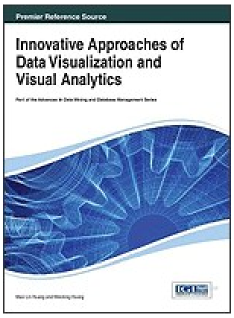

Innovative Approaches of Data Visualization and Visual Analytics

by Mao Lin Huang and Weidong Huang (eds)
IGI Global. (c) 2014. Copying Prohibited.

Chapter 1: Aesthetics in Data Visualization—Case Studies and Design Issues

Heekyoung Jung,
University of Cincinnati
USA

Tanyoung Kim,
Georgia Institute of Technology
USA

Yang Yang

Cathal Gurrin,
Dublin City University
Ireland

Luis Carli,
University of São Paulo
Brazil

Marco Carnesecchi,
Università della Valle d'Aosta & Università di Siena
Italy

Antonio Rizzo,
Università di Siena
Italy

ABSTRACT

Data visualization has been one of the major interests among interaction designers thanks to the recent advances of visualization authoring tools. Using such tools including programming languages with Graphics APIs, websites with chart topologies, and open source libraries and component models, interaction designers can more effectively create data visualization harnessing their prototyping skills and aesthetic sensibility. However, there still exist technical and methodological challenges for interaction designers in jumping into the scene. In this article, the authors introduce five case studies of data visualization that highlight different design aspects and issues of the visualization process. The authors also discuss the new roles of designers in this interdisciplinary field and the ways of utilizing, as well as enhancing, visualization tools for the better support of designers.

INTRODUCTION: DATA VISUALIZATION DESIGN

Nowadays we are flooded with data of diverse kinds due to the increasing computational capability and accessibility. Specifically, in addition to public data available on the Internet (e.g., census, demographics, environmental data), data pertaining personal daily activities are now more easily collected, for example, through mobile devices that can log people's running distances and time or their manual record of nutrition consumption. Due to such expanded sources of data, there appear new applications that involve data collection, visualization, exploration, and distribution in daily contexts. These applications do not only display static information but also let users navigate the data in forms of interactive visualizations. This emerging trend has brought both opportunities and challenges to interaction designers to develop new approaches to designing data-based applications.

Conveying information has been one of main functions of graphic and communication design since the analog printing era. The focus of information design is the communicative and aesthetic presentation of structured data as in an example of subway route map. In treating the increasing volumes of unprocessed data accessible either from public media or personal devices, the approaches of information design are now more diverse with the influence of other disciplines (Pousman & Stasko, 2007). Specifically, unlike information design in a traditional and confined manner, data visualization starts with the data that did not proceed through structuring and exist often in large volumes and complicated formats (Manovich, 2010). Thus, data visualization requires designers to obtain diverse knowledge and skill sets in addition to their visual aesthetic senses. The new requirements include human visual perception and cognition, statistics, and computational data mining.

Moreover, as data visualization has been more broadly applied to end-user services, interaction and experiential qualities need to be more critically considered. Those qualities of data application do not only rely on the usability of data perception or task-based navigation but also build up on aesthetics that affords engaging and exploratory data navigation. The latter have been remained as a less investigated area than the former due to the strong disciplinary tradition of data visualization in computer science and cognitive science.

Furthermore, visualizations are presented, used, and shared in diverse contexts from science labs to online journalism sites, to personal mobile devices. This means that data visualization has become a truly interdisciplinary field of research and practice by weaving informatics, programming, graphic design, and even media art (Vande Moere & Purchase, 2011).

In what follows, we overview current design approaches and tools for data visualization, then introduce five case studies for further discussion of design issues, process and tools in regard to aesthetics in data visualization.

CURRENT TOOLS AND PROBLEMS IN AESTHETICS OF DATA VISUALIZATION

In recent years visualization scientists, mostly from computer science, have suggested many visualization-authoring tools with a hope of expanding the creators of visualizations and the contexts of their use. We suppose these tools are largely categorized into three kinds—1.) a standalone programming language and its Integrated Development Environment (IDE) such as Adobe Flash ActionScript (Adobe) and Processing (Processing 2), 2.) online or installation-required programs that provide visualizations of given chart topologies, such as ManyEyes (IBM) and Tableau Public(Tableau, 2013), and 3.) libraries, toolkits or component model architecture integrated with existing programming language, such as d3.js (Bostock, 2012) and gRaphaël.js(Baranovsky) for web documents.

These tools certainly open new spaces in which designers can apply visualization techniques with less effort and can exert their aesthetic expressions. However, throughout the entire process from data acquisition to visualization, there appear challenging aspects for the specific goal—*the aesthetic and interactive qualities* of visualization, and for the specific group of authors—*interaction designers*.

Online tools or visualization applications provide a set of chart templates as a means of presenting complex data into perceivable information. However, the subtle variations of graphic and interactive design attributes are not fully considered in the existing visualization tools, resulting in excluding designers who wants to have the full freedom of aesthetics and expressiveness. Thus, aesthetic consideration is limited to selecting colors or symbols at a later phase of visualization. Visualization libraries, which are provided and shared by altruistic and enthusiastic visualization experts, have expanded interaction designers' job. However, the initial learning curves include the experiences of programming to some extent. Regardless of old-school graphic designers, interaction designers who do not have extensive computer science knowledge may feel difficulty in first facing the libraries and toolkits. The lack of the fundamental knowledge and tactical coding tips in the mother programing language and the library may results in the abandon of aesthetic expression.

In sum, the tools have limitations in aesthetic expression due to either the lack of expressive freedom or the requirements of computer science and programming knowledge. The more significant issue is that the aesthetic concern pertains to the mere surface of visualization, in other words, "making pretty appearance" with the previously processed and well-formatted dataset. We acknowledge that the visually pleasing appearance is a critical aspect of data visualization, especially when it comes to the job of graphic and interaction designers. However we argue that the aesthetic consideration of data visualization goes far beyond that, covering the wider process from the process of *obtaining* and *organizing* data, to *composing* and *narrating* meaningful messages from the given data, and *distributing* the visualizations for the audience's access (Segel & Heer, 2010). In this sense, when we discuss aesthetics of data visualization in this article, it is not only about the look and feel of data representation. Instead, aesthetics should be approached from this holistic perspective. Considering the situations and aesthetics of use is a significant part of the emerging research and practice agendas in interaction and experience design. However, we argue that in the field of data visualization, the concept of aesthetics is still remained as a look of visualization techniques.

Here we introduce the issues of designing visualizations focusing on the aesthetic values throughout the entire process. Unlike general interface design projects, data visualization may not be directly simulated in wireframe forms without a broader picture of how data are actually collected and organized. For example, multiple datasets can be layered in one frame through different modes. In this case, interactive and temporal attributes such as animation and transition become critical aspects, which should result in guiding users' data navigation in easier manners. However, current prototyping approaches are limited in fully supporting such dynamic navigation to some extent. On one hand, designers have used print-based mock-ups of static visualization images; they can produce these with Microsoft Excel, which is limited in demonstrating interactive aspects of data visualization. On the other hand, a programming-b approach can build high fidelity prototypes with interactivity by loading actual data. However, apparently it demands designers too much time and effort to learn a new skillset of coding. Even though designers are willing to learn programming, they often get exhausted in choosing appropriate languages and tools for their particular project; they are initially expected to understand the strengths, shortcomings, and compatibilities of all available tools. Due to these constraints of existing tools and process, interaction designers have not been fully involved in data visualization research and practices although they have much potential to contribute with their *storytelling ability*, *aesthetic sensibility,* and *logical thinking ability*.

CASE STUDIES OF AESTHETICS IN DATA VISUALIZATION

Motivated by the problems and constraints with the current design tools and approaches for data visualization, we propose three issues for designing aesthetic data visualization: data gathering, data representation, and data navigation. Through the discussion of these issues, we hope to explore better ways to utilize and improve visualization tools for designers when they exert their tacit knowledge and aesthetic sensibility. In what follows, we introduce five case studies of design projects in which data visualization plays a core part. Each project is described and analyzed according to the three issues:

- **Data Collection:** How did the designer access and collect the data?
- **Data Representation:** Why did the designer choose the functional forms and the aesthetic styles to represent the collected data?
- **Data Navigation:** How did the designer make the visualizations interactive to navigate the data?

In addition to these issues, we also discuss the process of each case and the tools used in it.

- **Design Process:** In which order did the designer consider and execute the three issues (data collection, representation, and navigation)? How did each step of design process mutually influence with one another?
- **Design Tools:** What visualization tools (libraries or languages) did the designer use? How did the tools influence each step of design process in both positive and negative ways?

Finally, we summarize all case studies by reflecting on challenges and accomplishments from them.

Case Study 1: Visual Representations of Online Banking Transactions

This project is to support online banking customers' financial managements of their income and expenditure through interactive data visualization, which is alternative to typical monthly tabular report. Web-based tools (i.e., JavaScript libraries, Adobe Flash) enable various types of interactive data visualization. This project especially aims to design prototypes of different visual representations—including different layouts, navigation structure and interfaces—according to different customer motivations. The designs of visualizations were evaluated based on real life user scenarios and iteratively refined to a final prototype.

Data Collection and Representation

We first analysed the current visual representations of transaction data offered to customers on the online banking service. It is currently displayed in a table of four columns: 1) date of payment, 2) description of income or expenditure item, 3) amount value and 4) value date. Each row represents a single occurrence of transaction and it is possible to sort their order by clicking on the header of each column. The current online representation allows customers only a little more interaction than paper-based report. In our redesign we focused on exploring more interactive data navigation and manipulation afforded by web technology. In Tufte's words this means letting users understand the data by *having it represented* instead of reading its analysis (Tufte, 2001). To explore possible visual representations we collected income and expenditure reports over the period of a year from around 100 anonymous users of the service who had consented to share their data. Specific research questions include: which type of visualization is most appropriate to different customer profiles? How to support customers reviewing payment history, comparing expenditures in different categories, and eventually managing their income and expenditure from their previous report?

The first prototype was a paper-based sketch (in Figure 1) only with a few changes from the existing table report: 1) income and expenditure are presented in separate columns, 2) the time interval can be specified in input sections and calendar pickers, 3) multiple accounts marked with different colours can be switched by different tabs. The significant change in this version from the tabular report is the separation between income and expenditure and a colour scale to represent different accounts. Figure 2 shows another representation of the transaction data: there is a spatial and chromatic separation between income and expenditure items. The two sliced segments of the inner circle represent the total amount of income (green) and expenditure (red). User can retrieve details of a selected transaction by clicking each of the outer arcs. The angle of each arc is proportional to the amount of money. This visual form allows comparative representation of single transactions and the overall balance at the same time.

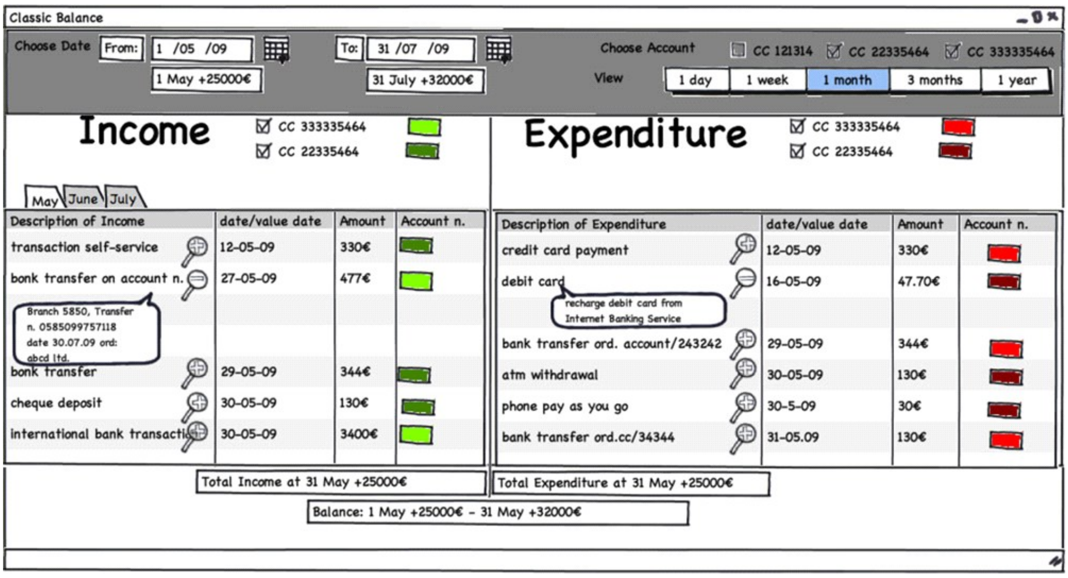

Figure 1: The classic balance view

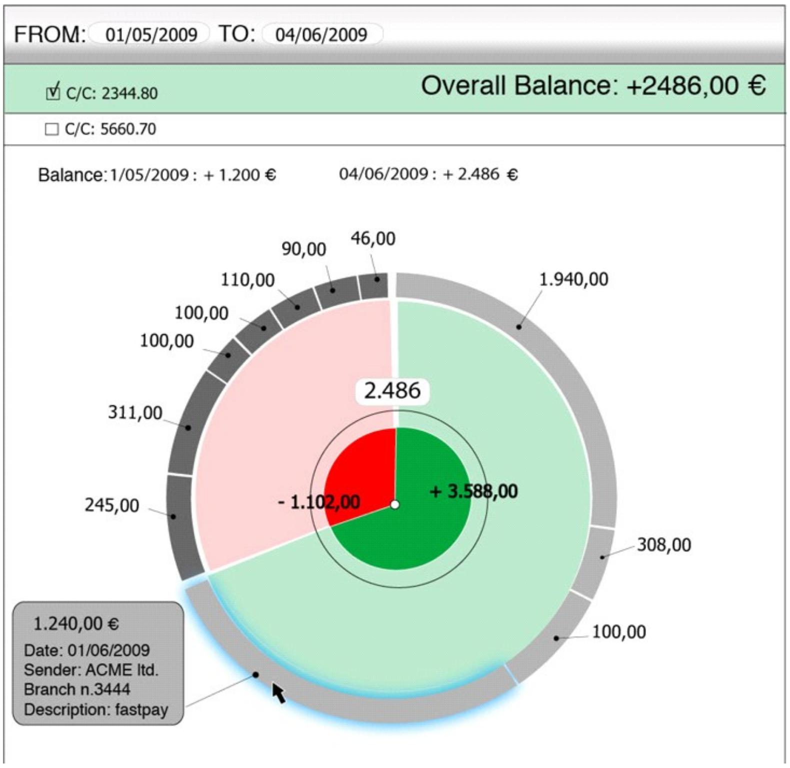

Figure 2: A clockwise view

Data Navigation (Embedding Prototypes in Real Life Scenario)

We mapped the sample balance data onto the initial prototypes above, but it was difficult to evaluate actual use in terms of interactive navigation without an activity scenario. Therefore we redesigned the online banking service page with more different visual representations of the sample data (Figure 3). Then we came up with tasks 1) to retrieve a particular payment transaction to a fictional travel agency from the report and 2) to make a new payment to the same recipient. The steps for each task consists of accessing the payments archive by choosing one of the visual representations on the main widget. Once the users access the visual report they have to find the particular transaction either by scrolling through the whole history of the month or by querying the name of the agency that received the payment in the search box.

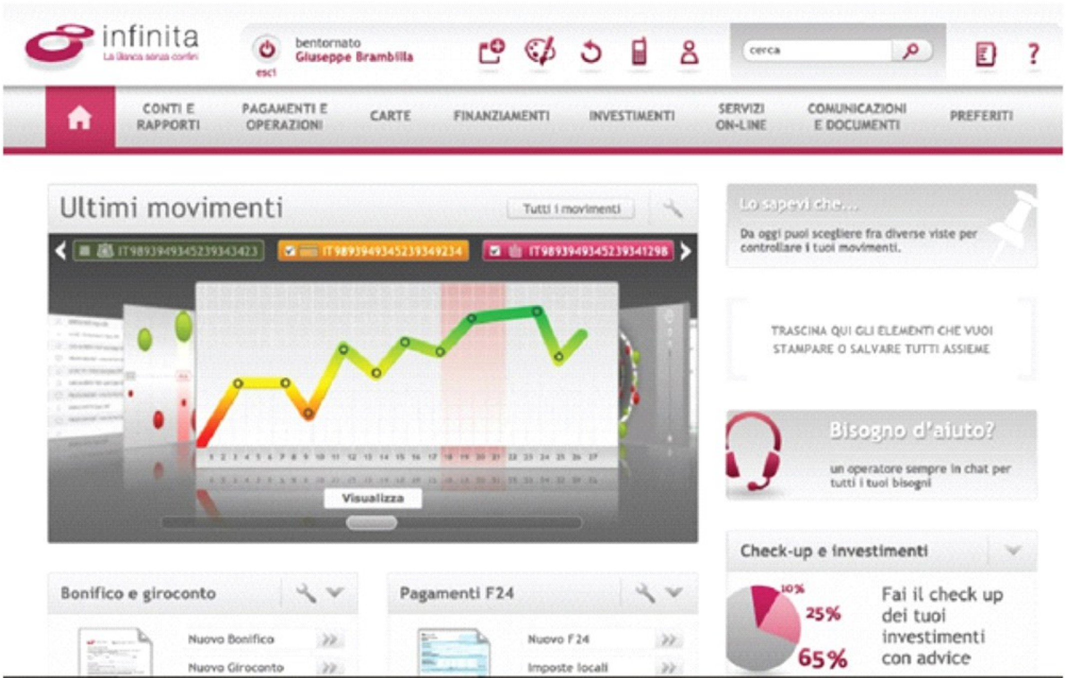

Figure 3: The mock-up of the web page of the banking service

Design Process and Tools

The prototypes presented above were designed using Balsamiq (Balsamiq Studios, 2013) mock-ups for the static visual representations and

Adobe Flash for the interactive visualization. The animations for the scenario were created using Actionscript. The choice of graphic-oriented tool instead of data visualization libraries had an important impact on the project with more freedom of choice in terms of encoding techniques and faster iterations of various visual representations. However, forms and styles of data representation are not only determined by the structure of data or design tools but also closely related to user goals for effective financial management, especially scenario-based iterative process. By using scenario-based process, design concepts and navigation steps were specified through quick and easy prototyping test. Our initial sketches (Figure 1) were barely distinguished from a typical tabular report except for a spatial and a colour separation between income and expenditure items. However, the following steps were to design prototypes that are drastically different from a table such as in the examples in Figures 2, 3 and 4. At first these new visual representations did put users in a situation where they had to explore the visualization and interpret what the visual elements mean. They then expressed their judgment on the aesthetics of the representation while trying to retrieve information and perform tasks according to given scenarios. After both the most and least similar prototypes to the original table were tested on users, we implemented them into web service in order to test visual representations as a part of the service ecosystem where individual transaction data is collected and retrieved. According to Distributed Cognition theories, the users can benefit from visual elements since external representations allow people to perceive data much faster and to cognitively process longer than they can in their mind internally (Kirsh, 2010). Users' performance in visual search can be more accurate with aesthetically aligned layouts, which is coherent with recent studies (Salimun, Purchase, Simmons, & Brewster, 2010). These assumptions are confirmed in the test we run on the web service UI. In fact, users preferred a rich and aesthetically attractive visualization (Figure 5) to one that they consider familiar (Figure 1) when the former is more functional than the latter to complete the task they are engaged in.

Figure 4: Comparing two bank accounts

Figure 5: Retrieving an expense using search

Case Study 2: Visualization of Weather Changes over Years in Multiple Cities

This project is the visualization of the weather records over five years in nine cities. To support time-based comparison we came up with a unified viewing mode, which overlays macro patterns of temperature changes from multiple years and still allows the access to details of daily data from the overview (Figure 6).

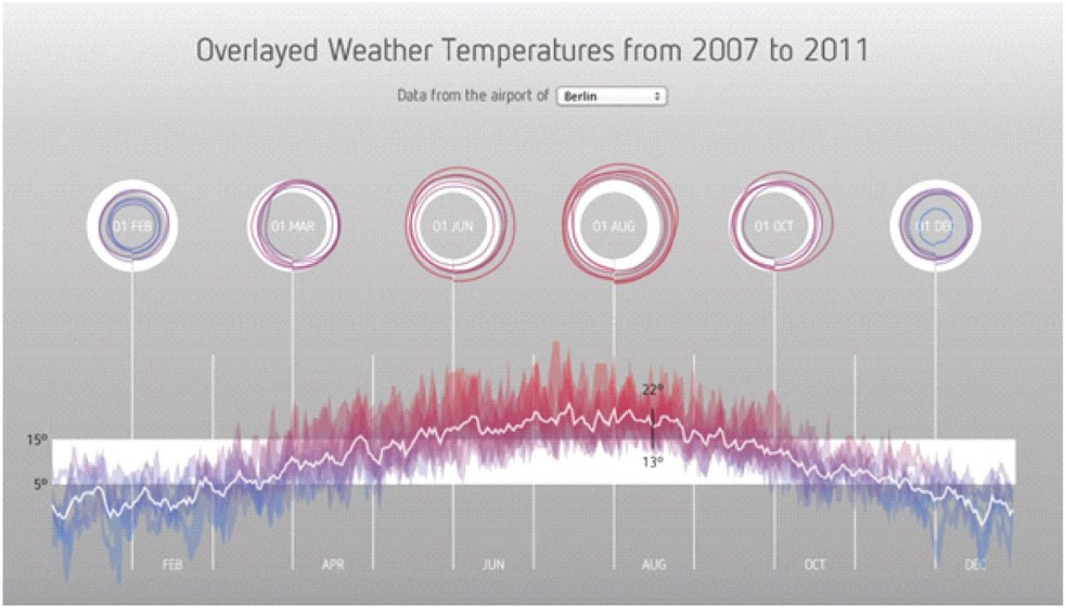

Figure 6: The overview of the weather visualization

Data Collection: Daily Data of All Years and Hourly Data of the Selected Few Days

We collect the weather records from the Weather Underground (2013) that provides the past five years' hourly-basis temperature data of the major capital cities. A simple URL returns the data that can be saved in CSV file format (Weather Undergound, 2007). Then we made a custom JavaScript to automate the process of accessing, downloading and aggregating the data over the targeting duration and places. Unfortunately, the site blocked this repetitive URL request, so this process was not completely automatically performed. Due to this constrains, we had to reconsider the volume of the dataset; finally we decided to collect day-by-day information from 2007 to 2011. Additionally, we collected hour-by-hour information only for six days in each year.

Data Representation: Timeline for Yearly View and Radials for Daily Data

The view of visualization is divided largely into two parts. One part is for the yearly view where the datasets of daily lowest and highest temperature from each year is plotted on one timeline of one year. The five datasets, each of which means a single year, are overlaid on the timeline. The other part is the daily view where the datasets of the six selected days from each year is plotted in a radial form. Each closed curves around the circle represents the hourly temperature change of the day. Same as the overlaid timeline view, the each of five closed curves represent one dataset of a day's hourly temperature change.

In the yearly view timeline, the area between the highest and lowest temperature is filled with a gradient set of colors to provide a sense to scale the differences of temperature in one dataset of one city, as well as within the entire datasets of all cities. The strokes of the radial day view are also painted with the same gradient scheme. Vertical grids that create the divisions between the months are also used to make connections between the yearly view and the daily view above by showing the position of the highlighted date on the yearly timeline. On the *Y* axis, we set a temperature range and filled it with white with the text of the two temperatures. This area works as a "base zone" of temperature, which helps users investigate the bounds of temperature change. This zone is also displayed as a white circle in each daily view above the year view; the inner bound is set on the minimal temperature, the outer bound on the maximum temperature. In general, our design requirements follow the idea of good "data-ink ratio" on the visualization (Tufte, 2011).

Data Navigation: Switching Datasets and Details on Demand

Our goal was to add interaction to the cases in which displaying visual elements as data representation or labels would make the visualization somewhat illegible. We included four possible interactive features in the visualization: 1) a conventional HTML select menu to change the city, 2) mouse hovering on the year view to display detailed information as text (the mean value of the maximum temperatures of the last five years, and the one of the minimum temperatures), 3) mouse hovering on the five circles of day view to display the time of the day at the point of mouse cursor, and the average temperature of the last five year's pointed time, and 4) dragging the white basis zone in year view, which prompts the change of radius and thickness of the associated circles in the day views.

Design Process and Tools

We started by sketching several possible techniques of data representation with pen and paper. Sketching allowed us envision the visualization forms and the layout. However, hand-drawing sketches are not appropriate to imagine the look when the data are actually populated. Thus, based on the initial forms from the sketches, we started coding and plotted a small portion of real data. We developed this visualization using SVG, JavaScript and the d3.js library (Bostock, Ogievetsky, & Heer 2011). The methods and abstractions that the d3.js provides are easy to use to encode data into graphic elements. We first code the viewing modules separately, which made testing different sizes and positions of the modules easier when we arranged them.

To seek the optimal forms and visually pleasing styles for the visualization, we kept iterating the code with the small parts of the real data. For the cases that we wanted to test various forms, we made the code more parametric, by creating and connecting variables through mathematical operations. (Figure 7 & 8). For developing the interaction we followed a very similar process to the one used for defining the forms and styles of the visualization: sketching, coding, and testing.

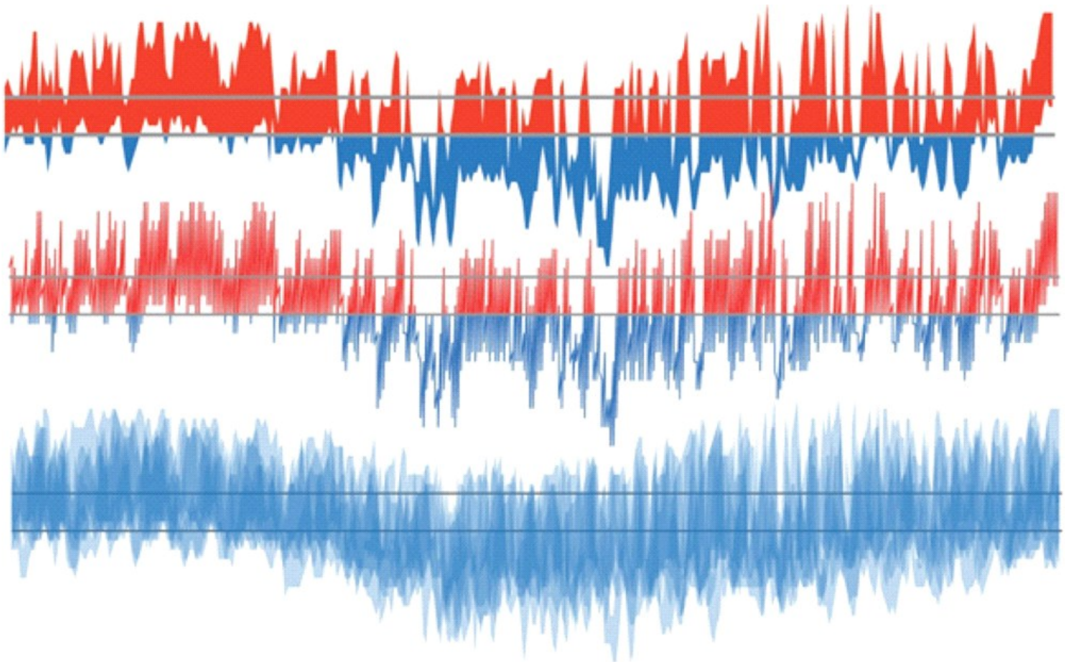

Figure 7: Some variations developed during the iteration of the year view

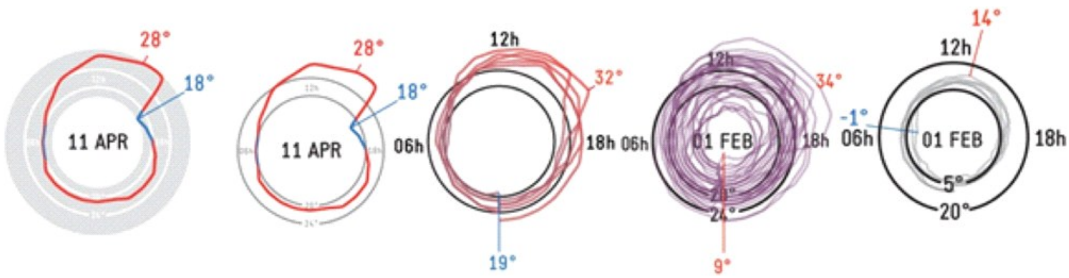

Figure 8: Some variations developed during the iteration of the day view

Case Study 3: Visualization of Daily Nutrition Consumption

This project is to support browsing and recording one's daily nutrition consumption through interactive data visualization and application. Especially with a focus on mobile health management, this project explored design solutions for small screen based browsing and recording nutritional facts of food items. Mobile devices can support convenient recording and monitoring nutrient intake at any time, which can be critical in health management (Andersson, Rosenqvist, & Sahrawi, 2007). However, there are still challenges in visualizing nutrition entries within a small screen user interface, such as displaying a large amount of data, interactively switching modes of contextual and local data, among others. At the same time, every food consists of various nutrients such as carbohydrates, proteins, and fats, which are hard to be understood, particularly when presented using only numbers. In consideration of these issues, we focused on finding metaphors for cognitive and embodied visual form in order to facilitate small touch screen information visualization and navigation.

Data Visualization

We first surveyed existing mobile and web nutrition management applications to understand specific design issues and the functional requirements in terms of menu structure or interaction modes. Nutrition facts of different foods are hard to be read and kept track of, particularly, when presented as numbers in a table. There is a need for exploring different ways of displaying a large amount of complex information based on the understanding of a person's cognitive processes. Based on the survey of existing applications, we learned that "time" is a pivotal element which people use to record and navigate their daily nutrition consumption. Then we sketched various types of wireframes to visualize nutrition intake over a timeline, for example, in line graphs, bar graphs, bubble charts and pie charts. A set of design requirements are specified below in terms menu structure and navigation of the application:

- Support tracing all food items taken in a day and comparison of nutrition intakes for recommended levels.
- Use food item icons instead of text for quick review and intuitive understanding.
- Provide rich information using preattentive visual elements such as colors and symbols for quick overview.
- Display two modes of information—overall food entries and specific nutrition components for each food, and support dynamic navigation between the modes.

Data Navigation

By analyzing selected applications, we were aware of the lack of motivation to record all food consumed and the cumbersomeness of browsing the data. This means that ease of use or efficiency of use is critical, but at the same time such design criteria are not enough. Some extra values are required to motivate and constantly engage people with a new type of application. Based on the objectives and requirements discussed above, we designed Food Watch, a mobile application for browsing and recording nutritional facts of food items. Food Watch visualizes nutritional composition of different food items (e.g. carbohydrates, proteins, fats) in pictorial elements and pie charts for intuitive perception and navigation of information in a small screen interface. The circular shape in the center of the display was selected with two different metaphors in mind: a wristwatch and a dinner plate (Figure 9).

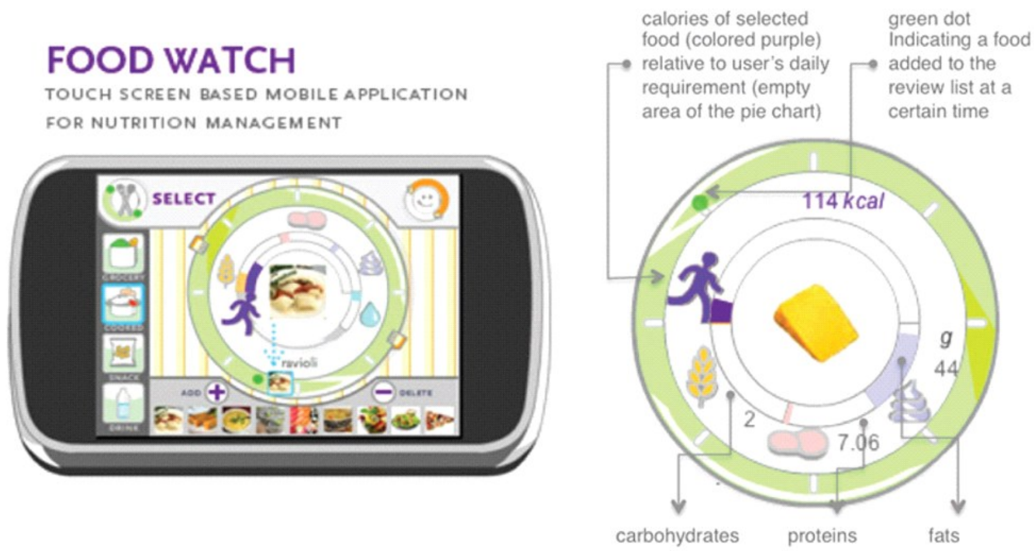

Figure 9: Food Watch Display Object and Visual Elements

Specifically, a plate metaphor is considered appropriate not only in terms of its everyday use for serving food, but also in that its image can be applied as display object (Ware, 2000) to embed pie charts of nutrition facts over its round shape. At the same time, the wristwatch metaphor provides a conceptual relation for recording and browsing daily nutrition intake as time-related information with its round shape. In this way, a set of food items can be browsed by turning a graphic plate and food can be dragged onto the plate for selection. Then, daily nutrition intakes can be recorded and browsed over the timeline of a graphic wristwatch in a different mode of interaction.

The interaction of the application consists of the three modes as specified below (Figure 10):

- Set up a user's profile by calculating personal daily nutrition requirements according to one's body factors.
- Browse different foods comparing their nutrition facts to the daily nutrition requirements calculated previously.
- Review all the foods added to the review list and their nutrition facts accumulated in a day.

Figure 10: Food Watch Interaction: 1) Set Up Profile (top left), 2) Browse Food Items (top right), 3) Select A Food Item (bottom left), and 4) Review Saved Food Items (bottom right)

Data Collection

The data was collected after we had made initial decisions on the overall design direction in terms of visualization form, interaction, and layout. The food database is built as a local xml file based on nutrition information collected from USDA Nutrition Data Laboratory (USDA, 2011). By using a local database we more focused on experimenting graphic and interactive design attributes in data visualization instead of real-time database connection. In addition, personal daily nutrition requirements are calculated according to one's body factors based on the formulation provided by WeightLossForAll.com. They are used to simulate a use case of the application by comparing nutrition facts of a selected food item to one's daily nutrition requirements.

Design Process and Tool

This study emphasizes the process of developing a form in order to illustrate how design intention comes into selecting specific shapes for particular aesthetic and functional purposes. The aim of using metaphors of a dinner plate and wristwatch for browsing and selecting food items was to simplify the visualization and navigation of nutrient data within a small touch screen interface. The shapes and functions of two existing physical objects (a dinner plate and a wristwatch), incorporated into a digital form, offer a visual and behavioral analogy for display and navigation of information. In this way a large amount of data is displayed in simple pictorial forms for quick overview of nutrition compositions of different food items while keeping details of data as well.

We used Flash Lite 3.1 (ActionScript 2.0) to simulate interactive visualization. Flash was a good choice to experiment various graphic shapes and symbolic icons to represent data in more familiar and engaging visual forms. However, dynamic data navigation in connection to the local XML database was not intuitive to code with ActionScript especially without expert programming knowledge. Specifically, the challenges are summarized into 1) browsing database by rotating a graphic object, 2) visualizing nutrition facts of a focused food item, 3) selecting food items by drag and drop, and 4) storing their nutrition facts for reviewing total daily consumption. We created so many small functions that had not been planned at the beginning. The overall design process was quite linear by starting from visual sketches, specification of interaction concepts, and building a database. However the implementation was really complicated, not proceeded step by step. We went through multiple sketches of interaction sequence and flowcharts in order to make sense of the relations of functions and variables that are used to store values from database and to draw graphic shapes from them.

Case Study 4: Personal Lifelog Visualization

A variety of life-logging devices with sensing technologies have been created and their applications provide us with the opportunity to track our lives accurately and automatically. In this context, there is a growing interest called the "Quantified Self Movement" driven by technologies that sense numerous aspects of an individual's life in detail. The idea of "quantified self" starts with tracking our daily activities such as location, mood, health factors, sleep patterns, photos, phone calls, and so on. Many technologies support the capturing of lifelogs. For example, ubiquitous smartphones allow us to record our life activities in previously unimaginable detail. These captured logs can be further used to infer interesting and useful insights about people, their communication patterns with others, and their interaction with environments.

However, current tools to manage such rich data archives are still in need of improvement in terms of storing and organizing. Since the captured material are voluminous and mixed in data types, it might be overwhelming and impractical to manually scan the full contents of these lifelogs, which eventually results in lots of "worn memories"—we write once but never read again. Besides, the raw data do not give much insight to users without additional semantic enrichment. Thus, a promising approach is to pre-process raw data to extract and aggregate useful information, and then apply visualization to the large-scale data archives as a means to target users. We argue that lifelog visualization is capable of displaying the sheer quantity of mixed multimedia contents in a meaningful way.

Data Collection: Smartphones with Various Sensors

We developed a lifelogging platform working on Android smartphones. A typical smartphone is equipped with many sensors to capture various sources of data (Table 1). The gathered data is first analyzed in the phone and then uploaded to the central server. At the server we perform additional semantic analysis and enrich the data into semantically rich "life streams"; by applying machine learning techniques, we extract semantics from raw sensor log streams; by grouping related sensed logs together, we reconstruct the data and design several use case scenarios. For example, by aggregating the logs from speaker, GPS, and Bluetooth together, we can detect whether test subject is at work or engaged in social activity. In this way, we can extract various semantic contexts that are meaningful to target users. We also run more processor intensive operations with the data such as face detection. Finally, the data were processed into a variety of formats such as JSON, XML, and CSV prior to the representation phase.

Table 1: Available sensors equipped on a user's smartphone

Sensor	Description
Acceleration	Physical movement of the user
GPS	Geographic location
Bluetooth	Social context
Wifi	Indoor location, location cache
Camera	Automatic photo/video capture
Speaker	Environmental sound/noise level

Data Representation & Navigation: Three Types of Visual Logs

Sellen and Whittaker (2010) summarized five functions of memory that lifelogs could potentially support, referred to as the 5 Rs: recollecting (recalling a specific piece of information or an experience), reminiscing (reliving past experiences for emotional or sentimental reasons), remembering (supporting memory or tasks such as showing up for appointments), reflecting (facilitating the reflections and reviews on past experiences) and retrieving (revisiting previously encountered digital items or information such as documents, email or web pages). According to these functions of lifelogs, our primary design goal is to support users' self-reflection, sharing, evoking thoughts, and reminiscence. We believe that the form and styles of visualization should be determined by the purpose of uses and context.

Since the data generated from multiple sensor sources are complex, we need a more systematic approach to exploring the match between user needs and data visualization. We first generate three scenarios of different user behaviors and contexts: visual diary, social interaction, and activities review. Then we characterize UI patterns based on these three contexts, and create wireframe and interfaces prototypes. Visual diary use case scenario help users to support recollecting, remembering, and reminiscing of their past experiences. Social interaction and activities review support more abstract representation of personal lifelogs to facilitate reflecting and retrieving functions.

1) Visual Diary

Sensing devices automatically capture thousands of photos, and many times more sensor readings per day (Kalnikaite, Steve, and Whittaker, 2012). Hence, grouping the sequences of related images into "events" is necessary in order to reduce complexity (Doherty & Smeaton, 2008). The visual design of the visualization is inspired by Squarified treemap pattern (Bruls,1999). Our visualization, called "Visual Diary," provides a summary of user's daily log as photographs, with emphasis on important events (Figure 11). Each grid represents an event and the size of the grid provides an immediate visual cue to the event's importance level. The position of each grid depicts the time sequence. At the same time, users are allowed to drill down (full photo stream and sensor log) inside each event.

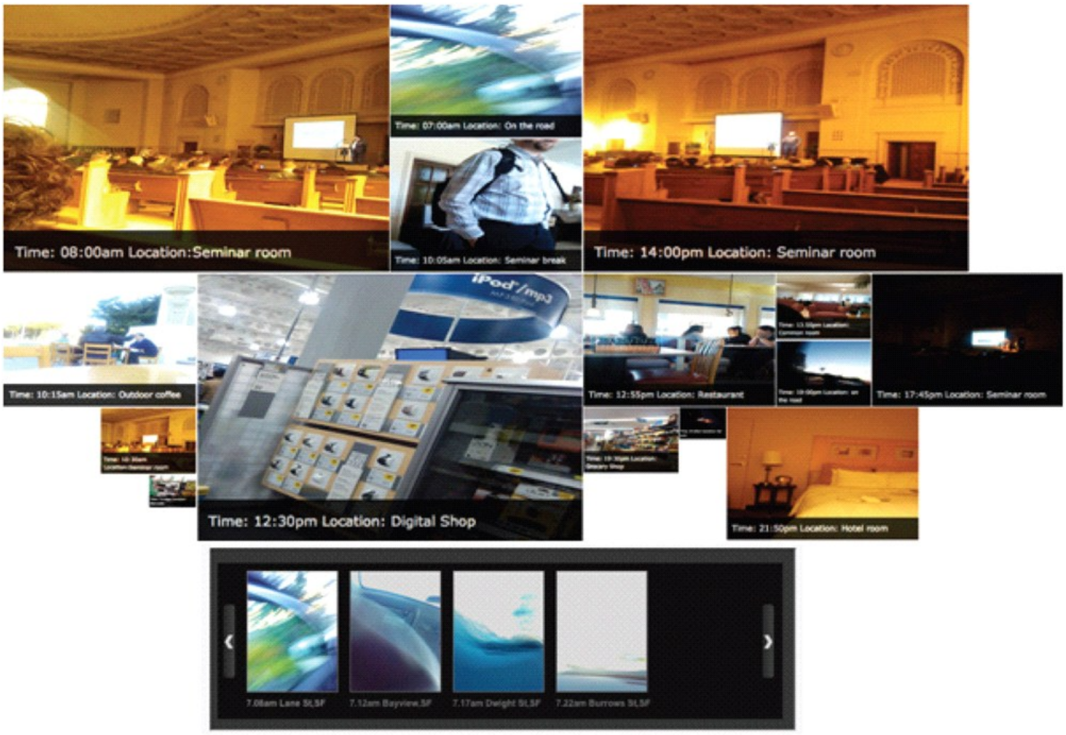

Figure 11: Interactive visual diary generated for one day, showing event segmentation

2) Social Interaction Radar

For the visualization of social contexts, we utilize the data from three embedded sensors (i.e., Bluetooth, Wi-Fi, and GPS) (Figure 12). With these datasets, we identify the social context of individual users over the course of a year. To support the different features of the multi-dimensional data, we adopted a Coxcomb visualization technique, which helps users to understand the whole and its individual parts simultaneously. This opens up new possibilities for rapid communication of complex constructs. Each concentric circle represents one type of the sensor data. This radar graph has three dimensions: 1) type of sensor, 2) social activity value, 3) time. Scrolling around the circle enables rapid exploration and comparison sensor values between different months.

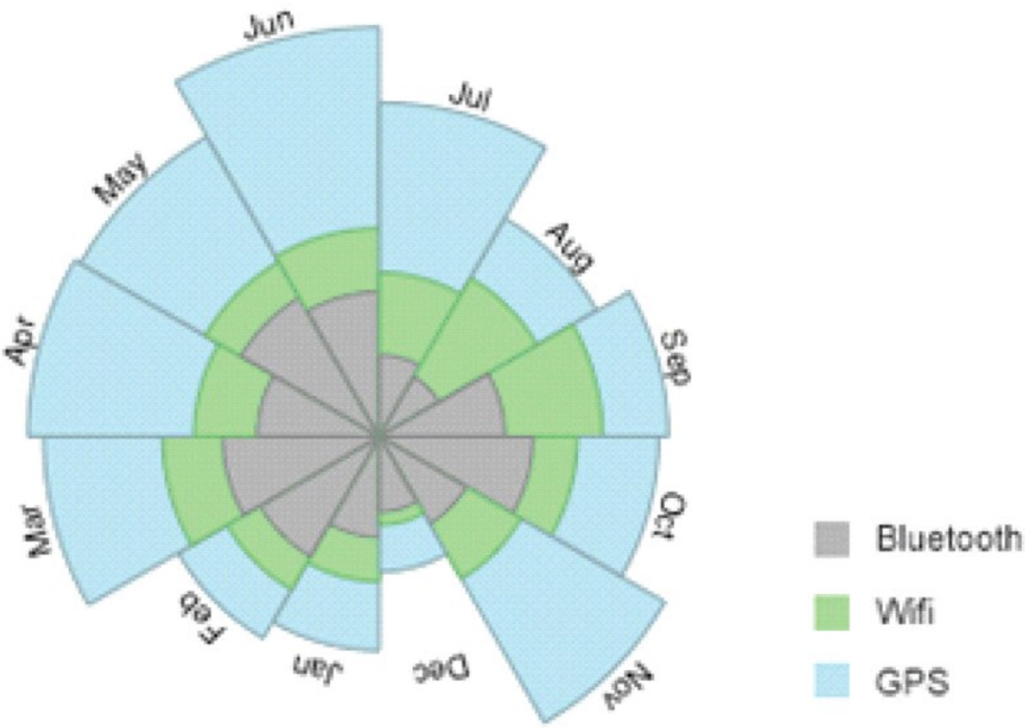

Figure 12: Social Interactive Multi-Dimension Radar Graph

3) Activity Calendar

Activity view allows users to gain a detailed understanding of their physical activities (Figure 13). We calculate the level of physical activity with the data from accelerometer and GPS sensors. We visualize the data in an annual calendar layout, with color-coding to present the activity intensity. A darker color indicates more activity involved in a given day. By investigating the activity pattern over the course of a full year, it is possible to detect a user's extreme days (i.e., the most active or the most quite days).

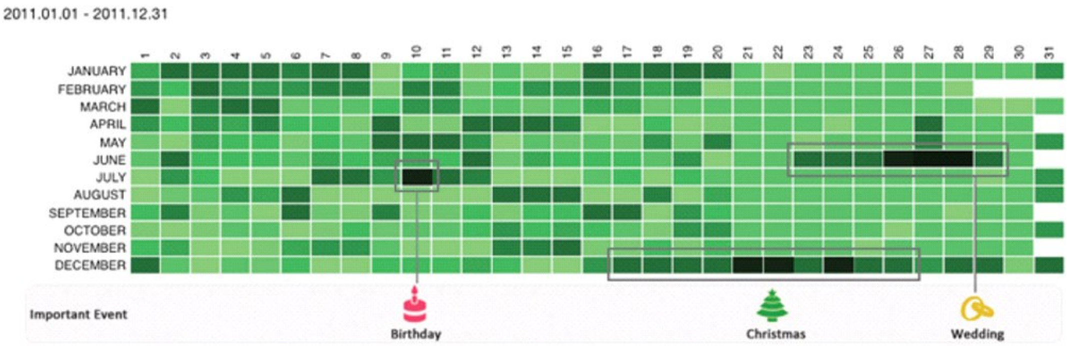

Figure 13: Yearly activity summary with a calendar view

Design Tool

When choosing the tools for visualization, we had to consider the objectives as well as the time constrains for creating interfaces. Our goal was to analyze the large data archive, and to create dynamic interactive visualizations that are later exhibited on users' web browser. Thus, the accessibility without plug-ins, the compatibility with web standards, and cross-browser support are our main concerns when choosing visualization tools. After the review of available tools, we chose open-source JavaScript-based toolkits—Protovis (Stanford Visual Group, 2010) and d3 (Bostock, 2012). It binds arbitrary data into DOM, and brings data to life using html, CSS, SVG that appear almost identically on different browsers and platforms. These independent toolkits make it possible to plot data in novel structure with rich user interaction. We also choose jQuery library for more dynamic interactive features. It is lightweight compared to other JavaScript frameworks and uses familiar CSS syntax that designer can learn and use easily. This combination of tools provides designers with powerful approaches to the aesthetic look and interactivity, which also makes the data manipulation customizable. Ultimately they empower designers with the freedom to focus on the aesthetics.

Case Study 5: Visualizations of Mobile Communication Data

This project is to use data visualization as a tool for exploratory data analysis (Shneiderman, 2002) to quickly discover insights for research/design opportunities. Mobile phones can help collect extensive data ranging from personal usage of the phone to inter-personal communication. In addition, due to their pervasive daily use, their various embedded sensors, and ubiquitous wireless technology infrastructure, mobile phones have received increased use as a new kind of research tool. Researchers in this field have primarily used data mining, machine learning, and other quantitative modeling methods, at which interaction designers are not typically trained. Such unmatched skillsets limit designers from being actively involved in data-driven research. In this challenging context, we suggest a new role for interaction designers in which they can apply and enhance their existing skills: designing information visualizations of the available data in a timely manner and deploying them to other researchers during the early phases of data-driven research. Using visualizations as a tool for *exploratory data analysis* (Shneiderman, 2002), researchers can find insights quickly to ground the next phase of research.

Data Collection

In late 2009 for more than one year, over 160 volunteers near Lake Geneva, Switzerland participated in a data collection campaign using mobile phones. We gave study volunteers smart phones equipped with special software that ran in the background and gathered data from embedded sensors continuously when the phone was turned on. The logged data were stored in the mobile phone and automatically uploaded to a database server on a daily basis when a known WLAN access point was detected. The data was subsequently made available to the public through Nokia Mobile Data Challenge (Nokia Research Center, 2012).

We categorize the collected data as follows:

- **Inter-Personal Communication Data:** Voice call logs and text message logs
- **Physical Proximity Data:** The Bluetooth IDs of mobile phones scanned within a physically close distance
- **Location Data:** GPS, WLAN access points, and cell network information (in the order of precision)
- **Media creation and Usage Data:** Logs of photo taking, video shooting, music play, web browsing, alarm setting, etc.

Design Process and Tools (Requirements for Exploratory Data Analysis)

The data set emerging from the mobile phone study was very large and varied. Performing exploratory analysis on this data would require a high degree of flexibility in supporting different subsets of the data extracted from the entire collection. We did not feel that any existing general-purpose visualization systems could provide all the different perspectives we desired on the data. Also, building a custom visualization system for the entire set appeared daunting and could possibly take a long time. Thus, we designed multiple, different interactive visualizations instead of a monolithic visual analytic system. Each visualization focuses on a different aspect of the data collection and is designed to best portray a particular aspect of the dataset. We assumed that the end-users of such visualizations would be researchers who want to explore massive datasets, especially related to mobile communication data, before they begin investigation with other sophisticated data analysis methods.

In designing these visualizations, our priority was to generate flexible datasets that are directly related to researchers' questions and easily modified to their ongoing requests. Initially, we extracted relevant data subsets from a database using simple SQL queries. We stored these data subsets in comma-separated-value formatted separated files to be linked and used in visualizations. To implement the visualizations, we primarily used Java-based Processing. The simple structure of the visualization source files allowed us to easily manage and modify datasets and visual design; we were able to quickly convert data formats, and generate new dimensions or datasets from the linked datasets; by simply editing several lines of source code in a Processing file, it was not difficult to change the size or color of the drawn visual elements. We also reused portions of visualization source code throughout multiple visualizations and modified it as necessary. This highly customized design process would hopefully make the visualizations versatile enough to support analysts' incremental demands during the analysis.

Data Visualization (An Individuals' Daily Life)

Due to space limitations, we select one example exemplifies the micro traits of single participants. This visualization focuses on data about individual participants in order to understand different mobile phone usage patterns depending on different temporal and social contexts. Better understanding the dissimilar lifestyles of individual participants would ground the design of personalized mobile services and applications. For the visualization design, we applied timeline-based visualization because our goal was to best support effective analysis with a minimal learning curve, not to invent new visualization techniques. That said, we believe these visualizations provide innovative designs for the visual analysis of communication and location data.

1) Dataset Datasets: Phone usage, Location, and Bluetooth Readings

Thanks to the wealth of data modalities tracked, the analysis of the data about a single user can inform extensive insights about his or her life. This visualization was created with multiple datasets from multiple users (Figure 14).

- Location data from the high fidelity GPS logs table containing latitude and longitude coordinates: The number of GPS entries is roughly equivalent to the frequency of physical movement.
- Bluetooth readings: Bluetooth detection data show the number of people appearing in a proximate physical distance and the frequency of appearance.
- Mobile phone usage data whose entries were parsed from five different log tables in the server: These tables include voice calls, text messages, web browsing, music play, and photo-shooting.

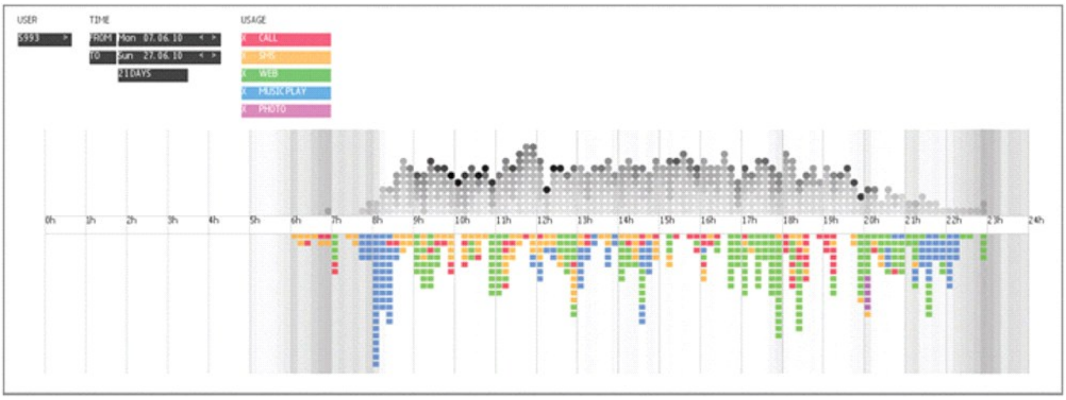

Figure 14: Integrated visualization of an individual user's phone usage logs (colored squared below x-axis), GPS-based moving status (gray-scale background), and Bluetooth encounters with other people (gray-scale circles above x-axis)

2) Timeline-Based Visualization with Multiple Elements

We processed the GPS logs in ten-minute intervals and represent them as the background of the timeline. The presence of more GPS entries results in a darker background. The same time interval is applied to the Bluetooth data, which are represented by the grey-scale dots above the horizontal 24-hour line. 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 presence. The squares below the bar are color-coded to represent the five different kinds of phone usage data. We added simple menus to enable analysts to select other participants and to select/de-select the categories of mobile phone usage. We also included a time selector, so they could adjust the time range of the data to any number of days within the sixteen weeks retrieved.

He has many more people using Bluetooth around him during the typical working hours, which might suggest that he is present in an office environment with business colleagues. He also exhibits a rough commuting time range between 6:30AM to 8AM. Between 12:10 PM to 12:30 PM, Bluetooth detection is evident by the two darker dots. Based upon consistency of occurrence, we infer that the other participants might be regular lunch friend(s).

Additionally, we examined visualizations of participants who were college students. We observed that students tended to have different patterns in terms of movement and Bluetooth detection (Figure 15). Both students shown did not exhibit a fixed short period of commuting pattern. Instead, they seemed to more randomly move around. One student had two to three mobile phones appear nearby during night time (roughly between 11PM to 7AM) (Figure 15-top). One of them might b a roommate, represented as a darkest dot, whereas the other dots might represent occasional visitors. The other student did not have a regular peak time in terms of the number of people nearby (Figure 15-bottom). She 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 distinguishable; for instance, the students used web browser SMS more than voice calls.

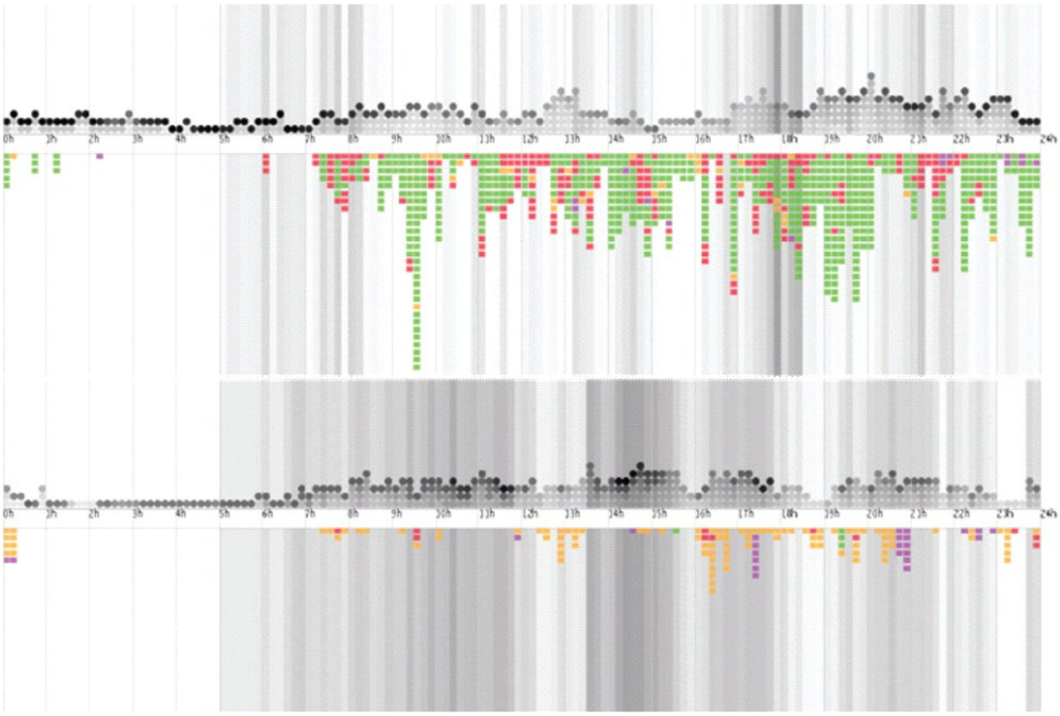

Figure 15: Visualization of integrated personal data from two college students

Summary and Reflection

In this chapter we introduced five case studies of data visualization projects. The first project is about visual representation of online bank transactions in order to improve user experience of managing multiple bank accounts and being more aware of their financial status. Traditionally design approaches to scientific visualization have been rather *data-centric* with focus on visual analysis of the data structure. However, with increasing end-user visualization applications, *task-centric* design approach should be more crucially considered in selecting forms of data representation and navigation interfaces that can best afford action-based user goals. The importance of a scenario lies in enabling a clear grasp of the contextual elements and in providing particular interaction paths (Rizzo & Bacigalupo, 2004). In this vein, this study illustrates that paper-based sketches with particular user scenarios can serve as an efficient design medium at early phases of design for quick iterations of initial ideas.

Moreover, as data visualization has been more broadly applied to end-user services, interaction and experiential qualities have become more critical design issues, not only in terms of usability of data perception and task-based navigation but also in terms of attractive and engaging representation and navigation of data. The second and the third case studies are more directly related to aesthetic forms and interface elements of data representation and navigation. In the weather visualization project, specific graph shapes, colors, and interfaces were iteratively tested and specified by coding (using d3.js). In the nutrition data application project, the main visual and interaction design directions were determined by metaphors from physical objects, and then simulated in Flash ActionScript. In both cases, programming played a significant role in stimulating design concepts and testing technical constraints in connection to database and interactive navigation. Although programming is still a barrier for many designers, it can provide more logical and consistent visual and interactive styles to multiple parts of a design system over iterative design process. Some common visualization tools and libraries (i.e., Flash ActionScript or d3.js) are efficient to demonstrate new interaction and interface concepts. However one of the big challenges is to manage overall layout and whole sequence of interaction as well as data collection and manipulation of the entire system of visualization applications. It would be really beneficial to create a new design process and tools that can better support designers’ systematic thinking and simulation of data ecosystem.

The last two case studies are about exploring new applications of data visualization. Due to increasing data capturing and processing power, we have more challenges as well as opportunities in making sense and use of such voluminous and complex data. The mobile lifelogs project explores different ways of representing personal data for various purposes of raising insights to personal activities and self-reflection. The mobile communication data project focuses on extracting research insights from data by using visualization as a rapid discovery tool for generating insights from large-scale data at an early phase of research. This visualization-based analysis of user-generated data can serve as a new research methodology, which is time-efficient and still people-centric by discovering traits about the people in the data. These approaches envision a new role for visualization with its strength in transforming ideas into visual artifacts in data-driven user research.

CONCLUSION

As discussed in the case studies, this design process is not always linear; when other requests from either users or designer themselves arise, iterations become necessary. Regardless of different design process and tools applied in each case study, they still share general issues to be discussed in terms of *1) data collection, 2) data representation,* and *3) data navigation*.

Data collection is about a whole system in which data is gathered and linked to visualization application. This issue is closely related to both user scenarios of data visualization (in terms of how data is provided, shared, and distributed) and technical challenges (in terms of how to retrieve and link data in proper formats).

Data representation is related to graphic forms through which overall structure of data is represented with its details. There are many standard forms of graphs such as bar chart, line chart, pie chart, etc. Beyond those graph forms, more exploratory forms of graphs are enabled and experimented thanks to advanced computing technology, including network diagram, tag cloud, bubble chart, direct visualization, which shows data as it is like in photo archive, etc. In addition, geographical maps are also frequently used to populate data onto familiar spatial coordinates for simple and efficient data perception. According to the increasing number of various forms of data representation, the criteria to select an efficient but still engaging forms is always a critical design consideration. The selection of an overall representation form proceeds to traditional aesthetic concerns in terms selection of colors, sizes, and layout of graphic shapes, which could be iteratively polished afterwards.

Data navigation, in comparison to data representation, is rather a less investigated aspect than the issue of data representation. There are a few options for standard interactivity including pulling out details-on-demand or browsing data by scrolling, panning, and zooming. While various forms of graphs and interfaces are used to represent data, interactive features to enter and navigate data are relatively limited in terms of diversity with similar interfaces in different applications. It is a great opportunity to explore new types of interactivity to navigate data, at the same time a huge challenge to simulate new interaction ideas with current design tools.

These three design issues—data collection, representation, and navigation—can be considered as significant building blocks that constitute an overall design process of data visualization, not necessarily put in a linear sequence. We expect the identification of the three topical design blocks could support interaction designers flexibly planning out a design process in consideration of iterative simulation and refinement of the connection of the three design issues.

REFERENCES

Adobe. (n.d.). ActionScript technology center. *Adobe Developer Connection.* Retrieved from http://www.adobe.com/devnet/actionscript.html.

Andersson, P., Rosenqvist, C., & Sahrawi, O. (2007). Mobile innovations in healthcare: customer involvement and the co-creation of value. *International Journal of Mobile Communications*, *5*(4), 371–388. doi:10.1504/IJMC.2007.012786

Balsamiq Studios. (2013). *Rapid wireframing tool.* Retrieved from http://www.balsamiq.com/.

Baranovsky, D. (n.d.). *GRaphael-Jevascript library.* Retrieved from http://g.raphaeljs.com/.

Bostock, M. (2012). *Data-driven documents.* Retrieved from http://d3js.org/.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309. doi:10.1109/TVCG.2011.185

Bruls, M., Huizing, K., & Wijk, J. (1999) Squarified treemaps, *In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization*. New Brunswick, NJ: IEEE Press.

Cross, N. (2011). *Design thinking: Understanding how designers think and work*. Oxford, UK: Berg Publishers.

Doherty, A. R., & Smeaton, A. F. (2008). Automatically segmenting lifelog data into events. In *Proceedings of WIAMIS*. New Brunswick, NJ: IEEE Press.

Goodman, E., Stolterman, E., & Wakkary, R. (2011). Understanding interaction design practice. In *Proceedings of Conference on Human Factors in Computing Systems,* 1061-1070. New York: ACM Press.

IBM. (n.d.). Many eyes. *Software Analytics.* Retrieved from http://www-958.ibm.com/.

Kalnikaite, V., & Whittaker, S. (2012). Recollection: How to design lifelogging tools that help locate the right information. *Human-Computer Interaction: The Agency Perspective Studies in Computational Intelligence*, 329-348. Berlin: Springer. doi:10.1007/978-3-642-25691-2_14

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*(4), 441–454. doi:10.1007/s00146-010-0272-8

Löwgren, J., & Stolterman, E. (2004). *Thoughtful Interaction Design: A Design Perspective on Information Technology*. Cambridge, MA: MIT Press.

Manovich, L. (2008). *Introduction to info-aesthetics.* Retrieved from http://goo.gl/NFLvy.

Manovich, L. (2010). *What is visualization?* Retrieved from http://www.datavisualisation.org/2010/11/levmanovichwhat- is-visualization/.

Nokia Research Center. (2012). *Nokia mobile data challenge.* Retrieved from http://research.nokia.com/page/12000.

Pousman, Z., & Stasko, J. T. (2007). Data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1145–1152. doi:10.1109/TVCG.2007.70541

Processing 2. (n.d.). *Processing.org.* Retrieved from http://processing.org/.

Rizzo, A., & Bacigalupo, M. (2004) Scenarios: heuristics for actions. In *Proceedings of XII European Conference on Cognitive Ergonomics.* York, UK: EACE.

Salimun, C., Purchase, H. C., Simmons, D. R., & Brewster, S. (2010). The effect of aesthetically pleasing composition on visual search performance. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries,* 422–431. New York: ACM Press.

Segel, E., & Heer, J. (2010). Narrative visualization: telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. doi:10.1109/TVCG.2010.179

Sellen, A. J., & Whittaker, S. (2010). Beyond total capture: A constructive critique of lifelogging. *Communications of the ACM*, *53*(5), 70–77. doi:10.1145/1735223.1735243

Shneiderman, B. (2002). Inventing discovery tools: Combining information visualization with data mining. *Information Visualization*, *1*(1), 5–12.

Stanford Visual Group. (2010). Protovis-A graphical approach to visualization. *Protovis.* Retrieved from http://mbostock.github.com/protovis/.

Tableau. (2013). Free data visualization software. *Tableau Public.* Retrieved from http://www.tableausoftware.com/public.

Tufte, E. R. (2001). *The visual display of quantitative information. Visual Explanations, 194–95*. Cockeysville, MD: PR Graphics.

USDA. (2011). Welcome to the USDA national nutrient database for standard reference. *National Agriculture Library.* Retrieved from http://ndb.nal.usda.gov/.

Vande Moere, A., & Purchase, H. (2011). On the role of design in information visualization. *Information Visualization*, *10*(4), 356–371. doi:10.1177/1473871611415996

Ware, C. (2000). *Information visualization: Perception for design (interactive technologies)*. New York: Morgan Kaufmann.

Weather Underground. (2007). *An example URL of daily weather data of london heathrow airport in 2007.* Retrieved from http://www.wunderground.com/history/airport/EGLL/2007/ 1/1/CustomHistory.html?dayend=31&monthend=12&yearend=2007& format=1.

Weather Underground. (2013). Welcome to weather underground! *Weather Forecasts & Reports.* Retrieved from http://www.wunderground.com.