

Is Your User Hunting or Gathering Insights? Identifying Insight Drivers Across Domains

Michael Smuc
Research Center KnowComm,
Danube University Krems
Dr.-Karl-Dorrek-Str. 30,
3500 Krems, Austria
+43 2732 893 2344

Eva Mayr
Research Center KnowComm,
Danube University Krems
Dr.-Karl-Dorrek Str. 30,
3500 Krems, Austria
+43 2732 893 2320

Hanna Risku
Research Center KnowComm,
Danube University Krems
Dr.-Karl-Dorrek Str. 30,
3500 Krems, Austria
+43 2732 893 2330

michael.smuc@donau-uni.ac.at eva.mayr@donau-uni.ac.at hanna.risku@donau-uni.ac.at

ABSTRACT

In recent years, using the number of insights to benchmark visual analytics tools has become a prominent method in the Infovis community. The insight methodology has become a frequently used instrument to measure the performance of tools that are developed for highly specialized purposes for highly specialized domain-experts. But some tools have a wider target group of experts with knowledge in different domains. The utility of the insight-method for other expert user groups without specific domain knowledge has been addressed to a far lesser extent. In a case study we give an illustration of how and where insights from experts with and without domain knowledge differ, and how these findings might enrich the evaluation of visualization tools designed for usage across different domains.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems – *human factors, human information processing*

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *evaluation/methodology, user-centered design*

General Terms

Human Factors

Keywords

insight methodology, domain experts, information visualization, visual analytics, evaluation, prior knowledge

1. INTRODUCTION

Ten years ago, Card, Mackinlay, and Shneiderman declared that “*the purpose of visualization is insight*” [1]. The idea to focus on insights for evaluation became more and more popular over the years, especially in the scientific communities interested in information visualization. For visual analytics tools, the insight methodology introduced by Sarayaia, North and Duca [15] reached a prominent status in benchmarking visualizations. This

methodology is widely used, even in contests for visual data analysis to measure the utility of the tools[13].

Many of these benchmarked tools have in common that they are especially designed for highly specialized domain experts and their highly specialized tasks – in domains such as psychotherapy [14], bioinformatics [15], meteorology [19], and so forth. However, typical test participants are not only experts with many years of experience but also novices and experts in training.

1.1 The Expert User Dilemma

The practical reasons for testing novices and treating in-training-experts as if they were real experts are manifold: Experts are rare, hard to get, and sometimes difficult to motivate, especially for long term iterative testing in non-scientific contexts. Another problem is that sometimes exponents of the target groups do not exist at the time of development. Especially when radically new visualizations are created, target groups often evolve in the future.

Another reason is that some applications are developed to support more than one domain. But what is the best way to deal with that situation with respect to usability and utility? Should we evaluate the tool for every possible domain? How can we give answers about the extendibility of the tool for other, currently unknown domains?

Our goal in the project DisCo was the development of a tool for a variety of users from different domains to analyze temporal data. Although our targeted users have many attributes in common (justifying the development of one common application), the tool should also support data analysis for different domains. So, we had different options to deal with the situation in this evaluation:

One option is to take the quick and dirty way. If the target group does not exist already or the target group is difficult to assemble, we can skip the insight benchmarking procedure and revert to general guidelines to evaluate the tool in the form of heuristics and design guidelines for visual analytics tools [21]. Unfortunately, most of these rule sets and guidelines are at an early stage, some of them are not developed for a detailed analysis or they lack transferability because they are too specialized for some visualizations.

Another option is to train novices or beginners in a domain to behave and think like experts. This is a time consuming affair and requires participants with a lot of time (they exist - let us call them students). Alongside arguments like “they will tell you what you taught them to tell” we rejected this option because we wanted to compare across domains – with the implication that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BELIV'10, April 10–11, 2010, Atlanta, GA, USA.
Copyright 2010 ACM 978-1-4503-0007-0...\$5.00.

the intense training sessions would have multiplied because they would have been required for each domain.

The option which looked most promising to us was to test real experts from different domains with datasets from their own as well as from other domains. But we found some methodological issues and theoretical questions. Evaluation by insights strongly depends on the categorisation of insights and the definition of what makes an insight and what should not be counted as an insight. In a definition following North's insight concept, an insight is a "non-trivial discovery about the data or, (...) a complex, deep, qualitative, unexpected and relevant assertion" [13]. With this definition, full of requirements of complexity, deepness and quality, only domain experts working in their own domain might be able to gather insights that can be counted.

Previous research concentrated on necessary prerequisites for insights: Insights on how the tool works and the application of prior knowledge [18] are decisive factors to gather more data insights. If an insight into the data activates prior knowledge, it is likely to generate further insights. Thereby, prior knowledge works like an insight driver and can lead to a plethora of insights. Furthermore, insights sometimes build up on previous insights. These prerequisites are usually subsumed as expertise. In the following section we will therefore take a closer look on what it is that makes an expert.

1.2 Experts' Ingredients

In the past decades, research on expert knowledge offered a rich basket of results on how experts and novices differ in a variety of domains, for example, in science/education, playing, music, and sport. Following Chi [5], "expertise, by definition, refers to the manifestation of skills and understanding resulting from the accumulation of a large body of knowledge". But how do these skills manifest, what is the basis for the deeper understanding of experts? In this section, we want to give a brief overview on some key findings in this field of research. In the following, we will describe how these findings might influence sampling and settings for user testing in the field of visual data analysis.

Much research on experts has been carried out in the area of problem solving, where the ability to rearrange and readapt knowledge representations turned out to be one of the crucial factors: Experts have a variety of strategies of searching through so-called "problem spaces" [12] and they search more effectively. They use a variety of *search heuristics* like backward- instead of forward-working searches [17][10] and hybrid breadth-in-depth searches. Another difference in the organisation of domain expertise is more procedural: In a study about representation differences in physics, experts worked rather *principle-based* while beginners categorize problems more on *situational criteria* or formula-based [7]. With rising expertise, deeper principles are more frequently used than shallow surface features. Chi [5] reports that novices do not differ in the number of statements used to describe biological systems and devices, but experts differed significantly in the number of statements referring to system behaviours and system functions. In summary, the experts' knowledge is not just a large unordered dump of facts, but rather a finely structured organization, where declarative and procedural mechanisms cooperate in an effective way [11].

Aside from these strategies, experts have superior skills in *encoding patterns* [3][4]. Their domain knowledge representation differs from those of novices: They perform better in memorization tasks by building larger chunks. Therefore, experts' knowledge seems to be better organized. On a lower level of processing, Chi and Koeske [6] found that experts show better skills in the detection of subtle *cues* in medical diagnostic tasks and they are better in analysing the *relations* among different cues. Furthermore, experts perceive differently and excel in finer *discrimination* skills, also in the field of scientific data visualizations [8]. To sum up, knowledge and problem spaces, search and recall strategies, and perceptive skills are some of the main ingredients that constitute expert minds in a general view.

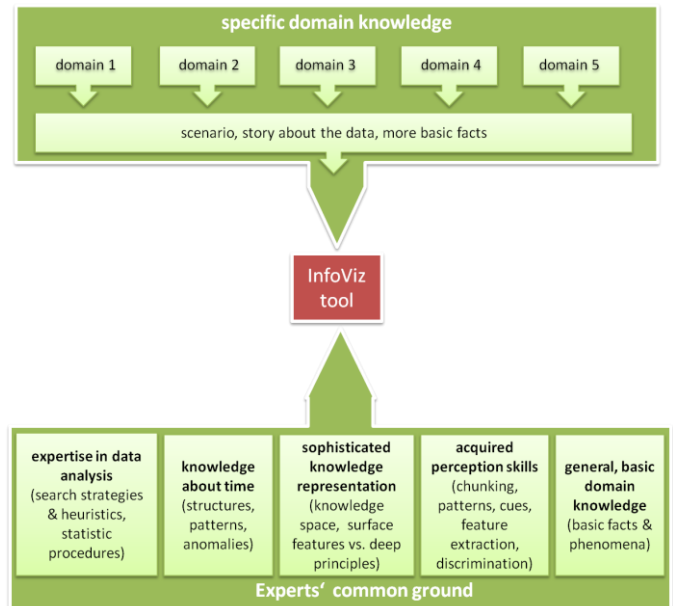


Figure 1: Different forms of expertise that might influence the usage of an InfoViz tool: Specific domain knowledge (top) and experts' common expertise (bottom).

Let us look at the intended target group in the project DisCo and narrow the boundaries of the experts' description further: In addition to the normal, generalized strengths of domain experts, our temporal data explorers have expertise in data analysis. They know the procedures, recommendations and pitfalls of statistical and visual data analysis and they possess a deep knowledge about temporal structures, patterns to look for and the peculiarities of the time domain.

Finally, the experts of our target group differ from each other in the deep knowledge and understanding in everyone's own specific domain. This kind of knowledge has the same mechanisms like any other expertise, but is full of declarative facts and procedures which are genuine to the domain.

Figure 1 gives an overview on many possible impacts of expertise when using an InfoViz tool. On the upper half, the experts' specific domain knowledge is pictured. To give an example from our study, financial experts know "that political changes will influence the USD-Euro exchange rates in a positive way for one or two days, but only if conservative parties

win” (Participant 9, see Table 1). We assume that most experts without specific financial expertise will not know these facts. But temporal explorers without financial expertise may also search for critical events (like political changes) and may apply some general, basic knowledge about the financial market, using the skills and knowledge illustrated in the lower part of Figure 1, called the experts’ common ground.

With this characterisation of temporal experts in mind, we want to present the evaluation of tools which should show good performance across domains. To make this evaluation possible we wanted to clarify some questions about domain-specific and cross-domain expertise: How influential is specific domain expertise when the insight performance is measured? Do we have to expect different numbers of insights for familiar and unfamiliar domains? Do the insights differ in quality for familiar and unfamiliar domains?

2. INSIGHT STUDY

Nine experts with many years of theoretical and practical expertise in the exploration and analysis of time-oriented data participated in the insight study. Most of them had specific domain expertise in the fields of business intelligence, management, education, or scientific data analysis. Table 1 gives an overview.

Table 1: Expertise of the participants of the insight study and the selected datasets (N=9).

	gender	expertise	datasets			
			economic turnover	traffic accident	education	finance
P1	f	education, traffic, management		●	●	
P2	f	education, management & controlling			●	
P3	f	education, business data analysis	●		●	●
P4	f	business data analysis, controlling	●			●
P5	m	education, scientific data analysis, controlling			●	
P6	m	education, computer science			●	
P7	m	business data analysis	●			●
P8	m	business data analysis, management	●			●
P9	m	business data analysis	●			●

To deal with this mixture of expertise in the sample, four datasets were collected to come up with an offer that suits each of the experts’ domain at least once. Additionally, to decrease the difference between the experts’ knowledge, we presented a short introduction about datasets, similar to Whiting et al. [20]. These dataset stories consisted of basic information (a short domain and

data description), quirks and peculiarities in the data set (special events, anomalies and exceptions as well as data acquisition errors). For example, the story about the financial dataset described the currency exchange rates for Euro – USD, the time interval and the range of the dataset. Also short- and long-term factors of influence were described in brief. The peculiarity that early exchange rates were synthesized and a few main events (like the date of the introduction of cash money) were also noted.

All participants had to solve six to seven tasks for each dataset with the aid of one of the two prototypes of visualization tools. These prototypes - Groove **Fehler! Verweisquelle konnte nicht gefunden werden.** and a multi-scale variant [16] - were assigned randomly to the datasets. Whereas Groove allows users to interactively fold and unfold time scales, the multi-scale shows all temporal granularities one below the other. The tasks consisted of basic visualization tasks like “name a value for a particular date” up to higher level tasks like “in which way does the current financial crisis influence the currency exchange rates” and open tasks like “which global and local trends do you see in the dataset”.

Thinking aloud, interaction logs, and viewing behaviour were recorded. The transcripts of the think aloud data were segmented and categorized in a similar manner to the categories proposed by Sarayaia, North and Duca [15] with a few noteworthy modifications: While they – as discussed above – only concentrated on deep, complex, and non-trivial insights, we also counted more shallow, elementary, and expectable discoveries as insights. Similar to the concept *knowledge building insights* proposed by Chang et al. [2], we defined insights as the understanding gained by an individual using a visualization tool (or parts thereof) for the purpose of data analysis, which is a gradual process towards discovering new knowledge (see [18] for examples and further discussion). This was necessary to ensure fair testing conditions across the experts, independent from their domain expertise. Otherwise, the non-experts might not have gathered many insights at all. Furthermore, if any hints for the use of prior knowledge are articulated, we also coded them.

3. RESULTS

To determine, whether differences exist between experts with and without domain knowledge, we compared the time the participants took to explore each data set and the associated number of data insights they generated.

The quantitative analysis shows no profound difference in the time the participants with and without expertise in a specific domain spent on each data set ($t = -0.29$, $df = 24$, $p > .05$). They gained a similar amount of insights into the data on the overview ($t = 1.44$, $df = 26$, $p > .05$) and on the detailed level ($t = 0.60$, $df = 26$, $p > .05$). Consequently, no differences between the average time needed for one data insight were observed ($t = -1.65$, $df = 24$, $p > .05$).

During the evaluation, we observed qualitative differences in the participants’ exploratory data analysis: Some participants explored the data set only descriptively, focusing on the flow of data mainly and only to a lesser extent on the meaning of their observations.

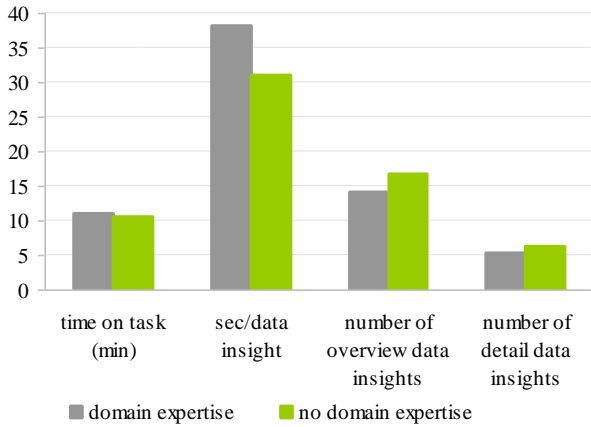


Figure 2: Differences in insight generation of participants with and without domain expertise.

For example, commenting the statistic with different data values marked in different shades of yellow to red:

"At noon there is a red belt." (P5)

Other participants focused more on the content and on the meaning as regards to the content.

"There is quite some breakfast business." (P8)

In [18] we proposed a method to analyse the generation of data insights dependent from prior knowledge and insights into the tool. Because the relations between insights play a major role, we called this method the *Relational Insight Organizer* (RIO). Figure 3 shows two examples for this kind of analysis, where the two different strategies to generate insights can be clearly distinguished.

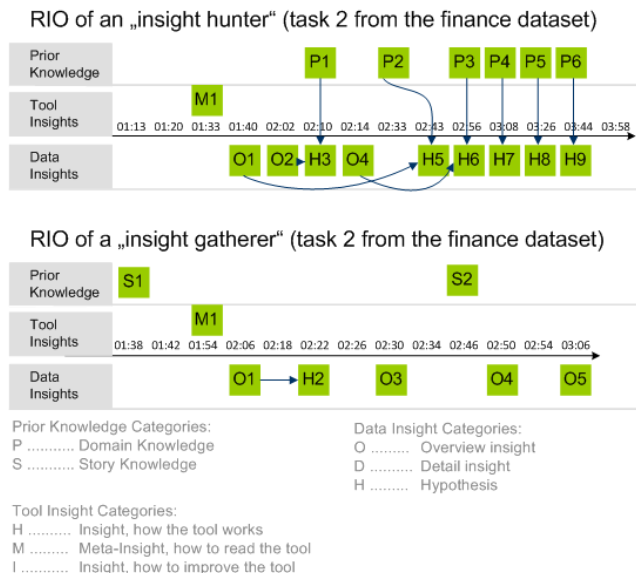


Figure 3: Relational Insight Organizer (RIO) for two different insight strategies.

Interestingly, this qualitative difference could not be explained through domain expertise. Only 50% of the domain experts but also 25% of the non-domain experts engaged in this more

interpretative behavior. The domain expertise showed no significant effect in our analysis ($CC = .25, p > .05$).

Rather than differentiating users according to their expertise level, a user typology on the basis of the type of insight (data characteristics vs. content / meaning) seems productive in this study. *Insight gatherers* concentrate on describing the characteristics of the data flow, whereas *insight hunters* use even the smallest domain information available to create novel interpretations and make as much sense of the data as possible (see Figure 3).

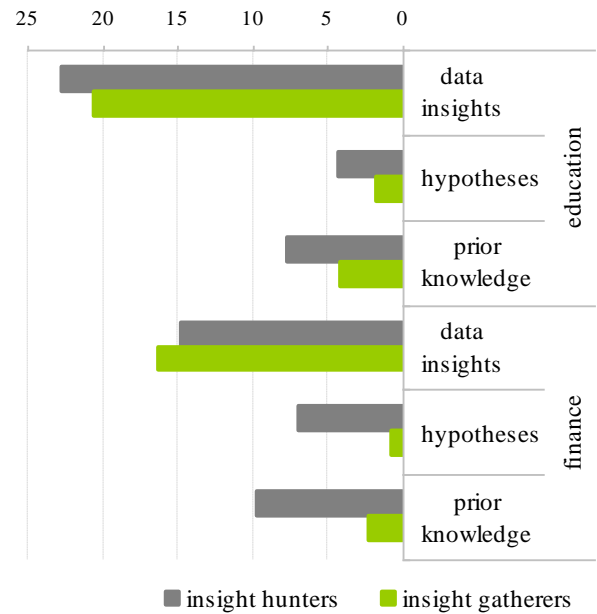


Figure 4: Number of data insights, hypothesis generation and use of prior knowledge for two different insight strategies¹.

As displayed in Figure 4 for two datasets, insight hunters have a similar amount of insights as insight gatherers ($t = 0.16, df = 14, p > .05$) but they differ highly from each other in the use of prior knowledge ($t = -4.87, df = 14, p < .001$) and the generation of hypotheses ($t = -3.80, df = 7.03, p < .01$): Insight hunters more often associate prior knowledge with the data insights gained to generate possible explanations for their observations of the data.

4. DISCUSSION

Differences between domain experts and novices were addressed already extensively in prior studies (see [5] for a review). We looked at the insights of experts in their own and in a foreign domain more detailed and found that a more differentiated view on domain expertise is necessary – at least for information visualizations this is the case.

Experts in the presented insight study did not differ from participants without domain-specific expertise in the number of

¹ Though a statistically significant difference on the number of hypotheses and the use of prior knowledge exists between participants with and without expertise as well, the effect size is much larger for the differences between the insight strategies.

insights generated or in the time needed for an insight. A qualitative analysis showed that some participants – the majority of them being participants with specific domain expertise, but also some without – engaged more deeply with the dataset at hand: They generated more hypotheses and used prior knowledge more frequently – they actively hunted insights. The other participants did stay on a descriptive level: They gathered insights, but did generate only some hypotheses based on their prior knowledge.

What do these findings implicate for insight methodology in the evaluation of information visualizations? One important consequence is that testing visualizations across domains can be done by selecting experts across domains as well. The experts are able to compensate for missing domain-knowledge with the help of their common ground on temporal data analysis and a story with background knowledge on the dataset.

Another important consequence regards the sampling in insight studies: Though most studies selected users with domain-specific expertise, we found that also without this domain-specific expertise users generate a similar amount of insights – given that these users have a general knowledge in data analysis (in our study: of temporal data) and given that insights are defined more widely; as all (even shallow) discoveries in the data. However, we must constrain our argument because in daily use experts will more likely engage in insight hunting and interpreting data than in insight gathering and describing data only. Therefore, for long-term evaluation in real-life settings our argument will not hold and real experts with domain-specific knowledge have to be sought as participants. Still, even in this user group of domain experts, only every second participant in our study did hunt for insights. To choose domain experts as participants for an evaluation is not a guaranteed success either. Consequently, we suggest being less restrictive in sampling, if the purpose of evaluation is a quick, less extensive evaluation of an early prototype or mock-up during iterative engineering based on the amount of insights only.

By analyzing the relation of insights to prior insights and prior knowledge, we differentiated between two types of users – insight hunters and insight gatherers. Identifying types of users could be relevant for other evaluations as well: You can more easily choose users with the most potential to further improve the tool (and hunters' insights are undoubtedly the best argument to sell). You can also identify groups of users which cannot fully exploit the potential of the tool and adapt the tool to meet their requirements in a better way.

In the insight study presented we gained interesting results which led us to question the role of expertise for insight methodology. It remains to be studied, whether the categorization of insight hunters and insight gatherers is exhaustive or a further differentiation is beneficial. We focused on the way users described their data insights to make the distinction. Further research should also take the users' behaviour, that is, interaction with the tool and visual work, into account to enrich the specification of insight hunters and gatherers. Prior research showed that experts have better discrimination skills [8] and apply better search strategies [10][17]. In our study we were not able to address this question, but it would be interesting to

expand this focus of research to insight gatherers and hunters as well.

From a methodological perspective, the insight approach can be seen as limited due to a self selection bias which insights are reported. For example, one could assume that experts would report rather elaborated than simplistic insights. Although we did not analyse the quality of insights in detail, we got the impression that experts also reported simplistic insights while analysing the think aloud protocols. In our view, this could be explained by two factors: First, the setting of the experiment facilitated the production of traceable insights due to the fact that the investigators were non-domain experts. Therefore, the participants tried to explain their insights in a rather comprehensible manner, without excluding basic steps. Another explanation could be that most of our participants were consultants or from the academic field with teaching experience, and they were therefore familiar to communicate their thoughts not only on a highly elaborated level. However, to answer this question more precisely, further research is needed.

Domain expertise is a field of research that already gained much attention. Overall, our findings are in line with prior research on this topic. By using participants with a common ground of expertise on temporal exploration and providing basic information on different datasets in stories, the participants' insights resemble those of other domain experts. Therefore, we would like to conclude that an expert is not an expert is not an expert. As evaluators of information visualization tools we always hunt for the best test users possible. But sometimes, it is sufficient to gather the second choice, whom you can get more easily.

5. ACKNOWLEDGMENTS

This research was conducted within the DisCo research project, supported by the program "FIT-IT Visual Computing" of the Federal Ministry of Transport, Innovation and Technology, Austria. Special thanks to our project partners Ximes Inc. for their help with the datasets and in finding participants as well as to the Department of Information and Knowledge Engineering at the University of Krems for the good cooperation and their conceptual and programming work.

6. REFERENCES

- [1] Card, S. K., Mackinlay, J. D., and Shneiderman, B. 1999. Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann.
- [2] Chang, R., Ziemkiewicz, C., Green, T. M., and Ribarsky, W. 2009. Defining insight for visual analytics. *IEEE Comput. Graph.* 29(2), 14-17.
- [3] Chase, W. and Simon, H. 1973a. Perception in chess. *Cognitive Psychol.* 4, 55-81.
- [4] Chase, W. and Simon, H. 1973b. The mind's eye in chess. In *Visual Information Processing*, W. G. Chase, Ed. Academic Press, New York, 215-281.
- [5] Chi, M. T. H. 2006. Laboratory methods for assessing experts' and novices' knowledge. In *The Cambridge Handbook of Expertise and Expert Performance*, K. A.

- Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffmann, Eds. Cambridge University Press, New York, 167-184.
- [6] Chi, M. T. H., and Koeske, R. 1983. Network representation of a child's dinosaur knowledge. *Dev. Psychol.* 19, 29-39.
- [7] Chi, M. T. H., Feltovich, P., and Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Sci.* 5, 121-152.
- [8] Hoffman, R. R., Trafton, G., and Roebber, P. 2006 *Minding the Weather: How Expert Forecasters Think*. MIT Press.
- [9] Lammarsch, T., Aigner, W., Bertone, A., Gärtner, J., Mayr, E., Miksch, S., and Smuc, M. 2009. Hierarchical Temporal Patterns and Interactive Aggregated Views for Pixel-based Visualizations. In *Proceedings of 13th International Conference on Information Visualisation, IV2009*. IEEE Computer Society Press, Washington, DC, 44-49.
- [10] Larkin, J. 1981. Cognition of learning physics. *Am. J. Phys.* 49, 534-541.
- [11] Larkin, J., McDermott, J., Simon, D., and Simon, H. 1980. Models of competence in solving physics problems. *Cognitive Sci.* 4, 317-348.
- [12] Newell, A. and Simon, H. A. 1972 *Human Problem Solving*. Prentice-Hall.
- [13] Plaisant, C., Fekete, J. D., and Grinstein, G. 2008. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE T. Vis. Comput. Gr.* 14(1), 120-134.
- [14] Rester, M., Pohl, M., Wiltner, S., Hinum, K., Miksch, S., Popow, C., and Ohmann, S. 2007. Evaluating an InfoVis technique using insight reports. In *Proceedings of the 11th International Conference Information Visualization, IV 2007*. IEEE Computer Society, Washington, DC, 693-700. DOI= <http://dx.doi.org/10.1109/IV.2007.44>.
- [15] Sarayaia, P. B., North, C., and Duca, K. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Trans. Vis. Comput. Graph.* 11, 443-456.
- [16] Shimabukuro, M., Flores, E. F., de Oliveira, M. C. F., Levkowitz, H. 2004. Coordinated views to assist exploration of spatio-temporal data: A case study. *Proceedings of the 2nd International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV04)*, IEEE CS Press, 107-117. DOI= <http://dx.doi.org/10.1109/CMV.2004.1319531>.
- [17] Simon, D. and Simon, H. 1978. Individual differences in solving physics problems. In *Children's Thinking: What Develops?* R. Siegler, Ed. Erlbaum, Hillsdale, NJ, 325-348.
- [18] Smuc, M., Mayr, E., Lammarsch, T., Aigner, W., Miksch, S., and Gärtner, J. 2009. To score or not to score? Tripling insights for participatory design. *IEEE Comput. Graph.* 29(3), 29-38.
- [19] Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., and Raymond, P. D. 2000. Turning pictures into numbers: Extracting and generating information from complex visualizations. *Int. J. Hum-Comput. St.* 53(5), 827-850.
- [20] Whiting, M. A., Haack, J., and Varley, C. 2008. Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software. In *Proceedings of the 2008 conference on BEYond time and errors: novel evaluation methods for Information Visualization*. ACM Press, New York, 8-15. DOI=<http://doi.acm.org/10.1145/1377966.1377977>.
- [21] Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., and Carpendale, S. 2006. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI workshop on BEYond time and errors: novel evaluation methods for information visualization*, ACM Press, New York, 1-6. DOI= <http://doi.acm.org/10.1145/1168149.1168162>.