A Validation Study of a Visual Analytics Tool with End Users

Heloisa Candello, Victor Fernandes Cavalcante, Alan Braz, and Rogério Abreu De Paula

IBM Research, Brazil
{heloisacandello,victorfc,alanbraz,ropaula}@br.ibm.com

Abstract. In this paper we describe an user evaluation that aimed to understand how a group of endusers interpret a visual analytics tool in the context of service delivery. It is common for service factories to have an organization devoted to handle incidents. Many incident management systems have strict controls on how fast incidents should be handled, often subjected to penalties when targets are not met. We call Time-Bounded Incident Management (TBIM) those systems, which require clearly defined incident resolution times. In our project, research scientists proposed a method and a visual representation named Workload Profile Chart (WPC) that had as primary goal to understand the area of incident management in a service delivery department. The objective of this visual representation is to help characterizing the performance of TBIM systems and diagnosing major issues such as resource and skill allocation problems, abnormal behavior, and incident characteristics. Researchers wanted to understand if end-users, the quality analysts (QAs), would comprehend the charts and would be able to use them to identify problems and propose effective improvement actions related to TBIM activities. The study was conducted with ten QAs of a service delivery department of a IT company based in Brazil. The data was analyzed using descriptive statistical and qualitative methods. As a result, participants were mainly guided by the axes titles and chart legends to interpret the visualizations, and not always understood what kind of data the chart was displaying. Those results served as insights of how QAs think when analyzing TBIM information in a service delivery department and what improvements in the visual representation tool may be proposed to facilitate their activity. At last we identified evidences of how to design better visual analytics tools based on participant's perceptions and interpretations of color differences and verbal information in chart labels and legend.

Keywords: visual analytics, service design, and user evaluation.

1 Introduction

Systems to deliver and provision services are the key engines of the new global economy where services are increasingly becoming the dominant mode of production. To meet those demands, more and more firms have established mass-scale, complex systems to deliver services using factory-like production methods, or service factories [1]. The modern call center is the typical example of mature implementation of the service factory model, while health, government, and IT services, among others, are just

A. Marcus (Ed.): DUXU 2014, Part IV, LNCS 8520, pp. 381–391, 2014.

starting to deliver services following this approach. It is common for service factories to have an organization devoted to handling incidents, which can be defined as: "any event which is not part of the standard operation of a service and which causes, or may cause, an interruption to or a reduction in, the quality of that service" [2]. They are, by definition, unpredictable and often demand rapid allocation of skilled resources for their resolution and bringing a service back to normal. For example, management of power failures or dangerous situations is an absolutely critical organization in the electric distribution sector. Many incident management systems have strict controls on how fast incidents should be handled, often subjected to penalties when targets are not met. We call Time-Bounded Incident Management (TBIM) those systems, which require clearly defined incident resolution times. Examples of TBIM systems include: fire and ambulance management, some call centers with required maximum resolution times, and most IT service delivery operations of IT outsourcing.

Nowadays, quality analysts know if incidents are handled on time formally and informally. Formally, by metrics available for each service pool; charts from official tools; characteristics of the incident report (e.g. severity); employees who acted incidents (level of expertise). Informally by chatting with people from the team (e.g. dispatchers) to know the root cause of incident and calling the customer to know informally if he is satisfied with the service. So that, many resources have to be consulted to know if the team is solving incidents on time from the internal point of view (e.g. employees are managing to solve the tickets) and for the costumer point of view (e.g. incidents solved according to the time agreement). Despite of being a cumbersome TBIM activity, it is also difficult to track how long the incident report will delay until it reach the right employee for solving. Our visual analytics tool, considers this data to clarify where might be lateness.

In this paper we describe a user evaluation study that aimed to understand how a group of end-users interpret this visual analytics tool in the context of service delivery.

2 Background

Information visualization is defined as "The use of computer-supported, interactive, visual representations of data to amplify cognition" [3]. Spencer [4] completes affirming that "visualization is a process of forming mental model of data, thereby gaining insight into that data". Visualization is a human cognitive activity, not something that a computer does in his views. Thereby, control of information is given over to viewers, not to editors, designers or decorators [5]. To unveil the mental model in every person's mind is a cumbersome activity. Therefore, user experience researchers and designers may dispose what will be formed in people's mind understanding better their contexts and practices in everyday life. Visualizations have their own purpose give insights on data to solve or clarify certain problems. And it posts a critical question: How best to transform the data into something that people can understand for optimal decision-making? [6].

User centered design approaches; such as user evaluation studies may enlighten this question. Hetzler and Turner [7] presented lessons learned from an observational study of an INfovis application. Ins-pire visualization tool uses statistical word patterns to characterize documents based on text content. The purpose of this tool was narrow the number of relevant documents before analysts read them. As a result, analysts found useful to see the main data grouped by similar visualizations and researchers learned that analysts use diverse type of tools to make their conclusions, so that Ins-pire should provide support to other tools and have other features such as to write reports while data analysis.

Toker et al [8] explores the relationship of cognitive abilities, personality, and attention patterns of users have an impact on using different visualization techniques. For doing so, they used eye-tracking studies to identify gaze patterns. Thirty-five subjects participated the study and accomplished three computer-based tasks for a three cognitive measures: verbal, visual and a perceptual speed. The cognitive measure with highest number of effects was Perceptual speed. Among their findings, they discovered that low perceptual speed users tended to access a visualization legend more than high perceptual speed users, suggesting that they should be specifically supported in terms of legend processing. Overall, their analysis show that some of the user characteristics affect user gaze behavior and it was possible to detect this through a variety of eye-tracking measures.

Afterward user studies and have a better understand of people's cognitions to interpret visualizations and their context. Designers may employ multimedia elements to built new or redesigned interfaces to improve user's interpretation of visual and verbal elements. According to Bertin [9]: "A graphic is no longer drawn' once and for all; it is constructed and reconstructed until it reveals all the relationships constituted by the interplay of data". In other words, "the best graphic operations are those carried out by the decision maker itself". He also created a visual grammar to make designers aware of visual elements characteristics when choosing them to compose a graph. In Tufte [10] words "When principles of design replicate principles of thought, the act of arranging information is becomes an act of insight. A more recent study [11] allows individuals may upload data, collaborate and generate visualizations at a large scale in a public website called ManyEyes. This tool also provides a wide range of visualizations types that may help designers to compare and choose appropriate visualizations for different contexts. Those studies have high relevance and serve as basis for future studies in the field. However, more studies are necessary to unveil ways of understanding the huge amount of data available nowadays. In doing so, helping designers to create more effective visual analytics tools.

3 The Visual Analytics Tool: Workload Profile Chart (WPC)

In our project, research scientists proposed a method and a visual representation named Workload Profile Chart that had as primary goal to understand the area of incident management in a service delivery department. The method preprocesses data corresponding to records of incidents (tickets) and plots the spreading of incidents on a log-log chart.

Particularly, the researchers were interested in how to evaluate the performance of a service pool in terms of its compliance with an established Service Level Agreement (SLA) as measured by its performance in those two key variables: the elapsed time since a ticket is reported until its assignment to a support analyst (assignment delay), and the time spent to have the ticket solved and closed by the analyst in charge to handle it (resolution time). Thus, the chart is comprised of the log of the assignment delay (X axis) and the log of resolution time (Y axis), both ranging between 0.1% and 1000% of the time allowance.

The objective of this visual representation was to help characterizing the performance of TBIM systems and diagnosing major issues such as resource and skill allocation problems, abnormal behavior, and ticket characteristics. Researchers wanted to understand if end-users, delivery quality analysts, would comprehend the charts and would be able to use them to identify problems and propose effective improvement actions related to TBIM activities.

The WPC is a density map resulting from the log-log plot and our inspection method takes one of these charts and systematically examines the concentration levels on different areas as defined in Figure 1. High or low concentration of tickets in a particular area corresponds to a set of specific characteristics likely to describe the reality of a service pools and, based on such concentration and an expected level used as reference (a baseline) for each area, a diagnostic is performed.

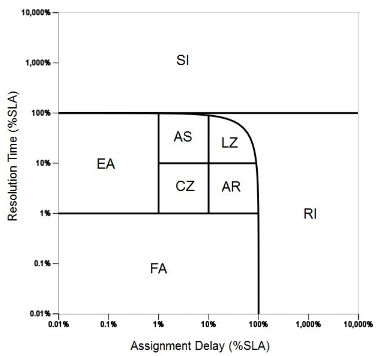


Fig. 1. Different areas of the WPC of the service pool

After analyzing dozens of WPCs, it was detected that high concentrations of tickets in different areas of a WPC are related to different kinds of problems and issues in a service pool. Accordingly, there are eight main areas of interest in a WPC, each of them corresponding to specific issues [12]:

- `Comfort Zone (CZ)`: this area contains tickets that are quickly assigned, resolved and closed. A high concentration of tickets in this area may indicate that the service pool has sufficient resources and most of the tickets have low difficulty to solve, meaning that analysts are typically working in their comfort zone.
- `False Alarms (FA)`: it corresponds to the area covering tickets whose resolution takes less than 1.0% of the SLA resolution time. These tickets are either very simple to solve or they are false alarms (tickets that should not exist) or duplicates. A high concentration of tickets in this area suggests that the performance of the service pool can be improved by automation or by the usage of better monitoring or filtering approaches.

- `Excess Availability (EA)`: tickets in this are quickly assigned but take some time to be solved. A high concentration of tickets in this area may indicate that there are often too many analysts available and ready to immediately start working on tickets as soon as they arrive and are dispatched. Such tickets are also very good candidates for automatic dispatching.
- `Adequate Resources (AR)`: this area contains tickets that meet their SLAs and to which assignments are not immediate but quickly resolved. A high concentration of tickets here may mean that resources are not always promptly available, so assignment takes some time. However, that does not compromise the SLA attainment because most tickets are easily solved without much work.
- `Adequate Skills (AS)`: this area comprises only tickets that meet their SLAs and that require a good amount of effort for their resolution. A high concentration of tickets in this area indicates that resolution tasks can be time consuming and that the SLAs cannot be made tighter without adding more skilled resources to the pool.
- `Limit Zone (LZ)`: this area comprises only tickets that meet their SLAs and that take some time both to be assigned and to be solved. So a high concentration in this area often means that the system is running close to its limit and therefore, resources and skills are being optimally used but with no slack. However, it means a high susceptibility to breakdown when unexpectedly heavy concentrations of incidents occur.
- `Resource Issues (RI)`: this area covers tickets that do not meet their SLA because they were assigned too late. A heavy concentration here often indicates that the service pool does not have enough support analysts or has dispatching issues.
- `Skills Issues (SI)`: this area covers tickets whose resolution took more than the agreed SLA time. A heavy concentration here usually means low skilled resources or a need to renegotiate SLAs.

4 Participants

Ten participants were randomly selected to attend the study. Eight did not know the project and two were familiar with the concept of the chart, but did not use the software to generate it before. All of them work at the same company and have a data analyst position. They have a diverse background in areas such: computing; engineering; math; project management and administration. As part of their everyday routine, they spend most of their time evaluating how their team is solving the incidents based on images and have a weekly meeting to report their analysis. In order to report their findings, they use at least seven to eight types of software. For instance, a software to see the customer metrics, other to generate the charts and tables and another to simulate and report the official data analysis. Usually, they are used to commercial software and prefer to run the analysis and charts on it, and then export to official company tools.

5 Method

The study was conducted with ten quality analysts of a service delivery department of a IT company based in Brazil. The experiment had two parts. First, participants followed a user case scenario [13,14] to explore the visual analytics tool and were encouraged to think aloud [15,16]. Second, they explained the visual representation to the test facilitator, imagining they were explaining it to a member of their own team. Participants also answered a semi-structured interview containing information of their job title and everyday work routine. Overall the study took 45 minutes long per participant. In this paper we are reporting the second part of the study, focused on chart interpretation. Participants were video recorded and signed a consent form allowing data for research proposes. Observation studies were done in situ. Researchers took notes during the user sessions. Additionally, video record files helped to go deep in details for data coding.

6 Data Analysis

The data was analyzed using descriptive statistical methods and qualitative methods. Basic statistical analysis was carried out to analyze the data from the semi-structured interviews. Tables and cross tabulation were applied to compare the results among participants and the use of the system. The restricted number of participants in the study was not enough to ensure the validity of the statistical analysis. Besides, these results did not give us enough evidence of essential chart interpretation trends. Writing on design evidence, Lawson [17] highlights that "we normally measure and express quantities by counting using a numerical system. This leads us to believe that all numbers behave in the same way and this is quite untrue". The same author emphasizes what designers really need is to have a feel for the meaning behind the numbers rather than precise methods of calculating them. (p.71). In agreement with Lawson ideas, a qualitative approach was applied in most of the process.

The data analysis was based on data transformation. Data transformation is a quantification of qualitative data. This involves creating qualitative codes and themes, and then counting the number of times they occur in the textual data. This enables researchers to compare quantitative results with qualitative data [18]. The transcriptions of the video observations, important notes taken during the fieldwork and suggestions given by participants while they were using the WPC were considered. Research questions were established to guide the analysis, such as:

- Does quality analysts understand the topic of the chart?
- Does quality analysts understand the regions of the chart proposed by researchers?
- Does quality analysts understand the structural elements of the chart (color, shapes, axis)?
- Which kind of decisions quality analysts can take based on the visual analytics tool results?
- What improvements can be made in the visual and interactive elements to assist quality analysts perform their data analysis?

These questions were kept in mind while the data was classified and codified. Additionally, issues were also rated as `Low` (one to three participants mentioned it), `Medium` (four to eight participants mentioned it) and `High` importance (over eight participants mentioned it).

7 Findings

Data was organized by categories (or themes) to help in the analysis. About 30 categories were identified in the whole study. The more frequent categories were associated to expectation of ways to add data and representations of their familiar visual analytics tools (seven participants). They also expect another type of chart, based on their previous experience with visual representations (six participants). Understanding was also a prevalence issue. And this category will be detailed here based on the questions previously described above on Section 6. We are also discussing possible improvements in the interface to provide better user experiences with the visual analytics tool and future decision make based on WPC.

7.1 Understanding

Participants were mainly guided by the axes titles and chart legends to interpret the visualizations, and not always understood what kind of data the chart was displaying. Users identified the log-log chart as a not intuitive representation. Six participants had difficulties to interpret it. They are used to decimal scale and not a logarithm scale. They would like to see the numbers not only the percentage on the chart. Three of those participants have their background on project management. Two participants were female, but more investigations are necessary to attest if gender and/or technical background may interfere in chart interpretation. One participant with difficulties on interpreting the chart commented: “I confess that I’m not understanding this chart here. There isn’t a type of chart that I have seen before. Our current graphics are limited. We don’t accept any value above 100% neither below 0%, and there is also no negative numbers. We just ignore them. I don’t understand the values above 100%, unless it is cumulative. But the percentage of resolution time, I can understand up to 100% of my SLA. How do I have points over 100%?”

Four participants understood what is the topic of the chart TBIM but had difficulties to understand the term Assignment Delay. They were not sure if it was the total time or delayed time to reach the dispatcher. They suggested to change the term for dispatcher efficiency, due to the dispatcher job is to assign the suitable ticket to the suitable analyst to solve it. Another participant stated: “ What is being measured is down time or full time? If it’s considering stopped time, does it measure only the time of break or full time? It should be measuring only one kind time.”

WPC was not a static chart. Participants could click on the screen, and they could see a tooltip with given information of that group of tickets position and the region acronym. In the interface, legend and chart were not always in the same view (depending on the resolution screen), which caused misinterpretations of the acronym in the chart. Moreover, not all participants interacted with the chart, and do not know why the chart was divided in parts, and also that each part had a name. (See Figure 2). Two participants stated that the regions (success and fail) from the legend where not identified in the chart.

Researchers named chart regions to help analysts to identify their team performance. In the user sessions, it was possible to validate the terms when participants explained in

Fig. 2. WPC chart and legend position inside the application interface

their words the meaning of each region. False Alarms (FA) was the most comprehensible term (seven people). Excess Availability (EA) and Resource Issues (RI) were the least comprehensible (eight people). Participants also had problems to distinguish between the regions Adequate Skills (AS) and Adequate Resources (AR). In their view if resources are people, so adequate resources means they have the right skill to solve incidents. Excess availability term generated doubts about what was in excess: resources or skills.

The idea to represent a group of ticket per square, density of tickets per one pixels of the chart, was also difficult to grasp. For users, each square was representing one ticket. Even with color legend displayed beside the chart. Perhaps, they interpreted in this way due to color scale used in the chart. Participants identified colors as level of severity of tickets (red is more severe than green) or time spent to solve the ticket (red took more time to solve than green).

Moreover to assist to understand the chart, and have a better analysis, participants would like to know the SLA agreed within the client, to be able to compare the chart generated to the baseline chart. Further, they also would like to see the ticket life-cycle in the same chart.

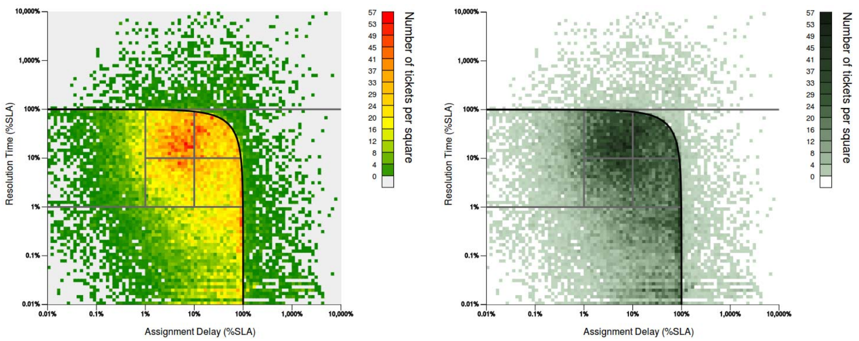
7.2 Design Improvements

Observation studies of user sessions assisted the team to have a picture of WPC chart and which improvements were needed to let quality analysts use it as a tool. We describe briefly visual component changes.

Color Scale. The use of color was misleading to understand concepts we were intent to communicate. Not all the participants understood the connection between the labels and the colors in the chart. Graph colors and values were very distinct to represent intensity. A monochromatic color scheme was considered a better choice for the aim to highlight density of tickets per square avoiding the cultural meaning carried by the colors in the previous chart. As Bertin [9] (p.85) affirms:

The use of color cannot be understood unless the notion of color (hue) is distinguished in a rigorous and definitive manner from the notion of value. They are two different sensations, which by nature, overlap. [. . .] A tone placed on a sheet of paper is therefore defined by two parameters: color (hue/tint) with the categories violet, blue, green [. . .] and value, defined by the percentage of black in the corresponding gray.

Therefore, we chose to vary the value of a green, instead of coloring the chart with different hues (see Figure 3). Iliinsky [19] agrees with Bertin that color is not ordered, is selective. Order can be achieved by a series of values from one of the scales of the spectrum (warm scale). The eyes mix the colors and order the components according to values. “. . . It is fantastic for labeling categories and terrible for representing quantities or ranks.” [19] p.7.



Color scale Green scale

Fig. 3. WPC charts before and after this study

Visual Configuration. Changes were made in the visual configuration of the tool to clarify what type of data the chart was showing. In agreement with the user studies, participants were guided by wording of titles and legends. Cleveland [20] earlier explored this issue with 55 subjects to understand how people perceive elements in a chart. In his study, participants perceive charts in two levels. First by the overall picture to see what is the subject of the chart and second by the elementary graphical perceptual task. Those sequences of actions were clear in our study.

With the aim to assist analysts to understand both pictures (overall and elementary) of the chart, a redesign of the tool was made. The main title was added to the chart, to help in external identification [9]. The components and legend (internal identification) was redesigned and placed to connect with visual elements in the chart. (see Figure 4). The current interface shows the legend and the chart side by side. Users can highlight the area in the chart with a click on the name of each region in the legend. Below in each region, users have questions to help them analyze the chart with the real number of tickets (decimal), not in log present in the region.

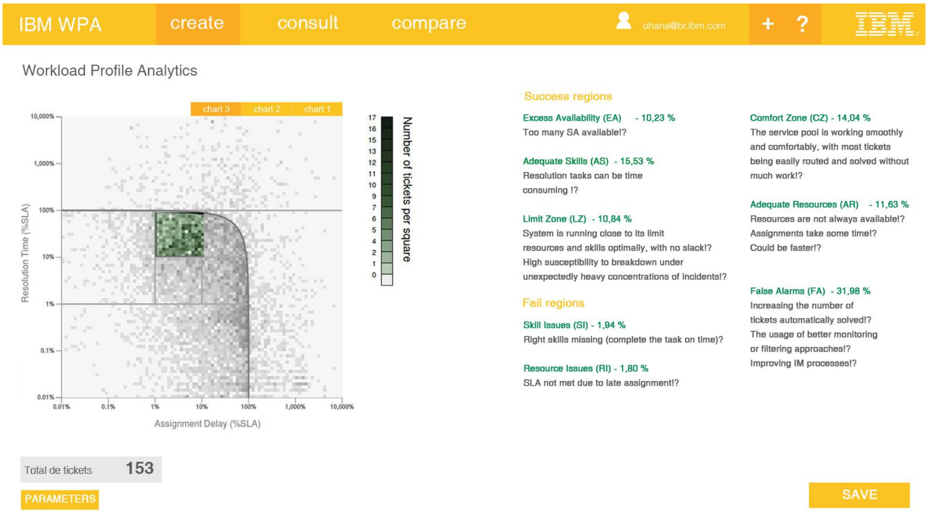


Fig. 4. Workload Profile Chart (WPC) proposed visual configuration

8 Discussion and Conclusions

Researchers created a log-log chart with the intent to find a visual representation that would couple amount of data to help understand TBIM. When users/quality analysts (QAs) tested it they need not only one data representation but wanted to explore more the database, switching between views of parameters that would help them to understand the problem and how to improve their performance pool. Therefore, multi-views of the chart would be necessary to satisfy their object. We are working for the WPC be visualized not only as a 2D chart, but a more interactive visual analytics tool where users can choose their set of parameters.

Although a more interactive tool might be the solution to suit this context, WPC showed information that is difficult to track in the daily basis activities of QAs along with the current tools they use, such as: time of assignment to solve the incident.

It is not easy to evaluate interactive visual analytics tools. So, designers must to think each kind of information should be shown in the screen, and each one should be unveiled when users click, or tap on the screen. It raises questions of what is relevant to the context? What are the primary information to be shown? And what is addiotional information, that users can explore deeply in the data set?

Those results served to have insights of how QAs think when analyzing TBIM information in a service delivery department and what improvements in the visual representation tool may be proposed to facilitate their activity. At last we identified evidences of how to design better visual analytics tools based on participant's perceptions and interpretations of color differences and verbal information in chart labels and legend.

References

1. de Souza, C.R.B., Pinhanez, C.S., Cavalcante, V.F.: Information Needs of System Administrators in Information Technology Service Factories. In: Proceedings of the 5th ACM Symposium on Computer Human Interaction for Management of Information Technology, CHIMIT 2011, pp. 3:1–3:10. ACM, New York (2011)
2. ITIL: Common ITIL Terms, http://itsm.the-hamster.com/itsm4.htm (accessed February 03, 2013)
3. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers Inc., San Francisco (1999)
4. Spence, R.: Information visualization. ACM Press books. Addison-Wesley, Harlow, England (2001)
5. Tufte, E.R.: Envisioning Information. Graphics Press, Cheshire (1990)
6. Ware, C.: Information Visualization - Perception for Design. Morgan Kaufmann, San Francisco (2004)
7. Hetzler, E.G., Turner, A.: Analysis Experiences Using Information Visualization. IEEE Computer Graphics and Applications 24(5), 22–26 (2004)
8. Toker, D., Conati, C., Steichen, B., Carenini, G.: Individual User Characteristics and Information Visualization: Connecting the Dots Through Eye Tracking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2013, pp. 295–304. ACM, New York (2013)
9. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. Esri Press, Redlands (2011)
10. Tufte, E.R.: Visual Explanations: Images and Quantities, Evidence and Narrative. Graphics Press, Cheshire (1997)
11. Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., McKeon, M.: Manyeyes: A site for visualization at internet scale. IEEE Transactions on Visualization and Computer Graphics 13(6), 1121–1128 (2007)
12. Cavalcante, V.F., Pinhanez, C.S., de Paula, R.A., Andrade, C.S., de Souza, C.R.B.: Data-driven analytical tools for characterization of productivity and service quality issues in IT service factories. Journal of Service Research: JSR (2013)
13. Cooper, A.: The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity. SAMS, Indianapolis (1999)
14. Sharp, H., Rogers, Y., Preece, J.: Interaction Design: Beyond Human-Computer Interaction, 2nd edn. Wiley, Indianapolis (2007)
15. Ericsson, K.A., Simon, H.: Protocol analysis: Verbal reports as data. MIT Press, Cambridge (1993)
16. Nielsen, J., Macks, R.L.: Usability Inspection Methods (1994)
17. Lawson, B.: How Designers Think: The Design Process Demystified, 4th edn. (2005)
18. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications (2009)
19. Iliinsky, N.: Choosing visual properties for successful visualizations, http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03323usen/YTW03323USEN.PDF (accessed February 03, 2013)
20. Cleveland, W.S., McGill, R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 79(387), 531–554 (1984)