International Workshop on Visual Analytics (2011)
S. Miksch and G. Santucci (Editors)

Exploring Complex Mobile Life through Lightweight Visualizations

Tanyoung Kim [1], Jan Blom [2] and John Stasko [1]

[1]Georgia Institute of Technology, Atlanta, GA, USA
[2]Nokia Research Center, Lausanne, Switzerland

Abstract

Smart phones can be used to collect various types of data from a user spanning call and text message logs to GPS coordinates, and to Bluetooth detection. These data are indicative of behavior across various socio-temporal-spatial contexts. For the purpose of inspecting the expected and discovering the unexpected from the data, we developed lightweight visualizations with a low cost. The interactive visualizations are customized with each data analysis task and evolved to support emerging diminutive inquiries during the inspection with themselves. By employing this analytic method at the early phase of data-related research projects, the insights can form a ground for subsequent product development activities focusing on engineering, design and market research.

Categories and Subject Descriptors (according to ACM CCS): H.2.8.h [Information Interfaces and Presentation]: Miscellaneous—

1. Introduction

In a paradigm that has become known as people-centric sensing [CEL*08], sensors in mobile phones and other wireless devices are used to collect large quantities of continuous data about their users. The data include various types of behaviors, from intra- to inter-personal level. Lausanne Data Collection Campaign [KBD*10] is a project of this kind, for which Nokia N95 smart phones were allocated to a heterogeneous sample of nearly 170 participants from Lake Geneva region and have been used over a period of one year. The data collection software runs on the background of the phones in a non-intrusive manner, yielding continuous data about the usage of the device and the spatio-temporal-social contexts of the behavior. The logged data are automatically uploaded every day to a database server and anonymized before being accessible to researchers. By the beginning of October 2010, 178 million entries were collected in the database, including 9M GPS entries, 83K unique GSM (Global Systems for Mobile communications) cell towers, 141K unique Bluetooth addresses, 16M Bluetooth encounters, 327K calls, and 146K text messages [ABKN11].

To explore the collected mobile communication data, we developed lightweight visualizations, which are distinguished from many other existing visual analytic systems mainly for their low-cost development and adaptability to various inquiries. Our analytic tools are employed at the early state of data-related research for system development or service design.

Starting with a discussion on the exploratory data analysis, we articulate the advantages of our visualizations over other analytic systems and the development process. We present a case study to show how interactive visualizations can be used to explore complex data sensed from mobile phones and to find insights about the users' behavior. Our contributions are:

- A case study exploring a dataset of mobile phone usage, geographic logs, and social interaction data.
- A novel method for developing and customizing visualizations for diverse analysis tasks with low cost.
- Examples of the benefits of an iterative development process, which readers can adapt to their own exploratory data analysis.

2. Exploratory data analysis

The voluminous data collected certainly has the potential as a valuable resource for understanding the complex and heterogeneous mobile life. By exploring the data, we aimed to find interesting facts about the mobile users and insights about their life.

DOI: 10.2312/PE/EuroVAST/EuroVA11/037-040

Our approach is explained as *exploratory data analysis* opposed to hypothesis testing [Shn02]. We did not set clear hypotheses that were useful for validating objective relationships and reducing generalized result s from the data. Instead, we started with open-ended questions such as "what is a participant X's daily mobile phone usage like?" and then searched for interesting patterns. In addition to some initial questions and conjectures, we hoped to discover other unexpected insights. Thus, our goal was to extract such useful outcomes focusing on quantity rather than quality, even if they were more miscellaneous than the results from a robust statistical data analysis. In addition, we aspired to derive insights from the data very rapidly.

To achieve our goal–finding quantity-oriented insights from data in a quick way–, the use of interactive visualization is a more promising method than hypothesis testing that aims to discover objective facts. In contrast to a hypotehsis-based statistical investigation, visual presentation assists by revealing trends, highlighting outliers, showing clusters, and exposing gaps [Shn02]. Bearing the strength of visual presentation in mind, we developed interactive visualizations and used them as analytic tools.

3. Lightweight visual analytic tools

3.1. Characteristics

To develop our tools we used Processing [Pro] and Adobe Flash. These programs are relatively easy to learn, support cross-platform development, and export outcomes to web applications that enable remote and collaborative analysis. Many open libraries (e.g., Prefuse/FLARE [Fla]) and APIs (e.g., GoogleMaps for Flash [Goo]) also strengthen the adaptability of these programs. A single researcher without significant programming experience developed one of the visualizations in a few days.

Because of this rapid development process, the visualization is certainly not a perfect visual design, it exhibits some usability issues, and it even has some non-critical bugs. However, it still provided analysts with basic interactive features such as zooming, filtering, showing details on demand, and navigation within the dataset Thus the visualizations function as discovery tools customized for the specific analysis tasks.

Our lightweight analytic tools are also versatile; changing several lines of codes can modify the visual representation. Interactive parts of one visualization are also easily duplicated in others. Datasets for visual representation are usually linked as external files, so it is effortless to change them or navigate within the datasets via the interactive menus.

3.2. Comparison with other systems

Existing visual analytic systems are often developed as stand-alone software that combines interactive information visualization and full-featured user interfaces for statistical analysis. Thus, it is obvious that developing such systems takes tremendous time and cost. Furthermore, despite the high cost, many are tailored to specific data types or purposes such as text documents [DZG*], social network [PS] and geographical data [AAW07].

Besides the tools for professional analysts, web-based freeware visualization applications such as Many Eyes [Man] and Tableau Public [Tab] have been widely used by journalists, educators, and even the general public. They allow people to use their own datasets, to visualize them, and to change the visualization formats. The generated visualizations are even interactive, thus basic data analysis is also possible.

However, as a trade off with the universal support, they often can be limiting to users; they permit only a single dataset in an organized spreadsheet per one visualization; a new dimension cannot be generated while processing the data; the data format cannot be converted to another. While making the visualizations described here, we were able to mine the loosely formatted raw data and generate a new dataset much faster than making a well-formatted spreadsheet beforehand. Each data entry is stored with a UNIX timestamp in the database, which can be trimmed and sorted more easily by the algorithm in the visualizations. Finally, although general tools support various types of visualizations, they do not open chances for users to create their own visualization designs. In contrast, our visualizations were initially created with the given tasks in mind, but were versatile enough to encompass multiple unrefined datasets and complex demands for visual tuning.

4. Case Studies: an individual's 24 hour

Because the data was so large and varied, we did not build a system to visualize the entire data collection at once. Instead, we used simple database queries to extract smaller, connected subsets of the data and we stored these datasets in spreadsheets and as comma-separated value files for further manipulation and use by the visualizations. Here we present a case study about representing a single participants' data. It is one example from a number of visualizations we created.

4.1. The data

In a table labeled "recordsâĂİ of our database, each entry is tagged with a unique record index, the corresponding participant's code, and a UNIX time stamp. The record index is linked to other tables such as call logs, GSM cells, GPS coordinates, music player usage, camera application, and Bluetooth detection. Thus, mining the data of a single participant from those multiple tables could give us extensive insight about device usage across everyday situations. The visualization allowed us to observe her patterns of mobile phone usage, locations, and social interaction with other

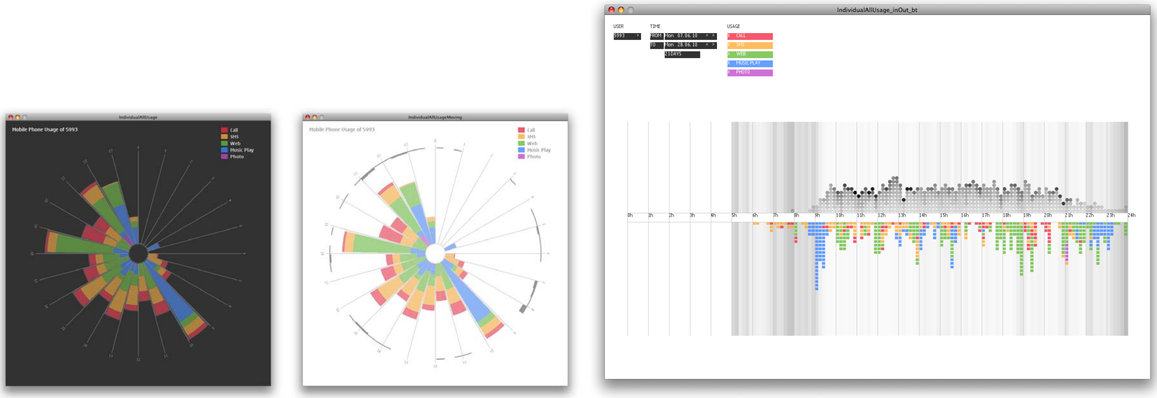

Figure 1: *(a) An hour-by-hour logs of mobile phone usage activities; (b) Representation of data in (a), integrated with GSM cell tower changes; (c) Integrated visualization of usage logs, GPS-based moving status, and Bluetooth encounters with other people*

people. Moreover, by employing demographic information from the pre-campaign survey, we were able to infer information about her general lifestyle and the context that formed any behavioral characteristics.

4.2. Evolution of the visualization

Visualizing usage data: With a user's 16-week (from Jun 7, 2010 to Sep 26, 2010) mobile usage data–making/receiving phone calls, making/receiving short messages, browsing websites, playing music, and taking photos–, we first wanted to find hour-based usage patterns (Fig. 1-(a)). We created a visualization in the form of a 24-hour clock. The five kinds of usage data are coded in different colors, and the number of arcs in each color represents the number of activities that occurred during the hour. The examination allowed us to raise follow-up questions and to formulate hypotheses regarding patterns and exceptions to those patterns. For instance, we observed that this person had browsed websites more frequently during a specific hour in the morning, which we inferred could be their commuting time.

Integrating location information: To investigate the new inference, we added another dataset, GSM cells. We assumed that changes in GSM cell towers signified location changes. Physical movement was represented as grey arcs around the circle: thicker arcs represent greater movement. However, the GSM cell tower data did not exhibit a clear pattern throughout a day, so we concluded the dataset was not sufficient to infer the movement (Fig. 1-(b)). Instead of GSM cell tower data, we used GPS coordinates.

Integrating Bluetooth detection and modifying the form of visualization: Through further development, we integrated Bluetooth detection data into the system. The data was used to infer the number of people detected in the proximity of the participant at any given time, as well as frequency of proximity encounters with any given user. We changed the visualization from a circle to a timeline, because it is more appropriate for the detailed presentation of time-based data and the scalability to cover multiple datasets. We processed the GPS and Bluetooth data in ten-minute intervals and represented it as the background of the timeline and the dots respectively. The darker the background was, the more GPS entries appeared. The number of GPS entries is roughly equivalent to the frequency of physical movement. The number of dots represents nearby people detected through Bluetooth, and the darkness of each dot is proportional to the frequency of the corresponding person's emergence (Fig. 1-(c)). We added simple menus that enable the analyst to select other participants and to turn on and off the categories of mobile phone usage. We also included the time selector, so the users can adjust the time range of the data as any number of days within the 16 weeks.

4.3. Insights of an individual's life

The subsequent version of the visualization provided more information and insights about the user, who identified himself as a full-time worker through the pre-campaign survey. There were more people with BlueTooth on around him during the typical working hours, which might indicate that he was present in a high tech office environment. He also exhibited a rough commuting time range between 7h30 to 9h. The darker two dots representing Bluetooth detection between 13h10 to 13h30, may be indicative of a repeatedly occurring lunch in a densely populated environment such as canteen. The current version did not indicate whether the two dots represent the same person yet, which should be something communicated in the next iteration of the visualization.

We also examined a visualization of another participant, who was a college student and had different patterns of

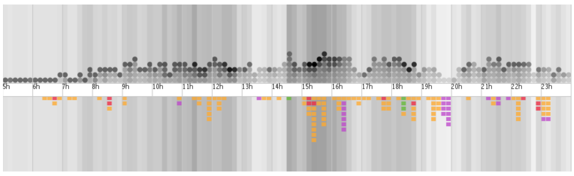

Figure 2: *A part of visualization of a college student*

movement and Bluetooth detection (Fig. 2). The student did not exhibit a fixed commuting pattern; further analysis of GPS data showed the places where he randomly walked around. He did not have a regular peak time in terms of the number of people nearby. He encountered less people but was around them more frequently than the office worker (i.e. less number of entire dots, but relatively more dark dots). The mobile usage pattern is also different; for instance, the student used SMS more than voice calls.

5. Discussion

This case study focused on a single user from a microscopic perspective. We also developed other visualizations for exploration from other perspectives. Some examples are multi-layered social networks presenting physical and virtual connections, temporal-geographical visualizations using an open map API, and mutuality of SMS communication. The main implication of analysis with these lightweight visualizations is that the findings can be the rationale for further consumer research, social-geographic research, network and mobility modeling, and privacy and security research.

Traditional market research can be used to assess trends of regional or age groups through quantitative methods such as surveys. In contrast to these cumulative summaries of consumers, the visual analysis of user-generated data here showed new insights about mobile users from a microscopic perspective. Visualizations can be used to detect multiple patterns of mobile phone usage depending on diverse temporal or geographical contexts. Through such exploratory analysis, we can formulate several hypotheses about behavior and social interactions of individuals studied. Later these hypotheses can be validated with statistical analysis or ethnographic research.

In addition to the benefits from the exploratory analysis where visualizations play a role as a supportive tool, the interactive visualizations themselves can be a useful source for the future product design. For instance, the case study in this paper, a visualization of a single user, could be a prototype user interface for a mobile application for personal information management.

6. Conclusions

Mobile phones can be used as viable measurement instruments to collect data pertaining to diverse usage behavior as well as temporal and geographical contexts. The data have opened opportunities to find interesting phenomena that could be the foundations for ubiquitous and smart mobile systems. The lightweight visual analytic tools we suggested in this paper stimulate analysts' intuition, ability of observation, and instinct of discovery. In doing so, they support the exploratory analysis that transforms data to meaningful information and insights.

References

[AAW07] Andrienko G., Andrienko N., Wrobel S.: Visual analytics tools for analysis of movement data. *SIGKDD Explor. Newsl. 9* (December 2007), 38–46. 2

[ABKN11] Aad I., Blom J., Kiukkonen N., Niemi V.: Nrc lausanne data collection campaign. In *INRIA Privacy Workshop* (Grenoble, 2011). 1

[CEL*08] Campbell A. T., Eisenman S. B., Lane N. D., Miluzzo E., Peterson R. A., Lu H., Zheng X., Musolesi M., Fodor K., Ahn G.-S.: The rise of people-centric sensing. *IEEE Internet Computing 12* (2008), 12–21. 1

[DZG*] Don A., Zheleva E., Gregory M., Tarkan S., Auvil L., Clement T., Shneiderman B., Plaisant C.: Discovering interesting usage patterns in text collections: integrating text mining with visualization. In *Proc. of CIKM '07*. 2

[Fla] prefuse: interactive information visualization toolkit. http://prefuse.org/. 2

[Goo] Google maps api for flash. http://code.google.com/apis/maps/documentation/flash/. 2

[KBD*10] Kiukkonen N., Blom J., Dousse O., Gatica-Perez D., Laurila J.: Towards rich mobile phone datasets: Lausanne data collection campaign. In *Proc. of ACM Int. Conf. on Pervasive Services (ICPS)* (2010). 1

[Man] Many eyes. http://www-958.ibm.com/software/data/cognos/manyeyes/. 2

[Pro] Processing.org. http://processing.org/. 2

[PS] Perer A., Shneiderman B.: Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proc. of CHI '08*. 2

[Shn02] Shneiderman B.: Inventing discovery tools: combining information visualization with data mining. *Information Visualization 1* (March 2002), 5–12. 2

[Tab] Tableau public. http://www.tableausoftware.com/public/. 2