage data and the trips data, respectively, with the goal of optimizing display space in mind.

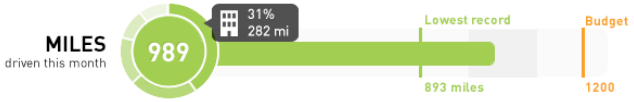

Figure 2. My redesign of the monthly mileage data showing total miles driven, breakdown of miles by trip category, lowest record, average miles driven (darker grey area), and budgeted miles for that month.

Figure 3. My redesign of the trips data as a list view instead of a grid view. It shows how many miles were driven in a day as a number and as a horizontal bar chart, intraday mileage data, duration, estimated cost of gas, mileage, and a trip detail icon.

When I tested these charts internally with three other employees, the main concern was that the charts were overwhelming and forced the viewer to separate the components and figure out what were the useful

takeaways. The reactions were along the lines of "Okay, that looks nice. Now what?" signifying that form factor was there, but not the substance.

In brainstorming types of data that could be useful for car drivers who want to save money and be environmentally friendly, we came up with ideas that were derived from the raw data that I had been trying to visualize: showing users ways to improve their fuel efficiency based on how they drive, suggesting alternative routes, modes of transportations, or gas stations that can save them money, and alerting them to driving patterns that are better for their car and for fellow drivers on the road. These are examples of what data intelligence could provide over just visualizing the raw data. Now that we have these rich data streams, the challenge becomes finding the right balance between visualizing raw data along with the insights from that data, so users quickly and easily understand how to act on their data, while being able to search for deeper information should they choose to.

Workshop Discussion

At the workshop, I hope to hear from other attendees their challenges and insights from working with data visualization, whether for personal or research purposes. As an academic community, I would like us to discuss methods for creating and evaluating the impact of personal data visualizations, especially for multiple streams of data, and also discuss where others see the future of personal data visualization.

Conclusion

As personal data becomes more multilayered and complex, the cognitive overload of analyzing visualizations increases to the point where users cannot keep track of all the data that is shown at once while determining what it is they are supposed to take away from the data. By changing the focus of visualizing personal data from visualizing in an engaging and optimized way, to also visualizing in the simplest and most relevant way, we can help users be more efficient, engaged, and enlightened in understanding their data.

Acknowledgements

Many thanks go to the QS community for facilitating discussions about data visualization and to individuals with whom I have had conversations about data visualization. I thank BodyTrack for allowing me to be a beta tester, and my team at MetroMile for continuously pushing for simple and meaningful personal data visualization.

References

[1] Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., et al. (2008). Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 54–63). ACM.

[2] Fan, C., Forlizzi, J., Dey, A. 2012. A Spark Of Activity: Exploring Information Art As Visualization For Physical Activity. *Proceedings of Ubiquitous Computing*. Ubicomp '10. ACM.

[3] Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 557–566). ACM.

[4] Lin, J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. *UbiComp 2006: Ubiquitous Computing*, 261–278.